Genome sequencing and assembly. The draft genome was generated at the U.S. Department of Energy Joint Genome Institute (JGI) using a combination of Illumina (Bennett 2004) and 454 technologies (Margulies et al. 2005). More specifically, an Illumina GAii shotgun library sequencing resulted in 62,371,762 reads adding up to 2,245.4 Mb. A 454 Titanium standard library produced 208,724 reads, and one paired-end 454 library with an average insert size of 9 kb generated 338,353 reads, resulting in 144.6 Mb of 454 data. Further description of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. The initial draft assembly contained 93 contigs in one scaffold. The 454 Titanium standard data and the 454 paired end data were assembled together with Newbler (version 2.3-PreRelease-6/30/2009). The Newbler consensus sequences were computationally shredded into 2 kb overlapping fake reads (shreds). Illumina sequencing data was assembled with VELVET, version 1.1.05 (Zerbino and Birney 2008), and the consensus sequence were computationally shredded into 1.5 kb shreds. The 454 Newbler consensus shreds, the Illumina VELVET consensus shreds and the read pairs in the 454 paired end library were integrated using parallel phrap, version 1.080812 (High Performance Software, LLC). The following finishing process was carried out: Illumina data was used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished); possible mis-assemblies were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher , or sequencing cloned bridging PCR fragments with subcloning; and gaps between contigs were closed by editing in Consed (Ewing and Green 1998; Ewing et al. 1998; Gordon, Abajian, and Green 1998), by PCR and by Bubble PCR (J-F Cheng, unpublished) primer walks. A total of 454 additional reactions and one shatter library were necessary to close gaps and to raise the quality of the finished sequence. The final assembly resulted in the 2,975,953 bp genome, and was based on 111.5 Mb of 454 draft data (~ 38.3x coverage) and 2,087.9 Mb of Illumina draft data (~716.4x coverage).

Genome annotation and data availability. Genes were identified using Prodigal (Hyatt et al. 2010), as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation with the JGI GenePRIMP pipeline (Pati et al. 2010). The predicted Open Reading Frames (ORFs) were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant (*nr*) database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to assert a product description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE (Lowe and Eddy 1997), RNAMMer (Lagesen et al. 2007), Rfam (Griffiths-Jones et al. 2003), TMHMM (Krogh et al. 2001), and signalP (Dyrløv Bendtsen et al. 2004). CRISPR Repeats were detected using CRISPR Repeat Tool 1.2 (Grissa and Vergnaud 2007). CRISPR-associated *Cas* genes were manually re-annotated using new classification scheme proposed in Makarova *et al*. (2011).

Supplementary Table 1. Annotation of CRISPR associated (Cas) genes according to (Makarova *et al.* 2011).

Locus tag	Annotation
Theba_2297	CRISPR-Cas system Type III, Cas5 group RAMP superfamily protein Csm4/Cmr3[a]
Theba_2298	CRISPR-Cas system Type III, predicted small subunit of Cmr complex, protein Cmr5
Theba_2299	CRISPR-Cas system Type III, Cas7 group RAMP superfamily protein Csm5/Cmr4
Theba_2300	CRISPR-Cas system Type III, Cas10 protein (inactivated Palm domain)[b]
Theba_2301	CRISPR-Cas system Type III, Cas7 group RAMP superfamily protein Csm3/Cmr6
Theba_2302	CRISPR-Cas system Type III, Cas7 group RAMP superfamily protein Cmr1
Theba_2304	CRISPR-Cas system subtype I-B, protein Cas2
Theba_2305	CRISPR-Cas system subtype I-B, protein Cas1
Theba_2306	CRISPR-Cas system subtype I-B, protein Cas4
Theba_2307	CRISPR-Cas system subtype I-B, protein Cas3
Theba_2308	CRISPR-Cas system subtype I-B, protein Cas5
Theba_2309	CRISPR-Cas system subtype I-B, protein Cas7
Theba_2310	CRISPR-Cas system subtype I-B, protein Cas8
Theba_2311	CRISPR-Cas system subtype I-B, protein Cas6
Theba_1605	CRISPR/Cas system-associated regulatory protein Csx1
Theba_1606	CRISPR/Cas system-associated regulatory protein Csx1 fused to HD nuclease and CYTH-like superfamily domains[c]
Theba_1607	DEAD/DEAH box helicase
Theba_1608	CRISPR-Cas system subtype III-B, Cas7 group RAMP superfamily protein Cmr6
Theba_1609	CRISPR-Cas system subtype III-B, predicted small subunit of Cmr complex, protein Cmr5
Theba_1610	CRISPR-Cas system subtype III-B, Cas7 group RAMP superfamily protein Cmr4
Theba_1611	CRISPR-Cas system subtype III-B, Cas5 group RAMP superfamily protein Cmr3
Theba_1612	CRISPR-Cas system subtype III-B, Cas10 protein
Theba_1613	CRISPR-Cas system subtype III-B, Cas7 group RAMP superfamily protein Cmr1
Theba_1614	CRISPR/Cas system-associated regulatory protein Csx1
Theba_1615	CRISPR-Cas system subtype I-B, protein Cas2
Theba_1618	CRISPR-Cas system subtype I-B, protein Cas1
Theba_1619	CRISPR-Cas system subtype I-B, protein Cas4
Theba_1620	CRISPR-Cas system subtype I-B, protein Cas3
Theba_1621	CRISPR-Cas system subtype I-B, protein Cas5
Theba_1622	CRISPR-Cas system subtype I-B, protein Cas7
Theba_1623	CRISPR-Cas system subtype I-B, protein Cas8

a) This is a variant of Type III also known as MTH326 module
b) Csx11 family with inactivated Palm domain.
c) Unique fusion

Supplementary Table 2. *Mesotoga prima* ORFs, predicted by Darkhorse to have been exchanged laterally with organisms outside the Thermotogales, that are found in clusters of two or more genes. This is probably an overestimate of the number of clusters, as many are separated from each other only by genes that did not have homologs in the *nr* database, and therefore could not be included in the DarkHorse analysis. 'Norm_LPI' refers to the normalized lineage probablity index reported by Darkhorse, and 'Cluster size' refers to the number of ORFs found in the cluster.

***Mesotoga prima* locus tag and annotation**	**norm_LPI**	**Top-scoring DarkHorse hit**	**Cluster size**
Theba_0044 Predicted dehydrogenases and related proteins	0.021	*Candidatus Caldiarchaeum subterraneum*	3
Theba_0045 Transposase and inactivated derivatives	0.626	*Finegoldia magna BVS033A4*	
Theba_0046 Predicted permease, DMT superfamily	0.587	*Desulfosporosinus sp. OT*	
Theba_0055 Permeases of the drug/metabolite transporter (DMT) superfamily	0.57	*Ornithinibacillus sp. TW25*	2
Theba_0056 NhaP-type Na+/H+ and K+/H+ antiporters	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_0060 Predicted metal-dependent hydrolase of the TIM-barrel fold	0.618	*Mahella australiensis 50-1 BON*	2
Theba_0061 hypothetical protein	0.557	*Chloroflexus aurantiacus J-10-fl*	
Theba_0063 Uroporphyrinogen-III decarboxylase	0.594	*Clostridium sp. D5*	6
Theba_0064 Uroporphyrinogen-III decarboxylase	0.517	*Dictyoglomus thermophilum H-6-12*	
Theba_0065 Uncharacterized metal-binding protein	0.587	*Desulfosporosinus sp. OT*	
Theba_0066 Vitamin B12 dependent methionine synthase, activation domain.	0.534	*Bilophila sp. 4_1_30*	
Theba_0067 Uroporphyrinogen-III decarboxylase	0.618	*Mahella australiensis 50-1 BON*	
Theba_0068 Domain of unknown function(DUF2779).	0.53	*Helicobacter bizzozeronii CIII-1*	
Theba_0078 Transcriptional regulator/sugar kinase	0.517	*Dictyoglomus thermophilum H-6-12*	7
Theba_0079 ABC-type sugar transport system, periplasmic component	0.568	*Desmospora sp. 8437*	
Theba_0080 ABC-type sugar transport systems, permease components	0.568	*Desmospora sp. 8437*	
Theba_0081 ABC-type sugar transport system, permease component	0.626	*Sulfobacillus acidophilus TPY*	
Theba_0082 Uncharacterized membrane protein	0.497	*Pseudonocardia dioxanivorans CB1190*	
Theba_0083 Signal transduction histidine kinase	0.571	*Bacillus megaterium WSH-002*	
Theba_0084 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	0.571	*Bacillus megaterium WSH-002*	
Theba_0095 Mannose-6-phosphate isomerase	0.026	*Methanosarcina mazei Go1*	6
Theba_0096 TIGR00341 family protein	0.026	*Methanosaeta concilii GP6*	
Theba_0097 Membrane transporters of cations and cationic drugs	0.576	*Halanaerobium praevalens DSM 2228*	
Theba_0098 Predicted acetamidase/formamidase	0.566	*Treponema phagedenis F0421*	
Theba_0099 Predicted integral membrane protein	0.532	*uncultured Desulfobacterium sp.*	
Theba_0100 Uncharacterized protein involved in an early stage of isoprenoid biosynthesis	0.535	*Serratia sp. AS12*	

Theba_0112 5"-nucleotidase/2",3"-cyclic phosphodiesterase and related esterases	0.531	*Rhodobacter sphaeroides WS8N*	2
Theba_0113 alkylhydroperoxidase AhpD family core domain	0.026	*Methanobacterium sp. AL-21*	
Theba_0127 Uncharacterized conserved protein	0.52	*Haliscomenobacter hydrossis DSM 1100*	5
Theba_0128 Predicted phosphoesterase or phosphohydrolase	0.025	*Methanocaldococcus vulcanius M7*	
Theba_0129 hypothetical protein	0.584	*Acetivibrio cellulolyticus CD2*	
Theba_0130 Uncharacterized conserved protein	0.594	*Clostridium cellulovorans 743B*	
Theba_0131 Domain of unknown function (DUF1905).	0.594	*Clostridium symbiosum WAL-14163*	
Theba_0138 Predicted helicase	0.607	*Microcoleus chthonoplastes PCC 7420*	4
Theba_0139 Predicted Zn peptidase	0.585	*Thermincola potens JR*	
Theba_0140 hypothetical protein	0.585	*Thermincola potens JR*	
Theba_0141 hypothetical protein	0.495	*Verrucosispora maris AB-18-032*	
Theba_0172 Transposase and inactivated derivatives	0.533	*Geobacter uraniireducens Rf4*	2
Theba_0173 PAS domain S-box	0.532	*Desulfurivibrio alkaliphilus AHT2*	
Theba_0189 Transposase	0.517	*Meiothermus silvanus DSM 9946*	4
Theba_0190 Penicillin V acylase and related amidases	0.535	*Desulfovibrio magneticus RS-1*	
Theba_0191 Predicted nucleotidyltransferases	0.577	*Ammonifex degensii KC4*	
Theba_0192 Uncharacterized conserved protein related to C-terminal domain of eukaryotic chaperone, SACSIN	0.526	*Anaerolinea thermophila UNI-1*	
Theba_0220 hypothetical protein	0.519	*Leadbetterella byssophila DSM 17132*	3
Theba_0221 hypothetical protein	0.532	*Syntrophobacter fumaroxidans MPOB*	
Theba_0222 hypothetical protein	0.519	*Mucilaginibacter paludis DSM 18603*	
Theba_0226 Peptidase M30.	0.571	*Bacillus coagulans 2-6*	2
Theba_0227 hypothetical protein	0.515	*Candidatus Solibacter usitatus Ellin6076*	
Theba_0231 Cysteine protease	0.519	*Mucilaginibacter paludis DSM 18603*	2
Theba_0233 Cytochrome c biogenesis protein	0.618	*Mahella australiensis 50-1 BON*	
Theba_0259 hypothetical protein	0.495	*Gordonibacter pamelaeae 7-10-1-b*	3
Theba_0260 Phage shock protein A (IM30), suppresses sigma54-dependent transcription	0.495	*Gordonibacter pamelaeae 7-10-1-b*	
Theba_0261 Formamidopyrimidine-DNA glycosylase	0.587	*Desulfosporosinus sp. OT*	
Theba_0408 Alpha-galactosidases/6-phospho-beta-glucosidases, family 4 of glycosyl hydrolases	0.579	*Thermoanaerobacter ethanolicus CCSD1*	2
Theba_0409 DNA-3-methyladenine glycosylase I	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_0435 ABC-type sugar transport systems, permease components	0.576	*Halanaerobium praevalens DSM 2228*	3
Theba_0436 ABC-type sugar transport system, permease component	0.566	*Spirochaeta smaragdinae DSM 11293*	
Theba_0437 ABC-type sugar transport system, periplasmic component	0.566	*Spirochaeta smaragdinae DSM 11293*	

Theba_0482 uracil-xanthine permease	0.594	*Clostridium botulinum BKT015925*	2
Theba_0483 Acetyltransferases, including N-acetylases of ribosomal proteins	0.579	*Thermoanaerobacter sp. X561*	
Theba_0488 hypothetical protein	0.525	*Pyramidobacter piscolens W5455*	2
Theba_0489 TRAP-type C4-dicarboxylate transport system, large permease component	0.525	*Pyramidobacter piscolens W5455*	
Theba_0491 Predicted deacylase	0.525	*Pyramidobacter piscolens W5455*	3
Theba_0492 Dipeptidase	0.525	*Dethiosulfovibrio peptidovorans DSM 11002*	
Theba_0493 Dipeptidase	0.525	*Dethiosulfovibrio peptidovorans DSM 11002*	
Theba_0504 Deoxyribodipyrimidine photolyase	0.583	*Heliobacterium modesticaldum Ice1*	3
Theba_0505 Uncharacterized conserved protein	0.026	*Methanosaeta concilii GP6*	
Theba_0506 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	0.57	*Ornithinibacillus sp. TW25*	
Theba_0554 malonyl CoA-acyl carrier protein transacylase	0.578	*Thermoanaerobacter tengcongensis MB4*	2
Theba_0555 Dioxygenases related to 2-nitropropane dioxygenase	0.578	*Thermoanaerobacter tengcongensis MB4*	
Theba_0586 Ribonuclease HI	0.626	*Anaerococcus hydrogenalis ACS-025-V-Sch4*	2
Theba_0587 D-aminopeptidase	0.594	*Clostridium difficile CD196*	
Theba_0591 FOG: WD40 repeat	0.607	*Arthrospira platensis NIES-39*	2
Theba_0592 Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.	0.574	*Natranaerobius thermophilus JW/NM-WN-LF*	
Theba_0595 Conserved protein/domain typically associated with flavoprotein oxygenases, DIM6/NTAB family	0.025	*Methanocella paludicola SANAE*	4
Theba_0596 Dipeptidase	0.525	*Aminomonas paucivorans DSM 12260*	
Theba_0597 magnesium Mg(2+) and cobalt Co(2+) transport protein (corA)	0.516	*Planctomyces brasiliensis DSM 5305*	
Theba_0598 PAS domain S-box/uncharacterized domain HDIG	0.535	*Desulfovibrio sp. FW1012B*	
Theba_0615 threonine synthase	0.025	*Methanococcus maripaludis X1*	2
Theba_0616 Transposase	0.585	*Desulfotomaculum kuznetsovii DSM 6115*	
Theba_0618 DNA replication protein	0.594	*Clostridium thermocellum JW20*	7
Theba_0619 Transposase and inactivated derivatives	0.594	*Clostridium thermocellum JW20*	
Theba_0620 Bacterial capsule synthesis protein PGA_cap.	0.563	*Brachyspira intermedia PWS/A*	
Theba_0621 PAS domain S-box/diguanylate cyclase (GGDEF) domain	0.526	*Anaerolinea thermophila UNI-1*	
Theba_0623 PAS domain S-box/uncharacterized domain HDIG	0.535	*Desulfovibrio sp. FW1012B*	
Theba_0625 hypothetical protein	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_0626 Uncharacterized protein conserved in bacteria	0.574	*Natranaerobius thermophilus JW/NM-WN-LF*	
Theba_0752 Methylase involved in ubiquinone/menaquinone biosynthesis	0.587	*Desulfosporosinus sp. OT*	4
Theba_0753 ABC-type sugar transport systems, permease components	0.573	*Paenibacillus sp. HGF5*	
Theba_0754 ABC-type sugar transport system, permease component	0.626	*Lachnospiraceae bacterium 3_1_57FAA_CT1*	

Theba_0755 ABC-type sugar transport system, periplasmic component	0.573	*Paenibacillus polymyxa M1*	
Theba_0771 histidinol phosphate phosphatase HisJ family	0.626	*Peptoniphilus sp. oral taxon 375 str. F0436*	2
Theba_0772 TIGR02453 family protein	0.533	*Geobacter bemidjiensis Bem*	
Theba_0854 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	0.531	*Methylovorus sp. MP688*	3
Theba_0855 PAS domain S-box	0.531	*Methyloversatilis universalis FAM5*	
Theba_0856 PAS domain S-box/uncharacterized domain HDIG	0.531	*Sideroxydans lithotrophicus ES-1*	
Theba_0896 ABC-type branched-chain amino acid transport systems, ATPase component	0.587	*Desulfosporosinus sp. OT*	2
Theba_0897 ABC-type branched-chain amino acid transport systems, ATPase component	0.587	*Desulfosporosinus sp. OT*	
Theba_0945 ABC-type sugar transport system, periplasmic component	0.518	*Oceanithermus profundus DSM 14977*	4
Theba_0946 ABC-type sugar transport systems, permease components	0.518	*Oceanithermus profundus DSM 14977*	
Theba_0947 ABC-type sugar transport system, permease component	0.518	*Oceanithermus profundus DSM 14977*	
Theba_0948 Uncharacterized protein conserved in bacteria	0.576	*Halothermothrix orenii H 168*	
Theba_0998 hypothetical protein	0.594	*Clostridium phytofermentans ISDg*	2
Theba_0999 urocanate hydratase	0.594	*Clostridium symbiosum WAL-14673*	
Theba_1002 Glycosyltransferase	0.563	*Carnobacterium sp. AT7*	2
Theba_1003 Fe-S oxidoreductase	0.594	*Clostridium sp. SY8519*	
Theba_1021 proline-specific peptidases, Bacillus coagulans-type subfamily	0.535	*Desulfovibrio sp. FW1012B*	7
Theba_1022 ABC-type sugar transport system, periplasmic component	0.594	*Clostridium bolteae ATCC BAA-613*	
Theba_1023 ABC-type sugar transport system, ATPase component	0.594	*Clostridium bolteae ATCC BAA-613*	
Theba_1024 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.594	*Clostridium bolteae ATCC BAA-613*	
Theba_1025 Sugar phosphate isomerases/epimerases	0.495	*Actinomyces sp. oral taxon 448 str. F0400*	
Theba_1026 Sugar phosphate isomerases/epimerases	0.515	*Coraliomargarita akajimensis DSM 45221*	
Theba_1027 Transposase and inactivated derivatives	0.626	*Finegoldia magna BVS033A4*	
Theba_1111 ABC-type multidrug transport system, ATPase and permease components	0.584	*Eubacterium limosum KIST612*	4
Theba_1112 ABC-type multidrug transport system, ATPase and permease components	0.619	*Thermoanaerobacterium xylanolyticum LX-11*	
Theba_1113 Transcriptional regulators	0.579	*Thermoanaerobacter wiegelii Rt8.B1*	
Theba_1114 putative toxin-antitoxin system toxin component, PIN family	0.579	*Thermoanaerobacter wiegelii Rt8.B1*	
Theba_1132 hypothetical protein	0.626	*Lachnospiraceae bacterium 3_1_57FAA_CT1*	2
Theba_1133 Uncharacterized conserved protein	0.583	*Oribacterium sinus F0268*	
Theba_1137 hypothetical protein	0.526	*Anaerolinea thermophila UNI-1*	2
Theba_1138 hypothetical protein	0.526	*Anaerolinea thermophila UNI-1*	
Theba_1146 Predicted nucleic acid-binding protein, contains PIN domain	0.517	*Pelodictyon phaeoclathratiforme BU-1*	8

Theba_1147 Uncharacterized conserved protein	0.554	*Fusobacterium sp. 3_1_5R*	
Theba_1148 Growth inhibitor	0.619	*Caldicellulosiruptor lactoaceticus 6A*	
Theba_1149 hypothetical protein	0.606	*Microcystis aeruginosa NIES-843*	
Theba_1150 Predicted periplasmic or secreted lipoprotein	0.553	*Candidatus Nitrospira defluvii*	
Theba_1151 Uncharacterised protein family (UPF0150).	0.517	*Thermus scotoductus SA-01*	
Theba_1152 Uncharacterized conserved protein	0.626	*Anaerococcus prevotii ACS-065-V-Col13*	
Theba_1153 hypothetical protein	0.597	*Eubacterium cylindroides T2-87*	
Theba_1157 Protein of unknown function (DUF2442).	0.517	*Pelodictyon phaeoclathratiforme BU-1*	2
Theba_1158 Cytotoxic translational repressor of toxin-antitoxin stability system	0.52	*Rhodothermus marinus SG0.5JP17-172*	
Theba_1160 hypothetical protein	0.626	*Lachnospiraceae bacterium 8_1_57FAA*	2
Theba_1161 Domain of unknown function (DUF1814).	0.626	*Peptoniphilus harei ACS-146-V-Sch2b*	
Theba_1167 Subtilisin-like serine proteases	0.574	*Natranaerobius thermophilus JW/NM-WN-LF*	7
Theba_1168 hypothetical protein	0.583	*Butyrivibrio proteoclasticus B316*	
Theba_1169 GDP-mannose 4,6-dehydratase	0.531	*Rhizobium etli CNPAF512*	
Theba_1170 S23 ribosomal protein.	0.52	*Paludibacter propionicigenes WB4*	
Theba_1171 Nucleoside-diphosphate-sugar epimerases	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_1172 Glycosyl transferase family 11.	0.519	*Prevotella sp. C561*	
Theba_1173 Glycosyltransferases involved in cell wall biogenesis	0.573	*Paenibacillus curdlanolyticus YK9*	
Theba_1187 Predicted periplasmic or secreted lipoprotein	0.533	*Geobacter uraniireducens Rf4*	4
Theba_1188 Uncharacterised protein family (UPF0150).	0.517	*Thermus scotoductus SA-01*	
Theba_1189 Uncharacterized conserved protein	0.584	*Faecalibacterium prausnitzii L2-6*	
Theba_1190 Predicted transcriptional regulator	0.525	*Anaerobaculum hydrogeniformans ATCC BAA-1850*	
Theba_1224 Predicted dehydrogenases and related proteins	0.517	*Dictyoglomus thermophilum H-6-12*	2
Theba_1225 ABC-type multidrug transport system, ATPase component	0.026	*Methanobacterium sp. AL-21*	
Theba_1259 Predicted CoA-binding protein	0.495	*Amycolatopsis mediterranei U32*	2
Theba_1260 Transposase	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_1285 hypothetical protein	0.554	*Fusobacterium sp. 3_1_5R*	4
Theba_1286 ABC-type sugar transport system, ATPase component	0.566	*Treponema denticola F0402*	
Theba_1287 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.566	*Spirochaeta sp. Buddy*	
Theba_1288 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.525	*Pyramidobacter piscolens W5455*	
Theba_1302 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	0.533	*Methylobacter tundripaludum SV96*	2
Theba_1303 Enterochelin esterase and related enzymes	0.707	*gamma proteobacterium NOR5-3*	

Theba_1309 hypothetical protein	0.594	*Clostridium saccharolyticum WM1*	2
Theba_1310 hypothetical protein	0.52	*Parabacteroides johnsonii DSM 18315*	
Theba_1314 Uncharacterized conserved protein	0.588	*Alkaliphilus metalliredigens QYMF*	12
Theba_1315 Uncharacterized protein, linocin/CFP29 homolog	0.611	*Haloplasma contractile SSD-17B*	
Theba_1316 isoleucyl-tRNA synthetase	0.587	*Desulfosporosinus sp. OT*	
Theba_1317 Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL homologs	0.525	*Aminomonas paucivorans DSM 12260*	
Theba_1318 Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homologs	0.525	*Aminomonas paucivorans DSM 12260*	
Theba_1319 Aerobic-type carbon monoxide dehydrogenase, middle subunit CoxM/CutM homologs	0.525	*Aminomonas paucivorans DSM 12260*	
Theba_1320 Predicted metal-dependent hydrolase	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_1321 Protein of unknown function (DUF1116).	0.525	*Aminomonas paucivorans DSM 12260*	
Theba_1323 Succinyl-CoA synthetase, alpha subunit	0.525	*Anaerobaculum hydrogeniformans ATCC BAA-1850*	
Theba_1324 Protein of unknown function (DUF2877).	0.525	*Aminomonas paucivorans DSM 12260*	
Theba_1325 Predicted acyl-CoA transferases/carnitine dehydratase	0.557	*Thermosinus carboxydivorans Nor1*	
Theba_1326 Predicted amidohydrolase	0.525	*Aminomonas paucivorans DSM 12260*	
Theba_1357 Putative protein-S-isoprenylcysteine methyltransferase	0.611	*Haloplasma contractile SSD-17B*	2
Theba_1358 5"-nucleotidase/2",3"-cyclic phosphodiesterase and related esterases	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_1368 Transcriptional regulators	0.573	*Paenibacillus mucilaginosus KNP414*	2
Theba_1369 Glutathione peroxidase	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_1412 Predicted dehydrogenases and related proteins	0.569	*Oceanobacillus iheyensis HTE831*	4
Theba_1413 Predicted dehydrogenases and related proteins	0.573	*Paenibacillus sp. HGF7*	
Theba_1414 Histidinol phosphatase and related hydrolases of the PHP family	0.566	*Spirochaeta coccoides DSM 17374*	
Theba_1415 Predicted transcriptional regulator, contains C-terminal CBS domains	0.57	*Ornithinibacillus sp. TW25*	
Theba_1420 3-phosphoshikimate 1-carboxyvinyltransferase	0.566	*Spirochaeta sp. Buddy*	3
Theba_1421 Uncharacterized protein conserved in bacteria	0.675	*butyrate-producing bacterium SS3/4*	
Theba_1423 Uncharacterized protein conserved in bacteria	0.675	*butyrate-producing bacterium SS3/4*	
Theba_1424 Sugar phosphate permease	0.525	*Aminomonas paucivorans DSM 12260*	4
Theba_1425 Transcriptional regulators	0.566	*Spirochaeta sp. Buddy*	
Theba_1426 NHL repeat./Redoxin.	0.607	*Microcoleus vaginatus FGP-2*	
Theba_1427 prepilin-type N-terminal cleavage/methylation domain	0.554	*Ilyobacter polytropus DSM 2926*	
Theba_1429 ABC-type sugar transport system, periplasmic component	0.566	*Treponema phagedenis F0421*	4
Theba_1430 ABC-type sugar transport systems, permease components	0.566	*Treponema phagedenis F0421*	
Theba_1431 ABC-type sugar transport system, permease component	0.566	*Treponema phagedenis F0421*	

Theba_1432 Cystathionine beta-lyases/cystathionine gamma-synthases	0.566	*Treponema phagedenis F0421*	
Theba_1434 Transcriptional regulator/sugar kinase	0.626	*Lachnospiraceae bacterium 3_1_57FAA_CT1*	5
Theba_1435 Phosphoheptose isomerase	0.53	*Campylobacter upsaliensis JV21*	
Theba_1436 hypothetical protein	0.566	*Treponema azotonutricium ZAS-9*	
Theba_1437 ADP-heptose synthase, bifunctional sugar kinase/adenylyltransferase	0.519	*Dyadobacter fermentans DSM 18053*	
Theba_1438 isoform II.	0.611	*Haloplasma contractile SSD-17B*	
Theba_1564 hypothetical protein	0.573	*Paenibacillus mucilaginosus KNP414*	3
Theba_1565 diguanylate cyclase (GGDEF) domain	0.517	*Chloroherpeton thalassium ATCC 35110*	
Theba_1566 diguanylate cyclase (GGDEF) domain	0.517	*Deinococcus deserti VCD115*	
Theba_1572 L-aminopeptidase/D-esterase	0.594	*Clostridium sp. HGF2*	2
Theba_1573 PAS domain S-box	0.531	*Dechlorosoma suillum PS*	
Theba_1585 TRAP-type C4-dicarboxylate transport system, large permease component	0.566	*Treponema phagedenis F0421*	3
Theba_1586 Predicted deacylase	0.525	*Pyramidobacter piscolens W5455*	
Theba_1587 Transposase	0.585	*Desulfotomaculum kuznetsovii DSM 6115*	
Theba_1589 Protein of unknown function (DUF2961).	0.531	*Sinorhizobium meliloti AK83*	2
Theba_1590 hypothetical protein	0.573	*Paenibacillus sp. Y412MC10*	
Theba_1592 ABC-type sugar transport system, permease component	0.531	*Sinorhizobium meliloti AK83*	6
Theba_1593 ABC-type sugar transport systems, permease components	0.563	*Abiotrophia defectiva ATCC 49176*	
Theba_1594 ABC-type sugar transport system, periplasmic component	0.563	*Abiotrophia defectiva ATCC 49176*	
Theba_1595 hypothetical protein	0.585	*Desulfotomaculum kuznetsovii DSM 6115*	
Theba_1596 hypothetical protein	0.584	*Dethiobacter alkaliphilus AHT 1*	
Theba_1597 looped-hinge helix DNA binding domain, AbrB family	0.584	*Dethiobacter alkaliphilus AHT 1*	
Theba_1599 Retron-type reverse transcriptase	0.571	*Bacillus pseudofirmus OF4*	2
Theba_1600 Uncharacterized conserved protein	0.534	*Pseudomonas fluorescens WH6*	
Theba_1605 CRISPR-associated protein (Cas_APE2256).	0.517	*Truepera radiovictrix DSM 17093*	3
Theba_1606 hypothetical protein	0.525	*Dethiosulfovibrio peptidovorans DSM 11002*	
Theba_1607 Superfamily II DNA helicase	0.579	*Thermoanaerobacter pseudethanolicus ATCC 33223*	
Theba_1618 CRISPR-associated endonuclease Cas1	0.553	*Sulfurihydrogenibium sp. YO3AOP1*	5
Theba_1619 CRISPR-associated protein Cas4	0.023	*Archaeoglobus veneficus SNP6*	
Theba_1620 CRISPR-associated helicase Cas3/CRISPR-associated endonuclease Cas3-HD	0.594	*Clostridium botulinum BKT015925*	
Theba_1621 CRISPR-associated protein Cas5, N-terminal domain	0.52	*Rhodothermus marinus SG0.5JP17-172*	
Theba_1622 CRISPR-associated protein Cas7/Cst2/DevR, subtype I-B/TNEAP	0.52	*Rhodothermus marinus SG0.5JP17-172*	

Theba_1643 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.579	*Thermoanaerobacter wiegelii Rt8.B1*	2
Theba_1644 ABC-type dipeptide transport system, periplasmic component	0.579	*Thermoanaerobacter wiegelii Rt8.B1*	
Theba_1651 Transcriptional regulators	0.563	*Weissella thailandensis fsh4-2*	4
Theba_1652 membrane complex biogenesis protein, BtpA family	0.535	*Erwinia billingiae Eb661*	
Theba_1653 membrane complex biogenesis protein, BtpA family	0.535	*Erwinia billingiae Eb661*	
Theba_1654 Sugar kinases, ribokinase family	0.569	*Lysinibacillus fusiformis ZC1*	
Theba_1656 ABC-type uncharacterized transport systems, ATPase components	0.563	*Brachyspira intermedia PWS/A*	3
Theba_1657 ABC-type uncharacterized transport system, permease component	0.675	*Clostridiales bacterium 1_7_47FAA*	
Theba_1658 Uncharacterized ABC-type transport system, permease component	0.526	*Anaerolinea thermophila UNI-1*	
Theba_1663 Retron-type reverse transcriptase	0.571	*Bacillus pseudofirmus OF4*	2
Theba_1664 Uncharacterized conserved protein	0.566	*Treponema primitia ZAS-2*	
Theba_1667 Predicted cobalamin binding protein	0.554	*Ilyobacter polytropus DSM 2926*	3
Theba_1668 diguanylate cyclase (GGDEF) domain	0.533	*Desulfohalobium retbaense DSM 5692*	
Theba_1669 Uncharacterized conserved protein	0.535	*Desulfovibrio fructosovorans JJ*	
Theba_1677 hypothetical protein	0.571	*Bacillus thuringiensis BMB171*	3
Theba_1678 Alkaline phosphatase	0.557	*Turicibacter sanguinis PC909*	
Theba_1679 Predicted multitransmembrane protein	0.535	*Desulfovibrio fructosovorans JJ*	
Theba_1689 Arylsulfatase A and related enzymes	0.535	*Escherichia fergusonii ECD227*	3
Theba_1690 anaerobic sulfatase-maturating enzyme	0.533	*Vibrio parahaemolyticus RIMD 2210633*	
Theba_1691 Transcriptional regulators	0.53	*Arcobacter sp. L*	
Theba_1693 Predicted transcriptional regulators	0.579	*Thermoanaerobacter italicus Ab9*	5
Theba_1694 ABC-type multidrug transport system, ATPase component	0.579	*Thermoanaerobacter italicus Ab9*	
Theba_1695 Predicted membrane protein	0.535	*Desulfovibrio aespoeensis Aspo-2*	
Theba_1696 diguanylate cyclase (GGDEF) domain	0.53	*Arcobacter sp. L*	
Theba_1697 Formate/nitrite family of transporters	0.585	*Desulfitobacterium hafniense DCB-2*	
Theba_1702 Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	0.531	*Agrobacterium tumefaciens F2*	6
Theba_1703 Uridine phosphorylase	0.027	*Thermococcus sp. 4557*	
Theba_1704 ABC-type sugar transport system, permease component	0.594	*Clostridium asparagiforme DSM 15981*	
Theba_1705 ABC-type sugar transport systems, permease components	0.594	*Clostridium leptum DSM 753*	
Theba_1706 ABC-type sugar transport system, periplasmic component	0.594	*Clostridium leptum DSM 753*	
Theba_1707 Transcriptional regulators	0.619	*Thermoanaerobacterium xylanolyticum LX-11*	
Theba_1720 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.579	*Thermoanaerobacter wiegelii Rt8.B1*	2

Theba_1721 ABC-type dipeptide transport system, periplasmic component	0.579	*Thermoanaerobacter wiegelii Rt8.B1*	
Theba_1728 Transcriptional regulators	0.563	*Weissella thailandensis fsh4-2*	4
Theba_1729 membrane complex biogenesis protein, BtpA family	0.535	*Erwinia billingiae Eb661*	
Theba_1730 membrane complex biogenesis protein, BtpA family	0.535	*Erwinia billingiae Eb661*	
Theba_1731 Sugar kinases, ribokinase family	0.569	*Lysinibacillus fusiformis ZC1*	
Theba_1733 ABC-type uncharacterized transport systems, ATPase components	0.563	*Brachyspira intermedia PWS/A*	3
Theba_1734 ABC-type uncharacterized transport system, permease component	0.675	*Clostridiales bacterium 1_7_47FAA*	
Theba_1735 Uncharacterized ABC-type transport system, permease component	0.526	*Anaerolinea thermophila UNI-1*	
Theba_1744 hypothetical protein	0.571	*Bacillus thuringiensis BMB171*	3
Theba_1745 Alkaline phosphatase	0.557	*Turicibacter sanguinis PC909*	
Theba_1746 Predicted multitransmembrane protein	0.535	*Desulfovibrio fructosovorans JJ*	
Theba_1756 Arylsulfatase A and related enzymes	0.535	*Escherichia fergusonii ECD227*	3
Theba_1757 anaerobic sulfatase-maturating enzyme	0.533	*Vibrio parahaemolyticus RIMD 2210633*	
Theba_1758 Transcriptional regulators	0.53	*Arcobacter sp. L*	
Theba_1760 Predicted transcriptional regulators	0.579	*Thermoanaerobacter italicus Ab9*	5
Theba_1761 ABC-type multidrug transport system, ATPase component	0.579	*Thermoanaerobacter italicus Ab9*	
Theba_1762 Predicted membrane protein	0.535	*Desulfovibrio aespoeensis Aspo-2*	
Theba_1763 diguanylate cyclase (GGDEF) domain	0.53	*Arcobacter sp. L*	
Theba_1764 Formate/nitrite family of transporters	0.585	*Desulfitobacterium hafniense DCB-2*	
Theba_1769 Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	0.531	*Agrobacterium tumefaciens F2*	6
Theba_1770 Uridine phosphorylase	0.027	*Thermococcus sp. 4557*	
Theba_1771 ABC-type sugar transport system, permease component	0.594	*Clostridium asparagiforme DSM 15981*	
Theba_1772 ABC-type sugar transport systems, permease components	0.594	*Clostridium leptum DSM 753*	
Theba_1773 ABC-type sugar transport system, periplasmic component	0.594	*Clostridium leptum DSM 753*	
Theba_1774 Transcriptional regulators	0.619	*Thermoanaerobacterium xylanolyticum LX-11*	
Theba_1781 Retron-type reverse transcriptase	0.571	*Bacillus pseudofirmus OF4*	3
Theba_1782 ABC-type dipeptide transport system, periplasmic component	0.535	*Klebsiella variicola At-22*	
Theba_1783 hypothetical protein	0.588	*Alkaliphilus oremlandii OhILAs*	
Theba_1790 Transcriptional regulator/sugar kinase	0.526	*Anaerolinea thermophila UNI-1*	2
Theba_1791 Entner-Doudoroff aldolase	0.564	*Lactobacillus casei LC2W*	
Theba_1794 hypothetical protein	0.526	*Anaerolinea thermophila UNI-1*	2
Theba_1795 Methylase involved in ubiquinone/menaquinone biosynthesis	0.497	*Pseudonocardia sp. P1*	
Theba_1811 methionine-S-sulfoxide reductase	0.024	*Haloarcula marismortui ATCC 43049*	2

Theba_1812 Amidases related to nicotinamidase	0.576	*Halanaerobium praevalens DSM 2228*	
Theba_1818 phosphoglycerate mutase, BPG-dependent, family 1	0.563	*Brachyspira intermedia PWS/A*	2
Theba_1819 Aminoglycoside N3"-acetyltransferase	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_1821 Putative NADH-flavin reductase	0.568	*Sporosarcina newyorkensis 2681*	2
Theba_1822 Predicted integral membrane protein	0.027	*Methanoregula boonei 6A8*	
Theba_1843 Zn-dependent hydrolases, including glyoxylases	0.611	*Haloplasma contractile SSD-17B*	7
Theba_1844 putative efflux protein, MATE family	0.594	*Clostridium tetani E88*	
Theba_1845 Asparaginase	0.027	*Thermococcus gammatolerans EJ3*	
Theba_1846 Transcriptional regulators	0.566	*Spirochaeta sp. Buddy*	
Theba_1847 ABC-type Fe3+ transport system, periplasmic component	0.566	*Spirochaeta sp. Buddy*	
Theba_1848 ABC-type Fe3+ transport system, permease component	0.566	*Spirochaeta sp. Buddy*	
Theba_1849 ABC-type spermidine/putrescine transport systems, ATPase components	0.495	*Mobiluncus mulieris FB024-16*	
Theba_1851 PAS domain S-box/uncharacterized domain HDIG	0.533	*Syntrophus aciditrophicus SB*	3
Theba_1852 hypothetical protein	0.576	*Halothermothrix orenii H 168*	
Theba_1853 Predicted acetyltransferase	0.576	*Halothermothrix orenii H 168*	
Theba_1862 Putative intracellular protease/amidase	0.557	*Roseiflexus sp. RS-1*	2
Theba_1863 X-Pro dipeptidyl-peptidase (S15 family).	0.531	*Pseudovibrio sp. JE062*	
Theba_1879 Acetyltransferases	0.573	*Paenibacillus mucilaginosus KNP414*	3
Theba_1880 Flavodoxin	0.594	*Clostridium botulinum Bf*	
Theba_1881 ABC-type multidrug transport system, ATPase and permease components	0.594	*Clostridium cellulovorans 743B*	
Theba_1924 hypothetical protein	0.531	*Hyphomicrobium sp. MC1*	2
Theba_1925 hypothetical protein	0.534	*Shewanella putrefaciens 200*	
Theba_1928 hypothetical protein	0.585	*Thermincola potens JR*	2
Theba_1929 Predicted permeases	0.611	*Haloplasma contractile SSD-17B*	
Theba_1934 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	0.526	*Anaerolinea thermophila UNI-1*	2
Theba_1935 FOG: GGDEF domain	0.534	*Pseudomonas mendocina NK-01*	
Theba_1944 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.566	*Spirochaeta caldaria DSM 7334*	6
Theba_1945 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.566	*Treponema phagedenis F0421*	
Theba_1946 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.566	*Treponema phagedenis F0421*	
Theba_1947 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.563	*Brachyspira intermedia PWS/A*	
Theba_1948 Uncharacterized conserved protein	0.571	*Bacillus megaterium WSH-002*	
Theba_1949 Predicted hydrolases or acyltransferases (alpha/beta hydrolase	0.57	*Ornithinibacillus sp. TW25*	

superfamily)

Theba_1953 lipoyltransferase and lipoate-protein ligase	0.587	*Desulfosporosinus sp. OT*	4
Theba_1954 pyruvate:ferredoxin (flavodoxin) oxidoreductase, homodimeric	0.587	*Desulfosporosinus sp. OT*	
Theba_1955 Uncharacterized conserved protein	0.533	*Geobacter uraniireducens Rf4*	
Theba_1956 N-acetylglucosamine-1-phosphate uridyltransferase (contains nucleotidyltransferase and I-patch acetyltransferase domains)	0.519	*Cytophaga hutchinsonii ATCC 33406*	
Theba_1967 hypothetical protein	0.525	*Pyramidobacter piscolens W5455*	2
Theba_1968 hypothetical protein	0.525	*Aminomonas paucivorans DSM 12260*	
Theba_1970 Predicted deacylase	0.576	*Halanaerobium praevalens DSM 2228*	3
Theba_1971 Dipeptidase	0.576	*Halanaerobium praevalens DSM 2228*	
Theba_1972 Dipeptidase	0.525	*Aminobacterium colombiense DSM 12261*	
Theba_1974 hypothetical protein	0.519	*Psychroflexus torquis ATCC 700755*	5
Theba_1975 Dipeptidase	0.532	*Desulfatibacillum alkenivorans AK-01*	
Theba_1976 hypothetical protein	0.525	*Pyramidobacter piscolens W5455*	
Theba_1977 TRAP-type C4-dicarboxylate transport system, large permease component	0.525	*Aminobacterium colombiense DSM 12261*	
Theba_1979 Predicted deacylase	0.525	*Aminobacterium colombiense DSM 12261*	
Theba_1981 haloacid dehalogenase superfamily, subfamily IA, variant 1 with third motif having Dx(3-4)D or Dx(3-4)E	0.557	*Turicibacter sp. HGF1*	2
Theba_1982 Inorganic pyrophosphatase	0.61	*Dehalogenimonas lykanthroporepellens BL-DC-9*	
Theba_1986 hypothetical protein	0.594	*Clostridium beijerinckii NCIMB 8052*	6
Theba_1987 Domain of unknown function (DUF1848).	0.619	*Thermoanaerobacterium xylanolyticum LX-11*	
Theba_1988 Uncharacterized protein conserved in bacteria	0.584	*Acetivibrio cellulolyticus CD2*	
Theba_1989 ABC-type antimicrobial peptide transport system, ATPase component	0.594	*Clostridium cellulovorans 743B*	
Theba_1990 ABC-type transport system, involved in lipoprotein release, permease component	0.611	*Haloplasma contractile SSD-17B*	
Theba_1991 Predicted cobalamin binding protein	0.594	*Clostridium cellulovorans 743B*	
Theba_2028 hypothetical protein	0.53	*Sulfurospirillum deleyianum DSM 6946*	3
Theba_2029 Bifunctional PLP-dependent enzyme with beta-cystathionase and maltose regulon repressor activities	0.578	*Tepidanaerobacter sp. Re1*	
Theba_2030 Glycosyltransferase	0.576	*Halothermothrix orenii H 168*	
Theba_2037 Predicted redox protein, regulator of disulfide bond formation	0.557	*Holdemania filiformis DSM 12042*	2
Theba_2038 hypothetical protein	0.026	*Methanobacterium sp. AL-21*	
Theba_2050 peptidase T	0.611	*Haloplasma contractile SSD-17B*	2
Theba_2051 Phosphoenolpyruvate carboxykinase (ATP)	0.583	*Clostridium lentocellum DSM 5427*	
Theba_2088 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.533	*Marinomonas sp. MWYL1*	8

Theba_2089 ABC-type sugar transport system, ATPase component	0.533	*Marinomonas sp. MWYL1*	
Theba_2090 Sugar phosphate isomerases/epimerases	0.566	*Treponema azotonutricium ZAS-9*	
Theba_2091 Predicted dehydrogenases and related proteins	0.566	*Treponema azotonutricium ZAS-9*	
Theba_2092 hypothetical protein	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_2093 hypothetical protein	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_2094 ABC-type multidrug transport system, ATPase component	0.568	*Planococcus donghaensis MPA1U2*	
Theba_2095 Transcriptional regulator	0.585	*Desulfitobacterium hafniense DCB-2*	
Theba_2097 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.568	*Solibacillus silvestris StLB046*	6
Theba_2098 ABC-type sugar transport system, ATPase component	0.531	*Paracoccus denitrificans SD1*	
Theba_2099 ribokinase	0.607	*Microcoleus vaginatus FGP-2*	
Theba_2100 membrane complex biogenesis protein, BtpA family	0.675	*Clostridiales bacterium 1_7_47FAA*	
Theba_2101 ABC-type ribose transport system, auxiliary component	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2102 Transcriptional regulators	0.533	*Aeromonas veronii B565*	
Theba_2105 lipid kinase, YegS/Rv2252/BmrU family	0.588	*Alkaliphilus metalliredigens QYMF*	3
Theba_2106 Predicted cobalamin binding protein	0.52	*Porphyromonas asaccharolytica PR426713P-I*	
Theba_2107 D-Lysine 5,6-aminomutase alpha subunit.	0.594	*Clostridium sp. SY8519*	
Theba_2125 large conductance mechanosensitive channel protein	0.533	*Methylophaga aminisulfidivorans MP*	2
Theba_2126 gamma-glutamyltranspeptidase	0.554	*Fusobacterium ulcerans ATCC 49185*	
Theba_2136 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.569	*Caldalkalibacillus thermarum TA2.A1*	4
Theba_2137 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.626	*Sulfobacillus acidophilus TPY*	
Theba_2138 hypothetical protein	0.026	*Methanobacterium sp. SWAN-1*	
Theba_2139 ABC-type Na+ efflux pump, permease component	0.626	*Lachnospiraceae bacterium 5_1_57FAA*	
Theba_2141 Predicted Zn-dependent proteases and their inactivated homologs	0.594	*Clostridium carboxidivorans P7*	2
Theba_2142 Predicted Zn-dependent proteases and their inactivated homologs	0.566	*Treponema denticola F0402*	
Theba_2144 heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	0.025	*Methanococcus maripaludis X1*	2
Theba_2145 Predicted transcriptional regulators	0.577	*Moorella thermoacetica ATCC 39073*	
Theba_2160 hypothetical protein	0.027	*Thermococcus sp. 4557*	2
Theba_2161 hypothetical protein	0.024	*Halorhabdus utahensis DSM 12940*	
Theba_2183 Major Facilitator Superfamily.	0.594	*Clostridium cellulovorans 743B*	7
Theba_2184 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.563	*Brachyspira intermedia PWS/A*	
Theba_2185 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.626	*Sulfobacillus acidophilus TPY*	
Theba_2186 ABC-type dipeptide/oligopeptide/nickel transport systems, permease	0.57	*Ornithinibacillus sp. TW25*	

components			
Theba_2187 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.57	*Ornithinibacillus sp. TW25*	
Theba_2188 ABC-type dipeptide transport system, periplasmic component	0.573	*Paenibacillus sp. Y412MC10*	
Theba_2189 Beta-lactamase class C and other penicillin binding proteins	0.594	*Clostridium carboxidivorans P7*	
Theba_2227 Predicted membrane-bound metal-dependent hydrolase (DUF457).	0.027	*Thermococcus onnurineus NA1*	3
Theba_2228 Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_2229 L-arabinose isomerase	0.535	*Escherichia coli UMNF18*	
Theba_2238 Transcriptional regulator	0.566	*Spirochaeta coccoides DSM 17374*	3
Theba_2239 plasma-membrane calcium-translocating P-type ATPase	0.626	*Lachnospiraceae bacterium 3_1_57FAA_CT1*	
Theba_2240 Transposase	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_2243 Predicted dehydrogenases and related proteins	0.535	*Escherichia coli NA114*	3
Theba_2244 Sugar phosphate isomerases/epimerases	0.535	*Escherichia coli WV_060327*	
Theba_2245 Threonine dehydrogenase and related Zn-dependent dehydrogenases	0.568	*Listeria monocytogenes J1-220*	
Theba_2252 ABC-type sugar transport system, periplasmic component	0.594	*Clostridium difficile QCD-63q42*	6
Theba_2253 hypothetical protein	0.675	*Clostridiales bacterium 1_7_47FAA*	
Theba_2254 ABC-type sugar transport system, ATPase component	0.531	*Paracoccus denitrificans SD1*	
Theba_2255 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.675	*Clostridiales bacterium 1_7_47FAA*	
Theba_2256 Nitrogen regulatory protein PII	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_2257 Predicted membrane-associated, metal-dependent hydrolase	0.584	*Acetivibrio cellulolyticus CD2*	
Theba_2259 Asparaginase	0.027	*Thermococcus sp. 4557*	3
Theba_2260 hypothetical protein	0.025	*Methanocella paludicola SANAE*	
Theba_2261 hypothetical protein	0.026	*Methanoplanus petrolearius DSM 11571*	
Theba_2264 Bacterial extracellular solute-binding proteins, family 3.	0.594	*Clostridium sticklandii DSM 519*	6
Theba_2265 Signal transduction histidine kinase	0.594	*Clostridium sticklandii DSM 519*	
Theba_2266 PAS domain S-box	0.594	*Clostridium sticklandii DSM 519*	
Theba_2267 PAS domain S-box/diguanylate cyclase (GGDEF) domain/uncharacterized domain HDIG	0.566	*Spirochaeta sp. Buddy*	
Theba_2268 PAS domain S-box	0.531	*Sideroxydans lithotrophicus ES-1*	
Theba_2269 Metal-dependent hydrolases of the beta-lactamase superfamily III	0.533	*Geobacter uraniireducens Rf4*	
Theba_2271 Predicted Fe-S oxidoreductases	0.626	*Lachnospiraceae bacterium 3_1_57FAA_CT1*	2
Theba_2272 Beta-lactamase class C and other penicillin binding proteins	0.611	*Haloplasma contractile SSD-17B*	
Theba_2275 Beta-lactamase class C and other penicillin binding proteins	0.606	*Cyanothece sp. PCC 7425*	2
Theba_2276 Uncharacterized conserved protein	0.534	*Pseudomonas fluorescens WH6*	

Theba_2278 FOG: EAL domain	0.566	*Spirochaeta sp. Buddy*	2
Theba_2279 PAS domain S-box/diguanylate cyclase (GGDEF) domain/uncharacterized domain HDIG	0.566	*Spirochaeta sp. Buddy*	
Theba_2281 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.566	*Spirochaeta smaragdinae DSM 11293*	13
Theba_2282 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.566	*Spirochaeta smaragdinae DSM 11293*	
Theba_2283 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.566	*Spirochaeta smaragdinae DSM 11293*	
Theba_2284 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.566	*Spirochaeta smaragdinae DSM 11293*	
Theba_2285 ABC-type dipeptide transport system, periplasmic component	0.566	*Spirochaeta smaragdinae DSM 11293*	
Theba_2286 ADP-ribosylglycohydrolase	0.583	*Heliobacterium modesticaldum Ice1*	
Theba_2287 Fucose permease	0.584	*Roseburia intestinalis M50/1*	
Theba_2288 ABC-type sugar transport system, periplasmic component	0.024	*Halalkalicoccus jeotgali B3*	
Theba_2289 Predicted amino acid racemase	0.578	*Tepidanaerobacter sp. Re1*	
Theba_2290 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_2291 ABC-type sugar transport system, ATPase component	0.024	*Halorhabdus tiamatea SARL4B*	
Theba_2292 hypothetical protein	0.519	*Bifidobacterium breve UCC2003*	
Theba_2293 Beta-galactosidase	0.519	*Bifidobacterium breve UCC2003*	
Theba_2322 DNA replication protein	0.626	*Ruminococcaceae bacterium D16*	2
Theba_2323 S23 ribosomal protein.	0.519	*Psychroflexus torquis ATCC 700755*	
Theba_2335 Retron-type reverse transcriptase	0.571	*Bacillus pseudofirmus OF4*	6
Theba_2336 putative efflux protein, MATE family	0.611	*Haloplasma contractile SSD-17B*	
Theba_2337 Transcriptional regulators	0.594	*Clostridium botulinum Bf*	
Theba_2338 FOG: Ankyrin repeat	0.002	*Trichomonas vaginalis G3*	
Theba_2339 aspartate racemase	0.534	*Idiomarina sp. A28L*	
Theba_2340 X-Pro dipeptidyl-peptidase (S15 family).	0.534	*Legionella pneumophila 130b*	
Theba_2350 Amino acid transporters	0.594	*Clostridium asparagiforme DSM 15981*	6
Theba_2351 hypothetical protein	0.52	*Haliscomenobacter hydrossis DSM 1100*	
Theba_2352 hypothetical protein	0.031	*Aciduliprofundum boonei T469*	
Theba_2353 Pirin-related protein	0.594	*Clostridium carboxidivorans P7*	
Theba_2354 PAS domain S-box	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2355 hypothetical protein	0.534	*Legionella longbeachae NSW150*	
Theba_2357 Transcriptional regulator	0.573	*Paenibacillus vortex V453*	2
Theba_2358 TRAP transporter, DctM subunit	0.566	*Spirochaeta sp. Buddy*	

Theba_2360 tripartite ATP-independent periplasmic transporter solute receptor, DctP family	0.531	*Oceanicola granulosus HTCC2516*	4
Theba_2361 dihydrodipicolinate synthase	0.619	*Caldicellulosiruptor hydrothermalis 108*	
Theba_2362 hypothetical protein	0.002	*Trichomonas vaginalis G3*	
Theba_2363 hypothetical protein	0.588	*Alkaliphilus metalliredigens QYMF*	
Theba_2366 ABC-type sugar transport system, periplasmic component	0.531	*Rhizobium etli CNPAF512*	10
Theba_2367 ABC-type sugar transport systems, permease components	0.52	*Ktedonobacter racemifer DSM 44963*	
Theba_2368 ABC-type sugar transport system, permease component	0.531	*Sinorhizobium medicae WSM419*	
Theba_2369 Putative glycosyl hydrolase of unknown function (DUF1680).	0.573	*Paenibacillus sp. HGF5*	
Theba_2370 Prephenate dehydratase	0.533	*Francisella sp. TX077308*	
Theba_2371 Predicted transcriptional regulators	0.568	*Listeria monocytogenes J1816*	
Theba_2372 Uncharacterized conserved protein	0.531	*Hyphomicrobium sp. MC1*	
Theba_2374 hypothetical protein	0.525	*Aminobacterium colombiense DSM 12261*	
Theba_2375 Short-chain alcohol dehydrogenase of unknown specificity	0.525	*Aminobacterium colombiense DSM 12261*	
Theba_2376 Acetoacetate decarboxylase	0.525	*Dethiosulfovibrio peptidovorans DSM 11002*	
Theba_2378 Fructose-2,6-bisphosphatase	0.501	*Streptomyces sp. AA4*	4
Theba_2379 Flavodoxins	0.026	*Methanobacterium sp. SWAN-1*	
Theba_2380 Uncharacterized conserved protein	0.52	*Haliscomenobacter hydrossis DSM 1100*	
Theba_2381 Uncharacterized protein related to plant photosystem II stability/assembly factor	0.533	*Geobacter uraniireducens Rf4*	
Theba_2412 Acetyltransferases	0.611	*Haloplasma contractile SSD-17B*	8
Theba_2413 Acylphosphatases	0.533	*Gallibacterium anatis UMN179*	
Theba_2414 Hydrolases of the alpha/beta superfamily	0.611	*Haloplasma contractile SSD-17B*	
Theba_2415 Predicted solute binding protein	0.566	*Spirochaeta caldaria DSM 7334*	
Theba_2416 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.563	*Brachyspira intermedia PWS/A*	
Theba_2417 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	0.566	*Treponema phagedenis F0421*	
Theba_2418 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.566	*Treponema phagedenis F0421*	
Theba_2419 oligopeptide/dipeptide ABC transporter, ATP-binding protein, C-terminal domain	0.566	*Spirochaeta caldaria DSM 7334*	
Theba_2424 Transcriptional regulators of sugar metabolism	0.525	*Dethiosulfovibrio peptidovorans DSM 11002*	7
Theba_2425 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.535	*Pantoea ananatis AJ13355*	
Theba_2426 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.535	*Serratia sp. AS12*	
Theba_2427 ABC-type sugar transport system, ATPase component	0.535	*Serratia sp. AS12*	
Theba_2428 ABC-type sugar transport system, periplasmic component	0.535	*Rahnella sp. Y9602*	

Theba_2429 D-xylulose kinase	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_2430 Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases	0.587	*Desulfosporosinus sp. OT*	
Theba_2449 Predicted acetyltransferase involved in intracellular survival and related acetyltransferases	0.556	*Acetonema longum DSM 6540*	2
Theba_2450 Uncharacterized conserved protein	0.533	*Halomonas sp. TD01*	
Theba_2462 Zn-dependent protease with chaperone function	0.532	*Bacteriovorax marinus SJ*	9
Theba_2463 hypothetical protein	0.519	*Capnocytophaga ochracea F0287*	
Theba_2464 Small-conductance mechanosensitive channel	0.568	*Solibacillus silvestris StLB046*	
Theba_2465 ABC-type proline/glycine betaine transport systems, permease component	0.535	*Rahnella sp. Y9602*	
Theba_2466 glycine betaine/L-proline transport ATP binding subunit	0.501	*Streptomyces sp. SirexAA-E*	
Theba_2467 ABC-type proline/glycine betaine transport systems, permease component	0.61	*Dehalogenimonas lykanthroporepellens BL-DC-9*	
Theba_2468 Periplasmic glycine betaine/choline-binding (lipo)protein of an ABC-type transport system (osmoprotectant binding protein)	0.535	*Escherichia coli PCN033*	
Theba_2469 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes	0.52	*Ktedonobacter racemifer DSM 44963*	
Theba_2470 hypothetical protein	0.594	*Clostridium difficile QCD-23m63*	
Theba_2474 Uncharacterized protein conserved in bacteria	0.526	*Anaerolinea thermophila UNI-1*	14
Theba_2475 ABC-type sugar transport system, permease component	0.573	*Paenibacillus sp. Y412MC10*	
Theba_2476 ABC-type sugar transport systems, permease components	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2477 hypothetical protein	0.573	*Paenibacillus sp. Y412MC10*	
Theba_2478 hypothetical protein	0.573	*Paenibacillus sp. Y412MC10*	
Theba_2480 ABC-type sugar transport systems, permease components	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2481 ABC-type sugar transport system, periplasmic component	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2482 ABC-type sugar transport system, periplasmic component	0.573	*Paenibacillus sp. Y412MC10*	
Theba_2483 Cellobiose phosphorylase	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2484 Beta-galactosidase	0.626	*Sulfobacillus acidophilus TPY*	
Theba_2485 Beta-glucosidase-related glycosidases	0.573	*Paenibacillus sp. HGF5*	
Theba_2486 Transcriptional regulators	0.573	*Paenibacillus vortex V453*	
Theba_2487 Beta-galactosidase	0.573	*Paenibacillus sp. Y412MC10*	
Theba_2489 Uncharacterized conserved protein	0.579	*Thermoanaerobacter wiegelii Rt8.B1*	
Theba_2494 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	0.532	*Syntrophobacter fumaroxidans MPOB*	3
Theba_2495 hypothetical protein	0.517	*Chlorobium limicola DSM 245*	
Theba_2496 hypothetical protein	0.626	*Lachnospiraceae bacterium 3_1_57FAA_CT1*	
Theba_2499 hypothetical protein	0.585	*Syntrophobotulus glycolicus DSM 8271*	10
Theba_2500 Lysophospholipase L1 and related esterases	0.557	*Turicibacter sanguinis PC909*	

Theba_2501 Uncharacterized protein conserved in bacteria	0.626	*Pseudoflavonifractor capillosus ATCC 29799*	
Theba_2502 Putative translation initiation inhibitor, yjgF family	0.571	*Bacillus cereus AH603*	
Theba_2503 Bifunctional PLP-dependent enzyme with beta-cystathionase and maltose regulon repressor activities	0.587	*Desulfosporosinus sp. OT*	
Theba_2504 dipeptidase, putative	0.594	*Clostridium symbiosum WAL-14163*	
Theba_2505 5"-nucleotidase/2",3"-cyclic phosphodiesterase and related esterases	0.626	*Pseudoflavonifractor capillosus ATCC 29799*	
Theba_2506 Protein related to penicillin acylase	0.531	*Cupriavidus necator N-1*	
Theba_2507 lipoate-protein ligase B	0.525	*Anaerobaculum hydrogeniformans ATCC BAA-1850*	
Theba_2508 lipoate synthase	0.626	*Sulfobacillus acidophilus TPY*	
Theba_2510 Short-chain dehydrogenases of various substrate specificities	0.526	*Anaerolinea thermophila UNI-1*	2
Theba_2511 Phytoene dehydrogenase and related proteins	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2513 N-formylmethionyl-tRNA deformylase	0.566	*Spirochaeta smaragdinae DSM 11293*	4
Theba_2514 Lysophospholipase L1 and related esterases	0.573	*Paenibacillus sp. HGF7*	
Theba_2515 Sugar kinases, ribokinase family	0.573	*Paenibacillus mucilaginosus KNP414*	
Theba_2516 hypothetical protein	0.534	*Idiomarina sp. A28L*	
Theba_2520 Uncharacterized conserved protein	0.52	*Haliscomenobacter hydrossis DSM 1100*	2
Theba_2521 hypothetical protein	0.571	*Bacillus sp. m3-13*	
Theba_2527 hypothetical protein	0.571	*Bacillus coagulans 2-6*	2
Theba_2528 hypothetical protein	0.534	*Idiomarina sp. A28L*	
Theba_2532 ABC-type sugar transport system, periplasmic component	0.497	*Rhodococcus erythropolis PR4*	6
Theba_2533 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.594	*Clostridium sp. SY8519*	
Theba_2534 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	0.531	*Collimonas fungivorans Ter331*	
Theba_2535 ABC-type sugar transport system, ATPase component	0.594	*Clostridium botulinum Bf*	
Theba_2536 dihydroxyacetone kinase, DhaK subunit	0.571	*Bacillus megaterium WSH-002*	
Theba_2537 dihydroxyacetone kinase, phosphoprotein-dependent, L subunit	0.557	*Megamonas hypermegale ART12/1*	
Theba_2539 Secreted and surface protein containing fasciclin-like repeats	0.026	*Methanohalobium evestigatum Z-7303*	3
Theba_2540 Phytoene dehydrogenase and related proteins	0.493	*Eggerthella sp. YY7918*	
Theba_2541 hypothetical protein	0.533	*Syntrophus aciditrophicus SB*	
Theba_2543 Penicillin V acylase and related amidases	0.553	*Parachlamydia acanthamoebae UV7*	3
Theba_2544 Sugar (pentulose and hexulose) kinases	0.571	*Bacillus megaterium WSH-002*	
Theba_2545 Transaldolase	0.619	*Thermoanaerobacterium thermosaccharolyticum DSM 571*	
Theba_2547 Transcriptional regulators	0.517	*Dictyoglomus thermophilum H-6-12*	2
Theba_2548 Predicted phosphosugar isomerases	0.026	*Methanobacterium sp. AL-21*	

Theba_2558 ABC-type branched-chain amino acid transport systems, ATPase component	0.584	*Eubacterium limosum KIST612*	4
Theba_2559 Predicted membrane protein	0.568	*Solibacillus silvestris StLB046*	
Theba_2560 Predicted membrane protein	0.571	*Bacillus selenitireducens MLS10*	
Theba_2561 EDD domain protein, DegV family	0.52	*Paludibacter propionicigenes WB4*	
Theba_2567 ABC-type spermidine/putrescine transport system, permease component II	0.566	*Treponema phagedenis F0421*	2
Theba_2568 ABC-type spermidine/putrescine transport system, permease component I	0.554	*Leptotrichia goodfellowii F0264*	
Theba_2574 pheromone shutdown-related protein TraB	0.57	*Ornithinibacillus sp. TW25*	2
Theba_2575 Predicted solute binding protein	0.566	*Spirochaeta caldaria DSM 7334*	
Theba_2589 Transcriptional regulators	0.569	*Caldalkalibacillus thermarum TA2.A1*	4
Theba_2590 ABC-type ribose transport system, auxiliary component	0.569	*Geobacillus sp. Y412MC61*	
Theba_2591 3-hexulose-6-phosphate synthase and related proteins	0.567	*Staphylococcus lugdunensis N920143*	
Theba_2592 Predicted sugar phosphate isomerase involved in capsule formation	0.573	*Paenibacillus sp. HGF7*	
Theba_2617 ABC-type sugar transport system, periplasmic component	0.497	*Micromonospora sp. L5*	6
Theba_2618 ABC-type sugar transport systems, permease components	0.569	*Geobacillus sp. C56-T3*	
Theba_2619 ABC-type sugar transport system, permease component	0.495	*Stackebrandtia nassauensis DSM 44728*	
Theba_2620 Cellulase (glycosyl hydrolase family 5).	0.531	*Methylobacterium sp. 4-46*	
Theba_2621 Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase	0.576	*Halothermothrix orenii H 168*	
Theba_2622 Beta-galactosidase/beta-glucuronidase	0.024	*Haloterrigena turkmenica DSM 5511*	
Theba_2633 Acetyltransferases	0.533	*Alcanivorax sp. DG881*	5
Theba_2634 Transglutaminase-like enzymes, putative cysteine proteases	0.026	*Methanoculleus marisnigri JR1*	
Theba_2635 Transposase	0.517	*Meiothermus silvanus DSM 9946*	
Theba_2636 Secreted and surface protein containing fasciclin-like repeats	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2637 hypothetical protein	0.576	*Halothermothrix orenii H 168*	
Theba_2646 Aspartate/tyrosine/aromatic aminotransferase	0.585	*Desulfotomaculum carboxydivorans CO-1-SRB*	3
Theba_2647 Predicted Fe-S oxidoreductases	0.027	*Thermococcus sibiricus MM 739*	
Theba_2648 ABC-type multidrug transport system, ATPase and permease components	0.571	*Bacillus megaterium WSH-002*	
Theba_2651 Arabinose efflux permease	0.587	*Desulfosporosinus sp. OT*	2
Theba_2652 Putative homoserine kinase type II (protein kinase fold)	0.611	*Haloplasma contractile SSD-17B*	
Theba_2654 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs	0.587	*Desulfosporosinus sp. OT*	3
Theba_2655 1-acyl-sn-glycerol-3-phosphate acyltransferase	0.611	*Haloplasma contractile SSD-17B*	
Theba_2656 hypothetical protein	0.517	*Dictyoglomus thermophilum H-6-12*	
Theba_2659 poly-gamma-glutamate synthase PgsB/CapB	0.533	*Francisella novicida FTG*	3

Theba_2660 hypothetical protein	0.576	*Halanaerobium praevalens DSM 2228*	
Theba_2661 Predicted membrane protein	0.594	*Clostridium cellulolyticum H10*	
Theba_2666 secondary thiamine-phosphate synthase enzyme	0.563	*Brachyspira intermedia PWS/A*	2
Theba_2667 Putative protein-S-isoprenylcysteine methyltransferase	0.626	*Sulfobacillus acidophilus TPY*	
Theba_2677 Malate/L-lactate dehydrogenases	0.594	*Clostridium sp. SY8519*	8
Theba_2678 Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases	0.564	*Lactobacillus salivarius ACS-116-V-Col5a*	
Theba_2679 Altronate dehydratase	0.016	*Vulcanisaeta moutnovskia 768-28*	
Theba_2680 Altronate dehydratase	0.533	*Gallibacterium anatis UMN179*	
Theba_2681 Transcriptional regulators	0.626	*Sulfobacillus acidophilus TPY*	
Theba_2682 Sugar phosphate isomerases/epimerases	0.594	*Clostridium bolteae ATCC BAA-613*	
Theba_2683 Phosphotransacetylase	0.534	*Bilophila sp. 4_1_30*	
Theba_2684 butyrate kinase	0.571	*Bacillus amyloliquefaciens TA208*	
Theba_2686 Uncharacterized conserved protein (DUF2088).	0.557	*Thermosinus carboxydivorans Nor1*	6
Theba_2687 ABC-type branched-chain amino acid transport systems, ATPase component	0.534	*Pseudomonas putida S16*	
Theba_2688 ABC-type branched-chain amino acid transport systems, ATPase component	0.587	*Desulfosporosinus sp. OT*	
Theba_2689 ABC-type branched-chain amino acid transport system, permease component	0.594	*Clostridium carboxidivorans P7*	
Theba_2690 Branched-chain amino acid ABC-type transport system, permease components	0.597	*Eubacterium cylindroides T2-87*	
Theba_2691 ABC-type branched-chain amino acid transport systems, periplasmic component	0.578	*Tepidanaerobacter sp. Re1*	
Theba_2709 Response regulator	0.571	*Bacillus mycoides Rock3-17*	2
Theba_2710 Acetyltransferases	0.569	*Brevibacillus brevis NBRC 100599*	
Theba_2714 Glycosyltransferases, probably involved in cell wall biogenesis	0.526	*Anaerolinea thermophila UNI-1*	3
Theba_2715 Predicted membrane protein	0.526	*Anaerolinea thermophila UNI-1*	
Theba_2716 Periplasmic protease	0.594	*Clostridium carboxidivorans P7*	
Theba_2724 Retron-type reverse transcriptase	0.571	*Bacillus pseudofirmus OF4*	2
Theba_2725 Transcriptional regulators	0.594	*Clostridium botulinum BKT015925*	

Supplementary Table 3. Functional classification of *Mesotoga prima* and *Kosmotoga olearia* genes, predicted to have been acquired through lateral gene transfer.

COG CATEGORY	Mesotoga LGT	Kosmotoga LGT
INFORMATION STORAGE AND PROCESSING	*99*	*58*
Translation; ribosomal structure and biogenesis	10	4
RNA processing and modification	0	0
Transcription	47	22
Replication; recombination and repair	42	32
Chromatin structure and dynamics	0	0
CELLULAR PROCESSES AND SIGNALING	*96*	*46*
Cell cycle control; cell division; chromosome partitioning	1	3
Nuclear structure	0	0
Defense mechanisms	24	11
Signal transduction mechanisms	33	12
Cell wall/membrane/envelope biogenesis	25	11
Cell motility	2	0
Cytoskeleton	0	0
Extracellular structures	0	0
Intracellular trafficking; secretion; and vesicular transport	0	0
Posttranslational modification; protein turnover; chaperones	11	9
METABOLISM	*292*	*135*
Energy production and conversion	31	24
Carbohydrate transport and metabolism	113	54
Amino acid transport and metabolism	88	18
Nucleotide transport and metabolism	7	4
Coenzyme transport and metabolism	10	14
Lipid transport and metabolism	13	7
Inorganic ion transport and metabolism	18	11
Secondary metabolites biosynthesis; transport and catabolism	12	3
POORLY CHARACTERIZED	*279*	*166*
General function prediction only	98	58
Function unknown	65	23
Not predicted	36	29
Not in COGs	80	56
TOTAL	**766**	**405**

Supplementary Table 4. Average Amino Acid Identity/Similarity between *M. prima* and selected *Thermotogales* genomes. The values were calculated from the regions of BLAST alignment between an *M. prima* ORF and its top-scoring BLASTP match in the corresponding Thermotogales genome. BLAST matches were required to have an E-value below 10^{-4} and to be aligned to at least 70% of the query.

Genome	**Identity**	**Similarity**
Fervidobacterium nodosum Rt17-B1	44.04	63.63
Kosmotoga olearia TBF 19.5.1	52.40	70.52
Marinitoga piezophila KA3	44.61	64.35
Petrotoga mobilis SJ95	43.74	63.74
Thermotoga lettingae TMO	44.99	64.51
Thermotoga maritima MSB8	44.59	63.96

References

Bennett S. 2004. Solexa Ltd. Pharmacogenomics 5:433–438.

Dyrløv Bendtsen J, Nielsen H, Heijne von G, Brunak S. 2004. Improved Prediction of Signal Peptides: SignalP 3.0. J. Mol. Biol. 340:783–795.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nuc. Acids Res. 32:1792–1797.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. Nuc. Acids Res. 31:439–441.

Grissa I, Vergnaud G. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nuc. Acids Res.:Web Server issue.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Guy L, Kultima JR, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26:2334–2335.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

Krogh A, Larsson B, Heijne von G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1. J. Mol. Biol. 305:567–580.

Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nuc. Acids Res. 35:3100–3108.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nuc. Acids Res. 25:955–964.

Makarova KS, Haft DH, Barrangou R, et al. 2011. Evolution and classification of the CRISPR-Cas systems. Nature Rev. Microbiol. 9:467–477.

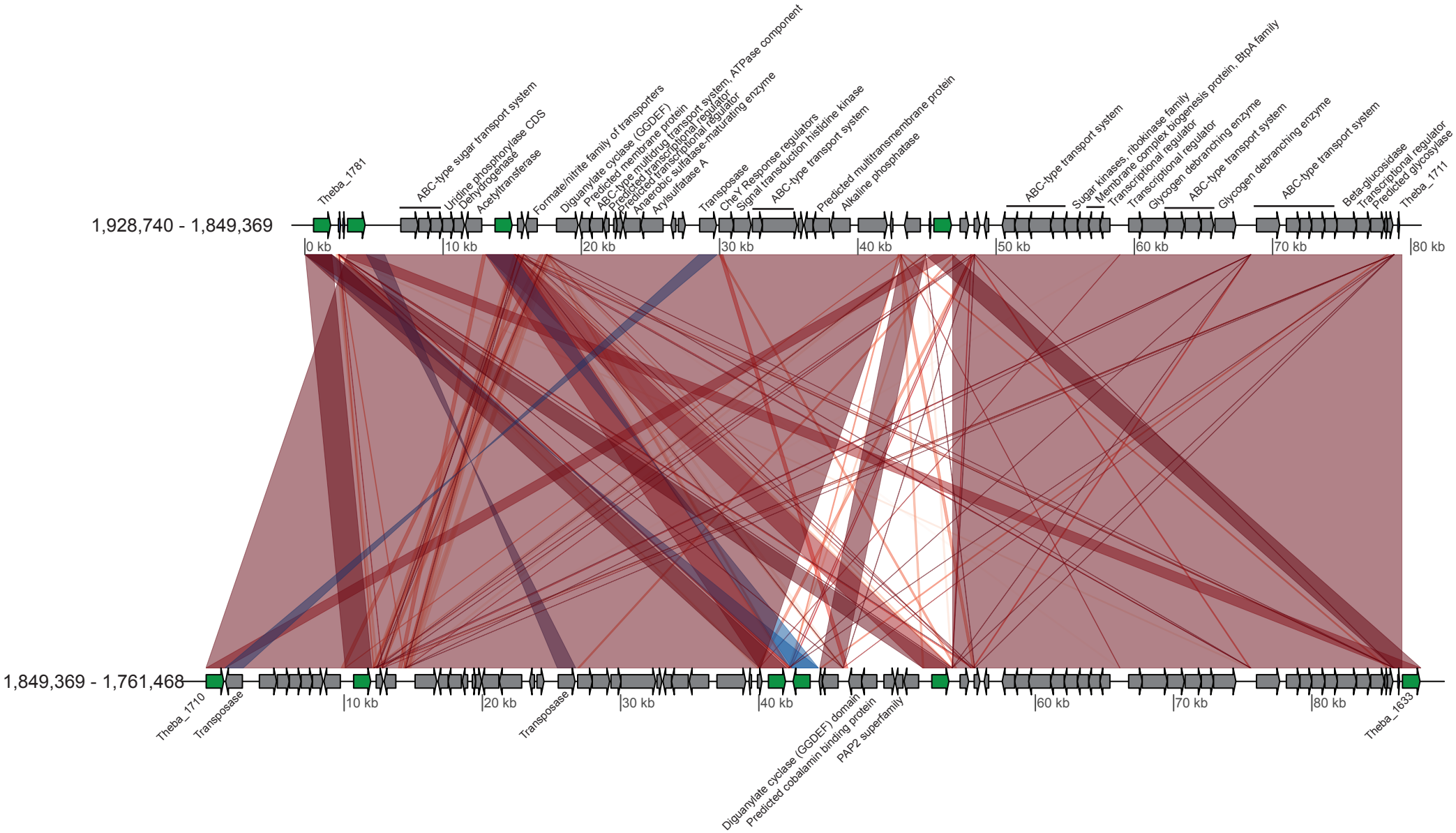
Margulies M, Egholm M, Altman WE, et al. 2005. Genome sequencing in microfabricated high-

density picolitre reactors. Nature 437:376–380.

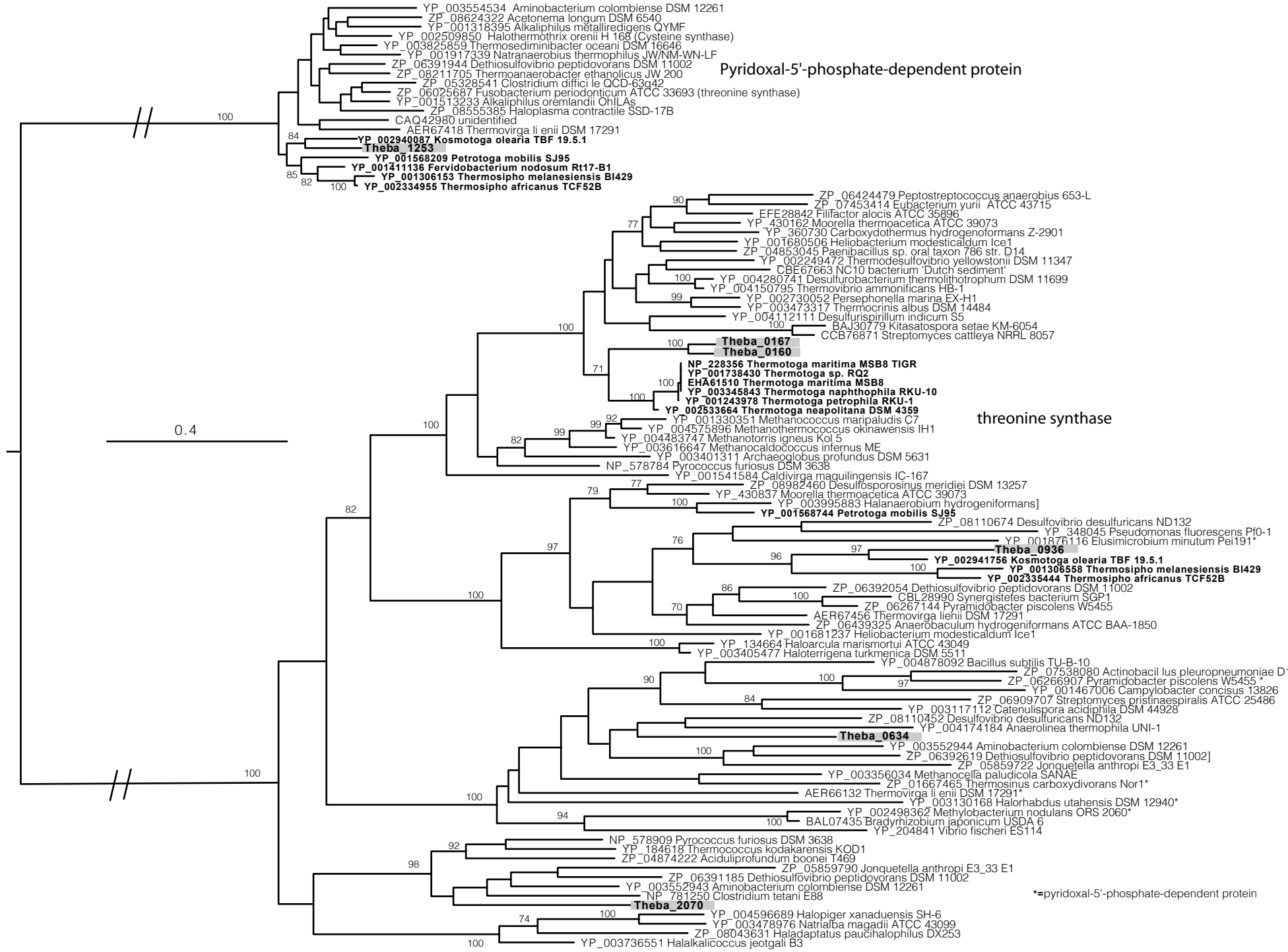
Markowitz VM, Chen I-MA, Palaniappan K, et al. 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nuc. Acids Res. 40:D115–22.

Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nature Methods 7:455–457.

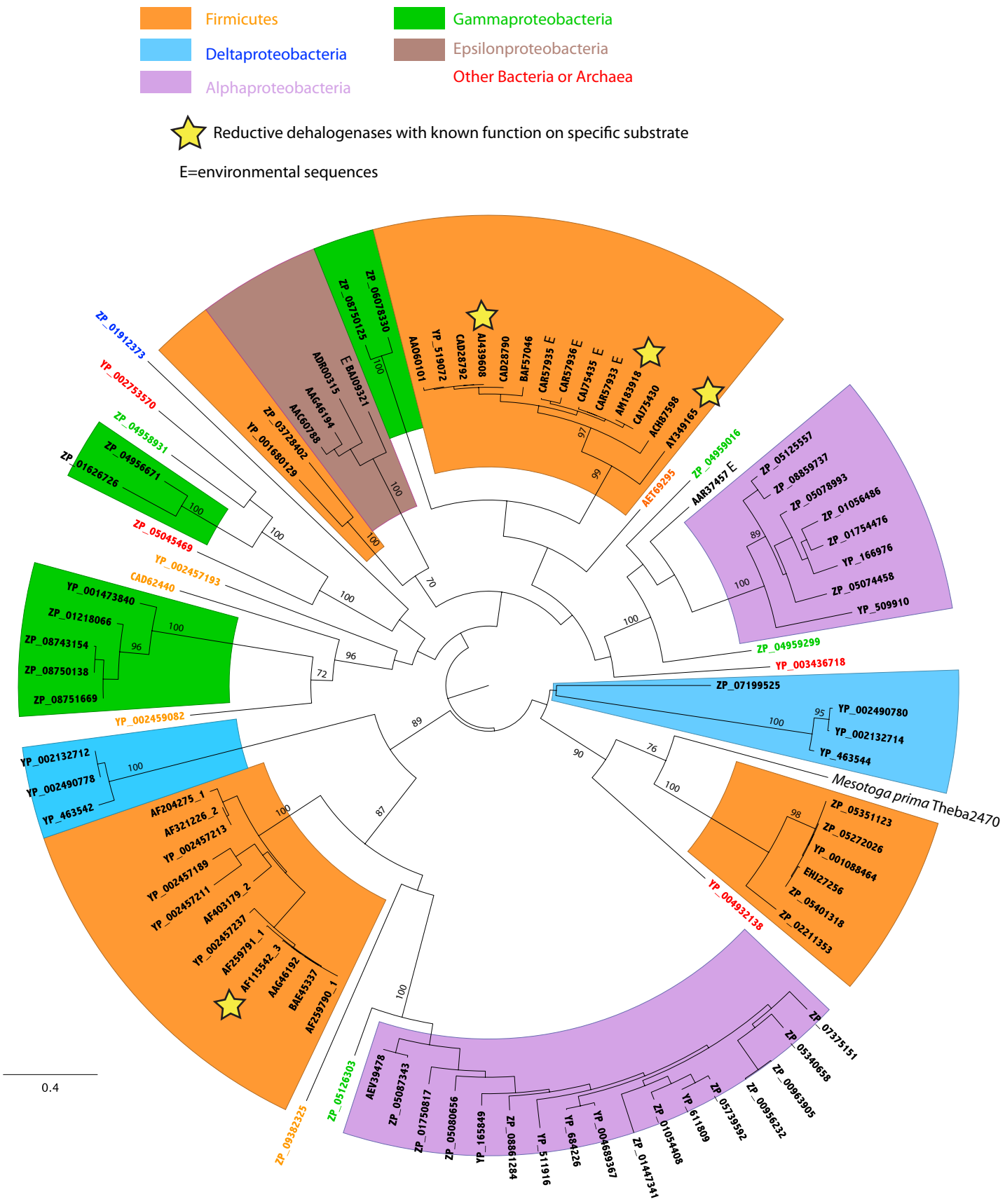
Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.



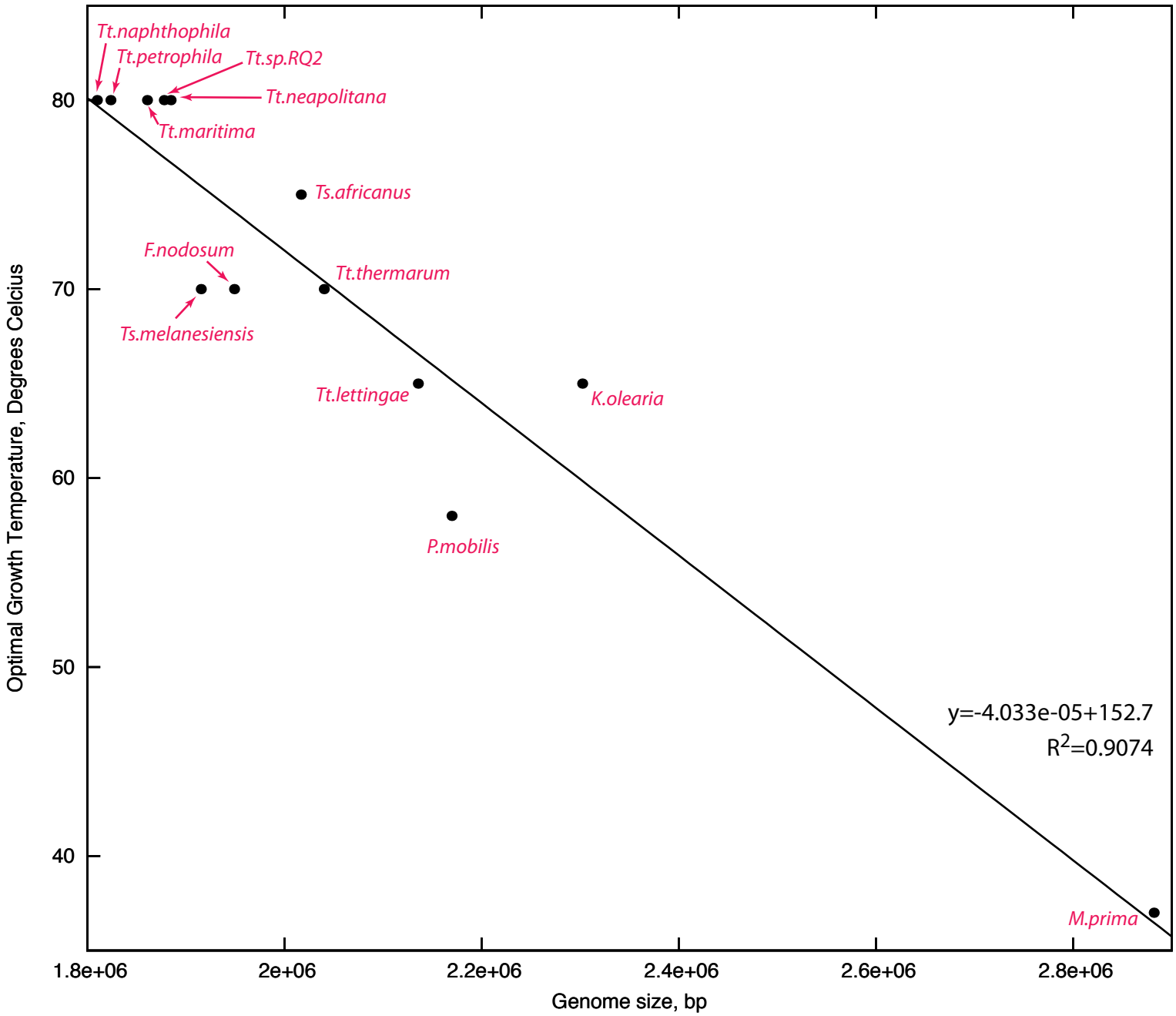
Supplemantary Fig. 1. Sequence alignment of the duplicated region. The comparison is based on BLASTN percent similarity values and drawn in genoPlotR (Guy, Kultima, and Andersson 2010). Direct comparisons are colored in red hues, and the reversed ones in blue hues. The annotations are shown for ORFs with predicted functions. The reverse transcriptase domains of the group II intron are given in green.



Supplementary Fig. 2. Evolution of threonine synthase genes in *M. prima.* The 100 top-scoring BLASTP matches to each of the *M. prima* homologs of threonine synthase gene were retrieved from nr database and aligned using the MUSCLE program (Edgar 2004). The alignment was edited by eye and identical or very similar sequences from closely related organisms were eliminated from further analysis. The maximum likelihood tree was reconstructed using the PhyML program (Guindon and Gascuel 2003), under WAG+Γ+I model, with 100 bootstrap replicates. Only bootstrap support values above 70% are shown. Thermotogales taxa are depicted in bold.



Supplementary Fig. 3. Evolutionary history of reductive dehalogenases. The 100 top-scoring BLASTP matches to each of the *M. prima* homologs were retrieved from nr database. Three additional sequences with known function on specific substrates were added to the dataset (accession numbers AY349165, AJ439608 and AM183918). The dataset was aligned using the MUSCLE program (Edgar 2004). Very divergent or partial sequences were removed from the alignment. The maximum likelihood tree was reconstructed using the PhyML program (Guindon and Gascuel 2003) under WAG+Γ+I model, with 100 bootstrap replicates. Only bootstrap support values above 70% are shown.



Supplementary Fig. 4. A correlation between genome size and optimal growth temperature among Thermotogales genomes. The correlation is based on data from *M. prima* and 12 other completed Thermotogales genomes present in GenBank. Linear regression analysis was performed in R ($p<0.001$, F-test). Abbreviations: *Tt, Thermotoga*; *Ts, Thermosipho*; *P, Petrotoga*; *M, Mesotoga*; *F, Fervidobacterium*; *K, Kosmotoga*.