# SUPPLEMENTARY MATERIALS

## Table of content

## Section S0: Overview of the experimental design

| | Controls | | | Replicates | | | | |
|---|---|---|---|---|---|---|---|---|
| | **C1** | **C2** | **C3** | **A** | **B** | **C** | **D** | **E** |
| | | | | | | | | |
| **Microcosms** | | | | | | | | |
| Bottle volume (L) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP buffer (g) | 500 | 500 | 500 | 500 | 500 | 500 | 500 | 500 |
| Substrate, unprinted office paper (g) | - | - | - | 5 | 5 | 5 | 5 | 5 |
| Inoculum (wet mass in g): stabilised thermophilic digestate | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| Headspace gas | $N_2$ | $N_2$ | $N_2$ | $N_2$ | $N_2$ | $N_2$ | $N_2$ | $N_2$ |
| Temperature (°C) | 55 | 55 | 55 | 55 | 55 | 55 | 55 | 55 |
| Incubation duration (days) | 120 | 120 | 120 | 60 | 120 | 120 | 120 | 120 |
| | | | | | | | | |
| **Analyses** | | | | | | | | |
| Biogas production and composition | X | X | X | X | X | X | X | X |
| pH | X | X | X | X | X | X | X | X |
| VFA, TOC, TIC | X | X | X | X | X | X | X | X |
| Isotopy | X | X | | X | X | X | X | X |
| | | | | | | | | |
| ARISA for bacteria | | | | X | Day 60 | | | |
| ARISA for archaea | | | | X | Day 60 | | | |
| | | | | | | | | |
| FISH | | | | Day 60 | | | | |
| | | | | | | | | |
| 16S pyrosequencing for bacteria | | | | Digestate Inoculum<br>Days 0, 60 | Days 60, 73 | | | |
| 16S pyrosequencing for archaea | | | | Digestate Inoculum<br>Days 0, 60 | Days 60, 73 | | | |
| | | | | | | | | |
| Pseudo 2D protein gel (offgel + chip) | | | | Day 60 | | | | |
| Metaproteomics | | | | Day 60 | | | | |

## Section S1: Protein analysis methods

The protein analyses were performed according to the procedure shown in **Figure S1**. The detailed procedure for each step is further decribed below.

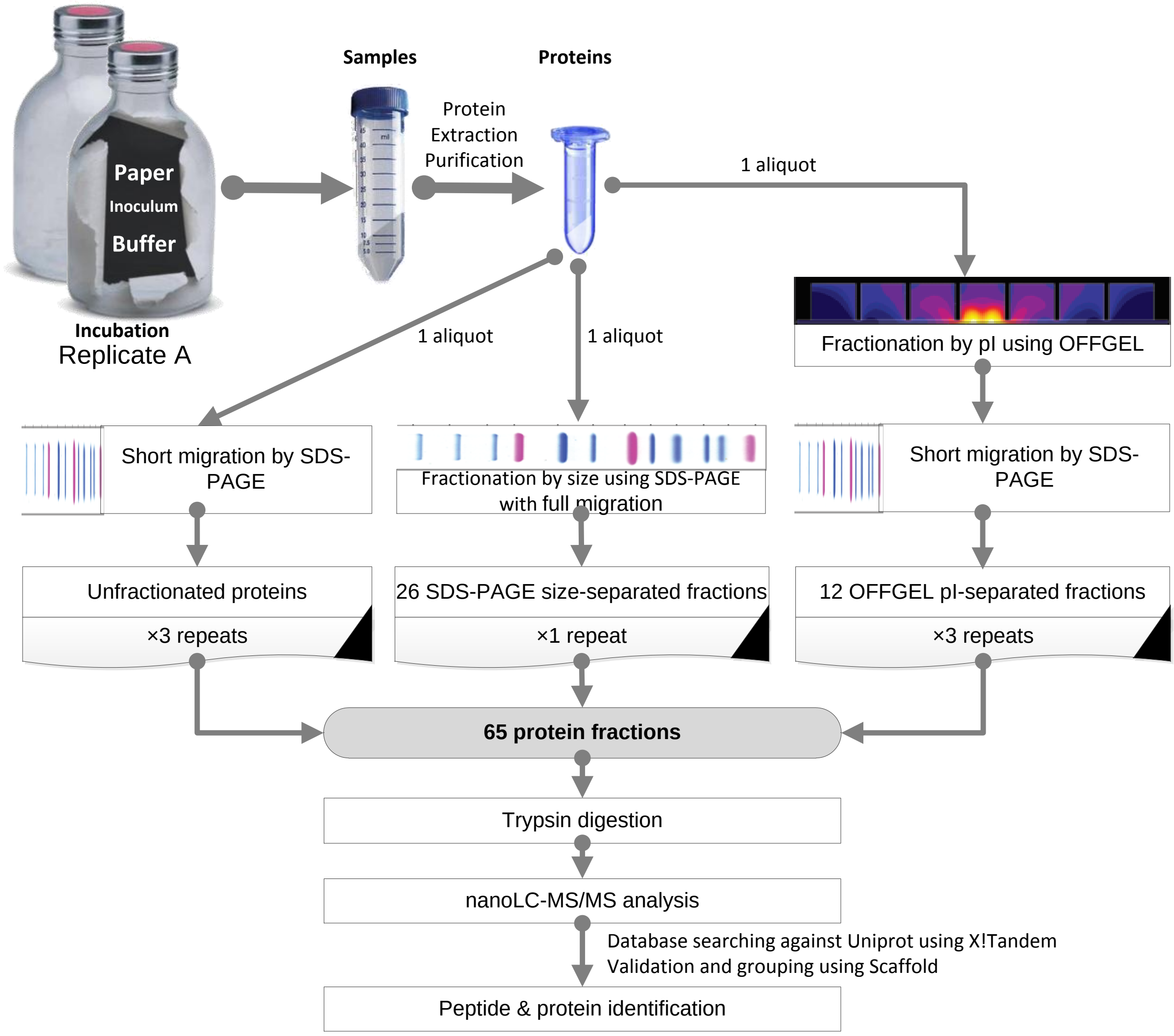

**Figure S1. Detailed workflow of protein analyses.**

### *Protein extraction and purification*

The 400 mL incubation medium collected at day 60 was separated into 50 mL aliquots in sterile screw cap tubes, centrifuged at 13100×g at 4$^{o}$C for 20 min and frozen at -80$^{o}$C. The protein extraction method was modified from (Wilmes and Bond, 2004) for better adaptation to samples containing solid waste residues. 15 mL of freshly prepared urea-thiourea-CHAPS (UTCHAPS) lysis buffer and five 3 mm diameter glass beads were directly added to one frozen pellet of about 4 g. The UTCHAPS lysis buffer consisted of 7 M Urea, 2 M thiourea, 4% (w/v) CHAPS (zwitterionic detergent, 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate), 10 mM Tris-1 mM EDTA, 50 mM dithiothreitol (DTT) and 1 mM PMSF (protease inhibitor, phenylmethylsulfonyl fluoride). The matrix was resuspended by vortexing and then placed on ice for 1.5 h with 30 s vortexing every 5 min. The matrix was mechanically lysed by 10 cycles of 30 s-bead beating and 2 min-cooling on ice. 0.1 mm Zirconia/Silica Beads (Biospec Products, Inc., USA) and 40 Hz vortex were used for bead beating. After lysis, the suspension was incubated on ice again for 30 min, settled down to remove beads, and then centrifuged at 10$^{o}$C successively at 200×g for 15 min, 5000×g for 15 min and 13100×g for 30 min to remove cell debris and solid waste matrix. The pellets from three cycles of above centrifugation were combined, resuspended in 4 mL of UTCHAPS by pipeting and pulse-vortexing, subjected to 5 cycles of 1 min-sonication on ice and 1 min-cooling on ice. The beads were washed with 1 mL of UTCHAPS. The washing effluent and the sonicated suspension were combined and centrifuged at 13100×g, 10$^{o}$C for 30 min. The supernatants from four cycles of above centrifugation were mixed, centrifuged at 22000×g, 10$^{o}$C for 30 min. The resulting supernatant (about 20 mL) remaining dark and turbid was retained as the crude protein extract.

The crude protein extract was purified by trichloroacetic acid (TCA)/acetone precipitation method. Briefly, one fifth volume of 100% aqueous TCA solution and a same volume of -20$^{o}$C cold acetone were added to the crude protein extract, to obtain a final TCA concentration of 10%. The mixture was incubated at -20$^{o}$C for 4 hr and centrifuged at 13100×g, 4$^{o}$C for 30 min. The supernatant was mixed again with one eighteenth volume of 100% aqueous TCA solution and half volume of -20$^{o}$C-cold acetone, incubated overnight at -20$^{o}$C and centrifuged at 13100×g, 4$^{o}$C for 45 min. The two pellets from the latter centrifugation were washed with 5 mL of -20$^{o}$C-cold acetone, incubated for 30 min at -20$^{o}$C and centrifuged at 22000×g, 4$^{o}$C for 10 min. The pellets were washed again with 5 mL -20$^{o}$C-cold acetone, centrifuged at 22000×g, 4$^{o}$C for 10 min and dried at room temperature. The dried protein pellets were suspended in a solution comprising 7 M urea and 2 M thiourea, by vortexing and mixing during 2 hr and the obtained solution was centrifuged at 22000×g, 10$^{o}$C for 45 min. The resulting supernatant was conserved as the purified protein. Protein content was quantified using a 2-D Quant Kit (GE Healthcare) and a High Sensitivity Protein 250 Kit (Agilent).

### *Fractionation with OFFGEL electrophoresis*

About 1 mg of the purified proteins was subjected to OFFGEL electrophoresis (Chenau et al., 2008; Hubner et al., 2008) for fractionation according to their pI. The Agilent 3100 OFFGEL Fractionator with a 12-well setup and the Agilent 3100 OFFGEL low resolution kit (pH 3-10) were used according to the standard protocol of the supplier. 12-cm-long IPG gel strips with a linear pH gradient ranging from 3 to 10 were rehydrated in the assembled device with OFFGEL buffer. The protein solution was diluted in OFFGEL buffer and loaded into wells. OFFGEL buffer consisted of urea, thiourea, DTT, glycerol and ampholyte. The parameters for fractionation were set according to default method OG12PR01, i.e., the sample was focused with a maximum current of 50 µA and typical voltages ranging from 500 to 4000 V until 20 kVh was reached after 15 h. The 12 recovered fractions (volumes between 50 and 200 µLwere used for downstream analyses.

### *OFFGEL electrophoresis combined with high-sensitivity on-chip protein detection*

An analytical 2D-GE-type analysis was realized by combining OFFGEL fractionation based on the isoelectric point followed by high-sensitivity on-chip electrophoresis. 2 µL fluorescent labeling dye included in the High Sensitivity Protein 250 Kit (Agilent part number 5067-1575) were added to about 100 µg sample protein and the mixture was incubated on ice for 30 min. Excess dye was quenched by the addition of 2 µL ethanolamine, followed by 10 min incubation on ice. The entire labeled protein was diluted in OFFGEL solution and fractionated on Agilent 3100 OFFGEL with the described above OG12PR01 method. The resulting 12 fractions were subsequently analyzed using the Agilent 2100 bioanalyzer and the Agilent High Sensitivity protein 250 assay, according to the standard protocol of the supplier. Briefly, 4 µL of each fractionated sample were mixed with 2 µL of buffer containing DTT and incubated at 95-100 $^{o}$C for 5 min. The wells and the channels on the chip were filled with both gel matrix and destaining solution. The samples and the ladder were loaded and the chip run was launched immediately afterwards.

### ***Fractionation with SDS-PAGE electrophoresis***

About 10 µg sample protein were separated by one-dimensional (1D) sodium dodecyl sulfate-polyacrylamide gel electrophoresis as previously described (Ibrahim *et al.*, 2007). Each 1D electrophoresis lane was cut into 26 gel pieces (2 mm in width).

To be compatible with the existing platform for protein identification using mass spectrometry, the protein in solution were embedded in SDS gels. Therefore, all the above-mentioned 12 OFFGEL fractions and a non-fractionated total protein sample (about 10 µg) were migrated in the same SDS-PAGE system for a very short time period. The top of each 1D electrophoresis lane, containg the proteins after the short time migration, was cut into 4 gel pieces (2 mm in width).

The gel pieces including 12 OFFGEL fractions, 26 SDS fractions and 1 original protein sample were digested in gel and the mass spectrometry analyses (MS/MS) were performed at the proteomic platform INRA-PAPPSO (Plateau d'Analyse Protéomique par Séquençage et Spectrométrie de Masse, Institut National de Recherche Agronomique, Jouy en Josas, France). The protein digestion and mass spectrometric analyses were performed twice independently for the original sample and for the OFFGEL fractions.

The above-mentioned preparation procedure and MS/MS analyses were applied once more independently to the same incubation sample and resulted in 12 analysed OFFGEL fractions and 1 analysed original protein sample. In total, the proteomic analyses thus generated 65 protein fraction datasets (**Figure S1**, orginal protein sample: 1×2+1; OFFGEL fractions: 12×2+12; SDS fractions: 26×1+0).

***Protein digestion***

In-gel digestion of the proteins was performed with the Progest system (Genomic Solution) according to the following protocol. Gel pieces were washed, first in two successive baths of (i) 10% acetic acid–40% ethanol and (ii) 100% acetonitrile (ACN) and, second, in two successive baths of (i) 25 mM $NH_4CO_3$ and (ii) 100% ACN. A total of 40 µl was used for each bath. Gel pieces were further incubated in 40 µl of 10 mM dithiothreitol in 25 mM $NH_4CO_3$ for 30 min at 55°C and in 30 µl of 50 mM iodoacetamide in 25 mM $NH_4CO_3$ for 45 min at room temperature for cysteine reduction and alkylation, respectively. Digestion was subsequently performed for 6 h at 37°C with 20 µl per gel piece of a solution prepared with125 ng modified trypsin (Promega) dissolved in 20% methanol and 20 mM $NH_4CO_3$. The peptides were extracted successively with (i) 20 µl of 0.5% trifluoroacetic acid–50% ACN and (ii) with 20 µl of 100% ACN. The resulting peptide extracts were dried in a vacuum centrifuge and suspended in 25 µl of 0.08% trifluoroacetic acid and 2% ACN prior to LC-MS/MS analyses.

### *Mass spectrometry analyses*

Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) analyses were performed with an Ultimate 3000 LC system (Dionex, Voisins le Bretonneux, France) connected by a nanoelectrospray interface to a linear ion trap mass spectrometer coupled to an Orbitrap detector (LTQ-Orbitrap, Thermo Fisher, Waltham, MA, USA; PAPPSO proteomic platform, INRA, Jouy-en-Josas). Four µl of tryptic peptide mixtures from each sample were loaded at a flow rate of 20 µL/min onto a precolumn (Pepmap C18, 0.3 × 5 mm, 100 Å, 5 µm; Dionex). After 4 minutes, the precolumn was connected to a separating Pepmap C18 nanocolumn (0.075 × 15 cm, 100 Å, 3 µm; Dionex) and the gradient was set to 300 nL/min. All peptides were separated in the nanocolumn using modified buffer elut-B with a linear gradient of acetonitrile from 2% to 36%, for 18 minutes. The eluting buffers were buffer elut-A: 0.1% formic acid, 2% acetonitrile and buffer elut-B: 0.1% formic acid, 80%

acetonitrile. The total run duration was 50 minutes, including the regeneration step. Ionization was performed at the liquid junction, with a spray voltage of 1.3 kV applied to a noncoated capillary probe (PicoTip Emiter, 10 μm ID; New Objective, Woburn, MA, USA). The peptide ions were analyzed using the *N*th-dependent method as follows: (1) full mass spectrometry scan (*m*/*z* 300–2000); (2) ZoomScan (scan of the 3 major ions); and (3) MS/MS on these 3 ions using standard peptide fragmentation parameters (*Qz*=0.25, activation time=30 ms, collision energy=40%).

### *Protein identification by spectral matching and validation by false discovery rate and probability algorithm*

The raw data produced on LTQ-Orbitrap mass spectrometer were first converted into mzXML format files with ReADW (http://sashimi.sourceforge.net) and, in a second step, protein identification was performed with X!Tandem software (X!Tandem CYCLONE 2010.12.01.1 [http://www.thegpm.org]) (Fenyö and Beavis, 2003) against Uniprot database (version January 2012, http://uniprot.org). The X!Tandem search parameters were as follows: trypsin specificity with one missed cleavage, fixed alkylation of cysteine, and variable oxidation of methionine. The mass tolerance was fixed to 10 ppm for precursor ions and 0.5 Da for fragment ions. For all proteins identified with a protein E-value of <0.01 in the first step, additional peptides were searched to reinforce identification, using similar parameters except that semitryptic peptides and protein N-terminal acetylations were accepted. All peptides identified with an E-value of <0.1 were conserved. All the obtained xml result files (total of 65, one for each fraction) were imported into the proteome software Scaffold 2.0 (Proteome Software, Inc.) to calculate peptide and protein probability. The filtering standard used to screen proteins for biological analysis was, protein probability > 95%, at least 2 unique peptides, the peptide probability of at least one unique peptide > 90%.

False discovery rate (FDR) was used to differentiate between true and false peptide identifications. False peptide sequences were assigned by using a decoy database (Käll *et al.*, 2008). The decoy database was generated from the original database by reversing protein sequences and then appending the obtained reversed sequences to the original database. Proteins with the reversed orientations were given a common identifier to facilitate their extraction. The creation of the decoy database was realized using the free public java software “Database Manager” (PAPPSO platform, INRA [http://moulon.inra.fr/PAPPSO]).

## Section S2: DNA analysis methods

### *DNA extraction and analyses*

DNA was extracted using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Inc.) according to the manufacturer's instructions. Several 16S rRNA based techniques were conducted to describe the microbial community, more precisely the fingerprinting technique automated ribosomal intergenic spacer analysis (ARISA), 16S rRNA gene tag pyrosequencing, and fluorescent *in situ* hybridization (FISH).

### *Automated Ribosomal Interspacer Analyses (ARISA)*

For community diversity fingerprint analyses by ARISA, the extracted DNA was amplified using primers 1389F and 71R to target the 16S-23S rRNA gene intergenic regions of archaeal populations, and primers ITSf and ITSReub to target those of bacterial populations. PCRs were performed in a thermal cycler (Mastercycler®ep realplex$^2$, Eppendorf AG, Germany) using TaKaRa Ex Taq™ mixture (TAKARA BIOTECHNOLOGY (Dalian Co., LTD). For archaea, the mixture was held at 95 °C for 3 min, followed by 35 cycles of 94 °C for 1 min, 57.4 °C for 1 min, 72 °C for 2 min and a final extension at 72 °C for 10 min. For bacteria, the PCR procedure was as follows: 94 °C for 3 min, followed by 30 cycles of 94 °C for 45 s, 55 °C for 1 min, 72 °C for 2 min and a final extension at 72 °C for 7 min. The PCR products were analyzed on a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA) using a DNA7500 LabChip (Agilent Technologies).

### *16S rRNA gene tag pyrosequencing*

For phylogenetic and diversity analyses by 16S rRNA gene tag pyrosequencing, the DNA was isolated as described above from 6 distinct samples: the industrial digestation sludge sample, the inoculum (the stabilized industrial digestate), microcosm A days 0 and 60, microcosm B days 60 and 73. The primer set 340F and 806R was chosen to target archaeal populations and primer set of 27F and 519R to target bacterial populations. Multiplexed bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) was performed using the Titanium platform (Roche Applied Science, Indianapolis, IN) as previously described (Kumar *et al.*, 2011) in a commercial facility (Research and Testing Laboratories, Lubbock, TX). Briefly, a single step PCR was used to amplify the 16S rRNA gene targeted regions as well as to introduce adaptor sequences and sample-specific 8-mer oligonucleotide tags into the DNA.

The reads were demultiplexed based on the barcode tag sequences. The reads were trimmed by removing the adaptors and barcode tag sequences, and by filtering-trimming according to the signal quality, using a 50 b sliding window and a 25 quality score threshold. Sequences containing homopolymers longer than 6 b were discarded as well as sequences shorter than 150 bp or longer than 600 bp. After alignment of the sequences against the ‘Silva_108_core’ database with PyNAST, putative chimeric sequences were removed using ChimeraSlayer. The remaining sequences were finally clustered into operational taxonomic units (OTUs) at the 97% sequence similarity threshold (species level) with uclust. These analyses were conducted within the virtual environment provided by the QIIME pipeline (Caporaso *et al.*, 2010) and homemade Python scripts. For each OTU totalizing more than 3 sequences, the taxonomic assignment was performed on the longest sequence with RDP classifier at a 0.8 confidence threshold. With QIIME, Silva taxonomy databases ‘Silva_RDP_taxa_mapping_species.txt’ and ‘Silva_RDP_taxa_mapping_genus.txt’ were used together with the reference sequence set ‘Silva_108_rep_set.fna’. In parallel, the taxonomic assignment was also analysed using the on-line RDP classifier (rdp.cme.msu.edu/classifier.jsp) and the ARB software. The results were manually curated for some dominant OTUs. The primer set targeting the archaeal population also amplified sequences attributed to bacteria. As indicated on Research and Testing Laboratories’ website, this was most probably due to the low relative abundance of the archaeal population compared to the bacterial one, as can be expected during methanization experiments.

The sff tag sequence files were deposited in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) as BioProject PRJNA182049 under the following accession numbers:

| **Sample** | **SRA accession primer set for the bacterial population** | **SRA accession primer set for the archaeal population** |
|---|---|---|
| Digestate (Dig) | SRR868692 | SRR868694 |
| Inoculum (Inoc) | SRR868690 | SRR868691 |
| Microcosm A, day 0 (A0) | SRR868688 | SRR868699 |
| Microcosm A, day 60 (A60) | SRR868686 | SRR868687 |
| Microscosm B, day 60 (B60) | SRR627748 | SRR627749 |
| Microcosm B, day 73 (B73) | SRR868682 | SRR868684 |

### *Fluorescent in situ hybridization (FISH)*

On pellets originating from 1 mL of samples, FISH was used to target *Coprothermobacter*, *Caldicellulosiruptor* and *Clostridium* members.

The 16S rRNA-targeting fluorescent probes used in the study are listed in **Table S2**. These oligonucleotide probes were labeled in 5’ with one dye, eitherfluoresceinisothiocyanate (FITC) or with one of the sulfoindocyanine dyes indocarbocyanine (Cy3) and indodicarbocyanine (Cy5). The probe specificities were confirmed using theARB database (http:/www.arb-home.de) (Ludwig *et al.*, 2004).

The fixation technique was modified from (Laloui-Carpentier *et al.*, 2006). Briefly, the pellets were resuspended in 200 µL phosphate-buffered saline (PBS, Sigma) and 600 µl 4% paraformaldehyde (Sigma). After incubation for 15 min at 4 $^{o}$C, the suspensions were centrifuged (13,800×g, 10 min) and resuspended in 500 µl 1×PBS and 500 µl 100% ethanol. Fixed cells were stored at -20 $^{o}$C. A conventional FISH protocol was followed (Manz *et al.*, 1992). An inverted Zeiss confocal laser scanning microscope (CLSM, LSM510-META) equipped with three lasers (Argon 488 nm, Helium –Neon 543 nm, Helium–Neon 633 nm) was used to visualize probe positive signals.

## Table S1. Primers used in the study

| Primer and probe name | Targeted phylogenetic groups | Sequences (5'-3') | Experiment type | Targeted site (*E. coli* 16S rRNA gene positions) | References |
|---|---|---|---|---|---|
| ITSF | Bacteria | 5'-GTC GTA ACA AGG TAG CCG TA-3' | ARISA | 1494-1513, 16S | (Cardinale *et al.*, 2004) |
| ITSReub | Bacteria | 5'-GCC AAG GCA TCC ACC-3' | ARISA | 37-23, 23S | (Cardinale *et al.*, 2004) |
| 1389F (Arch1406-1389r) | Archaea | 5'-ACG GGC GGT GTG TGC AAG-3' | ARISA | 1351-1368, 16S [a] | (Qu *et al.*, 2009) |
| 71R | Archaea | 5'-TCG CAG CTT RSC ACG YCC TTC-3' | ARISA | 78-57, 23S [a] | (Qu *et al.*, 2009) |
| 27F | Bacteria | 5'-AGA GTT TGA TCC TGG CTC AG-3' | 16S pyrosequencing | 8-27, 16S | (Lane, 1991) |
| 519R | Bacteria | 5'-GWA TTA CCG CGG CKG CTG-3' | 16S pyrosequencing | 536-519, 16S | (Lane, 1991) |
| 340F | Archaea | 5'-CCC TAY GGG GYG CAS CAG-3′ | 16S pyrosequencing | 321-338, 16S [a] | (Gantner *et al.*, 2011) |
| 806R | Archaea | 5'- GGA CTA CVS GGG TAT CTA AT -3' | 16S pyrosequencing | 760-741, 16S [a] | (Bates *et al.*, 2011) |

[a]: In the region of *Methanothermobacter thermautotrophicus* str. Delta H, accession number: NC_000916.

**Table S2. Fluorescently labeled rRNA-targeting probes used for FISH in the study and recommended hybridization conditions.**

| Probe name | Sequence (5'-3') | Targeted site (16S rRNA gene *E. coli* position) | Formamide % | Targeted microorganisms | Reference |
|---|---|---|---|---|---|
| Arch915 | GTG CTC CCC CGC CAA TTC CT | 915-934 | 20-30 | Most *Archaea* | (Padmasiri *et al.*, 2007) |
| Eub I | GCT GCC TCC CGT AGG AGT | 338 –355 | 0-50 | Most *Bacteria* | (Amann *et al.*, 1990) |
| Eub II | GCA GCC ACC CGT AGG TGT | 338 –355 | 0-50 | *Planctomycetales* | (Daims *et al.*, 1999) |
| Eub III | GCT GCC ACC CGT AGG TGT | 338 –355 | 0-50 | *Verrucomicrobiales* | (Daims *et al.*, 1999) |
| Copro929 | CUC CGC CGC UUG UGC GGA | 929-946 | 20-30 | *Coprothermobacter spp.* | The present study |
| Cal1415 | UCA GGU GUU GCU GAC UCU | 1415-1432 | 20-30 | *Caldicellulosiruptor spp.* | The present study |
| UCL284 | AAC CTC TCA GTT CGG CTA CC | 284-303 | 20-30 | *Clostridium Acetivibrio group*<br>*(including Clostridum thermocellum)* | (Li *et al.*, 2009) |

## Section S3: Incubation data

**Figure S2** shows the temporal evolution of physiochemical parameters during anaerobic incubation of office paper at 55°C with a thermophic inoculum. It indicates a general profile of batch anaerobic digestion, i.e. serial steps of hydrolysis, acidogenesis, acetogenesis and methanogenesis. The reaction rate of methanogenesis was lower than hydrolysis and acidogenesis. As a result, during the first 18 days, volatile fatty acids (VFA) were produced in high amount (about 360 mg-C/L) and the pH decreased from 7.5 to 5.8 and the production of methane was delayed. Afterwards, methanogenesis gradually accelerated until around day 60. At day 60, when samples were collected for protein analyses, 62% of the carbon in office paper had been degraded into gaseous methane and carbon dioxide and soluble carbon. At this time point, VFA were still remaining in the liquid phase (35-124 mg-C/L) and were mainly composed of acetate, propionate and lactate. After day 74, the VFA concentration decreased below 20 mg-C/L. Nevertheless, the TOC concentration remained constant at around 100 mg-C/L. At the end of incubation (day 113), 67% of carbon in office paper was either solubilised (converted to TOC) or transformed into biogas. As carbon used for cell-growth is usually below 5% under anaerobic conditions (Gerardi, 2003), this observation suggest that a significant part of the carbon contained in office paper could not be metabolized.

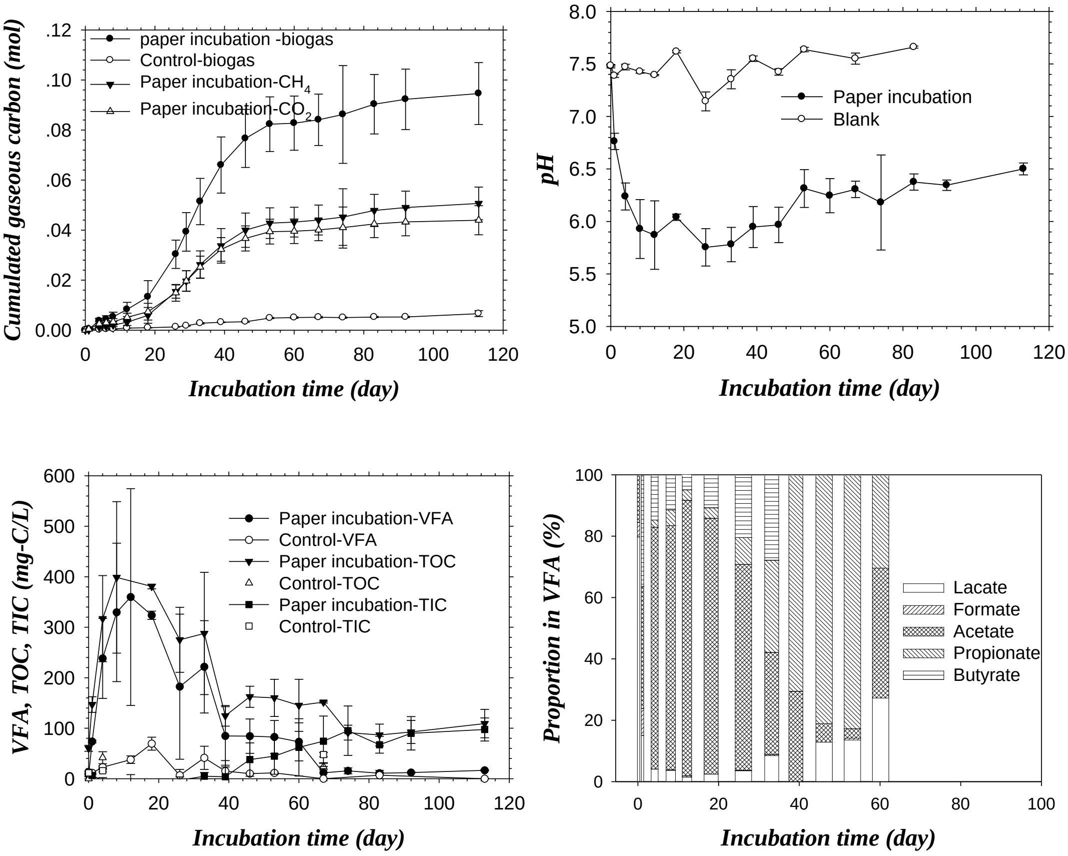

**Figure S2. Temporal evolutions for biogas production, pH, VFA, TOC and TIC in office-paper-containing and control batch incubations.**

VFA: Volatil fatty acids; TOC: Total organic carbon; TIC: Total iorganic carbon; mg-C/L: mg of carbon per liter

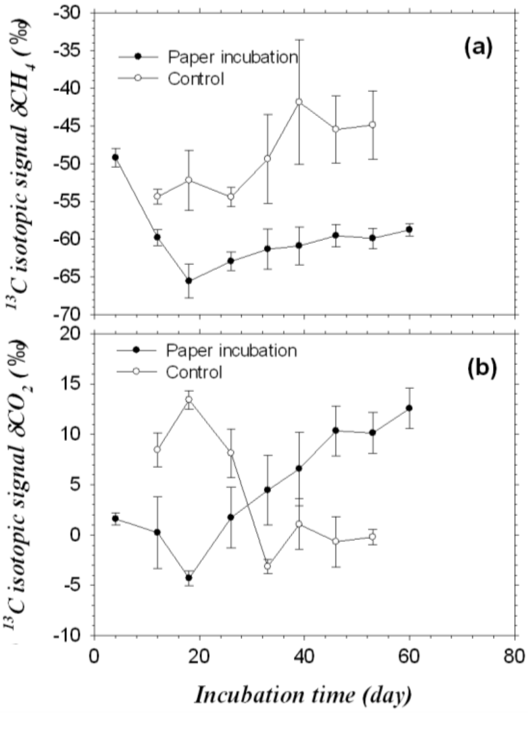

**Figure S3. Temporal evolution of the $^{13}$C isotopic signal for methane (δ$CH_4$) and $CO_2$ (δ$CO_2$) from the biogas.**

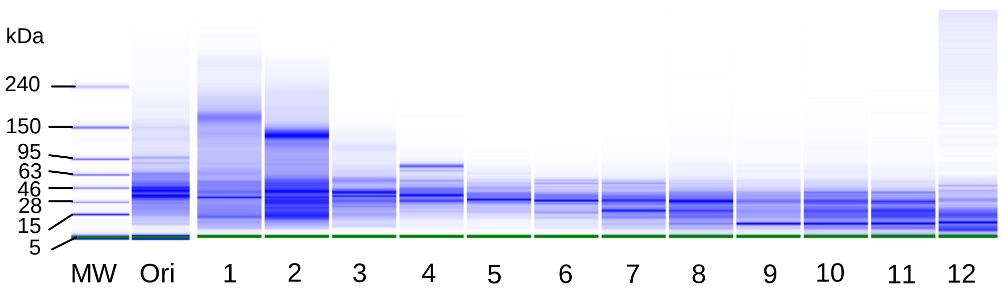

**Figure S4: Comparison of the OFFGEL-separated protein fractions with the non fractioned protein sample, by migration on high-sensitivity protein chips.**

MW: molecular weight marker – Ori: non fractioned protein sample – 1-12: the 12 OFFGEL fractions with pI ranging from 3 to 10 (pseudo-2D gel). Green pseudo-bands: internal molecular weight marker.

**Table S3: Comparison of the 3 protein analyses strategies in terms of analysed protein amounts and total number of obtained spectra.**

| | **Strategy 1** | **Strategy 2** | | **Strategy 3** |
|---|---|---|---|---|
| | **MW separation, TOTAL** | **OFFGEL separation, TOTAL** | **OFFGEL separation, each of 12 fractions** | **Unfractioned** |
| Initial sample concentration | 1 µg/µL | 1 µg/µL | | 1 µg/µL |
| Concentration of the loaded sample | 1 µg/µL | | 0.2 µg/µL (after focusing) | 1 µg/µL |
| Sample volume loaded on the gel | 10 µL | | 10 µL | 10 µL |
| Protein amount loaded on the gel | 10 µg | 24 µg (12×2 µg) | ~2 µg | 10 µg |
| Number of spectra (average of the technical replicates when relevant) | 75,415 | 196,744 | | 17,139 |
| Number of spectra / µg loaded protein | 7,542 | 8,197 | | 1,714 |
| Number of assigned spectra / µg loaded protein | 603 | 443 | | 98 |

## Table S4: Obtained and assigned spectrum numbers according to the separation strategies and to the technical replicates.

Strategy 1: separation according to the molecular weight by SDS-PAGE – Strategy 2: separation according to the pI by OFFGEL fractionation – Strategy 3: non fractioned protein sample.

| | **Fractions** | **Spectra obtained** | **Spectra assigned** | **Percentage of spectra assigned** |
|---|---|---|---|---|
| **Strategy 3**<br>**No separation**<br>**3x(1 fraction)** | Ori_1 | 16215 | 1244 | 7.67% |
| | Ori_2 | 18751 | 981 | 5.23% |
| | Ori_3 | 16451 | 706 | 4.29% |
| | **Total of the replicates 1,2,3** | **51417** | **2931** | **5.70%** |
| **Strategy 2**<br>**pI (OFFGEL)**<br>**3x(12 fractions)** | OFFGEL_1 | 162006 | 11211 | 6.92% |
| | OFFGEL_2 | 208778 | 8534 | 4.09% |
| | OFFGEL_3 | 219449 | 12115 | 5.52% |
| | **Total of the OFFGEL replicates 1,2,3** | **590233** | **31860** | **5.40%** |
| **Strategy 1**<br>**MW (SDS-PAGE)**<br>**1x(26 fractions)** | **SDS-PAGE** | **75415** | **6027** | **7.99%** |
| | ***Total (65 fractions)*** | ***717065*** | ***40818*** | ***5.69%*** |

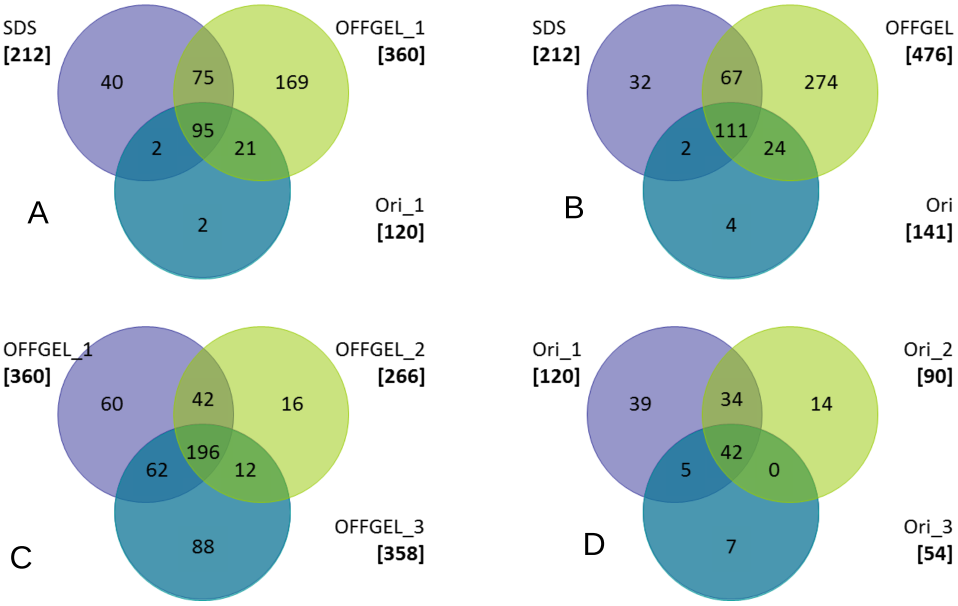

**Figure S5: Venn diagrams of the numbers of non-redundant protein groups identified according to the separation strategies and to the technical replicates.**

A: comparison of the 3 strategies based on 1 technical replicate for each strategy.

B: comparison of the 3 strategies based on all the available replicates.

C: comparison of the technical replicates from strategy 2, separation according to the pI by OFFGEL fractionation – D: comparison of the technical replicates from strategy 3, no separation.

**Figure S6: Presence/absence of the non-redundant protein groups identified in the different protein fractions obtained from strategies 1 and 2.**

A: Strategy 1, separation according to the molecular weight by SDS-PAGE, 26 fractions – B: Strategy 2, separation according to the pI by OFFGEL fractionation, 12 fractions; the data are shown for one technical replicate only (OFFGEL_1). Each line corresponds to a fraction. Each column corresponds to a protein group. The identification of the protein group in each fraction is materialized by green color.

**Table S5. Comparison of the functions related to polysaccharide hydrolysis across different biological systems and studies.**

| Cazy family | Known activities[a] | Present study[k] | Termite[b,d,i] | Human[h,i] | C.T. Avicel-grown[c,e,g,j] | C.T. cellobiose-grown[c,f,g,j] |
|---|---|---|---|---|---|---|
| GH1 | β-glucosidase; β-galactosidase; β-mannosidase; β-glucuronidase; β-D-fucosidase; phlorizin hydrolase; exo-β-1,4-glucanase; 6-phospho-β-galactosidase; 6-phospho-β-glucosidase; strictosidine β-glucosidase; lactase; amygdalin β-glucosidase; prunasin β-glucosidase; raucaffricine β-glucosidase; thioglucosidase; β-primeverosidase; isoflavonoid 7-O-β-apiosyl-β-glucosidase; hydroxyisourate hydrolase; β-glycosidase |  | 1 | (gene) |  |  |
| GH3 | β-glucosidase; xylan 1,4-β-xylosidase; β-N-acetylhexosaminidase; glucan 1,3-β-glucosidase; glucan 1,4-β-glucosidase; exo-1,3-1,4-glucanase; α-L-arabinofuranosidase. | 7 | 1 | (gene) |  |  |
| GH4 | maltose-6-phosphate glucosidase; α-glucosidase; α-galactosidase; 6-phospho-β-glucosidase; α-glucuronidase. Some enzymes prefer phosphorylated substrates over non-phosphorylated substrates. GH4 enzymes were the first glycosidases shown to demonstrate an absolute requirement for NAD+ and a divalent metal ion and in some instances reducing environments for catalytic activity. | 2 | 2 | (gene) |  |  |
| GH5 | Formerly known as cellulase family A. chitosanase; β-mannosidase; Cellulase; glucan 1,3-β-glucosidase; licheninase; glucan endo-1,6-β-glucosidase; mannan endo-β-1,4-mannosidase; endo-β-1,4-xylanase; cellulose β-1,4-cellobiosidase; endo-β-1,6-galactanase; β-1,3-mannanase; xyloglucan-specific endo-β-1,4-glucanase; mannan transglycosylase | 1 | 2 | (gene) | A3DDM7→<br>P10477↓<br>A3DHG6↓<br>Q05332 ↓<br>P04956 ↓<br>Q06853 | A3DDM7→<br>P10477↓<br>A3DHG6↓<br>Q05332 ↓<br>P04956 ↓<br>Q06853<br>***A3DHB4*** |
| GH8 | chitosanase; cellulase; licheninase; endo-1,4-β-xylanase; reducing-end-xylose releasing exo-oligoxylanase | 1 | (gene) | (gene) | A3DC29↓* | A3DC29↓* |
| GH9 | Endoglucanase; cellobiohydrolase;β-glucosidase | 11 | (gene) |  | A3DCH1 ↑<br>Q06852↑*<br>A3DD30↑*<br>A3DCH2 →<br>Q8VV73 →<br>Q70DK3 →<br>Q1H8Q0 ↓<br>P26224 →<br>A3DJ31→<br>A3DCJ2→<br>P0C2S4<br>***Q9L3J5***<br>***Q9AJF8***<br>***Q1JR83*** | A3DCH1 ↑<br>Q06852↑*<br>A3DD30↑*<br>A3DCH2 →<br>Q8VV73→<br>Q70DK3 →<br>Q1H8Q0 ↓<br>P26224 →<br>A3DJ31→<br>A3DCJ2→<br>P0C2S4 |
| GH10 | endo-1,4-β-xylanase; endo-1,3-β-xylanase | 2 | 1 | (gene) | O32374 ↓<br>P10478 ↓ | O32374↓<br>P10478 ↓<br>***Q70DK4***<br>***P51584*** |
| GH11 | xylanase | 2 | (gene) |  | A3DJP0↓* | A3DJP0↓* |
| GH13 | α-amylase; pullulanase; cyclomaltodextrin glucanotransferase; cyclomaltodextrinase; trehalose-6-phosphate hydrolase; oligo-α-glucosidase; maltogenic amylase; neopullulanase; α-glucosidase; maltotetraose-forming α-amylase; isoamylase; glucodextranase; maltohexaose-forming α-amylase; maltotriose-forming α-amylase; branching enzyme; trehalose synthase; 4-α-glucanotransferase; maltopentaose-forming α-amylase; amylosucrase; sucrose phosphorylase; malto-oligosyltrehalose trehalohydrolase; isomaltulose synthase; amino acid transporter. | 26 | 2 | 8 |  |  |
| GH16 | Lichenase |  | (gene) | (gene) | ***A3DBX3*** |  |
| GH18 | Chitinase |  | (gene) | (gene) | Q59326 → | Q59326→ |
| GH26 | Mannanase |  | (gene) | (gene) | A3DBE4↓<br>***Q9REC7*** | A3DBE4↓ |
| GH42 | β-galactosidase |  | 1 | (gene) |  |  |
| GH43 | β-xylosidase; β-1,3-xylosidase; β-L-arabinofuranosidase; arabinanase; xylanase; galactan 1,3-β-galactosidase |  | (gene) | (gene) |  | ***A3DEX4***<br>*A3DBC8*<br>*A3DHB4* |

| Cazy family | Known activities[a] | Present study[k] | Termite[b,d,i] | Human[h,i] | C.T. Avicel-grown[c,e,g,j] | C.T. cellobiose-grown[c,f,g,j] |
|---|---|---|---|---|---|---|
| GH44 | Endoglucanase; xyloglucanase | 11 | (gene) | | P71140 ↑* | P71140 ↑* |
| GH48 | endoglucanase; chitinase; cellobiohydrolase; endo-processive cellulases; [reducing end] cellobiohydrolase | 7 | | | A3DH67↑* | A3DH67↑* |
| GH53 | endo-β-1,4-galactanase | | (gene) | (gene) | A3DFA2→ | A3DFA2→ |
| GH74 | endoglucanase; oligoxyloglucan reducing end-specific cellobiohydrolase; xyloglucanase | | (gene) | | Q70DK5 ↓ | Q70DK5 ↓ |
| GH88 | d-4,5 unsaturated β-glucuronyl hydrolase | | 1 | (gene) | | |
| GH94 | cellobiose phosphorylase; cellodextrin phosphorylase; chitobiose phosphorylase; cyclic β-1,2-glucan synthase | 2 | 8 | 1 | | |
| CBM3 | cellulose-binding; sometimes chitin-binding. Previously known as cellulose-binding domain family III (CBD III) | 20 | | | Q06851 ↑*<br>Q1H8Q0<br>A3DJ31<br>A3DCJ2 | Q06851 ↑*<br>Q1H8Q0<br>A3DJ31<br>A3DCJ2 |
| CBM4 | Binding to xylan, β-1,3-glucan, β-1,3-1,4-glucan, β-1,6-glucan and amorphous cellulose but not with crystalline cellulose. Previously known as cellulose-binding domain family IV (CBD IV) | | 1 | (gene) | | |
| CBM6 | Binding to cellulose, amorphous cellulose, β-1,4-xylan, β-1,3-glucan, β-1,3-1,4-glucan, and β-1,4-glucan. | 2 | (gene) | (gene) | A3DJP0↓*<br>A3DHG6↓ | A3DJP0↓*<br>A3DHG6↓<br>***A3DEX4*** |
| CBM9 | found so far only in xylanases. The cellulose-binding function has been demonstrated in one case | 2 | | | | |
| CBM13 | Binding to mannose, xylan, GalNAc | | | (gene) | A3DHG6↓ | A3DHG6↓ |
| CBM20 | starch-binding function | | | 1 | | |
| CBM30 | Binding to cellulose. Formerly known as X7 modules | 11 | (gene) | | A3DD30↑* | A3DD30↑* |
| CBM32 | Binding to galactose and lactose, polygalacturonic acid, LacNAc. Formerly known as X56 modules. | 1 | (gene) | (gene) | A3DDM7→* | A3DDM7→* |
| CBM35 | Binds to xylan and the interaction is calcium dependent, binding to decorated soluble mannans and mannooligosaccharides and β-galactan. | | | | A3DBE4↓ | A3DBE4↓ |
| CBM42 | Binding to arabinofuranose | | | | | *A3DBC8*<br>*A3DHB4* |
| CBM44 | Binding to cellulose and xyloglucan | 11 | | | A3DD30↑* | A3DD30↑* |
| CBM48 | glycogen-binding function, appended to GH13 modules | | | 1 | | |
| GT35 | glycogen or starch phosphorylase | | | 5 | | |
| | CBM_X containing proteins | | 8 | 1 | | |
| | Cell surface anchor | | | | **A3DDE2** | |
| | Cellulosome enzyme, dockerin type I | | | | **A3DD46** | |
| | Cellulosome anchoring protein, cohesin region | | | | | ***P71143*** |

C.T.: *Clostridium thermocellum* (C.T.); termite: termite hindgut; human: human gut microbiota.

[a] Reference: (Cantarel *et al.*, 2009)

[b] Reference: (Warnecke *et al.*, 2007)

[c] Reference: (Gold and Martin, 2007)

[d] Number of expressed modules, as demonstrated by a metaproteomic analysis of the clarified gut fluid.

[e] Proteins in bold font and underlined are those found only in the Avicel sample.

[f] Proteins in bold italic font and underlined are those found only in the cellobiose sample.

[g] Proteins marked with * are those also detected in the present study.

[h] Reference:(Gill *et al.*, 2006; Verberkmoes *et al.*, 2009)

[i] The families marked as "(gene)" exist based on genomic analyses but were not detected during proteomic analyses

[j] Quantitative change on Avicel compared to cellobiose; ↑: increase; ↓: decrease; →: same level.

[k] Number of unique peptides.

## Table S6. Identified redundant proteins related to monosaccharide catabolism and gluconeogenesis.

| Number in Figure S7 | Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|
| (1) | Glucokinase, ROK family | 2.7.1.2 | 1 | B9MPB7_ANATD<br>E4Q7G4_CALH1<br>E4S583_CALKI<br>E4SGQ0_CALK2<br>G2PX09_9FIRM | 21 | 3 |
| (2) | Phosphoglucose isomerase | 5.3.1.9 | 2 | D9TIR0_CALOO<br>E4Q4B6_CALOW<br>E4QEC8_CALH1<br>E4S9X7_CALKI<br>E4SGH6_CALK2<br>G2PU14_9FIRM<br>PGMI_ANATD | 11 | 9 |
| | | | | A3DBX9_CLOTH<br>C7HJC1_CLOTM<br>D1NL73_CLOTM<br>E6UMA2_CLOTL | 274 | 5 |
| (3) | Phosphofructokinase | 2.7.1.11 | 6 | D9TFE9_CALOO<br>E4Q7L2_CALH1<br>E4S4D5_CALKI<br>E4SDF3_CALK2<br>G2PZ48_9FIRM | 22 | 8 |
| | | | | D9TKQ2_CALOO<br>E4Q160_CALOW | 88 | 6 |
| | | | | A3DCA6_CLOTH<br>C7HIP5_CLOTM<br>D1NLL1_CLOTM<br>E6ULI0_CLOTL | 22 | 4 |
| | | | | B5Y8B9_COPPD | 1 | 4 |
| | | | | B5Y8V1_COPPD | 1 | 3 |
| | | | | F1TI51_9CLOT | 516 | 2 |
| (4) | Fructose-1,6-bisphosphate aldolase | 4.1.2.13 | 5 | B9MPI4_ANATD<br>D9TIR2_CALOO<br>E4Q4B8_CALOW<br>E4QEC6_CALH1<br>E4S9X9_CALKI<br>E4SGH4_CALK2<br>G2PU12_9FIRM | 91 | 8 |
| | | | | B5Y768_COPPD | 46 | 6 |
| | | | | A3DCA8_CLOTH<br>C7HIP6_CLOTM<br>D1NLL3_CLOTM | 131 | 4 |

| Number in Figure S7 | Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|
| | | | | E6ULH9_CLOTL | | |
| | | | | A1HR25_9FIRM | 46 | 4 |
| | | | | B9MNC8_ANATD<br>D9TH04_CALOO<br>E4Q6M6_CALOW<br>E4QA14_CALH1<br>E4S4W0_CALKI<br>E4SCS5_CALK2<br>G2PV45_9FIRM | 8 | 2 |
| (5) | Triosephosphate isomerase | 5.3.1.1 | 1 | A3DBQ1_CLOTH<br>C7HFX7_CLOTM<br>D1NKZ7_CLOTM<br>E6UMQ3_CLOTL | 932 | 4 |
| (6) | Glyceraldehyde-3-phosphate dehydrogenase | 1.2.1.12 | 10 | A3DBP9_CLOTH<br>C7HFX5_CLOTM<br>D1NKZ6_CLOTM<br>E6UMQ5_CLOTL | 2346 | 11 |
| | | | | B9MS50_ANATD<br>E4Q1K9_CALOW<br>E4QA76_CALH1<br>E4S8C1_CALKI<br>E4SBP1_CALK2<br>G2PTW1_9FIRM | 3800 | 11 |
| | | | | B5Y8H9_COPPD | 3800 | 11 |
| | | | | E7EUT4_HUMAN<br>F6UI74_CALJA<br>G1QVR0_NOMLE<br>G3R288_GORGO<br>P04406_HUMAN<br>Q2TSD0_HUMAN<br>Q5RAB4_PONAB | 408 | 3 |
| | | | | A4J8U1_DESRM | 718 | 2 |
| | | | | F3ZYD0_MAHA5 | 3800 | 2 |
| | | | | B8BQU2_THAPS | 2346 | 2 |
| | | | | A4XKV3_CALS8 | 3800 | 2 |
| | | | | E1KG11_9FIRM | 1139 | 2 |
| | | | | F1T8C8_9CLOT | 2346 | 2 |
| (7) | Phosphoglycerate kinase | 2.7.2.3 | 3 | B9MS49_ANATD<br>D9TKB5_CALOO<br>E4Q1K8_CALOW<br>E4QA77_CALH1<br>E4S8C0_CALKI | 932 | 8 |
| | | | | A3DBQ0_CLOTH<br>C7HFX6_CLOTM<br>E6UMQ4_CLOTL | 654 | 4 |

| Number in Figure S7 | Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|
| | | | | B5Y8I0 _COPPD | 985 | 2 |
| (6/7) | Glyceraldehyde-3-phosphate ferredoxin oxidoreductase | 1.2.7.5 | 2 | B5Y6W0_COPPD | 64 | 8 |
| | | | | B9MQI2_ANATD<br>D9TJ88_CALOO<br>E4Q577_CALOW<br>E4QDF8_CALH1<br>E4S4G6_CALKI<br>E4SG03_CALK2<br>G2PXR6_9FIRM | 16 | 2 |
| (8) | Phosphoglycerate mutase | 5.4.2.1 | 0 | Non detected | | |
| (9) | Enolase | 4.2.1.11 | 4 | B9MS47_ANATD<br>E4S8B8_CALKI<br>E4SBP4_CALK2<br>G2PTV8_9FIRM | 3002 | 15 |
| | | | | A3DBQ5_CLOTH<br>C7HFY1_CLOTM<br>D1NL00_CLOTM<br>E6UMQ0_CLOTL | 461 | 8 |
| | | | | B5Y9E0_COPPD | 461 | 3 |
| | | | | B8I4U1_CLOCE<br>F1T8F4_9CLOT | 18 | 2 |
| (10) | Pyruvate kinase | 2.7.1.40 | 3 | B5Y8W1_COPPD | 1 | 4 |
| | | | | A4XKI1_CALS8 | 26 | 3 |
| | | | | A4XKU4_CALS8<br>E4QA86_CALH1<br>E4S8B2_CALKI | 10 | 2 |
| (11) | Glucose dehydrogenase/gluconolactonase | 1.1.1.47; 1.1.3.4; 1.1.3.5; 1.1.99.10; 1.1.5.2; 3.1.1.17 | 0 | Non detected | | |
| (12) | Gluconate dehydratase | 4.2.1.39 | 0 | Non detected | | |
| (13) | 2-keto-3-deoxygluconate aldolase | 4.1.2.- | 0 | Non detected | | |
| (14) | Glyceraldehyde dehydrogenase (Archaeal) | 1.2.7.5 | 0 | Non detected | | |
| (15) | Glycerate kinase | 2.7.1.165 | 0 | Non detected | | |
| (16) | 2-keto-3-deoxygluconate kinase | 2.7.1.45 | 0 | Non detected | | |
| (17) | 2-keto-3-deoxy-6-phosphogluconate aldolase | 4.1.2.14 | 0 | Non detected | | |
| (18) | Glucose-6-phosphate dehydrogenase/6-phosphogluconolactonase | 1.1.1.49; 3.1.1.31 | 0 | Non detected | | |
| (19) | Gluconate kinase | 2.7.1.12 | 0 | Non detected | | |
| (20) | 6-phosphogluconate dehydratase | 4.2.1.12 | 0 | Non detected | | |
| (21) | 6-phosphogluconate dehydrogenase (decarboxylase) | 1.1.1.44 | 0 | Non detected | | |
| (22) | Ribulose-5-phosphate 3-epimerase | 5.1.3.1 | 0 | Non detected | | |
| (23) | Transketolase N-terminal | 2.2.1.1 | 2 | E4Q4K1_CALOW<br>E4SEK4_CALK2 | 111 | 5 |
| | C-terminal | | | D9TFY6_CALOO | 202 | 4 |
| (24) | Phosphoenolpyruvate synthase | 2.7.9.2 | 0 | Non detected | | |

| Number in Figure S7 | Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|
| (25) | Fructose-1,6-bisphosphatase | 3.1.3.11 | 0 | Non detected | | |
| (26) | Ribose-5-phosphate isomerase | 5.3.1.6 | 1 | B5Y8G5_COPPD | 495 | 2 |
| (27) | Transaldolase | 2.2.1.2 | 3 | A4XL36_CALS8 | 549 | 7 |
| | | | | D1PQS7_9FIRM | 549 | 3 |
| | | | | B5Y9V1_COPPD | 549 | 2 |
| (28) | Xylose isomerise | 5.3.1.5 | 2 | A4XGJ4_CALS8 | 19 | 2 |
| | | | | D9TH03_CALOO | 11 | 2 |
| (29) | Xylulokinase | 2.7.1.17 | 1 | B9MPD2_ANATD<br>E4SGN3_CALK2 | 41 | 5 |
| (30) | Phosphoglucomutase | 5.4.2.2 | 1 | A4XLT8_CALS8<br>B9MKI8_ANATD<br>D9TJF9_CALOO<br>E4Q392_CALOW<br>E4Q808_CALH1<br>E4SAT6_CALKI<br>E4SDT1_CALK2<br>G2PWN0_9FIRM | 8 | 3 |
| (31) | Deoxyribose-phosphate aldolase | 4.1.2.4 | 1 | DEOC_COPPD | 1139 | 2 |
| (32) | Pyruvate phosphate dikinase | 2.7.9.1 | 6 | D9TFE9_CALOO<br>E4Q7L2_CALH1<br>\|E4S4D5_CALKI<br>E4SDF3_CALK2<br>G2PZ48_9FIRM | 22 | 8 |
| | | | | D9TKQ2_CALOO<br>E4Q160_CALOW | 88 | 6 |
| | | | | B5Y8B9_COPPD | 1 | 4 |
| | | | | A3DCA6_CLOTH<br>C7HIP5_CLOTM<br>D1NLL1_CLOTM<br>E6ULI0_CLOTL | 22 | 4 |
| | | | | B5Y8V1_COPPD | 1 | 3 |
| | | | | F1TI51_9CLOT | 516 | 2 |
| Nucleoside diphosphate (UDP) sugar related | N-acetylglucosamine-6-phosphate deacetylase | 3.5.1.25 | 1 | B5Y679_COPPD | 1 | 5 |
| | Glucosamine--fructose-6-phosphate aminotransferase (isomerizing) | 2.6.1.16 | 3 | D9TGM7_CALOO | 196 | 5 |
| | | | | A3DEL5_CLOTH<br>C7HIR5_CLOTM<br>D1NMS0_CLOTM<br>E6UNI4_CLOTL | 196 | 3 |
| | | | | B5Y682_COPPD | 1 | 2 |
| | Phosphoglucosamine mutase | 5.4.2.10 | 1 | A4XH45_CALS8<br>B9MMU5_ANATD<br>D9TGM8_CALOO<br>E4Q679_CALOW<br>E4QAN5_CALH1<br>E4S7B2_CALKI | 46 | 2 |

| Number in Figure S7 | Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|
| | | | | E4SD78_CALK2<br>G2PVP6_9FIRM | | |
| | Phosphate acetyltransferase | 2.3.1.8 | 1 | B9MJT4_ANATD<br>D9TK27_CALOO<br>E4Q213_CALOW<br>E4Q9J1_CALH1<br>E4S9B2_CALKI<br>E4SBF4_CALK2<br>G2PUC7_9FIRM | 17 | 3 |
| | UDP-N-acetylglucosamine 1-carboxyvinyltransferase | 2.5.1.7 | 1 | A3DHV1_CLOTH<br>C7HGF5_CLOTM<br>D1NP71_CLOTM<br>E6UU19_CLOTL | 496 | 2 |
| Adenosine diphosphosugar (ADP) related | Glucose-1-phosphate adenylyltransferase | 2.7.7.27 | 1 | B5Y9J8_COPPD | 410 | 2 |
| Others | L-fucose isomerase related protein | | 1 | A4XHV8_CALS8 | 83 | 4 |
| | Sugar-phosphate isomerase, RpiB/LacA/LacB family | 5.3.1.26 | 1 | A4XIR8_CALS8<br>B9MPJ5_ANATD<br>D9TIS4_CALOO<br>E4Q4C9_CALOW<br>E4QEB5_CALH1<br>E4S9Z4_CALKI<br>E4SGG3_CALK2<br>G2PU00_9FIRM | 495 | 2 |
| | Phosphoglucomutase | 5.4.2.2 | 1 | A4XLT8_CALS8<br>B9MKI8_ANATD<br>D9TJF9_CALOO<br>E4Q392_CALOW<br>E4Q808_CALH1<br>E4SAT6_CALKI<br>E4SDT1_CALK2<br>G2PWN0_9FIRM | 8 | 3 |

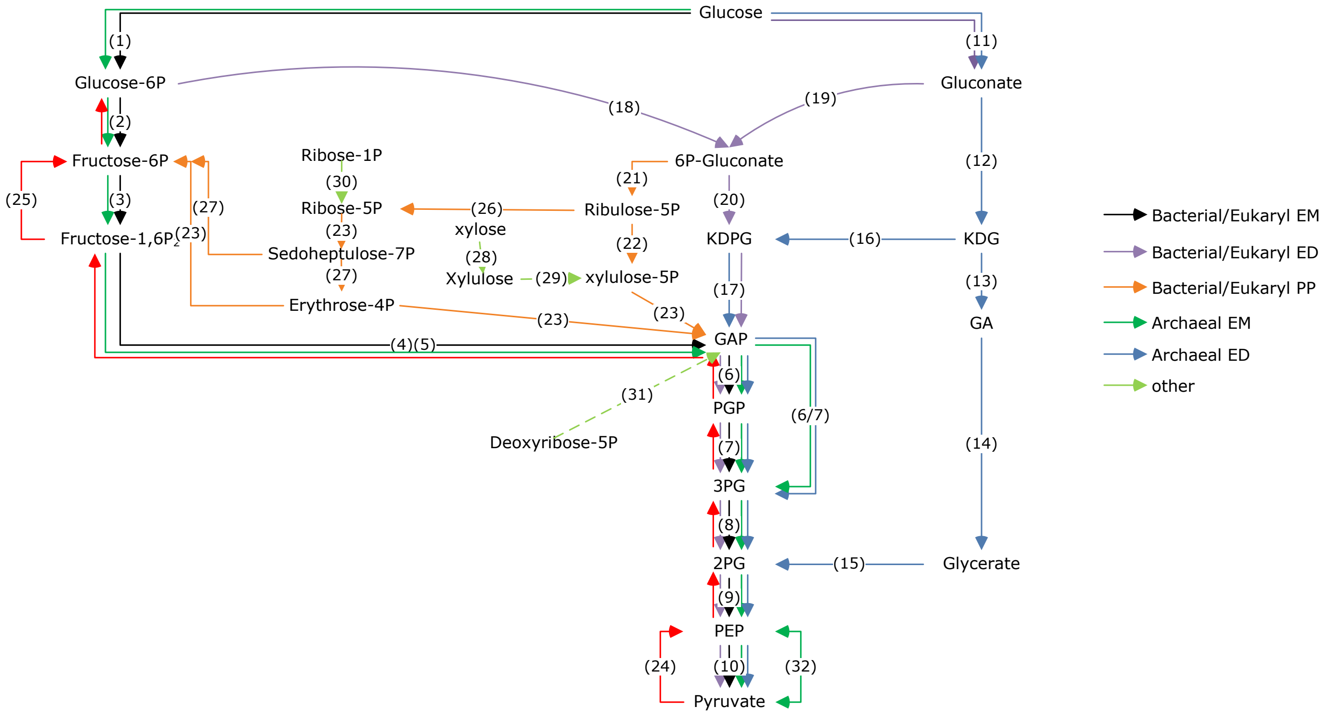

**Figure S7. Glucose catabolism and gluoconeogenesis: pathways and enzymes.**

— Bacterial/eukaryal Embden-Meyerhof (EM) pathway; — Bacterial/eukaryal Entner-Doudorofff (ED) pathway; — Bacterial/eukaryal pentose phosphate pathway (PPP); — Archaeal EM pathway; — Archaeal ED pathway; — other related pathways . **Abbreviations**: GA, glyceraldehyde; GAP, glyceraldehyde 3-phosphate; KDG, 2-keto-3-deoxygluconate; KDPG, 2-keto-3-deoxy-6-phosphogluconate; PEP, Phosphoenolpyruvate; PGP, 2,3-bisphosphoglycerate; 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate. **Enzymes**: (1) Glucokinase ; (2) Phosphoglucose isomerase ; (3) Phosphofructokinase; (4) Fructose-1,6-bisphosphate aldolase; (5) Triosephosphate isomerase; (6) Glyceraldehyde-3-phosphate dehydrogenase; (7) Phosphoglycerate kinase; (6/7) Glyceraldehyde-3-phosphate ferredoxin oxidoreductase and NAD+-dependent GAPDH; (8) Phosphoglycerate mutase ; (9) Enolase; (10) Pyruvate kinase ; (11) Glucose dehydrogenase/gluconolactonase; (12) Gluconate dehydratase; (13) 2-keto-3-deoxygluconate aldolase; (14) Glyceraldehyde dehydrogenase; (15) Glycerate kinase; (16) 2-keto-3-deoxygluconate kinase; (17) 2-keto-3-deoxy-6-phosphogluconate aldolase; (18) Glucose-6-phosphate dehydrogenase/6-phosphogluconolactonase; (19) Gluconate kinase; (20) 6-phosphogluconate dehydratase; (21) 6-phosphogluconate dehydrogenase (decarboxylase); (22) Ribulose-5-phosphate 3-epimerase; (23) Transketolase; (24) Phosphoenolpyruvate synthase ; (25) Fructose-1,6-bisphosphatase; (26) Ribose-5-phosphate isomerase; (27) Transaldolase; (28) Xylose isomerise; (29) Xylulokinase; (30) Phosphoglucomutase; (31) Deoxyribose-phosphate aldolase; (32) Pyruvate phosphate dikinase. Modify according to (Verhees *et al.*, 2003) and KEGG.

# Table S7. Identified redundant proteins related to pyruvate metabolism.

| Number in Figure S8 | Name | | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|---|
| (1) | Pyruvate:flavodoxin/flavodoxin oxidoreductase | Alpha subunit | 1.2.7.1 | 3 | E4SAA7_CALKI<br>G2PTM8_9FIRM | 26 | 8 |
| | | | | | B5Y9Q8_COPPD | 10 | 5 |
| | | | | | A3DI15_CLOTH<br>C7HDF6_CLOTM<br>D1NNZ5_CLOTM<br>E6URH3_CLOTL | 28 | 3 |
| | | Beta subunit | | 2 | B9MQN3_ANATD<br>D9TJE0_CALOO<br>E4Q5J4_CALOW<br>E4QD30_CALH1<br>E4SAA8_CALKI<br>E4SFV1_CALK2<br>G2PTM7_9FIRM | 45 | 5 |
| | | | | | B5Y9Q7_COPPD | 26 | 5 |
| | | Gamma subunit | | 4 | B9MQN0_ANATD<br>D9TJD7_CALOO<br>E4Q5J1_CALOW<br>E4SAA5_CALKI<br>E4SFV4_CALK2<br>G2PTN0_9FIRM | 47 | 7 |
| | | | | | B9MKE5_ANATD<br>E4SAI7_CALKI<br>E4SB95_CALK2<br>G2PWI7_9FIRM | 8 | 5 |
| | | | | | B5Y9R0_COPPD | 21 | 5 |
| | | | | | A3DI13_CLOTH<br>C7HDF8_CLOTM<br>D1NNZ7_CLOTM<br>E6URH1_CLOTL | 47 | 3 |
| | | nifJ family | | 2 | A0PZX5_CLONN<br>B1BCS7_CLOBO | 135 | 2 |
| | | | | | C4V109_9FIRM<br>E0NX90_9FIRM | 262 | 2 |
| (2) | pyruvate dehydrogenase E1 component | | 1.2.4.1 | 1 | B5Y8T9_COPPD | 36 | 8 |
| (3) | dihydrolipoamideacetyltransferase component of pyruvate dehydrogenase complex | | 2.3.1.12 | 1 | B5Y8U4_COPPD | 1 | 2 |
| (4) | Oxaloacetate decarboxylase | | 4.1.1.3 | 0 | Non detected | | |
| (5) | Malate dehydrogenase (Oxaloacetate-decarboxylating) | | 1.1.1.38 | 1 | A3DCA3_CLOTH<br>C7HIP2_CLOTM | 107 | 3 |

| Number in Figure S8 | Name | | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|---|
| | | | | | D1NLK8_CLOTM<br>E6ULI3_CLOTL | | |
| (6) | Malate dehydrogenase | | 1.1.1.37 | 2 | A5D2J2_PELTS | 3 | 4 |
| | | | | | A3DCA4_CLOTH<br>C7HIP3_CLOTM<br>D1NLK9_CLOTM,<br>E6ULI2_CLOTL | 9 | 3 |
| (7) | Citrate (Si)-synthase | | 2.3.3.1 | 1 | A3DJU4_CLOTH<br>C7HEE1_CLOTM<br>D1NR78_CLOTM<br>E6UTN8_CLOTL | 14 | 2 |
| (8) | Aconitase | | 4.2.1.3 | 1 | A4XHI4_CALS8<br>B9MNX1_ANATD<br>E4Q3K5_CALOW<br>E4Q7J4_CALH1<br>E4SAV2_CALKI<br>E4SGS7_CALK2<br>G2PWP6_9FIRM | 56 | 5 |
| (9) | Isocitrate dehydrogenase | | 1.1.1.41 | 1 | A4XHI5_CALS8 | 168 | 2 |
| (10) | 2-oxoacid:ferredoxin oxidoreductase | alpha | 1.2.7.3 | 1 | B5Y690_COPPD | 1 | 2 |
| | | beta | | 1 | B5Y691_COPPD | 3 | 3 |
| (11) | Succinyl-CoA synthase | alpha | 6.2.1.5 | 1 | Q2V0P6_9FIRM | 58 | 9 |
| | | beta | | 1 | A5D2I3_PELTS<br>Q2V0P7_9FIRM | 10 | 5 |
| (12) | Succinic dehydrogenase | | 1.3.99.1 | 0 | Non detected | | |
| (13) | Fumarase | | 4.2.1.2 | 1 | A5D2I1_PELTS | 228 | 3 |

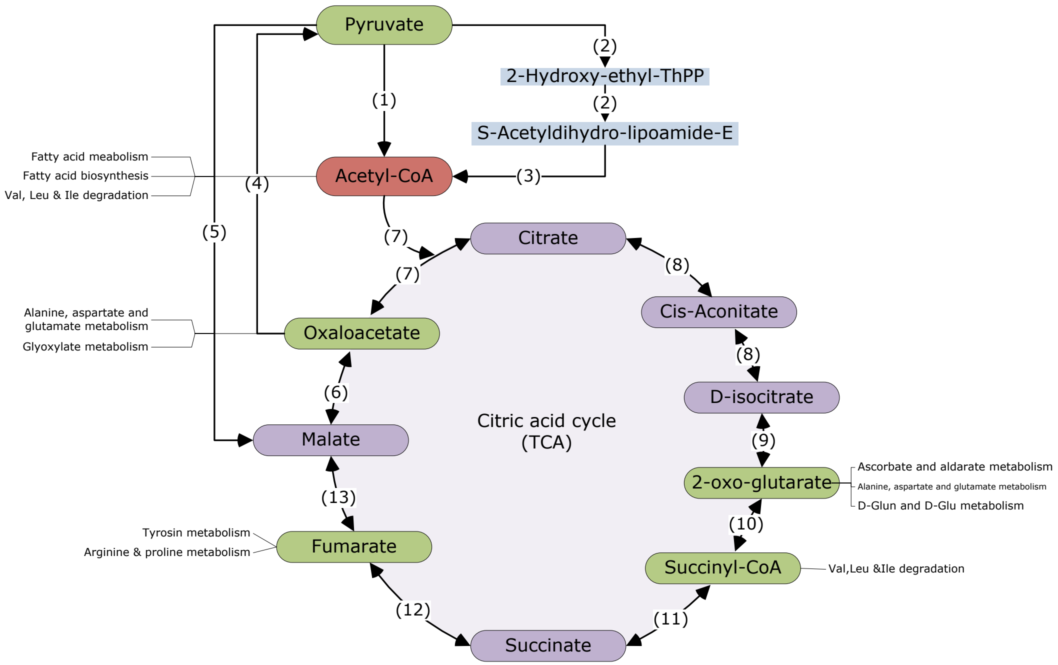

Figure S8. Pyruvate metabolism: pathways and enzymes

**Enzymes**: (1) Pyruvate:flavodoxin/flavodoxin oxidoreductase; (2) pyruvate dehydrogenase E1 component; (3) dihydrolipoamideacetyltransferase component of pyruvate dehydrogenase complex; (4) Oxaloacetate decarboxylase; (5) Malate dehydrogenase (Oxaloacetate-decarboxylating); (6) Malate dehydrogenase; (7) Citrate synthase; (8) Aconitase; (9) Isocitrate dehydrogenase; (10) 2-oxoacid:ferredoxin oxidoreductase; (11) Succinyl-CoA synthase; (12) Succinic dehydrogenase; (13) Fumarase.

# Table S8. Identified redundant proteins related to fermentation and acetogenesis.

| Number in Figure S9 | Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|
| (1) | Phosphate butyryltransferase | 2.3.1.19 | 1 | B5YA15_COPPD | 164 | 6 |
| (2) | Acyl-CoA dehydrogenase | 1.3.8.1 | 2 | A5D1D8_PELTS | 885 | 2 |
| | | | | B9MRM8_ANATD<br>D9TKT4_CALOO<br>E4Q127_CALOW<br>E4S757_CALKI<br>E4SC89_CALK2<br>G2PT10_9FIRM | 194 | 2 |
| (3) | Vinylacetyl-CoA delta-isomerase | 5.3.3.3 | 1 | A1HPH9_9FIRM | 152 | 2 |
| (4) | Enoyl-CoA hydratase/isomerase | 4.2.1.17 | 1 | A1HPI9_9FIRM | 54 | 2 |
| (5) | 4-aminobutyrate aminotransferase | 2.6.1.19 | 1 | B5Y9D6_COPPD | 1 | 2 |
| (6) | Acetyl-CoA acetyltransferase | 2.3.1.9 | 2 | D7CJV6_SYNLT | 194 | 2 |
| | | | | D7CLS5_SYNLT | 302 | 3 |
| (7) | NADPH-dependent butanol dehydrogenase | 1.1.1.- | 2 | B5Y9C1_COPPD | 1 | 4 |
| | | | | B5Y9G2_COPPD | 2 | 2 |
| (8) | Propionyl-CoA carboxylase | 2.1.3.1 | 4 | A3DDA5_CLOTH<br>C7HCY2_CLOTM<br>D1NME9_CLOTM<br>E6URX6_CLOTL | 916 | 12 |
| | | | | B5Y8Z4_COPPD | 252 | 6 |
| | | | | A4XKN1_CALS8<br>B9MRW2_ANATD<br>D9TKK1_CALOO<br>E4Q1D6_CALOW<br>E4QAU9_CALH1<br>E4S7R9_CALKI<br>E4SBW8_CALK2<br>G2PTG5_9FIRM | 252 | 2 |
| | | | | B1L7Z1_THESQ<br>G4FDA8_THEMA<br>Q9WZH5_THEMA | 916 | 2 |
| (9) | Acyl-CoA dehydrogenase | 1.3.8.7;<br>1.3.1.- | 0 | Non detected | | |
| (10) | Lactoylglutathione lyase | 5.1.99.1 | 1 | A4XGA4_CALS8<br>B9MP68_ANATD<br>D9TH33_CALOO<br>E4Q6P6_CALOW<br>E4S6A1_CALKI<br>G2PWY8_9FIRM | 17 | 2 |
| (11) | Acetate kinase | 2.7.2.1 | 3 | A4XL40_CALS8<br>B9MJT3_ANATD<br>E4Q9J2_CALH1<br>E4SBF5_CALK2 | 743 | 7 |

| Number in Figure S9 | Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|---|
| | | | | B5Y7W0_COPPD | 743 | 4 |
| | | | | C7HC04_CLOTM<br>D1NIY5_CLOTM<br>E6UP68_CLOTL<br>ACKA_CLOTH | 743 | 3 |
| (12) | Formamidase | 3.5.1.49 | 1 | B5Y871_COPPD | 2 | 6 |
| (13) | L-lactate dehydrogenase | 1.1.1.27 | 3 | D9TFN5_CALOO<br>E4Q426_CALOW<br>E4Q783_CALH1<br>E4SB24_CALK2 | 1142 | 3 |
| | | | | A3DCA4_CLOTH<br>C7HIP3_CLOTM<br>D1NLK9_CLOTM<br>E6ULI2_CLOTL | 9 | 3 |
| (14) | Methylglyoxal synthase | 4.2.3.3 | 1 | A4XKM4_CALS8<br>B9MRV5_ANATD<br>D9TKK8_CALOO<br>E4QAV6_CALH1<br>E4S7R1_CALKI<br>E4SBX5_CALK2<br>G2PTF8_9FIRM | 184 | 4 |
| (15) | Glyoxylate reductase | 1.1.1.79 | 1 | 9TH02_CALOO<br>E4Q6M4_CALOW<br>E4QA16_CALH1<br>E4S4V8_CALKI<br>G2PV47_9FIRM | 14 | 5 |
| (16) | Phosphate acetyltransferase | 2.3.1.8 | 1 | B9MJT4_ANATD<br>D9TK27_CALOO<br>E4Q213_CALOW<br>E4Q9J1_CALH1<br>E4S9B2_CALKI<br>E4SBF4_CALK2<br>G2PUC7_9FIRM | 17 | 3 |
| (17) | Iron-containing alcohol dehydrogenase | 1.1.1.1 | | E4SE14_CALK2 | 19 | 3 |
| | | | | A3DBL3_CLOTH<br>C7HFU0_CLOTM<br>D1NKW1_CLOTM<br>E6UMU0_CLOTL | 90 | 3 |
| (18) | Pyruvate carboxyltransferase | 4.1.3.4 | 1 | A4XHI0_CALS8<br>D9TIF9_CALOO<br>E4Q3K1_CALOW<br>E4Q7U1_CALH1<br>E4SAW9_CALKI<br>G2PX53_9FIRM | 79 | 4 |

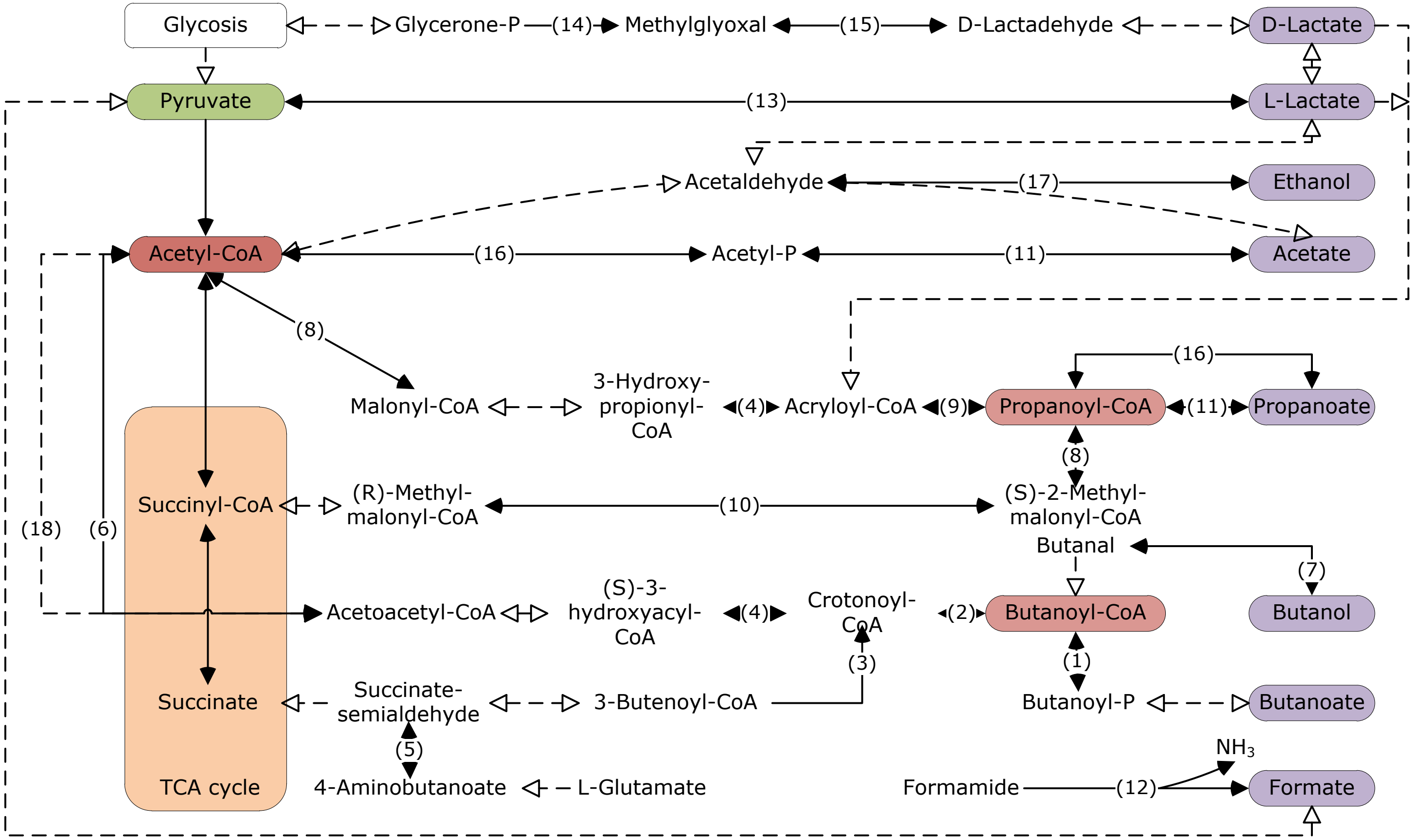

Figure S9. Fermentation and acetogenesis: pathways and enzymes

**Enzymes**: (1) Phosphate butyryltransferase; (2) Acyl-CoA dehydrogenase; (3) Vinylacetyl-CoA delta-isomerase; (4) Enoyl-CoA hydratase/isomerase; (5) 4-aminobutyrate aminotransferase; (6) Acetyl-CoA acetyltransferase; (7) NADPH-dependent butanol dehydrogenase; (8) propionyl-CoA carboxylase; (9) acyl-CoA dehydrogenase; (10) lactoylglutathione lyase; (11) Acetate kinase; (12) Formamidase; (13) L-lactate dehydrogenase; (14) methylglyoxal synthase; (15) glyoxylate reductase; (16) Phosphate acetyltransferase; (17) Iron-containing alcohol dehydrogenase; (18) Pyruvate carboxyltransferase.

## Table S9. Identified redundant proteins related to the methanogenis and the acetyl-CoA pathways.

| Number in Figure 4 | Protein name | Gene name | EC No | UniRef50 ID. | Protein and organism IDs* | Peptide number | Protein coverage (%) |
|---|---|---|---|---|---|---|---|
| (01) | Tungsten-containing formylmethanofuran dehydrogenase 2 subunit C | fwdC | 1.2.99.5 | UniRef50_O27600 | O74031_METWO | 5 | 30 |
| | | | | | O27600_METTH | 3 | 30 |
| (02) | Formylmethanofuran--tetrahydromethanopterin formyltransferase | ftr | 2.3.1.101 | UniRef50_Q57766 | P21348_METTH | 2 | 11 |
| (04) | 5,10-methenyltetrahydromethanopterin hydrogenase | hmd | 1.12.98.2 | UniRef50_Q02394 | P81221_METTW<br>Q50758_METTF | 4 | 17 |
| (04) | F420-dependent methylenetetrahydromethanopterin dehydrogenase | mtd | 1.5.99.9 | UniRef50_Q50501 | Q50501_METTH<br>P55300_METTM | 4 | 24 |
| (04) | F420-dependent methylenetetrahydromethanopterin dehydrogenase | mtd | 1.5.99.9 | UniRef50_A2SPK1 | A3CSZ5_METMJ | 2 | 11 |
| (05) | 5,10-methylenetetrahydromethanopterin reductase | mer | 1.5.99.11 | UniRef50_O27784 | O27784_METTH<br>Q50744_METTM | 7 | 32 |
| (06) | Tetrahydromethanopterin S-methyltransferase subunit H | mtrH | 2.1.1.86 | UniRef50_A6UWI0 | O27224_METTH | 9 | 61 |
| (06) | Tetrahydromethanopterin S-methyltransferase subunit H | mtrH | 2.1.1.86 | UniRef50_A6UWI0 | P80187_METTM | 3 | 45 |
| (06) | Tetrahydromethanopterin S-methyltransferase subunit A | mtrA | 2.1.1.86 | UniRef50_Q2FRN2 | O27227_METTH<br>P80184_METTM | 2 | 18 |
| (07) | Methyl-coenzyme M reductase subunit alpha | mcrA, mrtA | 2.8.4.1 | UniRef50_P07962 | O27232_METTH<br>P11558_METTM | 15 | 40 |
| (07) | Methyl-coenzyme M reductase I subunit beta | mcrB | 2.8.4.1 | UniRef50_Q60390 | P11560_METTM | 11 | 53 |
| (07) | Methyl-coenzyme M reductase I, beta subunit | mcrB | 2.8.4.1 | UniRef50_Q58252 | A3CT50_METMJ | 7 | 32 |
| (07) | Methyl-coenzyme M reductase I, beta subunit | mcrB | 2.8.4.1 | UniRef50_Q58252 | P07955_METBF | 2 | 8.1 |
| (07) | Methyl-coenzyme M reductase, alpha subunit | mcrA, mrtA | 2.8.4.1 | UniRef50_D5EAD6 | A3CT46_METMJ | 2 | 6.5 |
| (09) | Coenzyme F420 hydrogenase subunit alpha | frhA | 1.12.98.1 | UniRef50_P19496 | P19496_METTH | 9 | 28 |
| (09) | Coenzyme F420 hydrogenase subunit beta | frhB, fruB, frcB | 1.12.98.1 | UniRef50_P19499 | P19499_METTH | 4 | 29 |
| (09) | Coenzyme F420 hydrogenase subunit gamma | frhG, fruG | 1.12.98.1 | UniRef50_P80491 | A3CXJ9_METMJ | 2 | 9.9 |
| (10) | CoB-CoM heterodisulfide reductase iron-sulfur subunit A | hdrA | 1.8.98.1 | UniRef50_O27434 | O27434_METTH<br>Q50756_METTM | 4 | 7.6 |
| (10) | CoB--CoM heterodisulfide reductase iron-sulfur subunit C | hdrC | 1.8.98.1 | UniRef50_O27906 | O27906_METTH<br>Q50754_METTM | 2 | 21 |
| (10) | F420-non-reducing hydrogenase iron-sulfur subunit D | mvhD | 1.12.99.- | UniRef50_Q00405 | Q50781_METTH<br>P60238_METTM | 2 | 32 |
| (12) | Acetate kinase | ackA | 2.7.2.1 | UniRef50_Q46BI1 | A4XL40_CALS8<br>B9MJT3_ANATD<br>E4Q9J2_CALH1<br>E4SBF5_CALK2 | 7 | 21 |
| (12) | Acetate kinase | ackA | 2.7.2.1 | UniRef50_Q46BI1 | B5Y7W0_COPPD | 4 | 16 |

| Number in Figure 4 | Protein name | Gene name | EC No | UniRef50 ID. | Protein and organism IDs* | Peptide number | Protein coverage (%) |
|---|---|---|---|---|---|---|---|
| (12) | Acetate kinase | ackA | 2.7.2.1 | UniRef50_Q46BI1 | C7HC04_CLOTM<br>D1NIY5_CLOTM<br>E6UP68_CLOTL<br>O52594_CLOTH | 3 | 13 |
| (13) | phosphate acetyltransferase | pta | 2.3.1.8 | UniRef50_Q1MRP1 | B9MJT4_ANATD<br>D9TK27_CALOO<br>E4Q213_CALOW<br>E4Q9J1_CALH1<br>E4S9B2_CALKI<br>E4SBF4_CALK2<br>G2PUC7_9FIRM | 3 | 16 |
| (14) | Corrinoid/iron-sulfur protein large subunit | acsC | 2.1.1.- | UniRef50_Q07340 | Q07340_MOOTH<br>Q2RJ74_MOOTA | 2 | 3.6 |
| (14) | Carbon monoxide dehydrogenase, catalytic subunit | cooSI | 1.2.99.2 | UniRef50_F8ADS8 | F8ADS8_THEID<br>Q3AE44_CARHZ | 2 | 2.5 |
| (16) | Monomethylamine methyltransferase mtmB1 | mtmB | 2.1.1.- | UniRef50_P58866 | P58969_METMA | 3 | 7.4 |
| (21) | Formate--tetrahydrofolate ligase (formyltetrahydrofolate synthetase, FTHFS) | fhs | 6.3.4.3 | UniRef50_Q3ZX40 | G6GCN2_9FIRM<br>G7W7G1_9FIRM | 8 | 13 |
| (21) | Formate--tetrahydrofolate ligase (FTHFS) | fhs | 6.3.4.3 | UniRef50_Q3ZX40 | B8FGU7_DESAA | 4 | 12 |
| (21) | Formate--tetrahydrofolate ligase (FTHFS) | fhs | 6.3.4.3 | UniRef50_Q3ZX40 | F7NGT7_9FIRM | 3 | 13 |
| (21) | Formate--tetrahydrofolate ligase (FTHFS) | fhs | 6.3.4.3 | UniRef50_B5Y9A5 | B5Y9A5_COPPD | 2 | 4.9 |
| (22) | Pyridoxal-phosphate-dependent serine hydroxymethyltransferase (SHMT) | glyA | 2.1.2.1 | UniRef50_Q214H7 | B5Y8G6_COPPD | 2 | 8.9 |

[*] **For *Caldicellulosiruptor* species**: 9FIRM: *C. lactoaceticus*, except F7NGT7_9FIRM, *Acetonema longum*; ANATD: *C. becscii*; CALH1: *C. hydrothermalis*; CALK2: *C. kronotskyensis*; CALKI: *C. kristjanssonii*; CALS8: *C. saccharolyticus*; CALOO: *C. obsidiansis*; CALOW: *C. owensensis*. **For *Clostridium* species**: CLOTH: *C. thermocellum* (strain ATCC 27405 / DSM 1237); CLOTM: *C. thermocellum* (strain YS); CLOTL: *C. thermocellum* (strain DSM 1313 / LMG 6656 / LQ8); **For other bacterial species**: DESAA: *Desulfatibacillum alkenivorans* - CARHZ: *Carboxydothermus hydrogenoformans*; COPPD: *Coprothermobacter proteolyticus*; MOOTA: *Moorella thermoacetica* (strain ATCC 39073); MOOTH: *Moorella thermoacetica* (*Clostridium thermoaceticum*); THEID: *Thermodesulfatator indicus*. **For methanogenic archaeal species**: METBF: *M. barkeri*; METMA: *M. mazei*;METMJ*: M. marisnigri*; METTF: *M. thermoformicicum*; METTH: *M. thermoautotrophicum* (strain Delta H); METTM: *M. marburgensis*; METTW: *M. thermoautotrophicus* (strain Winter) METWO: *M. wolfei*.

## Table S10: Identified redundant proteins related to ammonia assimilation, amino acid synthesis and biodegradation.

| Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|
| Glutamate synthase | 1.4.1.13 | 4 | A4XGZ2_CALS8<br>A4XH19_CALS8 | 144 | 10 |
| | | | B9MLC1_ANATD<br>E4QDP2_CALH1<br>E4S9M5_CALKI<br>E4SEN1_CALK2<br>G2PU54_9FIRM | 22 | 6 |
| | | | B9MPK7_ANATD<br>D9TIT6_CALOO<br>E4QCE3_CALH1<br>E4QEA3_CALH1<br>E4SA06_CALKI<br>E4SGF1_CALK2<br>G2PTQ8_9FIRM | 103 | 3 |
| | | | A4XGZ5_CALS8<br>A4XH22_CALS8<br>B9MN96_ANATD<br>D9TGY0_CALOO<br>E4Q6B1_CALOW<br>E4QAJ5_CALH1<br>E4S7F6_CALKI<br>E4SCW0_CALK2<br>G2PV67_9FIRM | 16 | 2 |
| NAD-specific glutamate dehydrogenase | 1.4.1.2 | 1 | B5Y6Y1_COPPD | 376 | 9 |
| NADP-specific glutamate dehydrogenase | 1.4.1.4 | 1 | A3DCD3_CLOTH<br>C7HHQ3_CLOTM<br>D1NLN9_CLOTM<br>E6UUG6_CLOTL | 2207 | 13 |
| Glutamine synthetase | 6.3.1.2 | 1 | E4QEA4_CALH1 | 884 | 10 |
| Glutamine amidotransferase, class-II | | 1 | A4XGZ1_CALS8<br>A4XH18_CALS8 | 66 | 3 |
| Acetylglutamate kinase (L-arginine biosynthesis; N(2)-acetyl-L-ornithine from L-glutamate: step 2/4.) | 2.7.2.8 | 1 | A4XJN8_CALS8 | 53 | 2 |
| N-acetyl-gamma-glutamyl-phosphate reductase (L-arginine biosynthesis; N(2)-acetyl-L-ornithine from L-glutamate: step 3/4.) | 1.2.1.38 | 1 | B9MR76_ANATD | 347 | 2 |
| Argininosuccinate synthase (L-arginine biosynthesis; L-arginine from L-ornithine and carbamoyl phosphate: step 2/3) | 6.3.4.5 | 1 | A4XKG4_CALS8<br>B9MRP5_ANATD<br>E4Q144_CALOW<br>E4QBA0_CALH1<br>E4S775_CALKI | 1473 | 2 |

| Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|
| | | | E4SC72_CALK2<br>G2PT27_9FIRM | | |
| Ketol-acid reductoisomerase<br>(L-isoleucine biosynthesis; L-isoleucine from 2-oxobutanoate: step 2/4. L-valine biosynthesis; L-valine from pyruvate: step 2/4.) | 1.1.1.86 | 3 | A3DIE1_CLOTH<br>C7HGY3_CLOTM<br>D1NNE4_CLOTM<br>E6UPQ1_CLOTL | 1288 | 12 |
| | | | B9MNV1_ANATD<br>E4SGV9_CALK2 | 1288 | 11 |
| | | | B8I1T8_CLOCE<br>F1TFC7_9CLOT | 1288 | 2 |
| Branched-chain amino acid aminotransferase<br>(L-isoleucine biosynthesis; L-isoleucine from 2-oxobutanoate: step 4/4. L-leucine biosynthesis; L-leucine from 3-methyl-2-oxobutanoate: step 4/4. L-valine biosynthesis) | 2.6.1.42 | 1 | A3DDR1_CLOTH<br>C7HET2_CLOTM<br>D1NIG3_CLOTM<br>E6UQH3_CLOTL | 25 | 6 |
| Diaminopimelate epimerase<br>(Lysine biosynthesis) | 5.1.1.7 | 1 | E4S991_CALKI<br>G2PUJ6_9FIRM | 22 | 4 |
| Aspartate kinase<br>(L-lysine biosynthesis via DAP pathway; tetrahydrodipicolinate from L-aspartate: step 1/4. L-methionine biosynthesis via de novo pathway; L-homoserine from L-aspartate: step 1/3. L-threonine biosynthesis; L-threonine from L-aspartate: step 1/5.) | 2.7.2.4 | 1 | A4XI70_CALS8<br>E4Q476_CALOW<br>E4QE56_CALH1<br>E4S3W8_CALKI<br>E4SEY5_CALK2<br>G2PT39_9FIRM | 41 | 10 |
| Dihydrodipicolinate synthase<br>(L-lysine biosynthesis via DAP pathway; (S)-tetrahydrodipicolinate from L-aspartate: step 3/4.) | 4.2.1.52 | 1 | A4XJU0_CALS8<br>B9MRD1 ANATD<br>D9TL06_CALOO<br>E4Q6T1_CALOW<br>E4QBU9_CALH1<br>E4S681_CALKI<br>E4SF91_CALK2<br>G2PYX8_9FIRM | 86 | 3 |
| Diaminopimelate decarboxylase<br>(L-lysine biosynthesis via DAP pathway; L-lysine from DL-diaminopimelate: step 1/1.) | 4.1.1.20 | 1 | A4XI14_CALS8<br>D9TJ50_CALOO<br>E4Q545_CALOW<br>E4QDI9_CALH1<br>E4S4J7_CALKI<br>G2PXU8_9FIRM | 10 | 2 |
| 3-isopropylmalate dehydratase large subunit<br>(L-leucine biosynthesis; L-leucine from 3-methyl-2-oxobutanoate: step 2/4.) | 4.2.1.33 | 1 | B9MLV4_ANATD<br>D9TG27_CALOO<br>E4Q4W1_CALOW<br>E4QDA0_CALH1<br>E4S9H8_CALKI<br>G2PUH9_9FIRM | 333 | 3 |

| Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|
| Dihydroxy-acid dehydratase<br>(Valine, leucine and isoleucine biosynthesis) | 4.2.1.9 | 1 | E4S9E7_CALKI<br>G2PUP0_9FIRM | 498 | 4 |
| Aspartate-semialdehyde dehydrogenase<br>(Glycine, serine and threonine metabolism and lysine biosynthesis) | 1.2.1.11 | 1 | B9MRD2_ANATD<br>E4Q6T2_CALOW<br>E4QBU8_CALH1<br>E4SF90_CALK2 | 664 | 2 |
| Homoserine dehydrogenase | 1.1.1.3 | 1 | A4XI72_CALS8 | 32 | 4 |
| Ornithine carbamoyltransferase<br>(Arginine and proline metabolism) | 2.1.3.3 | 1 | B5Y9F2_COPPD | 278 | 2 |
| Cysteine synthase<br>(Cysteine metabolism) | 2.5.1.47 | 5 | A3DGI2_CLOTH<br>C7HI89_CLOTM<br>D1NPL2_CLOTM<br>E6UQ45_CLOTL | 2433 | 5 |
| | | | B9MRG2_ANATD<br>E4QBR9_CALH1 | 2433 | 4 |
| | | | A4XK88_CALS8<br>E4Q8L3_CALH1<br>E4SA35_CALKI<br>G2PVE7_9FIRM | 2433 | 3 |
| | | | D4J1P2_BUTFI | 2433 | 2 |
| | | | E4Q6S2_CALOW | 294 | 10 |
| O-acetylhomoserine/O-acetylserine sulfhydrylase<br>(Cysteine metabolism) | 2.5.1.49 | 1 | A4XJT1_CALS8 | 294 | 3 |
| MaoC domain protein dehydratase | 1.4.3.4 | 1 | A1HPI7_9FIRM | 4 | 2 |
| Thermostable monoacylglycerol lipase (Mglp)<br>(Serine hydrolysis) | 3.1.1.23 | 1 | B5Y789_COPPD | 1 | 3 |
| Alanine--glyoxylate aminotransferase 1<br>(Alanine and aspartate metabolism and glycine, serine and threonine metabolism) | 2.6.1.44 | 1 | B5Y5X4_COPPD | 1 | 9 |
| Adenylosuccinate synthetase<br>(Alanine, aspartate and glutamate metabolism) | 6.3.4.4 | 2 | B5Y8A7\|B5Y8A7_COPPD | 1 | 2 |
| | | | A4XML6_CALS8<br>B9MLK0_ANATD<br>E4Q9B7_CALH1<br>G2PY08_9FIRM | 3 | 2 |
| Ketoisovalerate oxidoreductase subunit VorA (VOR) (2-oxoisovalerate oxidoreductase alpha chain) (2-oxoisovalerateferredoxin reductase subunit alpha)<br>(Valine, leucine and isoleucine degradation) | 1.2.7.7 | 2 | B5Y8N6_COPPD | 50 | 4 |
| | | | B5Y8N7_COPPD | 1 | 3 |
| Ketoisovalerate oxidoreductase subunit VorB (VOR) (2-oxoisovalerate oxidoreductase beta chain) (2-oxoisovalerate ferredoxinreductase subunit beta) | 1.2.7.7 | 1 | B5Y8N5_COPPD | 1 | 5 |

| Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|
| (Valine, leucine and isoleucine degradation) | | | | | |
| Glycine reductase complex component B subunit gamma (Selenoprotein PB gamma) | 1.21.4.2 | 1 | B5Y746_COPPD | 111 | 9 |
| Glycine reductase complex component B subunits alpha and beta (Selenoprotein PB alpha/beta) | 1.21.4.2 | 1 | B5Y743_COPPD | 107 | 6 |
| Urocanate hydratase<br>(Amino-acid degradation; L-histidine degradation into L-glutamate; N-formimidoyl-L-glutamate from L-histidine: step 2/3.) | 4.2.1.49 | 1 | B5Y9A8_COPPD | 1286 | 6 |
| 2-isopropylmalate synthase<br>(Valine, leucine and isoleucine biosynthesis) | 2.3.3.13 | 2 | E4Q7V5_CALH1 | 241 | 9 |
| | | | A4XIL9_CALS8<br>B9MNV4_ANATD<br>D9TIE9_CALOO<br>E4Q3I9_CALOW<br>E4Q7V3_CALH1<br>E4S436_CALKI<br>E4SGV6_CALK2<br>G2PX64_9FIRM | 139 | 4 |
| Acetolactate synthase large subunit | 2.2.1.6 | 1 | D9TIV8_CALOO<br>E4Q4G3_CALOW<br>E4QE08_CALH1 | 535 | 4 |
| S-adenosylmethionine synthase<br>(S-adenosyl-L-methionine biosynthesis; S-adenosyl-L-methionine from L-methionine: step 1/1.) | 2.5.1.6 | | A4XHT_CALS8<br>B9MQ10_ANATD<br>D9TIW8_CALOO<br>E4Q4N6_CALOW<br>E4QE02_CALH1<br>E4S9D8_CALKI<br>E4SGB5_CALK2<br>G2PUN4_9FIRM | 1490 | 4 |
| Methylthioribose-1-phosphate isomerase<br>(Cysteine and methionine metabolism) | 5.3.1.23 | | A4XKS3_CALS8<br>B9MS06_ANATD<br>E4Q1H8_CALOW<br>E4QAA9_CALH1<br>E4S883_CALKI<br>E4SBS8_CALK2<br>G2PTS6_9FIRM | 29 | 2 |
| | | | A3DGQ2_CLOTH<br>C7HJ23_CLOTM<br>D1NPR1_CLOTM<br>E6UQP8_CLOTL | 4 | 12 |
| | | | A3DCG1_CLOTH<br>C7HHT0_CLOTM<br>D1NLR2_CLOTM | 4 | 7 |

| Name | E.C. number | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|
| | | | E6UUE1_CLOTL | | |
| | | | A3DI06_CLOTH<br>C7HDG5_CLOTM<br>D1NP04_CLOTM<br>E6URG4_CLOTL | 5 | 6 |
| | | | B5Y6Q7_COPPD | 1 | 5 |
| | | | D7CPT7_SYNLT | 1 | 2 |
| | | | A3DI47_CLOTH<br>C7HDC3_CLOTM<br>D1NNW3_CLOTM<br>E6UPK7_CLOTL | 4 | 2 |
| | | | A3DKE2_CLOTH<br>A3DKE3_CLOTH<br>A3DKE4_CLOTH<br>C7HJJ6_CLOTM<br>C7HJJ7_CLOTM<br>C7HJJ8_CLOTM<br>D1NKK6_CLOTM<br>D1NKK7_CLOTM<br>D1NKK8_CLOTM<br>E6ULT4_CLOTL<br>E6ULT5_CLOTL<br>E6ULT6_CLOTL | 12 | 2 |

## Table S11. Identified redundant proteins related to proteolysis.

| Name | MEROPS family | Number of identified non-redundant protein groups | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|---|
| Protease | S8A | 1 | B5Y6Q5_COPPD | 1 | 14 |
| M18 family aminopeptidase | M18 | 1 | B5Y7J5_COPPD | 1 | 7 |
| TldD/PmbA family protein, putative | U62 | 1 | B5Y6N5_COPPD | 1 | 6 |
| LmbIH | U62 | 1 | B5Y6N6_COPPD | 1 | 5 |
| Pyrrolidone-carboxylate peptidase | C15 | 1 | B5Y5X6_COPPD | 1 | 3 |
| Proline dipeptidase | M24 | 2 | B5Y8M3_COPPD | 1 | 3 |
| | | | B5Y692_COPPD | 10 | 2 |
| TldD protein | U62 | 1 | B5Y9Y2_COPPD | 1 | 3 |
| Peptidase T | M20 | 1 | B5Y9V8_COPPD | 586 | 3 |
| Intracellular protease 1 | C56 | 1 | B5Y9N6_COPPD | 1 | 2 |
| DO serine protease | S1B | 1 | B5YA04_COPPD | 1 | 2 |
| Propeptide, PepSY amd peptidase M4 | M4 | 1 | A3DHE1_CLOTH<br>C7HE90_CLOTM<br>D1NRL0_CLOTM<br>E6USR7_CLOTL | 7 | 8 |
| Peptidase S1 and S6, chymotrypsin/Hap | S1;S6 | 1 | A3DEY9_CLOTH<br>C7HFB4_CLOTM<br>D1NMX1_CLOTM<br>E6UMZ2_CLOTL | 6 | 7 |
| Intracellular protease, PfpI family | C56 | 1 | A3DJR3_CLOTH<br>C7HEH1_CLOTM<br>D1NRA8_CLOTM<br>E6UTK9_CLOTL | 451 | 5 |
| M18 family aminopeptidase | M18 | 1 | A3DDD2_CLOTH<br>C7HD12_CLOTM<br>D1NMH8_CLOTM<br>E6URU7_CLOTL | 88 | 4 |
| PgdS peptidase<br>Cysteine peptidase | C40 | 1 | A3DD11_CLOTH<br>C7HDK6_CLOTM<br>D1NI23_CLOTM<br>E6USM3_CLOTL | 4 | 3 |
| Oligoendopeptidase F | M3 | 1 | E4Q8X3_CALH1 | 1 | 3 |

| **Name** | **MEROPS family** | **Number of identified non-redundant protein groups** | **Identified protein ID** | **UniRef50 cluster size** | **Maximum number of assigned unique peptides** |
|---|---|---|---|---|---|
| M18 family aminopeptidase | M18 | 1 | A4XLU2_CALS8 | 8 | 2 |
| Peptidase M29 aminopeptidase II | M29 | 1 | E4Q2D7_CALOW<br>E4Q8Z4_CALH1<br>E4S9S4_CALKI | 100 | 2 |
| Heat shock protein HslVU, ATPase subunit HslU | AAA | 1 | B9MM30_ANATD<br>E4Q5B3_CALOW<br>E4QCV0_CALH1<br>E4S8P3_CALKI<br>E4SE70_CALK2<br>G2PUY2_9FIRM | 123 | 2 |
| ATPase AAA-2 domain protein | AAA | 1 | A4XM10_CALS8<br>E4Q3N5_CALOW | 44 | 2 |
| ATP-dependent Clp protease proteolytic subunit | S14 | 1 | F0DPE2_9FIRM<br>F6B8J8_DESCC | 25 | 2 |
| Archaean proteasome, beta component. Threonine peptidase | T01A | 1 | Q46G14_METBF | 10 | 2 |

## Table S12. Identified redundant proteins related to transport systems.

| Category | Name | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|
| Carbohydrate-specific ABC transporter | Extracellular solute-binding protein, family 1 | A3DE73_CLOTH<br>C7HBZ6_CLOTM<br>D1NIX7_CLOTM<br>E6UP76_CLOTL | 4 | 12 |
| | Extracellular solute-binding protein family 1 | G2PXK4_9FIRM | 8 | 5 |
| | Extracellular solute-binding protein, family 1 (Precursor) | A4XHC6_CALS8<br>D9TIQ5_CALOO<br>E4Q4B1_CALOW | 8 | 3 |
| | Extracellular solute-binding protein family 1 | E4Q988_CALH1<br>E4S6H5_CALKI | 6 | 3 |
| | D-xylose ABC transporter, periplasmic substrate-binding protein (EC 3.6.3.17) | A4XME8_CALS8<br>E4S7X0_CALKI | 18 | 3 |
| | Periplasmic binding protein/LacI transcriptional regulator | B9MMN7_ANATD<br>E4Q5R5_CALOW<br>E4QCD6_CALH1<br>E4SE02_CALK2 | 5 | 3 |
| | Extracellular solute-binding protein, family 1 | E4QED3_CALH1 | 8 | 2 |
| | Extracellular solute-binding protein family 5 (Precursor) | E4S411_CALKI | 3 | 2 |
| | Sugar-binding transport (ABC transporter) ATP-binding protein (CymD protein) | A4XLY9_CALS8<br>B9MKN8_ANATD<br>D9TFC9_CALOO<br>E4Q3E5_CALOW<br>E4Q7N2_CALH1<br>E4S4F5_CALKI<br>E4SDH3_CALK2<br>G2PZ28_9FIRM | 1617 | 2 |
| | ABC transporter related protein | E4S7W9_CALKI | 11 | 2 |
| | Maltose ABC transporter, periplasmic maltose-binding protein | B5Y6P4_COPPD | 1 | 7 |
| | Sugar ABC transporter, ATP-binding protein | B5Y8U9_COPPD | 1 | 3 |
| | Periplasmic binding protein/LacI transcriptional regulator | A8F3L8_THELT | 1 | 2 |

| Category | Name | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|
|  | Extracellular solute-binding protein family 1 (Precursor) | E1R2D9_SPISS | 3 | 2 |
| Peptide-specific ABC transporter | Bacterial extracellular solute-binding protein, family 5 | B5Y6V2_COPPD | 2 | 18 |
|  | Bacterial extracellular solute-binding protein, family 5 | B5Y6V3_COPPD | 2 | 2 |
|  | Oligopeptide/dipeptide ABC transporter, ATP-binding protein | B5Y897_COPPD | 30 | 4 |
|  | Oligopeptide transport ATP-binding protein AppF | B5Y898_COPPD | 376 | 3 |
|  | ABC transporter related protein (ABC-type antimicrobial peptide transport system, ATPase component) | A3DCU6_CLOTH<br>C7HFE1_CLOTM<br>D1NM47_CLOTM<br>E6UT84_CLOTL | 230 | 3 |
|  | ABC transporter related protein (ABC-type antimicrobial peptide transport system, ATPase component) | A3DH88_CLOTH<br>C7HDX4_CLOTM<br>D1NQ96_CLOTM<br>E6US63_CLOTL | 102 | 2 |
|  | Extracellular solute-binding protein, family 5 | A4XIB1_CALS8 | 1 | 7 |
|  | Extracellular ligand-binding receptor | A4XKU3_CALS8 | 19 | 3 |
|  | Extracellular solute-binding protein family 3 | B9ML21_ANATD<br>E4SF04_CALK2 | 8 | 2 |
|  | ABC transporter related protein | G2PZD6_9FIRM | 4 | 2 |
|  | Extracellular solute-binding protein family 5 | A9BFQ4_PETMO | 1 | 2 |
| Other ABC transporter related | ABC transporter related protein | A4XLW7_CALS8<br>B9MKL6_ANATD<br>D9TFA7_CALOO<br>E4Q3C1_CALOW<br>E4Q7Q4_CALH1<br>E4SDQ3_CALK2 | 548 | 2 |
|  | ABC transporter related protein | E4S7W6_CALKI<br>G2PV36_9FIRM | 34 | 2 |
| Ion channels and secondary transport | TrkA-N domain protein (K transport) | A4XFW7_CALS8 | 8 | 2 |
|  | Ferritin Dps family protein (Iron ion transport) | B9ML80_ANATD<br>E4Q488_CALOW<br>E4QE43_CALH1<br>E4SAP1_CALKI | 14 | 2 |

| Category | Name | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|
| | | E4SEW4_CALK2<br>G2PT55_9FIRM | | |
| | Phosphate uptake regulator, PhoU | A4XL29_CALS8<br>B9MJS2_ANATD<br>D9TK39_CALOO<br>E4Q201_CALOW<br>E4Q9K2_CALH1<br>E4S8X2_CALKI<br>E4SBG6_CALK2<br>G2PUB2_9FIRM | 9 | 3 |
| | Hemerythrin (Inorganic ion transport; oxygen transport) | B5YA48_COPPD | 2 | 2 |
| | Outer membrane pore protein | D6GAC5_9ENTR<br>B5XNZ9_KLEP3<br>A6TBT2_KLEP7<br>B9VMB2_KLEPN<br>C8T735_KLEPR<br>D3RLP3_KLEVT | 1746 | 3 |
| Phosphotransferase (PTS) stystem related | HPr kinase/phosphorylase (HPrK/P) | D9TI35_CALOO<br>E4SH73_CALK2 | 30 | 2 |
| | PTS system IIB component | B5Y677_COPPD | 1 | 2 |
| Pyrophosphate-energized proton pump | K(+)-insensitive pyrophosphate-energized proton pump 2 | D7CPG5_SYNLT | 6 | 3 |
| F-type ATPase | ATP synthase subunit alpha | A4XKX2_CALS8<br>D9TK95_CALOO<br>E4QA57_CALH1<br>E4S8R6_CALKI<br>E4SBM2_CALK2<br>G2PTY0_9FIRM | 2211 | 7 |
| | ATP synthase subunit beta | A4XKX0_CALS8<br>D9TK97_CALOO<br>E4Q1M6_CALOW<br>E4QA59_CALH1<br>E4S8R4_CALKI<br>E4SBM4_CALK2 | 3850 | 9 |
| V-type ATPase | V-type ATP synthase beta chain | A3DHP1_CLOTH<br>C7HHA5_CLOTM | 1058 | 7 |

| Category | Name | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|
| | | D1NNL5_CLOTM<br>E6UTX3_CLOTL | | |
| | V-type ATP synthase beta chain | D9PXH9_METTM<br>VATB_METTH | 1058 | 6 |

Note: **Transport systems (Table S12).** 21 clusters of ATP-binding cassette (ABC) transporters were retrieved, including 10 carbohydrate-specific and 8 peptide-specific ones involved in the transport of hydrolyzed sugars and peptides (carbon and nitrogen metabolism respectively). Other enzymes retrieved included the 6 with ion transport activity for K, Fe, phosphate, O2, etc., and 2 phosphotransferase system (PTS) components presenting sugar:hydrogen symporter activity.

## Table S13. Identified redundant proteins related to hydrogenase and related electron transfer proteins.

| Name | E.C. number | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|
| Hydrogenase, | | D9TKL8_CALOO | 64 | 12 |
| Hypothetical hydrogenase subunit (Fe-only) | | A5D0Q2_PELTS | 82 | 9 |
| Ferredoxin | | A3DC99_CLOTH<br>C7HIN8_CLOTM<br>D1NLK4_CLOTM<br>E6ULI7_CLOTL | 6 | 2 |
| Ferredoxin | | B5Y702_COPPD | 14 | 2 |
| NADH dehydrogenase (Quinone) | 1.6.99.5 | B9MRU4_ANATD<br>E4SBY6_CALK2 | 88 | 9 |
| NADH:ubiquinone oxidoreductase, NADH-binding (51kD) hydrogenase subunit | 1.6.99.5 | B5Y703_COPPD | 66 | 2 |
| NADH dehydrogenase (Quinone) | 1.6.99.5 | E6VU71_DESAO | 85 | 2 |
| NADH:ubiquinone oxidoreductase, NADH-binding 51 kD subunit | 1.6.99.5 | A5D0Q3_PELTS | 66 | 4 |
| NADH:ubiquinone oxidoreductase 24kD formate dehydrogenase subunit | 1.6.5.3 | A5CYU8_PELTS | 9 | 2 |
| NADH dehydrogenase (Ubiquinone), 24 kDa hydrogenase subunit | 1.6.5.3 | A4XKL0_CALS<br>B9MRU1_ANATD<br>D9TKM2_CALOO<br>E4Q190_CALOW<br>E4QAX0_CALH1<br>E4S7P7_CALKI<br>E4SBY9_CALK2<br>G2PTE4_9FIRM | 277 | 3 |

## Table S14. Identified redundant proteins related to cofactor and coenzyme synthesis.

| Category | Name | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|
| Thiamine | Thiamine-phosphate pyrophosphorylase | B5Y6I1_COPPD | 3 | 3 |
| Riboflavin | 6,7-dimethyl-8-ribityllumazine synthase | A3DBL9_CLOTH<br>C7HFU6_CLOTM<br>D1NKW7_CLOTM<br>E6UMT4_CLOTL | 280 | 2 |
| Pyridoxine | Putative uncharacterized protein (Pyridoxal biosynthesis protein) | G2PZD2_9FIRM<br>B9MKY9_ANATD<br>E4Q791_CALH1<br>E4SB33_CALK2<br>E4S415_CALKI<br>D9TFM6_CALOO<br>E4Q417_CALOW<br>A4XIB5_CALS8 | 304 | 2 |
| Coenzyme CoA | Type III pantothenate kinase | A3DIK9_CLOTH<br>C7HBR9_CLOTM<br>D1NPH0_CLOTM<br>E6UQB2_CLOTL | 16 | 2 |
|  | Type III pantothenate kinase | B9MPA9_ANATD<br>D9TII0_CALOO<br>E4Q3V0_CALOW<br>E4Q7H3_CALH1<br>E4S9V8_CALKI<br>E4SGQ7_CALK2<br>G2PU31_9FIRM | 147 | 2 |
|  | 3-methyl-2-oxobutanoate hydroxymethyltransferase | B9ML78_ANATD<br>D9TFT8_CALOO<br>E4Q486_CALOW<br>E4QE46_CALH1<br>E4SAP3_CALKI<br>E4SEW6_CALK2<br>G2PT53_9FIRM | 347 | 2 |
|  | 3-methyl-2-oxobutanoate hydroxymethyltransferase | B5Y864_COPPD | 3 | 4 |

| Category | Name | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|---|
| | Acetolactate synthase | A3DIE0_CLOTH<br>C7HGY4_CLOTM<br>D1NNE5_CLOTM<br>E6UPQ0_CLOTL | 356 | 2 |
| | Acetolactate synthase | D9TIE5_CALOO | 356 | 3 |
| Coenzyme $F_{390}$ | Phenylacetate-CoA ligase (Coenzyme $F_{390}$ synthetase) | A3DD21_CLOTH<br>E6USL3_CLOTL<br>C7HDL6_CLOTM<br>D1NI33_CLOTM<br>A6VIF0_METM7<br>D7DQS7_METV3 | 532 | 2 |
| Biotin | 8-amino-7-oxononanoate synthase | B5Y9Z4_COPPD | 48 | 9 |

Note: **Cofactor and coenzyme synthesis (Table S14)**. Identified proteins were associated to the biosynthesis of thiamine, riboflavin and pyridoxine, coenzyme CoA and F390 as well as biotin metabolism.

**Table S15. Identified redundant proteins related to CAS genes.**

| Name | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|
| CRISPR-associated regulatory protein, DevR family | A3DHS4_CLOTH<br>C7HGH0_CLOTM<br>D1NP86_CLOTM<br>E6UU04_CLOTL | 4 | 8 |
| CRISPR-associated protein, Csh2 family | A3DKB8_CLOTH<br>D1NKI2_CLOTM | 12 | 6 |
| CRISPR-associated protein, Csd2 family | B8I207_CLOCE | 141 | 2 |
| CRISPR-associated protein, Csh2 family | G2PYK4_9FIRM | 10 | 2 |
| CRISPR-associated regulatory protein, DevR family | E4Q9T2_CALH1<br>E4S7V7_CALKI | 6 | 2 |

Note: **CAS genes (Table S15)**. CAS genes are CRISPR-associated genes. CRISPRs (Clustered Regularly Interspaced Short Palindromic Repreats) are present in some bacteria and many archaea. CRISPRs are non-coding transcribed genome regions consisting in short DNA repeats separated by highly variable intervening sequences (spacers). Their presence is associated to CAS genes. The CRISPR and CAS proteins have been proposed to function as a defense mechanism against microbial viruses or other mobile elements (Barrangou *et al.*, 2007).

.

## Table S16. Identified redundant proteins related to cell mobility.

| Name | Identified protein ID | UniRef50 cluster size | Maximum number of assigned unique peptides |
|---|---|---|---|
| Flagellin protein | B5Y7C7_COPPD | 1 | 8 |
| Flagellin domain protein | G2PVZ5_9FIRM | 262 | 8 |
| Flagellin domain protein | E4SAE3_CALKI | 262 | 6 |
| Flagellin domain protein | A4XK35_CALS8 | 76 | 2 |
| Flagellin domain protein | D9TJP3_CALOO | 262 | 2 |
| Flagellin domain protein | E4Q8G4_CALH1 | 5 | 2 |
| Flagellin domain protein | A3DHL0_CLOTH<br>C7HID0_CLOTM<br>D1NNP6_CLOTM<br>E6UTE5_CLOTL | 262 | 5 |
| Flagellin domain protein | D7CPM1_SYNLT | 1 | 3 |
| Flagellin domain protein | F6CQT2_DESK7 | 10 | 2 |
| Flagellin protein | E5E7U1_9BACI | 262 | 2 |
| Flagellin (Fragment) | B5KF86_9RHOB | 3 | 2 |
| Flagellin domain protein | A9BJ12_PETMO | 19 | 2 |
| Flagellin domain protein | C6PPI6_9CLOT | 17 | 2 |
| Flagellin domain protein | D5WVR7_BACT2 | 2 | 2 |
| Methyl-accepting chemotaxis protein | B9MK66_ANATD<br>E4Q2I4_CALOW<br>E4SA50_CALK<br>E4SCM7_CALK2<br>G2PVG3_9FIRM | 7 | 3 |
| Methyl-accepting chemotaxis sensory transducer | E4Q1X4_CALOW<br>E4Q9Z5_CALH1<br>E4SCR4_CALK2 | 4 | 2 |
| CheC, inhibitor of MCP methylation / FliN fusion protein | B9MM12_ANATD<br>D9TG88_CALOO<br>E4Q519_CALOW<br>E4QCW8_CALH1<br>E4S947_CALKI<br>E4SE88_CALK2<br>G2PUW4_9FIRM | 9 | 3 |
| Major pilin protein FimA | B5Y6M7_COPPD | 3 | 2 |

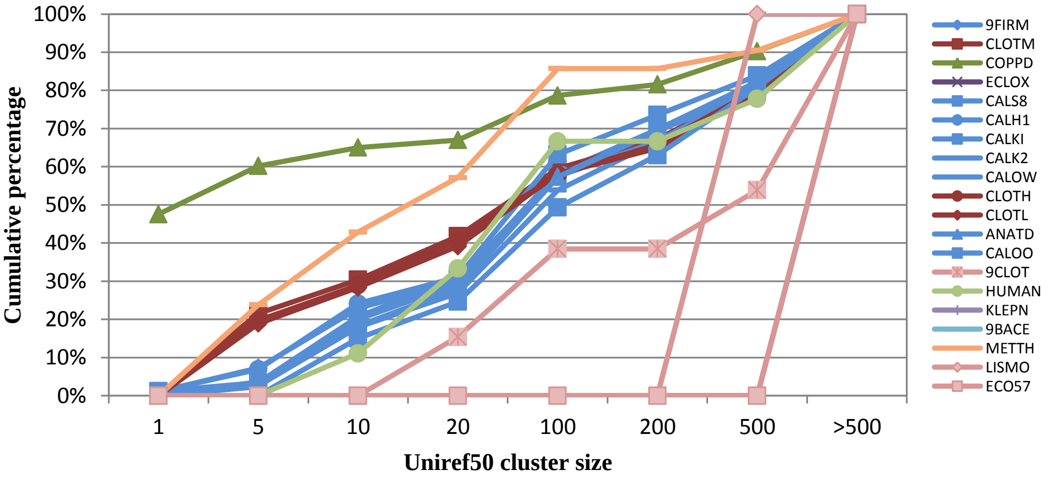

**Figure S10. Cumulative distribution of the identified proteins according to their UniRef50 cluster size, for the most retrieved species or strains.**

The size corresponds to the number of distinct proteins contained in a UniRef50 cluster (Suzek *et al.*, 2007), e.g. size one (X-axis) refers to clusters containing a single protein (very narrow biologic distribution). The higher sizes (X-axis) refer to UniRef50 clusters containing more distinct proteins (larger biologic distribution, the identified protein shares similarity with numerous other proteins from the database at a 50% identity threshold). As visible on the figure, half of the proteins identified and related to *Coprothermobacter proteolyticus* (COPPD) belong to UniRef50 clusters of size 1. Based on the figure, smaller size clusters were found for *Coprothermobacter proteolyticus* (COPPD), followed by *Methanothermobacter thermautotrophicus* (METTH), *Clostridium thermocellum* (CLOTM, CLOTH, CLOTL) and *Caldcellulosiruptor* spp (9FIRM, CALS8, CALH1, CALK1, CALJ2, CALOW, ANATD, CALOO) by decreasing order.

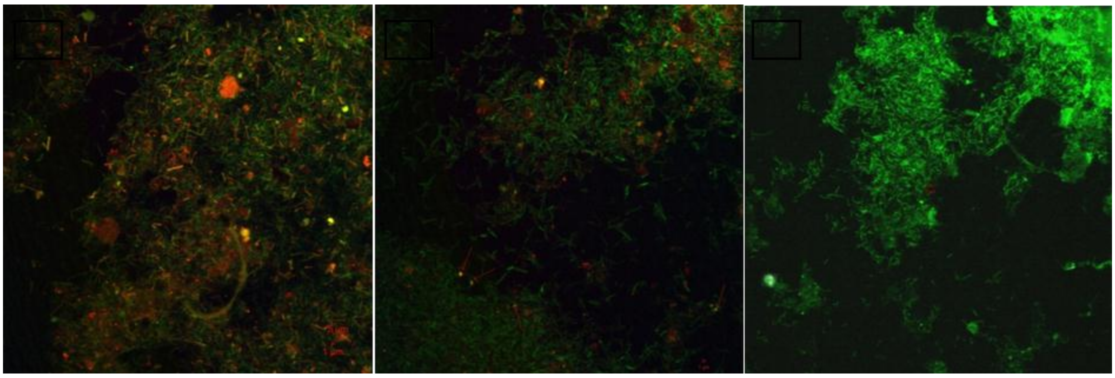

**Figure S11. FISH images of incubation samples hybridized with probes specific for various bacterial groups.**

(a) Green: probes Eub I, II, III-Cy3, targeting most bacteria – red: Copro925-Cy5, targeting *Coprothermobacter* species (b) green: probes Eub I, II, III-Cy3 – red: Cal1415-Cy5, targeting *Caldicellulosiruptor* species (c) green: UCL284, targeting a subgroup of the Clostridiales order including *Clostridium thermocellum*. More details concerning the FISH procedure and the probes in **Section S2** and **Table S2**.

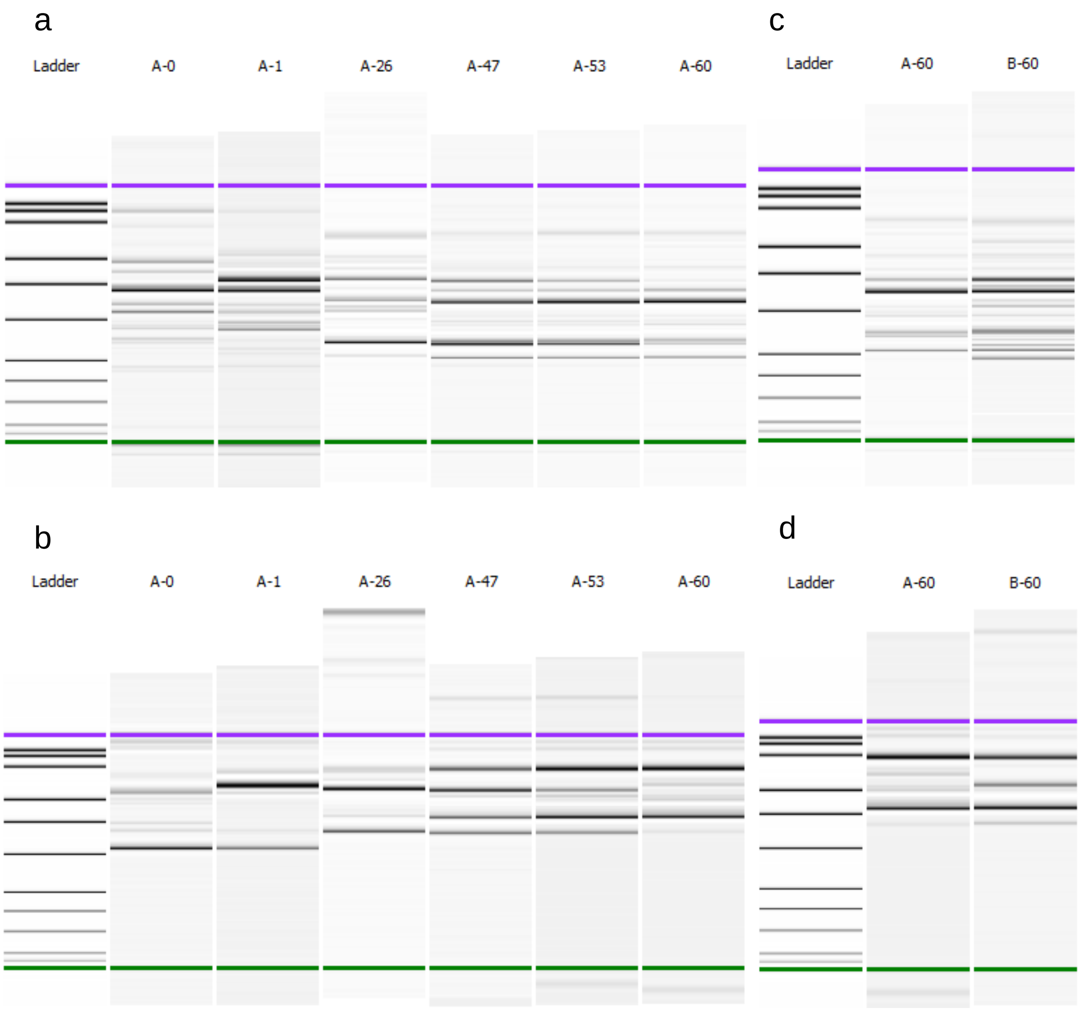

**Figure S12. Pseudo gels of ARISA profiles.**

**a**: Bacterial community dynamics for the incubation microcosm replicate A.

**b**: Archaeal community dynamics for the incubation microscosm replicate A.

For A and B, profiles were determined at days 0, 1, 26, 47, 53 and 60.

**c**: Bacterial communities at day 60, for the incubation microcosm replicates A and B.

**d**: Archaeal communities at day 60, for the incubation microcosm replicates A and B.

The microcosm A was used for metaproteomics analyses, at day 60.

Ladders (in bp): 1000, 850, 700, 500, 400, 300, 200, 150, 100, 50, 25.

Internal size standards: 1500 bp (violet pseudo band) – 15 bp (green pseudo band).

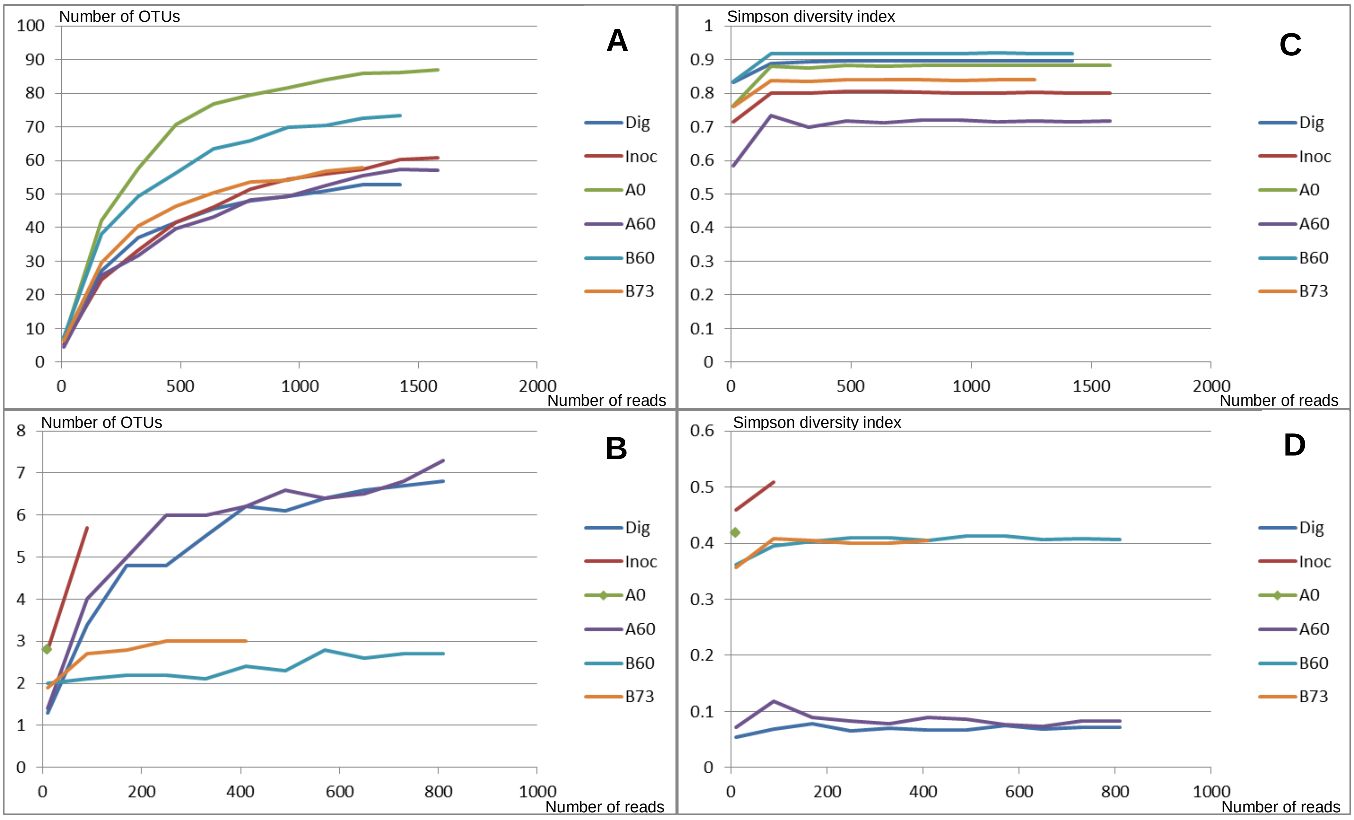

**Figure S13. Rarefaction curves based on the 16S tag pyrosequencing data.**

A. For sequences obtained using bacterial primers, number of OTU observed.

B. For sequences obtained using archaeal primers, number of OTU observed.

C. For sequences obtained using bacterial primers, Simpson diversity index.

D. For sequences obtained using archaeal primers, Simpson diversity index.

## Table S17. Taxonomic attribution of the redundant identified proteins, taxa ordered by decreasing number of identified redundant proteins.

| OrganismID | Organism | Identified protein number | Proportion compared to the total number of identified proteins (%) |
|---|---|---|---|
| 9FIRM | *Firmicutes(Caldicellulosiruptor lactoaceticus)* | 193 | 7.60 |
| CLOTM | *Clostridium thermocellum* (strain YS) | 160 | 6.30 |
| ECOLX | *Escherichia coli* | 114 | 4.49 |
| COPPD | *Coprothermobacter proteolyticus (strain ATCC 35245 / DSM 5265 / BT)* | 105 | 4.13 |
| CALS8 | *Caldicellulosiruptor saccharolyticus (strain ATCC 43494 / DSM 8903)* | 90 | 3.54 |
| CALH1 | *Caldicellulosiruptor hydrothermalis (strain DSM 18901 / VKM B-2411 / 108)* | 89 | 3.50 |
| CALKI | *Caldicellulosiruptor kristjanssonii (strain ATCC 700853 / DSM 12137 / I77R1B)* | 85 | 3.35 |
| CALK2 | *Caldicellulosiruptor kronotskyensis (strain DSM 18902 / VKM B-2412 / 2002)* | 82 | 3.23 |
| CALOW | *Caldicellulosiruptor owensensis (strain ATCC 700167 / DSM 13100 / OL)* | 79 | 3.11 |
| CLOTH | *Clostridium thermocellum (strain ATCC 27405 / DSM 1237)* | 79 | 3.11 |
| CLOTL | *Clostridium thermocellum (strain DSM 1313 / LMG 6656 / LQ8)* | 78 | 3.07 |
| ANATD | *Anaerocellum thermophilum (strain DSM 6725 / Z-1320) (Caldicelulosiruptor becscii)* | 75 | 2.95 |
| CALOO | *Caldicellulosiruptor obsidiansis (strain ATCC BAA-2073 / strain OB47)* | 73 | 2.87 |
| 9CLOT | *Clostridiaceae* | 30 | 1.18 |
| HUMAN | *Homo sapiens* | 28 | 1.10 |
| KLEPN | *Klebsiella pneumoniae* | 26 | 1.02 |
| 9BACE | *Bacteroides* | 24 | 0.94 |
| METTH | *Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H)* | 21 | 0.83 |
| LISMO | *Listeria monocytogenes* | 19 | 0.75 |
| ECO57 | *Escherichia coli O157:H7* | 17 | 0.67 |
| MOUSE | *Mus musculus* | 17 | 0.67 |
| SOLTU | *Solanum tuberosum* | 16 | 0.63 |
| METTM | *Methanothermobacter marburgensis (strain DSM 2133 / 14651 / NBRC 100331 / OCM 82 / Marburg)* | 13 | 0.51 |
| SHIFL | *Shigella flexneri* | 13 | 0.51 |
| NOMLE | *Nomascus leucogenys* | 11 | 0.43 |
| CLOBO | *Clostridium botulinum* | 10 | 0.39 |
| GORGO | *Gorilla gorilla gorilla* | 10 | 0.39 |
| 9TELE | *Teleostei* | 9 | 0.35 |
| CRAGI | *Crassostrea gigas* | 9 | 0.35 |
| MACMU | *Macaca mulatta* | 9 | 0.35 |
| PELTS | *Pelotomaculum thermopropionicum (strain DSM 13744 / JCM 10971 / SI)* | 9 | 0.35 |
| CLOPE | *Clostridium perfringens* | 8 | 0.31 |
| 9ENTR | *Enterobacteriaceae* | 7 | 0.28 |
| 9THEO | *Thermoanaerobacteriaceae* | 7 | 0.28 |
| PENJP | *Penaeus japonicus* | 7 | 0.28 |
| CALJA | *Callithrix jacchus* | 6 | 0.24 |
| ORYSJ | *Oryza sativa subsp. japonica* | 6 | 0.24 |
| STRPU | *Strongylocentrotus purpuratus* | 6 | 0.24 |
| THETN | *Thermoanaerobacter tengcongensis (strain DSM 15242 / JCM 11007 / NBRC 100824 / MB4)* | 6 | 0.24 |
| 9EUKA | *Eukaryota* | 5 | 0.20 |
| 9PERC | *Percomorpha* | 5 | 0.20 |
| BACAM | *Bacillus amyloliquefaciens* | 5 | 0.20 |
| BRAFL | *Branchiostoma floridae* | 5 | 0.20 |
| CLOCE | *Clostridium cellulolyticum (strain ATCC 35319 / DSM 5812 / JCM 6584 / H10)* | 5 | 0.20 |
| CLODI | *Clostridium difficile* | 5 | 0.20 |
| DESRM | *Desulfotomaculum reducens (strain MI-1)* | 5 | 0.20 |
| DICDI | *Dictyostelium discoideum* | 5 | 0.20 |
| GASAC | *Gasterosteus aculeatus* | 5 | 0.20 |
| MACFA | *Macaca fascicularis* | 5 | 0.20 |
| METMJ | *Methanoculleus marisnigri (strain ATCC 35101 / DSM 1498 / JR1)* | 5 | 0.20 |
| SYNLT | *Syntrophothermus lipocalidus (strain DSM 12680 / TGB-C1)* | 5 | 0.20 |
| XENLA | *Xenopus laevis* | 5 | 0.20 |
| 9ACAR | *Acari* | 4 | 0.16 |
| 9ARAC | *Arachnida* | 4 | 0.16 |
| 9BACI | *Bacillaceae* | 4 | 0.16 |
| 9BACT | *Bacteria* | 4 | 0.16 |
| 9NEOP | *Neoptera* | 4 | 0.16 |
| AMMDK | *Ammonifex degensii (strain DSM 10501 / KC4)* | 4 | 0.16 |
| BACOV | *Bacteroides ovatus* | 4 | 0.16 |
| CANFA | *Canis familiaris* | 4 | 0.16 |
| CARHZ | *Carboxydothermus hydrogenoformans (strain Z-2901 / DSM 6008)* | 4 | 0.16 |
| CELLD | *Cellulosilyticum lentocellum (strain ATCC 49066 / DSM 5427 / NCIMB 11756 / RHM5)* | 4 | 0.16 |
| CHICK | *Gallus gallus* | 4 | 0.16 |
| DANRE | *Danio rerio* | 4 | 0.16 |
| DAPPU | *Daphnia pulex* | 4 | 0.16 |
| DESCC | *Desulfotomaculum carboxydivorans (strain DSM 14880 / VKM B-2319 / CO-1-SRB)* | 4 | 0.16 |
| DICFS | *Dictyostelium fasciculatum (strain SH3)* | 4 | 0.16 |
| HORVD | *Hordeum vulgare var. distichum* | 4 | 0.16 |
| MAHA5 | *Mahella australiensis (strain DSM 15567 / CIP 107919 / 50-1 BON)* | 4 | 0.16 |
| MONDO | *Monodelphis domestica* | 4 | 0.16 |
| NEMVE | *Nematostella vectensis* | 4 | 0.16 |
| PERM5 | *Perkinsus marinus (strain ATCC 50983 / TXsc)* | 4 | 0.16 |

| | | | |
|---|---|---|---|
| PHYPA | *Physcomitrella patens subsp. patens* | 4 | 0.16 |
| POLPA | *Polysphondylium pallidum* | 4 | 0.16 |
| SHIDY | *Shigella dysenteriae* | 4 | 0.16 |
| THEET | *Thermoanaerobacter ethanolicus* | 4 | 0.16 |
| THEOJ | *Thermosediminibacter oceani (strain ATCC BAA-1034 / DSM 16646 / JW/IW-1228P)* | 4 | 0.16 |
| XENTR | *Xenopus tropicalis* | 4 | 0.16 |
| 9ASCI | *Ascidiacea* | 3 | 0.12 |
| 9MAXI | *Maxillopoda* | 3 | 0.12 |
| 9PERO | *Percoidei* | 3 | 0.12 |
| 9VEST | *Vetigastropoda* | 3 | 0.12 |
| AEDAE | *Aedes aegypti* | 3 | 0.12 |
| BACSU | *Bacillus subtilis* | 3 | 0.12 |
| BOVIN | *Bos taurus* | 3 | 0.12 |
| CIOIN | *Ciona intestinalis* | 3 | 0.12 |
| CLOC7 | *Clostridium cellulovorans (strain ATCC 35296 / DSM 3052 / OCM 3 / 743B)* | 3 | 0.12 |
| CLOK1 | *Clostridium kluyveri (strain NBRC 12016)* | 3 | 0.12 |
| CLOK5 | *Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680)* | 3 | 0.12 |
| CULQU | *Culex quinquefasciatus* | 3 | 0.12 |
| DICPU | *Dictyostelium purpureum* | 3 | 0.12 |
| HALDV | *Haliotis diversicolor* | 3 | 0.12 |
| HELAM | *Helicoverpa armigera* | 3 | 0.12 |
| HELER | *Heliocidaris erythrogramma* | 3 | 0.12 |
| HELTB | *Heliocidaris tuberculata* | 3 | 0.12 |
| HIRME | *Hirudo medicinalis* | 3 | 0.12 |
| HORSE | *Equus caballus* | 3 | 0.12 |
| LIMPO | *Limulus polyphemus* | 3 | 0.12 |
| LISIN | *Listeria innocua* | 3 | 0.12 |
| MAIZE | *Zea mays* | 3 | 0.12 |
| METSW | *Methanobacterium sp. (strain SWAN-1)* | 3 | 0.12 |
| MOLOC | *Molgula oculata* | 3 | 0.12 |
| PANTR | *Pan troglodytes* | 3 | 0.12 |
| SALSA | *Salmo salar* | 3 | 0.12 |
| SOLBR | *Solanum brevidens* | 3 | 0.12 |
| TAKRU | *Takifugu rubripes* | 3 | 0.12 |
| TEPAE | *Tepidanaerobacter acetatoxydans (strain DSM 21804 / JCM 16047 / Re1)* | 3 | 0.12 |
| THEXL | *Thermoanaerobacterium xylanolyticum (strain ATCC 49914 / DSM 7097 / LX-11)* | 3 | 0.12 |
| RAT | *RAT* | 2 | 0.08 |
| 9BASI | *Basidiomycota* | 2 | 0.08 |
| 9BIVA | *Bivalvia* | 2 | 0.08 |
| 9CNID | *Cnidaria* | 2 | 0.08 |
| 9ESCH | *Escherichia* | 2 | 0.08 |
| 9HYPO | *Hypocreales* | 2 | 0.08 |
| 9SOLN | *Solanum* | 2 | 0.08 |
| ACEAZ | *Acetohalobium arabaticum (strain ATCC 49924 / DSM 5501 / Z-7288)* | 2 | 0.08 |
| AILME | *Ailuropoda melanoleuca* | 2 | 0.08 |
| ANOGA | *Anopheles gambiae* | 2 | 0.08 |
| ANTPE | *Antheraea pernyi* | 2 | 0.08 |
| BACPN | *Bacillus subtilis subsp. spizizenii* | 2 | 0.08 |
| BACT2 | *Bacillus tusciae (strain DSM 2912 / NBRC 15312 / T2)* | 2 | 0.08 |
| BATDJ | *Batrachochytrium dendrobatidis (strain JAM81 / FGSC 10211)* | 2 | 0.08 |
| BOMMO | *Bombyx mori* | 2 | 0.08 |
| CLOAB | *Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787)* | 2 | 0.08 |
| CLOB0 | *Clostridium botulinum (strain H04402 065 / Type A5)* | 2 | 0.08 |
| CLOB1 | *Clostridium botulinum (strain ATCC 19397 / Type A)* | 2 | 0.08 |
| CLOB2 | *Clostridium botulinum (strain 230613 / Type F)* | 2 | 0.08 |
| CLOB6 | *Clostridium botulinum (strain 657 / Type Ba4)* | 2 | 0.08 |
| CLOBH | *Clostridium botulinum (strain Hall / ATCC 3502 / NCTC 13319 / Type A)* | 2 | 0.08 |
| CLOBJ | *Clostridium botulinum (strain Kyoto / Type A2)* | 2 | 0.08 |
| CLOBL | *Clostridium botulinum (strain Langeland / NCTC 10281 / Type F)* | 2 | 0.08 |
| CLOBM | *Clostridium botulinum (strain Loch Maree / Type A3)* | 2 | 0.08 |
| CLOBU | *Clostridium butyricum* | 2 | 0.08 |
| CLONN | *Clostridium novyi (strain NT)* | 2 | 0.08 |
| CRIGR | *Cricetulus griseus* | 2 | 0.08 |
| DESAP | *Desulforudis audaxviator (strain MP104C)* | 2 | 0.08 |
| DESK7 | *Desulfotomaculum kuznetsovii (strain DSM 6115 / VKM B-1805 / 17)* | 2 | 0.08 |
| DESRL | *Desulfotomaculum ruminis (strain ATCC 23193 / DSM 2154 / NCIB 8452 / DL)* | 2 | 0.08 |
| DICLA | *Dicentrarchus labrax* | 2 | 0.08 |
| DROER | *Drosophila erecta* | 2 | 0.08 |
| DROGR | *Drosophila grimshawi* | 2 | 0.08 |
| DROME | *Drosophila melanogaster* | 2 | 0.08 |
| DROMO | *Drosophila mojavensis* | 2 | 0.08 |
| DROPE | *Drosophila persimilis* | 2 | 0.08 |
| DROPS | *Drosophila pseudoobscura pseudoobscura* | 2 | 0.08 |
| DROSE | *Drosophila sechellia* | 2 | 0.08 |
| DROSI | *Drosophila simulans* | 2 | 0.08 |
| DROVI | *Drosophila virilis* | 2 | 0.08 |
| DROWI | *Drosophila willistoni* | 2 | 0.08 |
| DROYA | *Drosophila yakuba* | 2 | 0.08 |
| ESCFE | *Escherichia fergusonii* | 2 | 0.08 |
| HELZE | *Helicoverpa zea* | 2 | 0.08 |
| HETGA | *Heterocephalus glaber* | 2 | 0.08 |
| HEVBR | *Hevea brasiliensis* | 2 | 0.08 |
| HOLGL | *Holothuria glaberrima* | 2 | 0.08 |
| IXORI | *Ixodes ricinus* | 2 | 0.08 |

| | | | |
|---|---|---|---|
| KRYMA | *Kryptolebias marmoratus* | 2 | 0.08 |
| LISSE | *Listeria seeligeri* | 2 | 0.08 |
| LUMTE | *Lumbricus terrestris* | 2 | 0.08 |
| LYTPI | *Lytechinus pictus* | 2 | 0.08 |
| MEISD | *Meiothermus silvanus (strain ATCC 700542 / DSM 9946 / VI-R2)* | 2 | 0.08 |
| MELGA | *Meleagris gallopavo* | 2 | 0.08 |
| METBF | *Methanosarcina barkeri (strain Fusaro / DSM 804)* | 2 | 0.08 |
| METFV | *Methanothermus fervidus (strain ATCC 43054 / DSM 2088 / JCM 10308 / V24 S)* | 2 | 0.08 |
| METST | *Methanosphaera stadtmanae (strain DSM 3091)* | 2 | 0.08 |
| MIZYE | *Mizuhopecten yessoensis* | 2 | 0.08 |
| MONAL | *Monopterus albus* | 2 | 0.08 |
| MOOTA | *Moorella thermoacetica (strain ATCC 39073)* | 2 | 0.08 |
| NATTJ | *Natranaerobius thermophilus (strain ATCC BAA-1301 / DSM 18059 / JW/NM-WN-LF)* | 2 | 0.08 |
| OIKDI | *Oikopleura dioica* | 2 | 0.08 |
| ONCMY | *Oncorhynchus mykiss* | 2 | 0.08 |
| ONCTS | *Oncorhynchus tschawytscha* | 2 | 0.08 |
| ORNAN | *Ornithorhynchus anatinus* | 2 | 0.08 |
| ORNMO | *Ornithodoros moubata* | 2 | 0.08 |
| PAGMA | *Pagrus major* | 2 | 0.08 |
| PETMO | *Petrotoga mobilis (strain DSM 10674 / SJ95)* | 2 | 0.08 |
| PISOC | *Pisaster ochraceus* | 2 | 0.08 |
| PODCA | *Podocoryne carnea* | 2 | 0.08 |
| POERE | *Poecilia reticulata* | 2 | 0.08 |
| PONAB | *Pongo abelii* | 2 | 0.08 |
| PRUDU | *Prunus dulcis* | 2 | 0.08 |
| RANLE | *Rana lessonae* | 2 | 0.08 |
| SACKO | *Saccoglossus kowalevskii* | 2 | 0.08 |
| SARHA | *Sarcophilus harrisii* | 2 | 0.08 |
| SHEEP | *Ovis aries* | 2 | 0.08 |
| SHIBO | *Shigella boydii* | 2 | 0.08 |
| SPORE | *Sporisorium reilianum (strain SRZ2)* | 2 | 0.08 |
| STRFN | *Strongylocentrotus franciscanus* | 2 | 0.08 |
| STYPL | *Styela plicata* | 2 | 0.08 |
| TETNG | *Tetraodon nigroviridis* | 2 | 0.08 |
| THEBF | *Thermoanaerobacter brockii subsp. finnii (strain ATCC 43586 / DSM 3389 / AKO-1)* | 2 | 0.08 |
| THEIA | *Thermoanaerobacter italicus (strain DSM 9252 / Ab9)* | 2 | 0.08 |
| THEM3 | *Thermoanaerobacter mathranii (strain DSM 11426 / CIP 108742 / A3)* | 2 | 0.08 |
| THEMA | *Thermotoga maritima* | 2 | 0.08 |
| THEP3 | *Thermoanaerobacter pseudethanolicus (strain ATCC 33223 / 39E)* | 2 | 0.08 |
| THEPJ | *Thermincola potens (strain JR)* | 2 | 0.08 |
| THEPX | *Thermoanaerobacter sp. (strain X514)* | 2 | 0.08 |
| THESX | *Thermoanaerobacter sp. (strain X513)* | 2 | 0.08 |
| TRIAD | *Trichoplax adhaerens* | 2 | 0.08 |
| TRISP | *Trichinella spiralis* | 2 | 0.08 |
| TRIVU | *Trichosurus vulpecula* | 2 | 0.08 |
| USTMA | *Ustilago maydis (strain 521 / FGSC 9021)* | 2 | 0.08 |
| 9BACL | *Bacillales* | 1 | 0.04 |
| 9BILA | *Bilateria* | 1 | 0.04 |
| 9CHLR | *Chloroflexi* | 1 | 0.04 |
| 9CRYP | *Cryptophyta* | 1 | 0.04 |
| 9DIPT | *Diptera* | 1 | 0.04 |
| 9ENTE | *Enterococcaceae* | 1 | 0.04 |
| 9EUCA | *Eucarida* | 1 | 0.04 |
| 9EURY | *Euryarchaeota* | 1 | 0.04 |
| 9FUNG | *Fungi* | 1 | 0.04 |
| 9HEMI | *Hemiptera* | 1 | 0.04 |
| 9HYME | *Hymenoptera* | 1 | 0.04 |
| 9LABR | *Labroidei* | 1 | 0.04 |
| 9LIST | *Listeriaceae* | 1 | 0.04 |
| 9MOLL | *Mollusca* | 1 | 0.04 |
| 9NEOB | *Neobatrachia* | 1 | 0.04 |
| 9PEZI | *Pezizomycotina* | 1 | 0.04 |
| 9PIPI | *Pipidae* | 1 | 0.04 |
| 9POAL | *Poales* | 1 | 0.04 |
| 9PROT | *Proteobacteria* | 1 | 0.04 |
| 9RHOB | *Rhodobacterales* | 1 | 0.04 |
| 9SALA | *Salamandroidea* | 1 | 0.04 |
| 9SPHN | *Sphingomonadales* | 1 | 0.04 |
| 9THEM | *Thermotogaceae* | 1 | 0.04 |
| 9UROC | *Urochordata* | 1 | 0.04 |
| ACACA | *Acanthamoeba castellanii* | 1 | 0.04 |
| ACASC | *Acanthopagrus schlegelii* | 1 | 0.04 |
| ACIFV | *Acidaminococcus fermentans (strain ATCC 25085 / DSM 20731 / VR4)* | 1 | 0.04 |
| ACIHW | *Acidianus hospitalis (strain W1)* | 1 | 0.04 |
| ACITR | *Acipenser transmontanus* | 1 | 0.04 |
| ACREC | *Acromyrmex echinatior* | 1 | 0.04 |
| ACYPI | *Acyrthosiphon pisum* | 1 | 0.04 |
| AEDAL | *Aedes albopictus* | 1 | 0.04 |
| AGRIP | *Agrotis ipsilon* | 1 | 0.04 |
| AIPPU | *Aiptasia pulchella* | 1 | 0.04 |
| AJEC8 | *Ajellomyces capsulata (strain H88)* | 1 | 0.04 |
| AJECG | *Ajellomyces capsulata (strain G186AR / H82 / ATCC MYA-2454 / RMSCC 2432)* | 1 | 0.04 |
| AJECH | *Ajellomyces capsulata (strain H143)* | 1 | 0.04 |
| AJECN | *Ajellomyces capsulata (strain NAm1 / WU24)* | 1 | 0.04 |
| AJEDA | *Ajellomyces dermatitidis (strain ATCC 18188 / CBS 674.68)* | 1 | 0.04 |
| AJEDR | *Ajellomyces dermatitidis (strain ER-3 / ATCC MYA-2586)* | 1 | 0.04 |

| | | | |
|---|---|---|---|
| AJEDS | *Ajellomyces dermatitidis (strain SLH14081)* | 1 | 0.04 |
| ALIAD | *Alicyclobacillus acidocaldarius subsp. acidocaldarius (strain ATCC 27009 / DSM 446 / 104-1A)* | 1 | 0.04 |
| ALIAT | *Alicyclobacillus acidocaldarius (strain Tc-4-1)* | 1 | 0.04 |
| ALKMQ | *Alkaliphilus metalliredigens (strain QYMF)* | 1 | 0.04 |
| ALKOO | *Alkaliphilus oremlandii (strain OhILAs)* | 1 | 0.04 |
| AMOPA | *Amoebidium parasiticum* | 1 | 0.04 |
| ANAPL | *Anas platyrhynchos* | 1 | 0.04 |
| ANATU | *Anaerolinea thermophila (strain DSM 14523 / JCM 11388 / NBRC 100420 / UNI-1)* | 1 | 0.04 |
| ANDDA | *Andrias davidianus* | 1 | 0.04 |
| ANOCA | *Anolis carolinensis* | 1 | 0.04 |
| ANODA | *Anopheles darlingi* | 1 | 0.04 |
| ANSAN | *Anser anser anser* | 1 | 0.04 |
| APICC | *Apis cerana cerana* | 1 | 0.04 |
| APLCA | *Aplysia californica* | 1 | 0.04 |
| ARALL | *Arabidopsis lyrata subsp. lyrata* | 1 | 0.04 |
| ARATH | *Arabidopsis thaliana* | 1 | 0.04 |
| ARGTR | *Argiope trifasciata* | 1 | 0.04 |
| ARTBC | *Arthroderma benhamiae (strain ATCC MYA-4681 / CBS 112371)* | 1 | 0.04 |
| ARTGP | *Arthroderma gypseum (strain ATCC MYA-4604 / CBS 118893)* | 1 | 0.04 |
| ARTOA | *Arthrobotrys oligospora (strain ATCC 24927 / CBS 115.81 / DSM 1491)* | 1 | 0.04 |
| ARTOC | *Arthroderma otae (strain ATCC MYA-4605 / CBS 113480)* | 1 | 0.04 |
| ARTSS | *Arthromitus sp. (strain SFB-mouse-Japan)* | 1 | 0.04 |
| ARTSX | *Artemia sp.* | 1 | 0.04 |
| ASCSU | *Ascaris suum* | 1 | 0.04 |
| ASPCL | *Aspergillus clavatus (strain ATCC 1007 / CBS 513.65 / DSM 816 / NCTC 3887 / NRRL 1)* | 1 | 0.04 |
| ASPFC | *Neosartorya fumigata (strain CEA10 / CBS 144.89 / FGSC A1163)* | 1 | 0.04 |
| ASPFN | *Aspergillus flavus (strain ATCC 200026 / FGSC A1120 / NRRL 3357 / JCM 12722 / SRRC 167)* | 1 | 0.04 |
| ASPFU | *Neosartorya fumigata (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100)* | 1 | 0.04 |
| ASPKW | *Aspergillus kawachii (strain NBRC 4308)* | 1 | 0.04 |
| ASPNA | *Aspergillus niger (strain ATCC 1015 / CBS 113.46 / FGSC A1144 / LSHB Ac4 / NCTC 3858a / NRRL 328 / USDA 3528.7)* | 1 | 0.04 |
| ASPNC | *Aspergillus niger (strain CBS 513.88 / FGSC A1513)* | 1 | 0.04 |
| ASPTN | *Aspergillus terreus (strain NIH 2624 / FGSC A1156)* | 1 | 0.04 |
| PIG | *PIG* | 1 | 0.04 |
| BACA1 | *Bacillus atrophaeus (strain 1942)* | 1 | 0.04 |
| BACA2 | *Bacillus amyloliquefaciens (strain FZB42)* | 1 | 0.04 |
| BACAS | *Bacillus amyloliquefaciens (strain ATCC 23350 / DSM 7 / BCRC 11601 / NBRC 15535 / NRRL B-14393)* | 1 | 0.04 |
| BACF6 | *Bacteroides fragilis (strain 638R)* | 1 | 0.04 |
| BACFG | *Bacteroides fragilis* | 1 | 0.04 |
| BACFN | *Bacteroides fragilis (strain ATCC 25285 / NCTC 9343)* | 1 | 0.04 |
| BACFR | *Bacteroides fragilis (strain YCH46)* | 1 | 0.04 |
| BACLD | *Bacillus licheniformis (strain DSM 13 / ATCC 14580)* | 1 | 0.04 |
| BACLI | *Bacillus licheniformis* | 1 | 0.04 |
| BACNA | *Bacillus subtilis subsp. natto* | 1 | 0.04 |
| BACPZ | *Bacillus subtilis subsp. spizizenii (strain ATCC 23059 / NRRL B-14472 / W23)* | 1 | 0.04 |
| BACSE | *Bacteroides stercoris* | 1 | 0.04 |
| BACST | *Bacillus subtilis (strain BSn5)* | 1 | 0.04 |
| BACT6 | *Bacteroides helcogenes (strain ATCC 35417 / DSM 20613 / JCM 6297 / P 36-108)* | 1 | 0.04 |
| BACTN | *Bacteroides thetaiotaomicron (strain ATCC 29148 / DSM 2079 / NCTC 10582 / E50 / VPI-5482)* | 1 | 0.04 |
| BACTR | *Bacillus thermoglucosidasius* | 1 | 0.04 |
| BACUN | *Bacteroides uniformis* | 1 | 0.04 |
| BEABA | *Beauveria bassiana* | 1 | 0.04 |
| BIOAL | *Biomphalaria alexandrina* | 1 | 0.04 |
| BIOGL | *Biomphalaria glabrata* | 1 | 0.04 |
| BIOOB | *Biomphalaria obstructa* | 1 | 0.04 |
| BIOPF | *Biomphalaria pfeifferi* | 1 | 0.04 |
| BIOTE | *Biomphalaria tenagophila* | 1 | 0.04 |
| BOMMA | *Bombyx mandarina* | 1 | 0.04 |
| BOOMI | *Boophilus microplus* | 1 | 0.04 |
| BOTF4 | *Botryotinia fuckeliana (strain T4)* | 1 | 0.04 |
| BOTFB | *Botryotinia fuckeliana (strain B05.10)* | 1 | 0.04 |
| BOTFU | *Botryotinia fuckeliana* | 1 | 0.04 |
| BOTSH | *Botryllus schlosseri* | 1 | 0.04 |
| BRABE | *Branchiostoma belcheri* | 1 | 0.04 |
| BRUMA | *Brugia malayi* | 1 | 0.04 |
| BUTFI | *Butyrivibrio fibrisolvens* | 1 | 0.04 |
| BUTPB | *Butyrivibrio proteoclasticus (strain ATCC 51982 / DSM 14932 / B316)* | 1 | 0.04 |
| CALMQ | *Caldivirga maquilingensis (strain ATCC 700844 / DSMZ 13496 / JCM 10307 / IC-167)* | 1 | 0.04 |
| CALSA | *Caldocellum saccharolyticum* | 1 | 0.04 |
| CALSI | *Callinectes sapidus* | 1 | 0.04 |
| CAMDR | *Camelus dromedarius* | 1 | 0.04 |
| CAMFO | *Camponotus floridanus* | 1 | 0.04 |
| CARAU | *Carassius auratus* | 1 | 0.04 |
| CAVPO | *Cavia porcellus* | 1 | 0.04 |
| CEPAC | *Cephalosporium acremonium* | 1 | 0.04 |
| CERPY | *Cercopithecus pygerythrus* | 1 | 0.04 |
| CHAGB | *Chaetomium globosum (strain ATCC 6205 / CBS 148.51 / DSM 1962 / NBRC 6347 / NRRL 1970)* | 1 | 0.04 |
| CHATD | *Chaetomium thermophilum (strain DSM 1495 / CBS 144.50 / IMI 039719)* | 1 | 0.04 |
| CHLAA | *Chloroflexus aurantiacus (strain ATCC 29366 / DSM 635 / J-10-fl)* | 1 | 0.04 |
| CHLAD | *Chloroflexus aggregans (strain MD-66 / DSM 9485)* | 1 | 0.04 |
| CHLAE | *Chlorocebus aethiops* | 1 | 0.04 |

| | | | |
|---|---|---|---|
| CHLSY | *Chloroflexus aurantiacus (strain ATCC 29364 / DSM 637 / Y-400-fl)* | 1 | 0.04 |
| CHLVA | *Chlorella variabilis* | 1 | 0.04 |
| CLAPU | *Claviceps purpurea* | 1 | 0.04 |
| CLOAE | *Clostridium acetobutylicum (strain EA 2018)* | 1 | 0.04 |
| CLOB8 | *Clostridium beijerinckii (strain ATCC 51743 / NCIMB 8052)* | 1 | 0.04 |
| CLOBA | *Clostridium botulinum (strain Alaska E43 / Type E3)* | 1 | 0.04 |
| CLOBB | *Clostridium botulinum (strain Eklund 17B / Type B)* | 1 | 0.04 |
| CLOBK | *Clostridium botulinum (strain Okra / Type B1)* | 1 | 0.04 |
| CLOD6 | *Clostridium difficile (strain 630)* | 1 | 0.04 |
| CLODC | *Clostridium difficile (strain CD196)* | 1 | 0.04 |
| CLODR | *Clostridium difficile (strain R20291)* | 1 | 0.04 |
| CLOLD | *Clostridium ljungdahlii (strain ATCC 55383 / DSM 13528 / PETC)* | 1 | 0.04 |
| CLOP1 | *Clostridium perfringens (strain ATCC 13124 / NCTC 8237 / Type A)* | 1 | 0.04 |
| CLOPS | *Clostridium perfringens (strain SM101 / Type A)* | 1 | 0.04 |
| CLOSC | *Clostridium saccharolyticum* | 1 | 0.04 |
| CLOSD | *Clostridium sticklandii (strain ATCC 12662 / DSM 519 / JCM 1433 / NCIB 10654)* | 1 | 0.04 |
| CLOSR | *Clostridium stercorarium* | 1 | 0.04 |
| CLOSS | *Clostridium sp. (strain SY8519)* | 1 | 0.04 |
| CLOTE | *Clostridium tetani* | 1 | 0.04 |
| COCP7 | *Coccidioides posadasii (strain C735)* | 1 | 0.04 |
| COCPS | *Coccidioides posadasii (strain RMSCC 757 / Silveira)* | 1 | 0.04 |
| COLGM | *Colletotrichum graminicola (strain M1.001 / M2 / FGSC 10212)* | 1 | 0.04 |
| CORMC | *Corvus macrorhynchos* | 1 | 0.04 |
| CORMM | *Cordyceps militaris (strain CM01)* | 1 | 0.04 |
| CRAAR | *Crassostrea ariakensis* | 1 | 0.04 |
| CRISP | *Cricetidae sp.* | 1 | 0.04 |
| CTEID | *Ctenopharyngodon idella* | 1 | 0.04 |
| CULPP | *Culex pipiens pipiens* | 1 | 0.04 |
| CYAME | *Cyanidioschyzon merolae* | 1 | 0.04 |
| CYPCA | *Cyprinus carpio* | 1 | 0.04 |
| DANPL | *Danaus plexippus* | 1 | 0.04 |
| DEIDV | *Deinococcus deserti (strain VCD115 / DSM 17065 / LMG 22923)* | 1 | 0.04 |
| DEIML | *Deinococcus maricopensis (strain DSM 21211 / LMG 22137 / NRRL B-23946 / LB-34)* | 1 | 0.04 |
| DEIRA | *Deinococcus radiodurans (strain ATCC 13939 / DSM 20539 / JCM 16871 / LMG 4051 / NBRC 15346 / NCIMB 9279 / R1 / VKM B-1422)* | 1 | 0.04 |
| DENA2 | *Denitrovibrio acetiphilus (strain DSM 12809 / N2460)* | 1 | 0.04 |
| DERVA | *Dermacentor variabilis* | 1 | 0.04 |
| DESAA | *Desulfatibacillum alkenivorans (strain AK-01)* | 1 | 0.04 |
| DESAO | *Desulfovibrio aespoeensis (strain ATCC 700646 / DSM 10631 / Aspo-2)* | 1 | 0.04 |
| DESHD | *Desulfitobacterium hafniense (strain DCB-2 / DSM 10664)* | 1 | 0.04 |
| DESHY | *Desulfitobacterium hafniense (strain Y51)* | 1 | 0.04 |
| DESIS | *Desulfurispirillum indicum (strain ATCC BAA-1389 / S5)* | 1 | 0.04 |
| DESOH | *Desulfococcus oleovorans (strain DSM 6200 / Hxd3)* | 1 | 0.04 |
| DICT6 | *Dictyoglomus thermophilum (strain ATCC 35947 / DSM 3960 / H-6-12)* | 1 | 0.04 |
| DICTD | *Dictyoglomus turgidum (strain Z-1310 / DSM 6724)* | 1 | 0.04 |
| DOLTE | *Dolomedes tenebrosus* | 1 | 0.04 |
| DROAN | *Drosophila ananassae* | 1 | 0.04 |
| DRONO | *Dromaius novaehollandiae* | 1 | 0.04 |
| ECHGR | *Echinococcus granulosus* | 1 | 0.04 |
| ECO10 | *Escherichia coli O103:H2 (strain 12009 / EHEC)* | 1 | 0.04 |
| ECO1A | *Escherichia coli O111:H- (strain 11128 / EHEC)* | 1 | 0.04 |
| ECO24 | *Escherichia coli O139:H28 (strain E24377A / ETEC)* | 1 | 0.04 |
| ECO26 | *Escherichia coli O26:H11 (strain 11368 / EHEC)* | 1 | 0.04 |
| ECO44 | *Escherichia coli O44:H18 (strain 042 / EAEC)* | 1 | 0.04 |
| ECO45 | *Escherichia coli O45:K1 (strain S88 / ExPEC)* | 1 | 0.04 |
| ECO55 | *Escherichia coli (strain 55989 / EAEC)* | 1 | 0.04 |
| ECO5E | *Escherichia coli O157:H7 (strain EC4115 / EHEC)* | 1 | 0.04 |
| ECO5T | *Escherichia coli O157:H7 (strain TW14359 / EHEC)* | 1 | 0.04 |
| ECO7I | *Escherichia coli O7:K1 (strain IAI39 / ExPEC)* | 1 | 0.04 |
| ECO81 | *Escherichia coli O81 (strain ED1a)* | 1 | 0.04 |
| ECO8A | *Escherichia coli O8 (strain IAI1)* | 1 | 0.04 |
| ECO8N | *Escherichia coli O83:H1 (strain NRG 857C / AIEC)* | 1 | 0.04 |
| ECOAB | *Escherichia coli OR:K5:H- (strain ABU 83972)* | 1 | 0.04 |
| ECOBD | *Escherichia coli (strain B / BL21-DE3)* | 1 | 0.04 |
| ECOBR | *Escherichia coli (strain B / REL606)* | 1 | 0.04 |
| ECOBW | *Escherichia coli (strain K12 / MC4100 / BW2952)* | 1 | 0.04 |
| ECOCB | *Escherichia coli O55:H7 (strain CB9615 / EPEC)* | 1 | 0.04 |
| ECOD1 | *Escherichia coli (strain ATCC 33849 / DSM 4235 / NCIB 12045 / K12 / DH1)* | 1 | 0.04 |
| ECODH | *Escherichia coli (strain K12 / DH10B)* | 1 | 0.04 |
| ECOH1 | *Escherichia coli O78:H11 (strain H10407 / ETEC)* | 1 | 0.04 |
| ECOHS | *Escherichia coli O9:H4 (strain HS)* | 1 | 0.04 |
| ECOK1 | *Escherichia coli O1:K1 / APEC* | 1 | 0.04 |
| ECOKI | *Escherichia coli O18:K1:H7 (strain IHE3034 / ExPEC)* | 1 | 0.04 |
| ECOKO | *Escherichia coli (strain ATCC 55124 / KO11)* | 1 | 0.04 |
| ECOL6 | *Escherichia coli O6* | 1 | 0.04 |
| ECOLC | *Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks)* | 1 | 0.04 |
| ECOLI | *Escherichia coli (strain K12)* | 1 | 0.04 |
| ECOLU | *Escherichia coli O17:K52:H18 (strain UMN026 / ExPEC)* | 1 | 0.04 |
| ECOLW | *Escherichia coli (strain ATCC 9637 / CCM 2024 / DSM 1116 / NCIMB 8666 / NRRL B-766 / W)* | 1 | 0.04 |
| ECOS5 | *Escherichia coli O150:H5 (strain SE15)* | 1 | 0.04 |
| ECOSE | *Escherichia coli (strain SE11)* | 1 | 0.04 |
| ECOSM | *Escherichia coli (strain SMS-3-5 / SECEC)* | 1 | 0.04 |
| ECOUM | *Escherichia coli (strain UM146)* | 1 | 0.04 |
| ECOUT | *Escherichia coli (strain UTI89 / UPEC)* | 1 | 0.04 |

| | | | |
|---|---|---|---|
| ECTSI | *Ectocarpus siliculosus* | 1 | 0.04 |
| EMENI | *Emericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139)* | 1 | 0.04 |
| EMIHU | *Emiliania huxleyi* | 1 | 0.04 |
| ENTCL | *Enterobacter cloacae* | 1 | 0.04 |
| ENTCS | *Enterobacter cloacae (strain SCF1)* | 1 | 0.04 |
| EPICO | *Epinephelus coioides* | 1 | 0.04 |
| ESCF3 | *Escherichia fergusonii (strain ATCC 35469 / DSM 13698 / CDC 0568-73)* | 1 | 0.04 |
| ETHHY | *Ethanoligenens harbinense (strain DSM 18485 / JCM 12961 / CGMCC 1.5033 / YUAN-3)* | 1 | 0.04 |
| EUBGL | *Eubalaena glacialis* | 1 | 0.04 |
| EUBLK | *Eubacterium limosum (strain KIST612)* | 1 | 0.04 |
| EUBR3 | *Eubacterium rectale (strain ATCC 33656 / VPI 0990)* | 1 | 0.04 |
| EUBSP | *Eubacterium sp. (strain VPI 12708)* | 1 | 0.04 |
| EXODE | *Exophiala dermatitidis* | 1 | 0.04 |
| FENCH | *Fenneropenaeus chinensis* | 1 | 0.04 |
| FERBD | *Ferrimonas balearica (strain DSM 9799 / CCM 4581 / PAT)* | 1 | 0.04 |
| FILAD | *Filifactor alocis (strain ATCC 35896 / D40 B5)* | 1 | 0.04 |
| FUSOF | *Fusarium oxysporum (strain Fo5176)* | 1 | 0.04 |
| G_RAT | *#N/A* | 1 | 0.04 |
| GAEGA | *Gaeumannomyces graminis var. avenae* | 1 | 0.04 |
| GEOMG | *Geobacter metallireducens (strain GS-15 / ATCC 53774 / DSM 7210)* | 1 | 0.04 |
| GEOS0 | *Geobacillus sp. (strain Y4.1MC1)* | 1 | 0.04 |
| GEOS8 | *Geobacter sp. (strain M18)* | 1 | 0.04 |
| GEOSF | *Geobacter sp. (strain FRC-32)* | 1 | 0.04 |
| GEOSK | *Geobacter sulfurreducens (strain DL-1 / KN400)* | 1 | 0.04 |
| GEOSL | *Geobacter sulfurreducens (strain ATCC 51573 / DSM 12127 / PCA)* | 1 | 0.04 |
| GLOIN | *Glomus intraradices* | 1 | 0.04 |
| GLOMM | *Glossina morsitans morsitans* | 1 | 0.04 |
| GRYBI | *Gryllus bimaculatus* | 1 | 0.04 |
| GUITH | *Guillardia theta* | 1 | 0.04 |
| HAELO | *Haemaphysalis longicornis* | 1 | 0.04 |
| HALDH | *Haliotis discus hannai* | 1 | 0.04 |
| HALPG | *Halanaerobium praevalens (strain ATCC 33744 / DSM 2228 / GSL)* | 1 | 0.04 |
| HALSL | *Halanaerobium sp. (strain sapolanicus)* | 1 | 0.04 |
| HALTU | *Haliotis tuberculata* | 1 | 0.04 |
| HARSA | *Harpegnathos saltator* | 1 | 0.04 |
| HELMI | *Heliobacterium modesticaldum (strain ATCC 51547 / Ice1)* | 1 | 0.04 |
| HELMO | *Heliobacillus mobilis* | 1 | 0.04 |
| HELTI | *Helisoma trivolvis* | 1 | 0.04 |
| HELTR | *Helobdella triserialis* | 1 | 0.04 |
| HELVI | *Heliothis virescens* | 1 | 0.04 |
| HERA2 | *Herpetosiphon aurantiacus (strain ATCC 23779 / DSM 785)* | 1 | 0.04 |
| HIRBI | *Hirschia baltica (strain ATCC 49814 / DSM 5838 / IFAM 1418)* | 1 | 0.04 |
| HOMAM | *Homarus americanus* | 1 | 0.04 |
| HOMGA | *Homarus gammarus* | 1 | 0.04 |
| HOMVI | *Homalodisca vitripennis* | 1 | 0.04 |
| HUMGT | *Humicola grisea var. thermoidea* | 1 | 0.04 |
| HYDEC | *Hydractinia echinata* | 1 | 0.04 |
| HYDVU | *Hydra vulgaris* | 1 | 0.04 |
| HYPJQ | *Hypocrea jecorina (strain QM6a)* | 1 | 0.04 |
| HYPMO | *Hypophthalmichthys molitrix* | 1 | 0.04 |
| ISOGA | *Isochrysis galbana* | 1 | 0.04 |
| IXOPE | *Ixodes persulcatus* | 1 | 0.04 |
| IXOSC | *Ixodes scapularis* | 1 | 0.04 |
| KLEP3 | *Klebsiella pneumoniae (strain 342)* | 1 | 0.04 |
| KLEP7 | *Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578)* | 1 | 0.04 |
| KLEPR | *Klebsiella pneumoniae subsp. rhinoscleromatis* | 1 | 0.04 |
| KLEVT | *Klebsiella variicola (strain At-22)* | 1 | 0.04 |
| LACBS | *Laccaria bicolor (strain S238N-H82 / ATCC MYA-4686)* | 1 | 0.04 |
| LAGAC | *Lagenorhynchus acutus* | 1 | 0.04 |
| LAMGL | *Lama glama* | 1 | 0.04 |
| LAMJA | *Lampetra japonica* | 1 | 0.04 |
| LARCR | *Larimichthys crocea* | 1 | 0.04 |
| LATJA | *Lateolabrax japonicus* | 1 | 0.04 |
| LEUER | *Leucoraja erinacea* | 1 | 0.04 |
| LIPBO | *Liposcelis bostrychophila* | 1 | 0.04 |
| LIPST | *Lipomyces starkeyi* | 1 | 0.04 |
| LISIP | *Listeria ivanovii (strain ATCC BAA-678 / PAM 55)* | 1 | 0.04 |
| LISIV | *Listeria ivanovii* | 1 | 0.04 |
| LISM1 | *Listeria monocytogenes serotype 1/2a (strain 08-5578)* | 1 | 0.04 |
| LISM2 | *Listeria monocytogenes serotype 1/2a (strain 08-5923)* | 1 | 0.04 |
| LISMC | *Listeria monocytogenes serotype 4b (strain Clip81459)* | 1 | 0.04 |
| LISMF | *Listeria monocytogenes serotype 4b (strain F2365)* | 1 | 0.04 |
| LISMH | *Listeria monocytogenes serotype 4a (strain HCC23)* | 1 | 0.04 |
| LISML | *Listeria monocytogenes serotype 4c (strain L99)* | 1 | 0.04 |
| LISMM | *Listeria monocytogenes serotype 4a (strain M7)* | 1 | 0.04 |
| LISSS | *Listeria seeligeri serovar 1/2b (strain ATCC 35967 / DSM 20751 / CIP 100100 / SLCC 3954)* | 1 | 0.04 |
| LISW6 | *Listeria welshimeri serovar 6b (strain ATCC 35897 / DSM 20650 / SLCC5334)* | 1 | 0.04 |
| LITLI | *Littorina littorea* | 1 | 0.04 |
| LITVA | *Litopenaeus vannamei* | 1 | 0.04 |
| LOALO | *Loa loa* | 1 | 0.04 |
| LOLBL | *Loligo bleekeri* | 1 | 0.04 |
| LOLPE | *Loligo pealeii* | 1 | 0.04 |
| LOXAF | *Loxodonta africana* | 1 | 0.04 |

| | | | |
|---|---|---|---|
| LUMRU | *Lumbricus rubellus* | 1 | 0.04 |
| LYCSI | *Lycosa singoriensis* | 1 | 0.04 |
| LYGLI | *Lygus lineolaris* | 1 | 0.04 |
| LYMDI | *Lymantria dispar* | 1 | 0.04 |
| MACRS | *Macrobrachium rosenbergii* | 1 | 0.04 |
| MAGO7 | *Magnaporthe oryzae (strain 70-15 / ATCC MYA-4617 / FGSC 8958)* | 1 | 0.04 |
| MALGO | *Malassezia globosa (strain ATCC MYA-4612 / CBS 7966)* | 1 | 0.04 |
| MANSE | *Manduca sexta* | 1 | 0.04 |
| MARHT | *Marinithermus hydrothermalis (strain DSM 14884 / JCM 11576 / T1)* | 1 | 0.04 |
| MEGAM | *Megalobrama amblycephala* | 1 | 0.04 |
| MEGEL | *Megasphaera elsdenii* | 1 | 0.04 |
| MEGNO | *Megaptera novaeangliae* | 1 | 0.04 |
| MEIRD | *Meiothermus ruber (strain ATCC 35948 / DSM 1279 / VKM B-1258 / 21)* | 1 | 0.04 |
| MELLP | *Melampsora larici-populina (strain 98AG31 / pathotype 3-4-7)* | 1 | 0.04 |
| MERMT | *Meretrix meretrix* | 1 | 0.04 |
| MERUN | *Meriones unguiculatus* | 1 | 0.04 |
| MESAU | *Mesocricetus auratus* | 1 | 0.04 |
| MESMA | *Mesobuthus martensii* | 1 | 0.04 |
| METAQ | *Metarhizium acridum (strain CQMa 102)* | 1 | 0.04 |
| METAR | *Metarhizium robertsii (strain ARSEF 23 / ATCC MYA-3075)* | 1 | 0.04 |
| METB6 | *Methanoregula boonei (strain 6A8)* | 1 | 0.04 |
| METHJ | *Methanospirillum hungatei (strain JF-1 / DSM 864)* | 1 | 0.04 |
| METIK | *Methanotorris igneus (strain DSM 5666 / JCM 11834 / Kol 5)* | 1 | 0.04 |
| METM7 | *Methanococcus maripaludis (strain C7 / ATCC BAA-1331)* | 1 | 0.04 |
| METMA | *Methanosarcina mazei (strain ATCC BAA-159 / DSM 3647 / Goe1 / Go1 / JCM 11833 / OCM 88)* | 1 | 0.04 |
| METP4 | *Methanoplanus petrolearius (strain DSM 11571 / OCM 486 / SEBR 4847)* | 1 | 0.04 |
| METPE | *Methanosphaerula palustris (strain ATCC BAA-1556 / DSM 19958 / E1-9c)* | 1 | 0.04 |
| METS5 | *Metallosphaera sedula (strain ATCC 51363 / DSM 5348)* | 1 | 0.04 |
| METSL | *Methanobacterium sp. (strain AL-21)* | 1 | 0.04 |
| METTF | *Methanobacterium thermoformicicum* | 1 | 0.04 |
| METTW | *Methanothermobacter thermoautotrophicus (strain Winter)* | 1 | 0.04 |
| METV3 | *Methanococcus voltae (strain ATCC BAA-1334 / A3)* | 1 | 0.04 |
| METWO | *Methanobacterium wolfei* | 1 | 0.04 |
| MICSR | *Micromonas sp. (strain RCC299 / NOUM17)* | 1 | 0.04 |
| MISAN | *Misgurnus anguillicaudatus* | 1 | 0.04 |
| MONBE | *Monosiga brevicollis* | 1 | 0.04 |
| MOOTH | *Moorella thermoacetica* | 1 | 0.04 |
| MUSPF | *Mustela putorius furo* | 1 | 0.04 |
| MUSVI | *Mustela vison* | 1 | 0.04 |
| MYCGM | *Mycosphaerella graminicola (strain CBS 115943 / IPO323)* | 1 | 0.04 |
| MYLPI | *Mylopharyngodon piceus* | 1 | 0.04 |
| MYOLU | *Myotis lucifugus* | 1 | 0.04 |
| MYTGA | *Mytilus galloprovincialis* | 1 | 0.04 |
| MYXFH | *Myxococcus fulvus (strain ATCC BAA-855 / HW-1)* | 1 | 0.04 |
| NECH7 | *Nectria haematococca (strain 77-13-4 / ATCC MYA-4622 / FGSC 9596 / MPVI)* | 1 | 0.04 |
| NEOFI | *Neosartorya fischeri (strain ATCC 1020 / DSM 3700 / FGSC A1164 / NRRL 181)* | 1 | 0.04 |
| NEUCR | *Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987)* | 1 | 0.04 |
| NEUT8 | *Neurospora tetrasperma (strain FGSC 2508 / ATCC MYA-4615 / P0657)* | 1 | 0.04 |
| NEUT9 | *Neurospora tetrasperma (strain FGSC 2509 / P0656)* | 1 | 0.04 |
| NILLU | *Nilaparvata lugens* | 1 | 0.04 |
| OCEP5 | *Oceanithermus profundus (strain DSM 14977 / NBRC 100410 / VKM B-2274 / 506)* | 1 | 0.04 |
| OCTVU | *Octopus vulgaris* | 1 | 0.04 |
| ONCNE | *Oncorhynchus nerka* | 1 | 0.04 |
| OREMO | *Oreochromis mossambicus* | 1 | 0.04 |
| ORYSI | *Oryza sativa subsp. indica* | 1 | 0.04 |
| OSMMO | *Osmerus mordax* | 1 | 0.04 |
| OSTLU | *Ostreococcus lucimarinus (strain CCE9901)* | 1 | 0.04 |
| OSTTA | *Ostreococcus tauri* | 1 | 0.04 |
| PACLE | *Pacifastacus leniusculus* | 1 | 0.04 |
| PARBA | *Paracoccidioides brasiliensis (strain ATCC MYA-826 / Pb01)* | 1 | 0.04 |
| PARBP | *Paracoccidioides brasiliensis (strain Pb03)* | 1 | 0.04 |
| PARBR | *Paracoccidioides brasiliensis* | 1 | 0.04 |
| PAROL | *Paralichthys olivaceus* | 1 | 0.04 |
| PARUW | *Protochlamydia amoebophila (strain UWE25)* | 1 | 0.04 |
| PASDO | *Passer domesticus* | 1 | 0.04 |
| PELCD | *Pelobacter carbinolicus (strain DSM 2380 / Gra Bd 1)* | 1 | 0.04 |
| PELFU | *Pelteobagrus fulvidraco* | 1 | 0.04 |
| PENMA | *Penicillium marneffei* | 1 | 0.04 |
| PENMO | *Penaeus monodon* | 1 | 0.04 |
| PENMQ | *Penicillium marneffei (strain ATCC 18224 / CBS 334.59 / QM 7333)* | 1 | 0.04 |
| PERFL | *Perca fluviatilis* | 1 | 0.04 |
| PERFV | *Perca flavescens* | 1 | 0.04 |
| PHACI | *Phascolarctos cinereus* | 1 | 0.04 |
| PHYIT | *Phytophthora infestans (strain T30-4)* | 1 | 0.04 |
| PHYPO | *Physarum polycephalum* | 1 | 0.04 |
| PHYSP | *Phytophthora sojae (strain P6497)* | 1 | 0.04 |
| PINFU | *Pinctada fucata* | 1 | 0.04 |
| PLAFE | *Platichthys flesus* | 1 | 0.04 |
| PLAMG | *Placopecten magellanicus* | 1 | 0.04 |
| PODAN | *Podospora anserina (strain S / ATCC MYA-4624 / DSM 980 / FGSC 10383)* | 1 | 0.04 |
| POLGS | *Polymorphum gilvum (strain LMG 25793 / CGMCC 1.9160 / SL003B-26A1)* | 1 | 0.04 |

| | | | |
|---|---|---|---|
| PSEF5 | *Pseudomonas fluorescens (strain Pf-5 / ATCC BAA-477)* | 1 | 0.04 |
| PSEPR | *Pseudorasbora parva* | 1 | 0.04 |
| PUCGR | *Puccinia graminis* | 1 | 0.04 |
| PUCGT | *Puccinia graminis f. sp. tritici (strain CRL 75-36-700-3 / race SCCL)* | 1 | 0.04 |
| PYRLU | *Pyrocystis lunula* | 1 | 0.04 |
| PYRTT | *Pyrenophora teres f. teres (strain 0-1)* | 1 | 0.04 |
| RABIT | *Oryctolagus cuniculus* | 1 | 0.04 |
| RAMTT | *Ramlibacter tataouinensis (strain ATCC BAA-407 / DSM 14655 / LMG 21543 / TTB310)* | 1 | 0.04 |
| RHDSA | *Rhodomonas salina* | 1 | 0.04 |
| RHIAP | *Rhipicephalus appendiculatus* | 1 | 0.04 |
| RHOPS | *Rhodopseudomonas palustris (strain BisB5)* | 1 | 0.04 |
| RHOPX | *Rhodopseudomonas palustris (strain DX-1)* | 1 | 0.04 |
| RICCO | *Ricinus communis* | 1 | 0.04 |
| ROSCS | *Roseiflexus castenholzii (strain DSM 13941 / HLO8)* | 1 | 0.04 |
| ROSHA | *Roseburia hominis (strain DSM 16839 / NCIMB 14029 / A2-183)* | 1 | 0.04 |
| ROSS1 | *Roseiflexus sp. (strain RS-1)* | 1 | 0.04 |
| RUMGN | *Ruminococcus gnavus* | 1 | 0.04 |
| RUMHA | *Ruminococcus hansenii* | 1 | 0.04 |
| SCHCM | *Schizophyllum commune (strain H4-8 / FGSC 9210)* | 1 | 0.04 |
| SCHJA | *Schistosoma japonicum* | 1 | 0.04 |
| SCHMA | *Schistosoma mansoni* | 1 | 0.04 |
| SCLS1 | *Sclerotinia sclerotiorum (strain ATCC 18683 / 1980 / Ss-1)* | 1 | 0.04 |
| SERL3 | *Serpula lacrymans var. lacrymans (strain S7.3)* | 1 | 0.04 |
| SERL9 | *Serpula lacrymans var. lacrymans (strain S7.9)* | 1 | 0.04 |
| SERQU | *Seriola quinqueradiata* | 1 | 0.04 |
| SHIB3 | *Shigella boydii serotype 18 (strain CDC 3083-94 / BS512)* | 1 | 0.04 |
| SHIBS | *Shigella boydii serotype 4 (strain Sb227)* | 1 | 0.04 |
| SHIDS | *Shigella dysenteriae serotype 1 (strain Sd197)* | 1 | 0.04 |
| SHIF2 | *Shigella flexneri serotype X (strain 2002017)* | 1 | 0.04 |
| SHIF8 | *Shigella flexneri serotype 5b (strain 8401)* | 1 | 0.04 |
| SHISO | *Shigella sonnei* | 1 | 0.04 |
| SHISS | *Shigella sonnei (strain Ss046)* | 1 | 0.04 |
| SIGHI | *Sigmodon hispidus* | 1 | 0.04 |
| SILST | *Silicibacter sp. (strain TM1040)* | 1 | 0.04 |
| SINCH | *Siniperca chuatsi* | 1 | 0.04 |
| SOLIN | *Solenopsis invicta* | 1 | 0.04 |
| SOLSE | *Solea senegalensis* | 1 | 0.04 |
| SORMK | *Sordaria macrospora (strain ATCC MYA-333 / DSM 997 / K(L3346) / K-hell)* | 1 | 0.04 |
| SPECI | *Spermophilus citellus* | 1 | 0.04 |
| SPICH | *Spirochaeta caldaria (strain ATCC 51460 / DSM 7334 / H1)* | 1 | 0.04 |
| SPISS | *Spirochaeta smaragdinae (strain DSM 11293 / JCM 15392 / SEBR 4228)* | 1 | 0.04 |
| STECO | *Stenella coeruleoalba* | 1 | 0.04 |
| SUBDO | *Suberites domuncula* | 1 | 0.04 |
| SULAT | *Sulfobacillus acidophilus (strain TPY)* | 1 | 0.04 |
| SYNAS | *Syntrophus aciditrophicus (strain SB)* | 1 | 0.04 |
| SYNGF | *Syntrophobotulus glycolicus (strain DSM 8271 / FlGlyR)* | 1 | 0.04 |
| TALSN | *Talaromyces stipitatus (strain ATCC 10500 / CBS 375.48 / QM 6759 / NRRL 1006)* | 1 | 0.04 |
| THAPS | *Thalassiosira pseudonana* | 1 | 0.04 |
| THEAQ | *Thermus aquaticus* | 1 | 0.04 |
| THEEB | *Thermosynechococcus elongatus (strain BP-1)* | 1 | 0.04 |
| THELA | *Thermomyces lanuginosus* | 1 | 0.04 |
| THELT | *Thermotoga lettingae (strain ATCC BAA-301 / DSM 14385 / TMO)* | 1 | 0.04 |
| THEM7 | *Thermaerobacter marianensis (strain ATCC 700841 / DSM 12885 / JCM 10246 / 7p75a)* | 1 | 0.04 |
| THESQ | *Thermotoga sp. (strain RQ2)* | 1 | 0.04 |
| THESS | *Thermus scotoductus (strain ATCC 700910 / SA-01)* | 1 | 0.04 |
| THET2 | *Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039)* | 1 | 0.04 |
| THET8 | *Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)* | 1 | 0.04 |
| THETC | *Thermoanaerobacterium thermosaccharolyticum (strain ATCC 7956 / DSM 571 / NCIB 9385 / NCA 3814)* | 1 | 0.04 |
| THETG | *Thermus thermophilus (strain SG0.5JP17-16)* | 1 | 0.04 |
| THEVL | *Thermosynechococcus vulcanus* | 1 | 0.04 |
| THIHA | *Thielavia heterothallica (strain ATCC 42464 / BCRC 31852 / DSM 1799)* | 1 | 0.04 |
| THITE | *Thielavia terrestris (strain ATCC 38088 / NRRL 8126)* | 1 | 0.04 |
| TODPA | *Todarodes pacificus* | 1 | 0.04 |
| TOXCI | *Toxoptera citricida* | 1 | 0.04 |
| TRICA | *Tribolium castaneum* | 1 | 0.04 |
| TRIGR | *Tripneustes gratilla* | 1 | 0.04 |
| TRINI | *Trichoplusia ni* | 1 | 0.04 |
| TRIRC | *Trichophyton rubrum (strain ATCC MYA-4607 / CBS 118892)* | 1 | 0.04 |
| TRIRU | *Trichophyton rubrum* | 1 | 0.04 |
| TRISC | *Triakis scyllium* | 1 | 0.04 |
| TRIVH | *Trichophyton verrucosum (strain HKI 0517)* | 1 | 0.04 |
| TRURR | *Truepera radiovictrix (strain DSM 17093 / CIP 108686 / LMG 22925 / RQ-24)* | 1 | 0.04 |
| UNCRE | *Uncinocarpus reesii (strain UAMH 1704)* | 1 | 0.04 |
| UREUN | *Urechis unicinctus* | 1 | 0.04 |
| VEIPT | *Veillonella parvula (strain ATCC 10790 / DSM 2008 / JCM 12972 / Te3)* | 1 | 0.04 |
| VERA1 | *Verticillium albo-atrum (strain VaMs.102 / ATCC MYA-4576 / FGSC 10136)* | 1 | 0.04 |
| VEREI | *Verminephrobacter eiseniae (strain EF01-2)* | 1 | 0.04 |
| VOLCA | *Volvox carteri* | 1 | 0.04 |
| WHEAT | *Triticum aestivum* | 1 | 0.04 |
| WUCBA | *Wuchereria bancrofti* | 1 | 0.04 |
| XENBO | *Xenopus borealis* | 1 | 0.04 |

## Table S18. Identified proteins related to *Coprothermobacter proteolyticus*.

| COG Functions | Number of identified non-redundant protein groups | **Identified protein ID** | Name | **UniRef50 cluster size** | Specific to *C.proteolyticus* | Maximum number of unique peptides assigned | Note |
|---|---|---|---|---|---|---|---|
| INFORMATION STORAGE AND PROCESSING | | | | | | | |
| [J] Translation, ribosomal structure and biogenesis | 5 | B5Y932 | 50S ribosomal protein L10 | 1 | Yes | 4 | |
| | | B5Y938 | Elongation factor Tu 1 | 2399 | | 5 | |
| | | B5Y948 | Ribosomal subunit interface protein | 1 | Yes | 2 | |
| | | B5Y9E6 | Methionyl-tRNA synthetase | 1 | Yes | 2 | |
| | | B5Y9M7 | Translation elongation factor EF-G | 1 | Yes | 5 | |
| [A] RNA processing and modification | 0 | | No detected | | | 0 | |
| [K] Transcription | 0 | | No detected | | | 0 | |
| [L] Replication, recombination and repair | 0 | | No detected | | | 0 | |
| [B] Chromatin structure and dynamics | 0 | | No detected | | | 0 | |
| CELLULAR PROCESSES AND SIGNALING | | | | | | | |
| [D] Cell cycle control, cell division, chromosome partitioning | 1 | B5Y8D3 | Cell division protein ftsZ | 1 | Yes | 2 | |
| [Y] Nuclear structure | 0 | | No detected | | | 0 | |
| [V] Defense mechanisms | 0 | | No detected | | | 0 | |
| [T] Signal transduction mechanisms | 0 | | No detected | | | 0 | |
| [M] Cell wall/membrane/envelope biogenesis | 2 | B5Y682 | Glucosamine--fructose-6-phosphate aminotransferase-related protein | 1 | Yes | 2 | |
| | | B5Y8C5 | UDP-N-acetylmuramoylalanine--D-glutamate ligase | 1 | Yes | 2 | |
| [N] Cell motility | 1 | B5Y7C7 | Flagellin protein | 1 | Yes | 8 | |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| [Z] Cytoskeleton | 0 |  | No detected |  |  | 0 |  |
| [W] Extracellular structures | 0 |  | No detected |  |  | 0 |  |
| [U] Intracellular trafficking, secretion, and vesicular transport | 0 |  | No detected |  |  | 0 |  |
| [O] Posttranslational modification, protein turnover, chaperones | 6 | B5Y5X6 | Pyrrolidone-carboxylate peptidase | 1 | Yes | 3 | Peptidase (C15)* |
|  |  | B5Y6Q5 | Protease | 1 | Yes | 14 | Peptidase (S8A) |
|  |  | B5Y893 | 10 kDa chaperonin | 432 |  | 2 | Chaperones |
|  |  | B5Y894 | 60 kDa chaperonin | 87 |  | 10 | Chaperones |
|  |  | B5Y9G5 | Bacterioferritin comigratory protein, AhpC/TSA family | 10 |  | 5 | Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen |
|  |  | B5YA04 | DO serine protease | 1 | Yes | 2 | Peptidase (S1B) |
| METABOLISM |  |  |  |  |  |  |  |
| [C] Energy production and conversion | 21 | B5Y690 | 2-oxoacid:ferredoxin oxidoreductase subunit alpha | 1 | Yes | 2 |  |
|  |  | B5Y691 | 2-oxoacid:ferredoxin oxidoreductase subunit beta | 3 |  | 4 |  |
|  |  | B5Y6W0 | Tungsten-containing aldehyde ferredoxin oxidoreductase | 64 |  | 6 |  |
|  |  | B5Y702 | Ferredoxin | 14 |  | 5 |  |
|  |  | B5Y703 | NADH:ubiquinone oxidoreductase, nadh-binding (51 kd) subunit | 66 |  | 4 |  |
|  |  | B5Y7W0 | Acetate kinase | 447 |  | 3 |  |
|  |  | B5Y871 | Acetamidase/formamidase | 2 |  | 8 |  |
|  |  | B5Y8N5 | Ketoisovalerate oxidoreductase subunit VorB (VOR) (2-oxoisovalerate oxidoreductase beta chain) (2-oxoisovalerate ferredoxinreductase subunit beta) | 1 | Yes | 2 |  |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | B5Y8N6 | Ketoisovalerate oxidoreductase subunit VorA (VOR) (2-oxoisovalerate oxidoreductase alpha chain) (2-oxoisovalerateferredoxin reductase subunit alpha) | 50 |  | 2 |  |
|  |  | B5Y8N7 | Ketoisovalerate oxidoreductase subunit VorA (VOR) (2-oxoisovalerate oxidoreductase alpha chain) (2-oxoisovalerateferredoxin reductase subunit alpha) | 1 | Yes | 4 |  |
|  |  | B5Y8T9 | 2-oxo acid dehydrogenase subunit E1 | 36 |  | 4 |  |
|  |  | B5Y8U0 | 2-oxoisovalerate dehydrogenase subunit beta (Branched-chain alpha-keto acid dehydrogenase e1 component beta chain) (Bckdhe1-beta) | 15 |  | 2 |  |
|  |  | B5Y8U4 | Dihydrolipoyllysine-residue acetyltransferase component of pyruvatedehydrogenase complex (E2) (Dihydrolipoamideacetyltransferase component of pyruvate dehydrogenase complex) | 1 | Yes | 5 |  |
|  |  | B5Y9C1 | NADPH-dependent butanol dehydrogenase | 1 | Yes | 5 |  |
|  |  | B5Y9E2 | OorB subunit of 2-oxoglutarate:acceptor oxidoreductase | 2 |  | 5 |  |
|  |  | B5Y9G2 | NADH-dependent butanol dehydrogenase a (Bdh i) | 2 |  | 2 |  |
|  |  | B5Y9Q7 | Pyruvate synthase subunit PorB (Pyruvate oxidoreductasebeta chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit beta) | 26 |  | 6 |  |
|  |  | B5Y9Q8 | Pyruvate synthase subunit PorA (Pyruvate oxidoreductasealpha chain) (POR) (Pyruvic-ferredoxin oxidoreductase subunit alpha) | 10 |  | 2 |  |
|  |  | B5Y9R0 | Pyruvate/ketoisovalerate oxidoreductases common subunit gamma | 21 |  | 4 |  |
|  |  | B5Y9S3 | Superoxide reductase (SOR) | 5 |  | 6 | antioxidant |
|  |  | B5YA15 | Phosphate butyryltransferase | 164 |  | 5 |  |
| [G] Carbohydrate transport and metabolism | 16 | B5Y677 | PTS system IIB component | 1 | Yes | 2 |  |
|  |  | B5Y679 | N-acetylglucosamine-6-phosphate deacetylase | 1 | Yes | 5 |  |
|  |  | B5Y6P4 | Maltose ABC transporter, periplasmic maltose-binding protein | 1 | Yes | 7 | Sugar ABC transporter |
|  |  | B5Y6P8 | Amylopullulanase (Alpha-amylase/pullulanase) | 1 | Yes | 26 | Related to carbohydrate |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | hydrolysis |
| | | B5Y768 | Fructose-1,6-bisphosphate aldolase, class II | 46 | | 6 | |
| | | B5Y7V7 | Pyruvate, phosphate dikinase | 1120 | | 2 | |
| | | B5Y8B9 | 6-phosphofructokinase 1 | 1 | Yes | 4 | |
| | | B5Y8G5 | Ribose 5-phosphate isomerase B | 495 | | 2 | |
| | | B5Y8H9 | Glyceraldehyde-3-phosphate dehydrogenase, type I | 3800 | | 11 | |
| | | B5Y8I0 | Phosphoglycerate kinase | 985 | | 2 | |
| | | B5Y8V1 | 6-phosphofructokinase 2 | 1 | Yes | 3 | |
| | | B5Y8W1 | Pyruvate kinase | 1 | Yes | 4 | |
| | | B5Y9E0 | Enolase | 461 | | 3 | |
| | | B5Y9J8 | Glucose-1-phosphate adenylyltransferase | 1 | Yes | 2 | |
| | | B5Y9V1 | Probable transaldolase | 549 | | 2 | |
| | | B5YA06 | Putative uncharacterized protein | 1 | Yes | 2 | |
| [E] Amino acid transport and metabolism | 19 | B5Y5X4 | Alanine--glyoxylate aminotransferase 1 | 1 | Yes | 9 | |
| | | B5Y692 | Proline dipeptidase | 10 | | 2 | Peptidase (M24) |
| | | B5Y6V2 | Bacterial extracellular solute-binding protein, family 5 | 2 | Yes | 18 | Oligopeptide ABC transporter |
| | | B5Y6V3 | Bacterial extracellular solute-binding protein, family 5 | 2 | Yes | 2 | Oligopeptide ABC transporter |
| | | B5Y6Y1 | Glutamate dehydrogenase | 376 | | 9 | |
| | | B5Y767 | Succinyl-diaminopimelate desuccinylase | 9 | | 9 | |
| | | B5Y7J5 | M18 family aminopeptidase | 1 | Yes | 7 | Peptidase (M18) |
| | | B5Y852 | L-asparaginase, thermolabile | 1 | Yes | 2 | |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | B5Y897 | Oligopeptide transport ATP-binding protein AppD | 30 |  | 4 | Oligopeptide ABC transporter |
|  |  | B5Y898 | Oligopeptide transport ATP-binding protein AppF | 376 |  | 3 | Oligopeptide ABC transporter |
|  |  | B5Y8A8 | Arginine decarboxylase | 1 | Yes | 2 |  |
|  |  | B5Y8G6 | Pyridoxal-phosphate-dependent serine hydroxymethyltransferase | 2795 |  | 2 |  |
|  |  | B5Y8M3 | Proline dipeptidase | 1 | Yes | 3 | Peptidase (M24) |
|  |  | B5Y9A8 | Urocanate hydratase | 1282 |  | 6 |  |
|  |  | B5Y9D6 | 4-aminobutyrate aminotransferase | 1 | Yes | 5 |  |
|  |  | B5Y9F2 | Ornithine carbamoyltransferase | 278 |  | 2 |  |
|  |  | B5Y9Q5 | Aspartate aminotransferase A (Transaminase A) (AspAT) | 1 | Yes | 5 |  |
|  |  | B5Y9R7 | Proline dehydrogenase, beta subunit | 1 | Yes | 2 |  |
|  |  | B5Y9V8 | Peptidase T | 586 |  | 3 | Peptidase (M20) |
| [F] Nucleotide transport and metabolism | 6 | B5Y6U0 | Inosine-5'-monophosphate dehydrogenase | 2008 |  | 2 |  |
|  |  | B5Y7V3 | Deoxyribose-phosphate aldolase | 1139 |  | 2 |  |
|  |  | B5Y8A7 | Adenylosuccinate synthetase | 1 | Yes | 2 |  |
|  |  | B5Y8K9 | Putative ribose-phosphate pyrophosphokinase | 1 | Yes | 2 |  |
|  |  | B5Y8Q0 | Multifunctional protein: 5-nucleotidase 2,3-cyclic-nucleotide 2-phosphodiesterase | 2 | Yes | 28 |  |
|  |  | B5Y9A5 | Formate--tetrahydrofolate ligase | 1 | Yes | 2 |  |
| [H] Coenzyme transport and metabolism | 3 | B5Y6I1 | Thiamine-phosphate synthase | 3 |  | 4 |  |
|  |  | B5Y864 | 3-methyl-2-oxobutanoate hydroxymethyltransferase | 3 |  | 3 |  |
|  |  | B5Y9Z4 | Putative 8-amino-7-oxononanoate synthase/2-amino-3-ketobutyrate coenzyme A ligase | 48 |  | 9 |  |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [I] Lipid transport and metabolism | 1 | B5Y8Z4 | Methylmalonyl-CoA carboxyltransferase 12S subunit (Transcarboxylase 12S subunit) | 252 | | 6 | |
| [P] Inorganic ion transport and metabolism | 1 | B5YA48 | Hemerythrin | 2 | Yes | 2 | Oxygen transport |
| [Q] Secondary metabolites biosynthesis, transport and catabolism | 1 | B5Y6G9 | Pseudouridine-5'-phosphate glycosidase | 1 | Yes | 5 | Synthase for indigoidine blue pigment |
| POORLY CHARACTERIZED | | | | | | | |
| [R] General function prediction only | 10 | B5Y678 | Putative uncharacterized protein | 7 | | 6 | |
| | | B5Y683 | Oxidoreductase family, NAD-binding Rossmann fold protein | 1 | Yes | 2 | |
| | | B5Y6N5 | TldD/PmbA family protein, putative | 1 | Yes | 6 | Peptidase (U62), Microcin |
| | | B5Y6N6 | LmbIH | 1 | Yes | 5 | Peptidase (U62), Microcin |
| | | B5Y789 | Thermostable monoacylglycerol lipase (Mglp) | 1 | Yes | 3 | |
| | | B5Y8U9 | Sugar ABC transporter, ATP-binding protein | 1 | Yes | 3 | ABC transporter |
| | | B5Y8V0 | Basic membrane protein family protein | 1 | Yes | 8 | |
| | | B5Y9F3 | Putative uncharacterized protein | 40 | | 3 | |
| | | B5Y9N6 | Intracellular protease 1 (Intracellular protease I) | 1 | Yes | 2 | Peptidase (C56) |
| | | B5Y9Y2 | TldD protein | 1 | Yes | 3 | Peptidase (U62), Microcin |
| [S] Function unknown | 4 | B5Y6M7 | Major pilin protein FimA | 3 | | 2 | |
| | | B5Y6U2 | Rubrerythrin-related protein | 2 | Yes | 2 | Antioxidant |
| | | B5Y7L1 | Putative uncharacterized protein | 1 | Yes | 2 | |
| | | B5Y8J9 | Stage V sporulation protein S | 369 | | 2 | |
| COG non hits | 7 | B5Y6H1 | Surface layer protein | 1 | Yes | 5 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | B5Y6Q7 | Copper amine oxidase N-domain family | 1 | Yes | 5 | RTX toxin, Ig-like |
| | | B5Y6X3 | Putative uncharacterized protein | 1 | Yes | 2 | Virulence factor, pectin lyase fold, bacterioin |
| | | B5Y743 | Glycine reductase complex component B subunits alpha and beta (Selenoprotein PB alpha/beta) | 107 | | 6 | |
| | | B5Y746 | Glycine reductase complex component B subunit gamma (Selenoprotein PB gamma) | 111 | | 9 | |
| | | B5Y7R5 | Endoxylanase | 1 | Yes | 3 | |
| | | B5Y9H7 | Putative uncharacterized protein | 26 | | 3 | |

*The symbol between parentheses refers to the identifier of the Families of Proteolytic Enzymes, as listed in the database MEROPS of proteolytic enzymes (http: merops.sanger.ac.uk/index.shtml) (Rawlings *et al.*, 2012)

## Table S19. Identified proteins related to *Clostridium thermocellum*.

| COG Functions | Number of identified non-redundant protein groups | **Identified protein ID** | Name | **UniRef50 cluster size** | Specific to *C.thermocellum* | Maximal number of unique peptides | Note |
|---|---|---|---|---|---|---|---|
| INFORMATION STORAGE AND PROCESSING | | | | | | | |
| [J] Translation, ribosomal structure and biogenesis | 3 | A3DBI2 _CLOTH<br>C7HH27_CLOTM<br>D1NKS9_CLOTM<br>E6UN48_CLOTL | Asparagine--tRNA ligase | 32 | | 3 | |
| | | A3DDI0_CLOTH<br>C7HD62_CLOTM<br>D1NI84_CLOTM<br>E6UR48_CLOTL | Protein hfq | 513 | | 3 | |
| | | A3DJ00_CLOTH | Elongation factor Tu | 2366 | | 11 | |
| [A] RNA processing and modification | 0 | | No detected | | | 0 | |
| [K] Transcription | 3 | A3DDV0_CLOTH<br>C7HEX3_CLOTM<br>D1NIK3_CLOTM<br>E6UPZ4_CLOTL | RNA polymerase sigma factor | 32 | | 2 | |
| | | A3DIK9_CLOTH<br>C7HBR9_CLOTM<br>D1NPH0_CLOTM<br>E6UQB2_CLOTL | Type III pantothenate kinase | 16 | | 2 | |
| | | C7HHG8_CLOTM<br>D1NQZ3_CLOTM<br>E6USZ3_CLOTL | DNA-directed RNA polymerase subunit alpha | 1739 | | 3 | |
| [L] Replication, recombination and repair | 2 | A3DHS4_CLOTH<br>C7HGH0_CLOTM<br>D1NP86_CLOTM<br>E6UU04_CLOTL | CRISPR-associated regulatory protein, DevR family | 4 | Yes | 8 | |
| | | A3DKB8_CLOTH<br>D1NKI2_CLOTM | CRISPR-associated protein, Csh2 family | 12 | | 6 | |
| [B] Chromatin structure and dynamics | 0 | | No detected | | | 0 | |
| CELLULAR PROCESSES AND SIGNALING | | | | | | | |
| [D] Cell cycle control, cell division, chromosome partitioning | 1 | A3DCK4_CLOTH<br>C7HFZ0_CLOTM<br>D1NLX1_CLOTM<br>E6UTW0_CLOTL | Cell division protein FtsZ | 273 | | 2 | |
| [Y] Nuclear structure | 0 | | No detected | | | 0 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| [V] Defense mechanisms | 2 | A3DCU6_CLOTH<br>C7HFE1_CLOTM<br>D1NM47_CLOTM<br>E6UT84_CLOTL | ABC transporter related protein | 230 | | 3 | Direct the export and processing of small antibacterial peptides called bacteriocins. |
| | | A3DH88_CLOTH<br>C7HDX4_CLOTM<br>D1NQ96_CLOTM<br>\|E6US63_CLOTL | ABC transporter related protein | 102 | | 2 | Direct the export and processing of small antibacterial peptides called bacteriocins. |
| [T] Signal transduction mechanisms | 0 | | No detected | | | 0 | |
| [M] Cell wall/membrane/envelope biogenesis | 3 | A3DD11_CLOTH<br>C7HDK6_CLOTM<br>D1NI23_CLOTM<br>E6USM3_CLOTL | NLP/P60 protein (Precursor) | 4 | Yes | 3 | Peptidase (C40), Cell wall-associated hydrolases (invasion-associated proteins) |
| | | A3DEL5_CLOTH<br>C7HIR5_CLOTM<br>D1NMS0_CLOTM<br>E6UNI4_CLOTL | Glucosamine/fructose-6-phosphate aminotransferase, isomerizing | 196 | | 3 | |
| | | A3DHV1_CLOTH<br>C7HGF5_CLOTM<br>D1NP71_CLOTM<br>E6UU19_CLOTL | UDP-N-acetylglucosamine 1-carboxyvinyltransferase 1 | 496 | | 2 | |
| [N] Cell motility | 3 | A3DBG8_CLOTH<br>C7HH13_CLOTM<br>D1NKR6_CLOTM<br>E6UN62_CLOTL | Ig domain protein group 2 domain protein | 5 | | 5 | Related to carbohydrate hydrolysis, Ig-like |
| | | A3DEF9_CLOTH<br>C7HC84_CLOTM<br>E6UNR4_CLOTL | Twitching motility protein | 44 | | 2 | |
| | | A3DHL0_CLOTH<br>C7HID0_CLOTM<br>D1NNP6_CLOTM<br>E6UTE5_CLOTL | Flagellin domain protein | 262 | | 5 | |
| [Z] Cytoskeleton | 0 | | No detected | | | 0 | |
| [W] Extracellular structures | 0 | | No detected | | | 0 | |
| [U] Intracellular trafficking, secretion, and vesicular transport | 0 | | No detected | | | 0 | |
| [O] Posttranslational modification, protein turnover, chaperones | 3 | A3DEY9_CLOTH<br>C7HFB4_CLOTM<br>D1NMX1_CLOTM<br>E6UMZ2_CLOTL | Peptidase S1 and S6 chymotrypsin/Hap | 6 | | 7 | Peptidase (S1, S6) |
| | | A3DGV4_CLOTH<br>C7HH48_CLOTM<br>D1NPW4_CLOTM<br>E6UR92_CLOTL | Peroxiredoxin | 1579 | | 4 | Antioxidant |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | C7HHZ5_CLOTM<br>D1NQV1_CLOTM<br>E6USV2_CLOTL<br>CH60_CLOTH | 60 kDa chaperonin | 2696 |  | 10 | Chaperonin |
| METABOLISM |  |  |  |  |  |  |  |
| [C] Energy production and conversion | 14 | A3DBL3_CLOTH<br>C7HFU0_CLOTM<br>D1NKW1_CLOTM<br>E6UMU0_CLOTL | Iron-containing alcohol dehydrogenase | 90 |  | 3 |  |
|  |  | A3DC43_CLOTH<br>D1NLE7_CLOTM | Aldo/keto reductase | 128 |  | 3 |  |
|  |  | A3DC99_CLOTH<br>C7HIN8_CLOTM<br>D1NLK4_CLOTM<br>E6ULI7_CLOTL | Ferredoxin | 6 |  | 2 |  |
|  |  | A3DCA3_CLOTH<br>C7HIP2_CLOTM<br>D1NLK8_CLOTM<br>E6ULI3_CLOTL | Malic protein NAD-binding protein | 107 |  | 3 |  |
|  |  | A3DCA4_CLOTH<br>C7HIP3_CLOTM<br>D1NLK9_CLOTM<br>E6ULI2_CLOTL | L-lactate dehydrogenase 1 (Precursor) | 9 |  | 3 |  |
|  |  | A3DDG8_CLOTH<br>C7HD50_CLOTM<br>D1NI71_CLOTM<br>E6UR60_CLOTL | Putative uncharacterized protein | 16 |  | 3 |  |
|  |  | A3DGA9_CLOTH<br>C7HCF4_CLOTM<br>D1NK50_CLOTM<br>E6UPH1_CLOTL | Putative uncharacterized protein | 423 |  | 5 |  |
|  |  | A3DGB3_CLOTH<br>C7HCF0_CLOTM<br>D1NK54_CLOTM<br>E6UPH5_CLOTL | Rubrerythrin | 22 |  | 4 |  |
|  |  | VATB_CLOTH<br>C7HHA5_CLOTM<br>D1NNL5_CLOTM<br>E6UTX3_CLOTL | V-type ATP synthase beta chain | 1058 |  | 7 |  |
|  |  | A3DI13_CLOTH<br>C7HDF8_CLOTM<br>D1NNZ7_CLOTM<br>E6URH1_CLOTL | Pyruvate/ketoisovalerate oxidoreductase, gamma subunit | 47 |  | 3 |  |
|  |  | A3DI15_CLOTH<br>C7HDF6_CLOTM<br>D1NNZ5_CLOTM<br>E6URH3_CLOTL | Pyruvate flavodoxin/ferredoxin oxidoreductase domain protein | 28 |  | 3 |  |
|  |  | A3DJE3_CLOTH<br>C7HI13_CLOTM | Phosphoenolpyruvate carboxykinase [GTP] | 1673 |  | 3 |  |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | D1NQT3_CLOTM<br>E6USE3_CLOTL | | | | | |
| | | A3DJU4_CLOTH<br>C7HEE1_CLOTM<br>D1NR78_CLOTM<br>E6UTN8_CLOTL | Citrate synthase | 14 | | 2 | |
| | | C7HC04_CLOTM<br>D1NIY5_CLOTM<br>E6UP68_CLOTL<br>O52594 _CLOTH | Acetate kinase | 447 | | 3 | |
| [G] Carbohydrate transport and metabolism | 11 | A3DBP9_CLOTH<br>C7HFX5_CLOTM<br>D1NKZ6_CLOTM<br>\|E6UMQ5_CLOTL | Glyceraldehyde-3-phosphate dehydrogenase, type I | 2346 | | 11 | |
| | | A3DBQ0_CLOTH<br>C7HFX6_CLOTM<br>E6UMQ4_CLOTL | Phosphoglycerate kinase | 932 | | 4 | |
| | | A3DBQ1_CLOTH<br>C7HFX7_CLOTM<br>D1NKZ7_CLOTM<br>E6UMQ3_CLOTL | Triosephosphate isomerase | 932 | | 4 | |
| | | A3DBQ5_CLOTH<br>C7HFY1_CLOTM<br>D1NL00_CLOTM<br>E6UMQ0_CLOTL | Enolase | 461 | | 8 | |
| | | A3DBX9_CLOTH<br>C7HJC1_CLOTM<br>D1NL73_CLOTM,<br>E6UMA2_CLOTL | Glucose-6-phosphate isomerase | 274 | | 5 | |
| | | A3DC35_CLOTH<br>C7HGL0_CLOTM<br>E6ULX2_CLOTL | Glycosyltransferase 36 | 70 | | 3 | |
| | | A3DCA6_CLOTH<br>C7HIP5_CLOTM<br>D1NLL1_CLOTM,<br>E6ULI0 _CLOTL | Phosphofructokinase | 22 | | 4 | |
| | | A3DCA8_CLOTH<br>C7HIP6_CLOTM<br>D1NLL3_CLOTM<br>E6ULH9_CLOTL | Fructose-1,6-bisphosphate aldolase, class II | 131 | | 4 | |
| | | A3DE73_CLOTH<br>C7HBZ6_CLOTM<br>D1NIX7_CLOTM<br>E6UP76_CLOTL | Extracellular solute-binding protein family 1 (Precursor) | 4 | Yes | 12 | |
| | | A3DJP0_CLOTH<br>C7HJV5_CLOTM<br>D1NR31_CLOTM<br>D1NRP9_CLOTM<br>E6UTI4_CLOTL<br>E6UTI5_CLOTL | Glycoside hydrolase family 11 (Precursor) | 20 | | 2 | Related to carbohydrate hydrolysis, CBM06, GH11, CE4 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | O52779_CLOTM<br>O52780_CLOTM<br>O87118_CLOTM<br>O87119_CLOTM | | | | | |
| | | A3DJQ6_CLOTH<br>C7HEH9_CLOTM<br>D1NRB6_CLOTM<br>E6UTK1_CLOTL<br>Q93HT8_CLOTM | Glycosyltransferase 36 | 11 | | 2 | Related to carbohydrate GH94 |
| [E] Amino acid transport and metabolism | 8 | A3DCD3_CLOTH<br>C7HHQ3_CLOTM<br>D1NLN9_CLOTM<br>E6UUG6_CLOTL | Glutamate dehydrogenase | 2207 | | 13 | |
| | | A3DDD2_CLOTH<br>C7HD12_CLOTM<br>D1NMH8_CLOTM<br>E6URU7_CLOTL | Peptidase M18 aminopeptidase I | 88 | | 4 | Peptidase (M18) |
| | | A3DDQ5_CLOTH<br>C7HES5_CLOTM,<br>D1NIF6_CLOTM<br>E6UQI0_CLOTL | 3-dehydroquinate dehydratase, type II | 824 | | 2 | |
| | | A3DDR1_CLOTH<br>C7HET2_CLOTM<br>D1NIG3_CLOTM<br>E6UQH3_CLOTL | Branched-chain amino acid aminotransferase | 25 | | 6 | |
| | | A3DGI2_CLOTH<br>C7HI89_CLOTM<br>D1NPL2_CLOTM<br>E6UQ45_CLOTL | Cysteine synthase A | 2433 | | 5 | |
| | | A3DHE1_CLOTH<br>C7HE90_CLOTM<br>D1NRL0_CLOTM<br>E6USR7_CLOTL | Propeptide PepSY amd peptidase M4 (Precursor) | 7 | | 8 | Peptidase (M4), extracellulqr, membrane |
| | | A3DIE0_CLOTH<br>C7HGY4_CLOTM<br>D1NNE5_CLOTM<br>E6UPQ0_CLOTL | Acetolactate synthase, small subunit | 356 | | 2 | |
| | | A3DIE1_CLOTH<br>C7HGY3_CLOTM<br>D1NNE4_CLOTM<br>E6UPQ1_CLOTL | Ketol-acid reductoisomerase | 1288 | | 12 | |
| [F] Nucleotide transport and metabolism | 1 | A3DD87_CLOTH<br>C7HDU1_CLOTM<br>D1NMD2_CLOTM<br>E6URZ2_CLOTL | IMP dehydrogenase | 144 | | 9 | |
| [H] Coenzyme transport and metabolism | 2 | A3DD21_CLOTH<br>C7HDL6_CLOTM<br>D1NI33_CLOTM<br>E6USL3_CLOTL | Phenylacetate-CoA ligase | 532 | | 2 | |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | A3DBL9_CLOTH<br>C7HFU6_CLOTM<br>D1NKW7_CLOTM<br>E6UMT4_CLOTL | 6,7-dimethyl-8-ribityllumazine synthase | 280 |  | 2 |  |
| [I] Lipid transport and metabolism | 2 | A3DDA5_CLOTH<br>C7HCY2_CLOTM<br>D1NME9_CLOTM<br>E6URX6_CLOTL | Carboxyl transferase | 916 |  | 12 |  |
|  |  | A3DK85_CLOTH<br>C7HCQ8_CLOTM<br>D1NKE9_CLOTM<br>E6ULF9_CLOTL | Enoyl-[acyl-carrier-protein] reductase [NADH] | 830 |  | 2 |  |
| [P] Inorganic ion transport and metabolism | 0 |  | Non detected |  |  | 0 |  |
| [Q] Secondary metabolites biosynthesis, transport and catabolism | 0 |  | Non detected |  |  | 0 |  |
| POORLY CHARACTERIZED |  |  |  |  |  |  |  |
| [R] General function prediction only | 2 | A3DD30_CLOTH<br>C7HDM6_CLOTM<br>D1NI43_CLOTM<br>P71140_CLOTM<br>E6USK3_CLOTL | Glycoside hydrolase family 9 | 19 |  | 11 | Related to carbohydrate hydrolysis, GH9, GH44, CBM30, CBM44 |
|  |  | A3DJR3_CLOTH<br>C7HEH1_CLOTM<br>D1NRA8_CLOTM<br>E6UTK9_CLOTL | Intracellular protease, PfpI family | 451 |  | 5 | Peptidase (C56) |
| [S] Function unknown | 4 | A3DFI1_CLOTH<br>C7HFP7_CLOTM<br>D1NMY9_CLOTM<br>E6UNJ5_CLOTL | X-X-X-Leu-X-X-Gly heptad repeat-containing protein (Precursor) | 5 | Yes | 12 |  |
|  |  | A3DFJ4_CLOTH<br>C7HFN3_CLOTM<br>D1NN02_CLOTM<br>E6UNK9_CLOTL | Pyridoxamine 5'-phosphate oxidase-related FMN-binding protein | 39 |  | 2 |  |
|  |  | A3DFK3_CLOTH<br>D1NN10_CLOTM | Linocin_M18 bacteriocin protein | 4 | Yes | 8 |  |
|  |  | A3DG98_CLOTH<br>C7HCG5_CLOTM<br>D1NK39_CLOTM<br>E6UPG0_CLOTL | Putative uncharacterized protein | 6 |  | 2 | Membrane protease subunit |
| COG non hits | 18 | A3DCG1_CLOTH<br>C7HHT0_CLOTM<br>D1NLR2_CLOTM<br>E6UUE1_CLOTL | Copper amine oxidase-like domain-containing protein (Precursor) | 4 | Yes | 7 |  |
|  |  | A3DDD5_CLOTH<br>C7HD15_CLOTM<br>D1NMI1_CLOTM<br>E6URU4_CLOTL | Dockerin type 1 | 4 | Yes | 2 | Related to carbohydrate hydrolysis, CBM |
|  |  | A3DEP5_CLOTH | Putative uncharacterized protein | 4 | Yes | 5 |  |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | D1NJ41_CLOTM | (Precursor) | | | | |
| | | A3DF75_CLOTH<br>C7HJB0_CLOTM<br>D1NJH2_CLOTM<br>E6UMH2_CLOTL | Putative uncharacterized protein | 8 | Yes | 3 | |
| | | \|A3DGQ2_CLOTH<br>C7HJ23_CLOTM<br>D1NPR1_CLOTM<br>E6UQP8_CLOTL | Copper amine oxidase-like domain-containing protein (Precursor) | 4 | Yes | 12 | |
| | | A3DH67_CLOTH<br>C7HDZ8_CLOTM<br>D1NQ73_CLOTM<br>E6US41_CLOTL<br>P0C2S5_CLOTM | Endoglucanase SS | 90 | | 7 | Related to carbohydrate hydrolysis, GH48 |
| | | A3DHX1_CLOTH<br>O86999_CLOTM | Ig domain protein (Precursor) | 5 | Yes | 32 | Ig-like-peptidase |
| | | A3DI06_CLOTH<br>C7HDG5_CLOTM<br>D1NP04_CLOTM<br>E6URG4_CLOTL | Copper amine oxidase-like domain-containing protein (Precursor) | 5 | Yes | 6 | |
| | | A3DI47_CLOTH<br>C7HDC3_CLOTM<br>D1NNW3_CLOTM<br>E6UPK7_CLOTL | Copper amine oxidase-like domain-containing protein (Precursor) | 4 | Yes | 2 | |
| | | A3DJE_CLOTH<br>C7HI10_CLOTM<br>D1NQT6_CLOTM<br>E6USE6_CLOTL | S-layer domain-containing protein (Precursor) | 4 | Yes | 2 | |
| | | A3DKE2_CLOTH<br>A3DKE3_CLOTH<br>A3DKE4_CLOTH<br>C7HJJ6_CLOTM<br>C7HJJ7_CLOTM<br>C7HJJ8_CLOTM<br>D1NKK6_CLOTM<br>\|D1NKK7_CLOTM<br>D1NKK8_CLOTM<br>E6ULT4_CLOTL<br>E6ULT5_CLOTL<br>E6ULT6_CLOTL | Copper amine oxidase-like domain-containing protein (Precursor) | 12 | Yes | 2 | |
| | | A3DKE6_CLOTH<br>C7HJJ5_CLOTM<br>D1NKK9_CLOTM<br>E6ULT7_CLOTL | Putative uncharacterized protein | 5 | Yes | 2 | |
| | | C7HDC4_CLOTM<br>E6URK4_CLOTL | Type 3a cellulose-binding domain protein (Precursor) | 4 | Yes | 5 | Related to carbohydrate, CBM3 |
| | | C7HGD2_CLOTM<br>E6UUH9_CLOTL | Ig domain protein (Precursor) | 5 | Yes | 3 | Ig-like-peptidase |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | D1NHZ0_CLOTM<br>E6UU82_CLOTL<br>Q06851_CLOTH | Cellulosome anchoring protein cohesin region | 14 | Yes | 20 | Related to carbohydrate hydrolysis |
|  |  | D1NHZ1_CLOTM<br>E6UU83_CLOTL<br>Q06852_CLOTH | Cellulosome anchoring protein cohesin region | 9 |  | 6 | Related to carbohydrate hydrolysis |
|  |  | D1NJG7_CLOTM | S-layer domain protein | 4 | Yes | 7 |  |
|  |  | D1NMG1_CLOTM<br>E6URW4_CLOTL | Putative uncharacterized protein | 34 |  | 2 |  |

[*] CLOTH: *Clostridium thermocellum (strain ATCC 27405 / DSM 1237)*; CLOTM: *Clostridium thermocellum (strain YS)*; CLOTL: *Clostridium thermocellum (strain DSM 1313 / LMG 6656 / LQ8)*.

## Table S20. Identified proteins related to *Caldicellulosiruptor* spp.

| COG Functions | Number of identified non-redundant protein groups | **Identified protein ID** | Name | **UniRef50 cluster size** | Specific to *C. thermocellum* | Maximal number of unique peptides | Note |
|---|---|---|---|---|---|---|---|
| INFORMATION STORAGE AND PROCESSING | | | | | | | |
| [J] Translation, ribosomal structure and biogenesis | 11 | A4XHB0_CALS8 | Histidine--tRNA ligase | 2 | | 3 | |
| | | A4XHB1_CALS8<br>D9TIN1_CALOO<br>E4Q407_CALOW<br>E4SGK5_CALK2 | Aspartate--tRNA ligase | 10 | | 2 | |
| | | A4XHY9_CALS8<br>B9MQ85_ANATD<br>D9TJ22_CALOO<br>E4Q4U1_CALOW<br>E4QDU5_CALH1<br>E4S4M6_CALKI<br>E4SG60_CALK2<br>G2PXX6_9FIRM | Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B | 97 | | 2 | |
| | | A4XJH1_CALS8<br>B9MQM8_ANATD<br>D9TJD5_CALOO<br>E4Q5I9_CALOW<br>E4QD35_CALH1<br>E4SAA3_CALKI,<br>E4SFV6_CALK2<br>G2PTN2_9FIRM | Serine--tRNA ligase | 193 | | 2 | |
| | | A4XKS3_CALS8<br>B9MS06_ANATD<br>E4Q1H8_CALOW<br>E4QAA9_CALH1<br>E4S883_CALKI<br>E4SBS8_CALK2<br>G2PTS6_9FIRM | Methylthioribose-1-phosphate isomerase | 29 | | 2 | |
| | | A4XL64_CALS8<br>B9MR54_ANATD<br>D9TL75_CALOO<br>E4Q6E3_CALOW<br>E4S5N0_CALKI<br>E4SFG4_CALK2<br>G2PYA8_9FIRM | Polyribonucleotide nucleotidyltransferase | 235 | | 3 | |
| | | A4XM01_CALS8 | Elongation factor Ts | 134 | | 2 | |
| | | B9MQ84_ANATD<br>E4S4M7_CALKI<br>G2PYC3_9FIRM | Glutamyl-tRNA(Gln) amidotransferase subunit A | 115 | | 6 | |
| | | B9MQH1_ANATD<br>D9TJ70_CALOO<br>E4Q565_CALOW<br>E4QDG9_CALH1<br>E4SG14_CALK2 | Elongation factor Tu | 2366 | | 9 | |
| | | D9TIV1_CALOO<br>E4Q4F6_CALOW<br>E4S9F3_CALKI | Lysine--tRNA ligase | 543 | | 3 | |
| | | D9TJ69_CALOO | Elongation factor G | 2859 | | 9 | |
| [A] RNA processing and modification | 0 | | Non detected | | | | |
| [K] Transcription | 8 | G2PZF2_9FIRM<br>B9MQ36_ANATD<br>E4QDX5_CALH1<br>E4SG86_CALK2<br>E4S3Z6_CALKI<br>D9TIZ7_CALOO<br>E4Q4R7_CALOW<br>A4XHW4_CALS8 | RNA polymerase sigma factor | 32 | | 2 | |
| | | A4XKW1_CALS8 | Putative transcriptional regulator, AsnC | 50 | | 2 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | family | | | | |
| | | A4XLY8_CALS8 | Transcriptional regulator, CdaR | 8 | Yes | 5 | |
| | | A4XM62_CALS8<br>B9MM93_ANATD<br>D9THG3_CALOO<br>E4Q291_CALOW<br>E4Q997_CALH1<br>E4S6G5_CALKI<br>E4SHJ3_CALK2<br>G2PWT0_9FIRM | Transcriptional regulator, AbrB family | 22 | | 4 | |
| | | B9MPA9_ANATD<br>D9TII0_CALOO<br>E4Q3V0_CALOW<br>E4Q7H3_CALH1<br>E4S9V8_CALKI<br>E4SGQ7_CALK2<br>G2PU31_9FIRM | Type III pantothenate kinase | 147 | | 2 | |
| | | B9MPB7_ANATD<br>E4Q7G4_CALH1<br>E4S583_CALK<br>E4SGQ0_CALK2<br>G2PX09_9FIRM | Glucokinase, ROK family | 21 | | 3 | |
| | | B9MQK3_ANATD<br>D9TJA9_CALOO<br>E4Q598_CALOW<br>E4QDD5_CALH1<br>E4S482_CALKI<br>G2PXB3_9FIRM | Transcriptional regulator, DeoR family | 10 | | 3 | |
| | | E4Q843_CALH1<br>E4SAJ6_CALKI<br>E4SB85_CALK2<br>G2PWJ5_9FIRM | DNA-directed RNA polymerase subunit alpha | 1739 | | 5 | |
| [L] Replication, recombination and repair | 4 | A4XIT2_CALS8<br>D9TIU0_CALOO<br>E4Q4E5_CALOW<br>E4QE99_CALH1<br>E4SA74_CALKI<br>E4SGE7_CALK2<br>G2PTQ4_9FIRM | Histone family protein<br>DNA-binding protein | 1388 | | 5 | |
| | | A4XLD6_CALS8 | Protein RecA | 923 | | 3 | |
| | | E4Q9T2_CALH1<br>E4S7V7_CALKI | CRISPR-associated regulatory protein, DevR family | 6 | | 2 | |
| | | G2PYK4_9FIRM | CRISPR-associated protein, Csh2 family | 10 | | 2 | |
| [B] Chromatin structure and dynamics | 0 | | Non detected | | | | |
| CELLULAR PROCESSES AND SIGNALING | | | | | | | |
| [D] Cell cycle control, cell division, chromosome partitioning | 1 | A4XI09_CALS8<br>B9MQA6_ANATD<br>E4QDJ7_CALH1<br>E4S4K5_CALKI<br>E4SG39_CALK2<br>G2PXV5_9FIRM | Cell division protein FtsZ | 10 | | 5 | |
| [Y] Nuclear structure | 0 | | Non detected | | | | |
| [V] Defense mechanisms | 0 | | Non detected | | | | |
| [T] Signal transduction mechanisms | 1 | D9TI35_CALOO<br>E4SH73_CALK2 | HPr kinase/phosphorylase | 8 | Yes | 5 | |
| [M] Cell wall/membrane/envelope biogenesis | 5 | A4XGE7_CALS8 | Putative septation protein spoVG | 21 | | 2 | |
| | | A4XKF5_CALS8 | Transferase hexapeptide repeat containing protein | 32 | | 2 | |
| | | D9TGM7_CALOO | Glucosamine/fructose-6-phosphate aminotransferase, isomerizing | 196 | | 5 | |
| | | D9TJP6_CALOO | N-acetylneuraminate synthase | 99 | | 3 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | D9TKS8_CALOO | Nucleotide sugar dehydrogenase | 124 | | 7 | |
| [N] Cell motility | 10 | A4XIY9_CALS8 | Response regulator receiver protein | 173 | | 2 | |
| | | A4XK35_CALS8 | Flagellin domain protein | 76 | | 2 | |
| | | B9MK66_ANATD<br>E4Q2I4_CALOW<br>E4SA50_CALKI<br>E4SCM7_CALK2<br>G2PVG3_9FIRM | Methyl-accepting chemotaxis sensory transducer | 7 | Yes | 3 | |
| | | B9MM12_ANATD<br>D9TG88_CALOO<br>E4Q519_CALOW<br>E4QCW8_CALH1<br>E4S947_CALKI<br>E4SE88_CALK2<br>G2PUW4_9FIRM | CheC, inhibitor of MCP methylation / FliN fusion protein | 9 | | 3 | |
| | | D9TJP3_CALOO | Flagellin domain protein | 262 | | 2 | |
| | | E4Q1X4_CALOW<br>E4Q9Z5_CALH1<br>E4SCR4_CALK2 | Methyl-accepting chemotaxis sensory transducer | 4 | Yes | 2 | |
| | | E4Q3R8_CALOW | Twitching motility protein | 130 | | 2 | |
| | | E4Q8G4_CALH1 | Flagellin domain protein | 5 | | 2 | |
| | | E4SAE3_CALKI | Flagellin domain protein | 262 | | 6 | |
| | | G2PVZ5_9FIRM | Flagellin domain protein | 262 | | 8 | |
| [Z] Cytoskeleton | 0 | | Non detected | | | | |
| [W] Extracellular structures | 0 | | Non detected | | | | |
| [U] Intracellular trafficking, secretion, and vesicular transport | 0 | | Non detected | | | | |
| [O] Posttranslational modification, protein turnover, chaperones | 9 | CH60_CALS8 | 60 kDa chaperonin | 87 | | 14 | |
| | | DNAK_CALS8<br>D9TJX2_CALOO<br>E4Q2E4_CALOW<br>E4Q8Y7_CALH1<br>E4S9T1_CALKI<br>E4SD08_CALK2<br>G2PV98_9FIRM | Chaperone protein DnaK | 160 | | 12 | |
| | | A4XLV0_CALS8 | 2-alkenal reductase (Precursor) | 8 | Yes | 4 | |
| | | A4XM10_CALS8<br>E4Q3N5_CALOW | ATPase AAA-2 domain protein | 44 | | 2 | |
| | | B9MLY9_ANATD<br>D9TG66_CALOO<br>E4Q4Z6_CALOW<br>E4QD65_CALH1<br>E4S970_CALKI<br>E4SEB1_CALK2<br>G2PUU1_9FIRM | 60 kDa chaperonin | 87 | | 2 | |
| | | B9MM30_ANATD<br>E4Q5B3_CALOW<br>E4QCV0_CALH1<br>E4S8P3_CALKI<br>E4SE70_CALK2<br>G2PUY2_9FIRM | ATP-dependent protease ATPase subunit HslU | 123 | | 2 | |
| | | B9MQ15_ANATD<br>D9TIX3_CALOO<br>E4Q4P1_CALOW<br>E4QDZ7_CALH1<br>E4S9D3_CALKI<br>E4SGB0_CALK2<br>G2PUM8_9FIRM | Band 7 protein | 8 | Yes | 2 | |
| | | E4Q1T4_CALOW | Peroxiredoxin | 158 | | 4 | |
| | | E4SD93_CALK2 | SufBD protein | 13 | | 3 | |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| METABOLISM |  |  |  |  |  |  |  |
| [C] Energy production and conversion | 25 | A4XG86_CALS8 | Aldo/keto reductase | 721 |  | 4 |  |
|  |  | A4XHI4_CALS8<br>B9MNX1_ANATD<br>E4Q3K5_CALOW<br>E4Q7J4_CALH1<br>E4SAV2_CALKI<br>E4SGS7_CALK2<br>G2PWP6_9FIRM | Aldehyde ferredoxin oxidoreductase | 16 |  | 2 |  |
|  |  | A4XJY0_CALS8 | Acyl-CoA dehydrogenase domain protein | 194 |  | 2 |  |
|  |  | A4XKL0_CALS8<br>B9MRU1_ANATD<br>D9TKM2_CALOO<br>E4Q190_CALOW<br>E4QAX0_CALH1<br>E4S7P7_CALKI<br>E4SBY9_CALK2<br>G2PTE4_9FIRM | NADH dehydrogenase (Quinone) | 88 |  | 9 |  |
|  |  | A4XKL2_CALS8 | Pyruvate-flavodoxin oxidoreductase | 266 |  | 2 |  |
|  |  | ATPB_CALS8<br>D9TK97_CALOO<br>E4Q1M6_CALOW<br>E4QA59_CALH1<br>E4S8R4_CALKI<br>E4SBM4_CALK2 | L-lactate dehydrogenase (Precursor) | 1142 |  | 3 |  |
|  |  | A4XKX2_CALS8<br>D9TK95_CALOO<br>E4QA57_CALH1<br>E4S8R6_CALKI<br>E4SBM2_CALK2<br>G2PTY0_9FIRM | D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding | 14 |  | 5 |  |
|  |  | A4XKY2_CALS8 | Conserved carboxylase region | 221 |  | 3 |  |
|  |  | A4XL40_CALS8<br>B9MJT3_ANATD<br>E4Q9J2_CALH1<br>E4SBF5_CALK2 | Hydrogenase, Fe-only | 64 |  | 12 |  |
|  |  | B9MJT4_ANATD<br>D9TK27_CALOO<br>E4Q213_CALOW<br>E4Q9J1_CALH1<br>E4S9B2_CALKI<br>E4SBF4_CALK2<br>G2PUC7_9FIRM | O-acetylhomoserine/O-acetylserine sulfhydrylase | 294 |  | 10 |  |
|  |  | B9MKE5_ANATD<br>E4SAI7_CALKI<br>E4SB95_CALK2<br>G2PWI7_9FIRM | ABC transporter related protein | 34 |  | 2 |  |
|  |  | B9MNM5_ANATD<br>D9TI79_CALOO<br>E4Q8A0_CALH1<br>E4S8K5_CALKI<br>E4SH27_CALK2<br>G2PVH2_9FIRM | Pyruvate flavodoxin/ferredoxin oxidoreductase domain protein | 26 |  | 8 |  |
|  |  | B9MQI2_ANATD<br>D9TJ88_CALOO<br>E4Q577_CALOW<br>E4QDF8_CALH1<br>E4S4G6_CALKI<br>E4SG03_CALK2<br>G2PXR6_9FIRM | Succinyl-CoA ligase [ADP-forming] subunit alpha | 58 |  | 9 |  |
|  |  | B9MQN0_ANATD<br>D9TJD7_CALOO<br>E4Q5J1_CALOW<br>E4SAA5_CALKI<br>E4SFV4_CALK2<br>G2PTN0_9FIRM | Aldo/keto reductase | 721 |  | 4 |  |
|  |  | B9MQN3_ANATD<br>D9TJE0_CALOO<br>E4Q5J4_CALOW<br>E4QD30_CALH1<br>E4SAA8_CALKI | Glutamate synthase (NADPH), homotetrameric | 103 |  | 3 |  |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | E4SFV1_CALK2<br>G2PTM7_9FIRM | | | | | |
| | | B9MRM8_ANATD<br>D9TKT4_CALOO<br>E4Q127_CALOW<br>E4S757_CALKI<br>E4SC89_CALK2<br>G2PT10_9FIRM | Aldehyde ferredoxin oxidoreductase | 16 | | 2 | |
| | | B9MRU4_ANATD<br>E4SBY6_CALK2 | Pyruvate/ketoisovalerate oxidoreductase, gamma subunit | 47 | | 7 | |
| | | D9TFN5_CALOO<br>E4Q426_CALOW<br>E4Q783_CALH1<br>E4SB24_CALK2 | Acyl-CoA dehydrogenase domain protein | 194 | | 2 | |
| | | D9TH02_CALOO<br>E4Q6M4_CALOW<br>E4QA16_CALH1<br>E4S4V8_CALKI<br>G2PV47_9FIRM | NADH dehydrogenase (Quinone) | 88 | | 9 | |
| | | D9TKJ8_CALOO<br>E4Q1D9_CALOW | Pyruvate-flavodoxin oxidoreductase | 266 | | 2 | |
| | | D9TKL8_CALOO | L-lactate dehydrogenase (Precursor) | 1142 | | 3 | |
| | | E4Q3J2_CALOW<br>E4Q7V0_CALH1 | D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding | 14 | | 5 | |
| | | E4S7W6_CALKI<br>G2PV36_9FIRM | ABC transporter related protein | 34 | | 2 | ABC transporter |
| | | E4SAA7_CALKI<br>G2PTM8_9FIRM | 4Fe-4S ferredoxin iron-sulfur binding domain-containing protein | 32 | | 2 | |
| | | E4SE14_CALK2 | O-acetylhomoserine/O-acetylserine sulfhydrylase | 294 | | 10 | |
| [G] Carbohydrate transport and metabolism | 37 | A4XGJ4_CALS8 | Xylose isomerase domain protein TIM barrel | 19 | | 2 | Carbohydrate isomerization |
| | | A4XH45_CALS<br>B9MMU5_ANATD<br>D9TGM8_CALOO<br>E4Q679_CALOW<br>E4QAN5_CALH1<br>E4S7B2_CALKI<br>E4SD78_CALK2<br>G2PVP6_9FIRM | Phosphoglucosamine mutase | 46 | | 2 | |
| | | A4XHC6_CALS8<br>D9TIQ5_CALOO<br>E4Q4B1_CALOW | Extracellular solute-binding protein, family 1 (Precursor) | 8 | Yes | 3 | Membrane sugar ABC trasporter |
| | | A4XHV8_CALS8 | L-fucose isomerase like protein | 83 | | 4 | |
| | | A4XIR8_CALS8<br>B9MPJ5_ANATD<br>D9TIS4_CALOO<br>E4Q4C9_CALOW<br>E4QEB5_CALH1<br>E4S9Z4_CALKI<br>E4SGG3_CALK2<br>G2PU00_9FIRM | Sugar-phosphate isomerase, RpiB/LacA/LacB family | 495 | | 2 | |
| | | A4XKI1_CALS8 | Pyruvate kinase | 26 | | 3 | |
| | | A4XKM4_CALS8<br>B9MRV5_ANATD<br>D9TKK8_CALOO<br>E4QAV6_CALH1<br>E4S7R1_CALKI<br>E4SBX5_CALK2<br>G2PTF8_9FIRM | Methylglyoxal synthase | 184 | | 4 | |
| | | A4XKU9_CALS8<br>B9MS46_ANATD<br>E4Q1K5_CALOW<br>E4QA80_CALH1<br>E4SBP5_CALK2 | Uronate isomerase | 373 | | 2 | |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | A4XKV3_CALS8 | Glyceraldehyde-3-phosphate dehydrogenase, type I | 3800 |  | 2 |  |
|  |  | A4XKV5_CALS8 | Pyruvate, phosphate dikinase | 1120 |  | 2 |  |
|  |  | A4XL36_CALS8 | Probable transaldolase | 549 |  | 7 |  |
|  |  | A4XLY9_CALS8<br>B9MKN8_ANATD<br>D9TFC9_CALOO<br>E4Q3E5_CALOW<br>E4Q7N2_CALH1<br>E4S4F5_CALKI<br>E4SDH3_CALK2<br>G2PZ28_9FIRM | ABC transporter related protein | 1617 |  | 2 | Sugar ABC transporter |
|  |  | A4XME8_CALS8<br>E4S7X0_CALKI | D-xylose ABC transporter, periplasmic substrate-binding protein (Precursor) | 18 |  | 3 | Sugar xylose ABC transporter |
|  |  | A4XN28_CALS8<br>B9MLB5_ANATD<br>D9TFV3_CALOO<br>E4Q4G6_CALOW<br>E4QE34_CALH1<br>E4S9M6_CALKI<br>E4SEW2_CALK2<br>F7KFZ8_9FIRM | Glycoside hydrolase, family 4 | 45 |  | 2 | Carbohydrate hydrolysis |
|  |  | A4XLT8_CALS8<br>B9MKI8_ANATD<br>D9TJF9_CALOO<br>E4Q392_CALOW<br>E4Q808_CALH1<br>E4SAT6_CALKI<br>E4SDT1_CALK2<br>G2PWN0_9FIRM | Phosphoglucomutase | 8 | Yes | 3 |  |
|  |  | B9MMN7_ANATD<br>E4Q5R5_CALOW<br>E4QCD6_CALH1<br>E4SE02_CALK2 | Periplasmic binding protein/LacI transcriptional regulator (Precursor) | 5 | Yes | 3 | Sugar ABC transporter |
|  |  | B9MN93_ANATD<br>D9TGX7_CALOO<br>E4Q6A9_CALOW<br>E4QAK4_CALH1 | Glycoside hydrolase family 3 domain protein | 38 |  | 2 | Carbohydrate hydrolysis |
|  |  | B9MNC8_ANATD<br>D9TH04_CALOO<br>E4Q6M6_CALOW<br>E4QA14_CALH1<br>E4S4W0_CALKI<br>E4SCS5_CALK2<br>G2PV45_9FIRM | Ketose-bisphosphate aldolase | 8 | Yes | 2 |  |
|  |  | B9MPD2_ANATD<br>E4SGN3_CALK2 | Xylulokinase | 41 |  | 5 | Carbohydrate degradation |
|  |  | B9MPI4_ANATD<br>D9TIR2_CALOO<br>E4Q4B8_CALOW<br>E4QEC6_CALH1<br>E4S9X9_CALKI<br>E4SGH4_CALK2<br>G2PU12_9FIRM | Fructose-1,6-bisphosphate aldolase, class II | 91 |  | 8 |  |
|  |  | B9MS47_ANATD<br>E4S8B8_CALKI<br>E4SBP4_CALK2<br>G2PTV8_9FIRM | Enolase | 3002 |  | 15 |  |
|  |  | B9MS49_ANATD<br>D9TKB5_CALOO<br>E4Q1K8_CALOW<br>E4QA77_CALH1,<br>E4S8C0_CALKI | Triosephosphate isomerase | 932 |  | 8 |  |
|  |  | B9MS50_ANATD<br>E4Q1K9_CALOW<br>E4QA76_CALH1<br>E4S8C1_CALKI<br>E4SBP1_CALK2<br>G2PTW1_9FIRM | Glyceraldehyde-3-phosphate dehydrogenase, type I | 3800 |  | 11 |  |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | B9MS53_ANATD<br>E4S8C3_CALKI<br>E4SBN9_CALK2 | Pyruvate, phosphate dikinase | 1120 |  | 23 |  |
|  |  | D9TFE9_CALOO<br>E4Q7L2_CALH1<br>E4S4D5_CALKI<br>E4SDF3_CALK2<br>G2PZ48_9FIRM | Phosphofructokinase | 22 |  | 8 |  |
|  |  | D9TFY6_CALOO | Transketolase central region | 202 |  | 4 |  |
|  |  | D9TH03_CALOO | Xylose isomerase domain-containing protein TIM barrel | 11 |  | 2 | Carbohydrate isomerization |
|  |  | D9TIR0_CALOO<br>E4Q4B6_CALOW<br>E4QEC8_CALH1<br>E4S9X7_CALKI<br>E4SGH6_CALK2<br>G2PU14_9FIRM<br>Q44407_ANATD | Bifunctional phosphoglucose/phosphomannose isomerase | 11 |  | 9 |  |
|  |  | D9TKQ2_CALOO<br>E4Q160_CALOW | 6-phosphofructokinase | 88 |  | 6 |  |
|  |  | E4Q4K1_CALOW<br>E4SEK4_CALK2 | Transketolase domain-containing protein | 111 |  | 5 |  |
|  |  | E4Q988_CALH1<br>E4S6H5_CALKI | Extracellular solute-binding protein family 1 | 6 | Yes | 3 | Membrane sugar ABC transporter |
|  |  | E4QED3_CALH1 | Extracellular solute-binding protein family 1 (Precursor) | 8 | Yes | 2 | Membrane sugar ABC transporter |
|  |  | E4S6B0_CALKI<br>G2PWX9_9FIRM | Glycosyltransferase 36 | 8 | Yes | 2 | Carbohydrate hydrolysis |
|  |  | E4S7W9_CALKI | ABC transporter related protein | 11 |  | 2 | ABC transporter |
|  |  | E4SCW4_CALK2 | Glycoside hydrolase family 3 domain protein | 38 |  | 7 | Carbohydrate hydrolysis |
|  |  | G2PWE2_9FIRM | Glycoside hydrolase family 10 | 4 |  | 2 | Carbohydrate hydrolysis |
|  |  | G2PXK4_9FIRM | Extracellular solute-binding protein family 1 (Precursor) | 8 | Yes | 5 | Membrane sugar ABC transporter |
| [E] Amino acid transport and metabolism | 44 | A4XGZ2_CALS8<br>A4XH19_CALS8 | Glutamate synthase (NADPH), homotetrameric | 144 |  | 10 |  |
|  |  | A4XGZ5_CALS8<br>A4XH22_CALS8<br>B9MN96_ANATD<br>D9TGY0_CALOO<br>E4Q6B1_CALOW<br>E4QAJ5_CALH1<br>E4S7F6_CALKI<br>E4SCW0_CALK2<br>G2PV67_9FIRM | Glutamate synthase, alpha subunit domain protein | 16 |  | 2 |  |
|  |  | A4XGZ1_CALS8<br>A4XH18_CALS8 | Glutamine amidotransferase, class-II | 66 |  | 3 |  |
|  |  | A4XHI0_CALS8<br>D9TIF9_CALOO<br>E4Q3K1_CALOW<br>E4Q7U1_CALH1<br>E4SAW9_CALKI<br>G2PX53_9FIRM | Pyruvate carboxyltransferase | 79 |  | 4 |  |
|  |  | A4XHI5_CALS8 | Isocitrate dehydrogenase (NAD(+)) | 168 |  | 2 |  |
|  |  | A4XHR8_CALS8 | Branched-chain amino acid aminotransferase | 68 |  | 3 |  |
|  |  | A4XI14_CALS8<br>D9TJ50_CALOO<br>E4Q545_CALOW<br>E4QDI9_CALH1<br>E4S4J7_CALKI<br>G2PXU8_9FIRM | Diaminopimelate decarboxylase | 10 |  | 2 |  |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | A4XI70_CALS8<br>E4Q476_CALOW<br>E4QE56_CALH1<br>E4S3W8_CALKI<br>E4SEY5_CALK2<br>G2PT39_9FIRM | Aspartate kinase | 41 | | 10 | |
| | | A4XI72_CALS8 | Homoserine dehydrogenase | 32 | | 4 | |
| | | A4XIB1_CALS8 | Extracellular solute-binding protein, family 5 (Precursor) | 1 | Yes | 7 | Membrane oligopeptide ABC transporter |
| | | A4XIL9_CALS8<br>B9MNV4_ANATD<br>D9TIE9_CALOO<br>E4Q3I9_CALOW<br>E4Q7V3_CALH1<br>E4S436_CALKI<br>E4SGV6_CALK2<br>G2PX64_9FIRM | 2-isopropylmalate synthase/homocitrate synthase family protein | 139 | | 4 | |
| | | A4XJN8_CALS8 | Acetylglutamate kinase | 53 | | 2 | |
| | | A4XJT1_CALS8 | O-acetylhomoserine/O-acetylserine sulfhydrylase | 294 | | 3 | |
| | | A4XJU0_CALS8<br>B9MRD1_ANATD<br>D9TL06_CALOO<br>E4Q6T1_CALOW<br>E4QBU9_CALH1<br>E4S681_CALKI<br>E4SF91_CALK2<br>G2PYX8_9FIRM | Dihydrodipicolinate synthase | 86 | | 3 | |
| | | A4XK88_CALS8<br>E4Q8L3_CALH1<br>E4SA35_CALKI<br>G2PVE7_9FIRM | Cysteine synthase A | 2433 | | 3 | |
| | | A4XKG4_CALS8<br>B9MRP5_ANATD<br>E4Q144_CALOW<br>E4QBA0_CALH1<br>E4S775_CALKI<br>E4SC72_CALK2<br>G2PT27_9FIRM | Argininosuccinate synthase | 1473 | | 2 | |
| | | A4XKM9_CALS8 | Spermidine synthase | 13 | | 2 | |
| | | A4XKU3_CALS8 | Extracellular ligand-binding receptor (Precursor) | 19 | | 3 | Membrane amino acid transporter |
| | | A4XLU2_CALS8 | Peptidase M18, aminopeptidase I | 8 | Yes | 2 | Peptidase (M18) |
| | | A4XMY2_CALS8 | Phospho-2-dehydro-3-deoxyheptonate aldolase | 34 | | 4 | |
| | | B9ML21_ANATD<br>E4SF04_CALK2 | Extracellular solute-binding protein family 3 (Precursor) | 8 | Yes | 2 | Membrane amino acid ABC transporter |
| | | B9MLC1_ANATD<br>E4QDP2_CALH1<br>E4S9M5_CALKI<br>E4SEN1_CALK2<br>G2PU54_9FIRM | Ferredoxin-dependent glutamate synthase | 22 | | 6 | |
| | | B9MLV4_ANATD<br>D9TG27_CALOO<br>E4Q4W1_CALOW<br>E4QDA0_CALH1<br>E4S9H8_CALKI<br>G2PUH9_9FIRM | 3-isopropylmalate dehydratase large subunit | 333 | | 3 | |
| | | B9MNV1_ANATD<br>E4SGV9_CALK2 | Ketol-acid reductoisomerase | 1288 | | 11 | |
| | | B9MPK7_ANATD<br>D9TIT6_CALOO<br>E4QCE3_CALH1<br>E4QEA3_CALH1<br>E4SA06_CALKI | Glutamate synthase (NADPH), homotetrameric | 103 | | 3 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | E4SGF1_CALK2<br>G2PTQ8_9FIRM | | | | | |
| | | B9MQK7_ANATD<br>D9TJB3_CALOO<br>E4SFX8_CALK2 | Histidinol-phosphate aminotransferase 1 | 21 | | 2 | |
| | | B9MR76 _ANATD | N-acetyl-gamma-glutamyl-phosphate reductase | 10 | | 2 | |
| | | B9MRD2_ANATD<br>E4Q6T2_CALOW<br>E4QBU8_CALH1<br>E4SF90_CALK2 | Aspartate-semialdehyde dehydrogenase | 664 | | 2 | |
| | | B9MRG2_ANATD<br>E4QBR9_CALH1 | Cysteine synthase | 2433 | | 4 | |
| | | D9TIE5_CALOO | Acetolactate synthase, small subunit | 356 | | 3 | |
| | | D9TIV8_CALOO<br>E4Q4G3_CALOW<br>E4QE08_CALH1 | Acetolactate synthase, large subunit, biosynthetic type | 535 | | 4 | |
| | | E4Q2D7_CALOW<br>E4Q8Z4_CALH1<br>E4S9S4_CALKI | Peptidase M29 aminopeptidase II | 100 | | 2 | Peptidase (M29) |
| | | E4Q2J8_CALOW<br>E4Q8I6_CALH1 | Gamma-glutamyl phosphate reductase | 28 | | 4 | |
| | | E4Q5F3_CALOW | Pyruvate carboxyltransferase | 79 | | 7 | |
| | | E4Q6S2_CALOW | O-acetylhomoserine/O-acetylserine sulfhydrylase | 294 | | 10 | |
| | | E4Q7V5_CALH1 | 2-isopropylmalate synthase | 241 | | 9 | |
| | | E4Q8X3_CALH1 | Oligoendopeptidase F (Precursor) | 1 | Yes | 3 | Peptidase |
| | | E4QEA4_CALH1 | Glutamine synthetase, type I | 884 | | 10 | |
| | | E4S411_CALKI | Extracellular solute-binding protein family 5 (Precursor) | 3 | Yes | 2 | Membrane sugar ABC transporter |
| | | E4S991_CALKI<br>G2PUJ6_9FIRM | Diaminopimelate epimerase | 22 | | 4 | |
| | | E4S9E7_CALKI<br>G2PUP0_9FIRM | Dihydroxy-acid dehydratase | 498 | | 4 | |
| | | E4SGQ5_CALK2 | Aminotransferase class I and II | 22 | | 2 | |
| | | G2PZD6_9FIRM | ABC-type transporter, periplasmic subunit (Precursor) | 4 | Yes | 2 | Membrane ABC transporter |
| | | Q44408_ANATD | Probable phosphatase Athe_0620 | 34 | | 3 | |
| [F] Nucleotide transport and metabolism | 4 | A4XML6_CALS8<br>B9MLK0_ANATD<br>E4Q9B7_CALH1<br>G2PY08_9FIRM | Adenylosuccinate synthetase | 3 | | 2 | |
| | | A4XIS2_CALS8<br>B9MPJ9_ANATD<br>D9TIS8_CALOO<br>E4Q4D3_CALOW<br>E4QEB1_CALH1<br>E4S9Z8_CALKI<br>E4SGF9_CALK2<br>G2PTZ6_9FIRM | Ribose-phosphate pyrophosphokinase | 604 | | 2 | |
| | | D9TFR9_CALOO<br>E4QE65_CALH1<br>E4S3X7_CALKI<br>G2PZH1_9FIRM | Adenine phosphoribosyltransferase | 372 | | 2 | |
| | | E4Q2J3_CALOW<br>E4Q8J0_CALH1<br>\|E4SA58_CALKI<br>G2PVW6_9FIRM | Inosine-5'-monophosphate dehydrogenase | 2008 | | 6 | |
| [H] Coenzyme transport and metabolism | 3 | G2PZD2_9FIRM<br>B9MKY9_ANATD<br>E4Q791_CALH1<br>E4SB33_CALK2 | Putative uncharacterized protein | 304 | | 2 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | E4S415_CALKI<br>D9TFM6_CALOO<br>E4Q417_CALOW<br>A4XIB5_CALS8 | | | | | |
| | | A4XHT8_CALS8<br>B9MQ10_ANATD<br>D9TIW8_CALOO<br>E4Q4N6_CALOW<br>E4QE02_CALH1<br>E4S9D8_CALKI<br>E4SGB5_CALK2<br>G2PUN4_9FIRM | S-adenosylmethionine synthase | 1490 | | 4 | |
| | | B9ML78B_ANATD<br>D9TFT8_CALOO<br>E4Q486_CALOW<br>E4QE46_CALH1<br>E4SAP3_CALKI<br>E4SEW6_CALK2<br>G2PT53_9FIRM | 3-methyl-2-oxobutanoate hydroxymethyltransferase | 374 | | 2 | |
| [I] Lipid transport and metabolism | 3 | A4XKN1_CALS8<br>B9MRW2_ANATD<br>D9TKK1_CALOO<br>E4Q1D6_CALOW<br>E4QAU9_CALH1<br>E4S7R9_CALKI<br>E4SBW8_CALK2<br>G2PTG5_9FIRM | Carboxyl transferase | 252 | | 2 | |
| | | D9TKY2_CALOO<br>E4Q6V4_CALOW | 3-oxoacyl-[acyl-carrier-protein] synthase 2 | 315 | | 3 | |
| | | E4QD00_CALH1 | FMN-dependent NADH-azoreductase | 26 | | 2 | |
| [P] Inorganic ion transport and metabolism | 4 | A4XFW7_CALS8 | TrkA-N domain protein | 8 | Yes | 2 | |
| | | A4XGR4_CALS8<br>D9TIA8_CALOO<br>E4Q802_CALH1<br>E4S6B5_CALKI<br>E4SGZ6_CALK2<br>G2PWX5_9FIRM | Hemerythrin-like metal-binding protein | 9 | | 3 | |
| | | A4XL29_CALS8<br>B9MJS2_ANATD<br>D9TK39_CALOO<br>E4Q201_CALOW<br>E4Q9K2_CALH1<br>E4S8X2_CALKI<br>E4SBG6_CALK2<br>G2PUB2_9FIRM | Phosphate uptake regulator, PhoU | 9 | Yes | 3 | |
| | | B9ML80_ANATD<br>E4Q488_CALOW<br>E4QE43_CALH1<br>E4SAP1_CALKI<br>E4SEW4_CALK2<br>G2PT55_9FIRM | Ferritin Dps family protein | 14 | | 2 | |
| [Q] Secondary metabolites biosynthesis, transport and catabolism | 1 | A4XM78_CALS8 | Acetyl xylan esterase | 23 | | 4 | Carbohydrate esterase |
| POORLY CHARACTERIZED | | | | | | | |
| [R] General function prediction only | 10 | A4XH40_CALS8 | Oxidoreductase domain protein | 11 | | 2 | |
| | | A4XL44_CALS8 | Protein hfq | 13 | | 3 | |
| | | A4XL50_CALS8 | Amino acid-binding ACT domain protein | 6 | | 2 | |
| | | A4XLW7_CALS8<br>B9MKL6_ANATD<br>D9TFA7_CALOO<br>E4Q3C1_CALOW<br>E4Q7Q4_CALH<br>E4SDQ3_CALK2 | ABC transporter related protein | 548 | | 2 | ABC transporter, ATPase-AAA |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | B9MKL8_ANATD<br>E4Q7Q2_CALH1<br>E4S4P3_CALKI<br>E4SDQ1_CALK2 | Basic membrane lipoprotein (Precursor) | 21 | | 2 | |
| | | B9MLL8_ANATD | Appr-1-p processing domain protein | 15 | | 2 | |
| | | B9MNS6_ANATD<br>E4Q7W6_CALH1<br>E4S457_CALKI<br>E4SGX3_CALK2 | Small GTP-binding protein | 90 | | 2 | |
| | | E4Q4V8_CALOW | Aldo/keto reductase | 56 | | 2 | |
| | | E4S812_CALKI<br>E4SC32_CALK2<br>G2PY07_9FIRM | Aldo/keto reductase | 3 | | 2 | |
| | | E4SE63_CALK2 | Radical SAM domain protein | 17 | Yes | 5 | |
| [S] Function unknown | 5 | A4XIP9_CALS8<br>B9MN39_ANATD | Extradiol ring-cleavage dioxygenase, class III enzyme, subunit B | 7 | | 2 | |
| | | A4XJ14_CALS8 | DegV family protein | 9 | | 2 | |
| | | Y1314_CALS8 | UPF0210 protein Csac_1314 | 804 | | 2 | |
| | | A4XKU4_CALS8<br>E4QA86_CALH1<br>E4S8B2_CALKI | Putative uncharacterized protein | 10 | | 2 | |
| | | A4XLC4_CALS8 | Putative uncharacterized protein | 81 | | 2 | |
| COG non hits | 7 | A4XGA4_CALS8<br>B9MP68_ANATD<br>D9TH33_CALOO<br>E4Q6P6_CALOW<br>E4S6A1_CALKI<br>G2PWY8_9FIRM | Putative uncharacterized protein | 17 | | 2 | |
| | | A4XIF5_CALS8<br>P22534_CALSA | Glycoside hydrolase, family 48 (Precursor) | 90 | | 3 | Carbohydrate hydrolysis |
| | | A4XKH8_CALS8<br>B9MRQ9_ANATD<br>D9TKQ4_CALOO<br>E4Q158_CALOW<br>E4QB87_CALH1<br>E4S788_CALKI<br>E4SC59_CALK2<br>G2PTB2_9FIRM | Tryptophan RNA-binding attenuator protein | 103 | | 2 | |
| | | B9MQ80_ANATD<br>D9TJ19_CALOO<br>E4Q4T8_CALOW<br>E4QDV2_CALH1<br>E4SG65_CALK2 | Putative uncharacterized protein | 8 | Yes | 3 | |
| | | E4Q3S3_CALOW<br>E4Q7A9_CALH1 | Putative uncharacterized protein | 8 | Yes | 3 | |
| | | E4QC28_CALH1 | S-layer domain protein (Precursor) | 3 | Yes | 2 | |
| | | G2PVR9_9FIRM | S-layer domain-containing protein (Precursor) | 4 | Yes | 3 | |

*9FIRM: *Firmicutes* (including *Caldicellulosiruptor lactoaceticus)*; ANATD: *Caldicelulosiruptor becscii* (*Anaerocellum thermophilum*); CALH1: *Caldicellulosiruptor hydrothermalis*; CALK2: *Caldicellulosiruptor kronotskyensis*; CALKI: *Caldicellulosiruptor kristjanssonii*; CALS8: *Caldicellulosiruptor saccharolyticus*; CALSA: *Caldicellulosiruptor saccharolyticus*; CALOO: *Caldicellulosiruptor obsidiansis*; CALOW: *Caldicellulosiruptor owensensis.*

## References

Amann RI, Krumholz L, Stahl DA (1990). Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* **172:** 762-770.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S *et al* (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315:** 1709-1712.

Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME J* **5:** 908-917.

Bhaya D, Davison M, Barrangou R (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annual review of genetics 45: 273-297.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37:** D233-238.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* **7:** 335-336.

Cardinale M, Brusetti L, Quatrini P, Borin S, Puglia AM, Rizzi A *et al* (2004). Comparison of different primer sets for use in automated ribosomal intergenic spacer analysis of complex bacterial communities. *Applied and Environmental Microbiology* **70:** 6147-6156.

Chenau J, Michelland S, Sidibe J, Seve M. (2008) Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. *Proteome Science* **6**:9.

Criscuolo A, Gribaldo S (2010). BMGE (Block Mapping and Gathering with Entropy): selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology* **10**:210.

Daims H, Brühl A, Amann R, Schleifer K-H, Wagner M (1999). The domain-specific probe EUB338 is insufficient for the detection of all bacteria: development and evaluation of a more comprehensive probe set. *Systematic and Applied Microbiology* **22:** 434-444.

Fenyö D, Beavis RC (2003). A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Analytical Chemistry* **75:** 768-774.

Gantner S, Andersson AF, Alonso-Sáez L, Bertilsson S (2011). Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. *Journal of Microbiological Methods* **84:** 12-18.

Gerardi MH (2003). *The Microbiology of Anaerobic Digesters*. John Wiley & Sons, Inc.

Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS *et al* (2006). Metagenomic analysis of the human distal gut microbiome. *Science* **312:** 1355-1359.

Gold ND, Martin VJJ (2007). Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis. *Journal of Bacteriology* **189:** 6787-6795.

Guindon S, Gascuel O (2003). PhyML: “A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood”, *Systematic Biology* **52**:696-704.

Hubner NC, Ren S, Mann M. (2008) Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* **8**:4862-72.

Ibrahim M, Nicolas P, Bessieres P, Bolotin A, Monnet V, Gardan R (2007). A genome-wide survey of short coding sequences in streptococci. *Microbiology* **153:** 3631-3644.

Käll L, Storey JD, MacCoss MJ, Noble WS (2008). Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *Journal of Proteome Research* **7:** 29-34.

Kumar PS, Brooker MR, Dowd SE, Camerlengo T (2011). Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing. *PLoS ONE* **6:** e20956.

Laloui-Carpentier W, Li T, Vigneron V, Mazeas L, Bouchez T (2006). Methanogenic diversity and activity in municipal solid waste landfill leachates. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **89:** 423-434.

Lane DJ (1991). 16S/23S rRNA sequencing. In: Stakebrandt E and Goodfellow M (eds). *Nucleic Acid Techniques in Bacterial Systematics*. Wiley: Chichester. pp 115-175.

Letunic I, Bork P (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* **39:** W475-W478.

Li TL, Mazeas L, Sghir A, Leblon G, Bouchez T (2009). Insights into networks of functional microbes catalysing methanization of cellulose under mesophilic conditions. *Environmental Microbiology* **11:** 889-904.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al* (2004). ARB: a software environment for sequence data. *Nucleic Acids Research* **32:** 1363-1371.

Manz W, Amann R, Ludwig W, Wagner M, Schleifer K-H (1992). Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria: problems and solutions. *Systematic and Applied Microbiology* **15:** 593-600.

Padmasiri SI, Zhang JZ, Fitch M, Norddahl B, Morgenroth E, Raskin L (2007). Methanogenic population dynamics and performance of an anaerobic membrane bioreactor (AnMBR) treating swine manure under high shear conditions. *Water Research* **41:** 134-144.

Qu X, Mazéas L, Vavilin VA, Epissard J, Lemunier M, Mouchel J-M *et al* (2009). Combined monitoring of changes in delta$^{13}$$CH_4$ and archaeal community structure during mesophilic methanization of municipal solid waste. *FEMS Microbiology Ecology* **68:** 236-245.

Rawlings ND, Barrett AJ, Bateman A (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **40:** D343-D350.

Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23:** 1282-1288.

Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J *et al* (2009). Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal* **3:** 179-189.

Verhees CH, Kengen SWM, Tuininga JE, Schut GJ, Adams MWW, De Vos WM *et al* (2003). The unique features of glycolytic pathways in Archaea. *Biochemical Journal* **375:** 231-246.

Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT *et al* (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:** 560-U17.

Wilmes P, Bond PL (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environmental Microbiology* **6:** 911-920.