Chapter 1

General introduction

Carbon monoxide (CO) is a well-known toxic gas for many organisms. However, CO-oxidizing microbes (carboxydotrophs) can use CO as energy and/or carbon source for their growth (CO metabolism; Mörsdorf et al. 1992). The oxido-reductive enzyme catalyzing interconversion between CO and CO_2 is CO dehydrogenase (CODH; King and Weber 2007). For aerobic carboxydotrophs, CODH with Mo in its active site (Mo-CODH) is a hetero trimer (Dobbek et al. 1999). Their three subunits, namely small subunit (an iron–sulfur protein), medium subunit (a flavin adenine dinucleotide-binding protein), and large subunit (the catalytic, molybdenum cytosine dinucleotide-binding protein), are encoded by *coxS*, *coxM*, and *coxL* genes, respectively (Santiago et al. 1999; King and Weber 2007). On the other hand, for anaerobic, CODH with Ni in its active site (Ni-CODH; hereinafter referred to as CODH) is a homo dimer and it is encoded by *cooS* (Dobbek et al. 2001; Svetlitchnyi et al. 2004).

Thanks to the low redox potential of the CO/CO_2 couple (EO' [CO/CO_2] = -520 to -558 mV) (Grahame et al. 1995), anaerobic carboxydotrophs use CODH to couple CO oxidation to CO_2 with reduction of various terminal electron acceptors under anaerobic condition (Oelgeschläger et al. 2008). Their various types of CO metabolism have been characterized based on available terminal electron acceptors, such as sulfate to sulfide (sulfate-reducers), water to H_2 (hydrogenogens), Fe^{3+} to Fe^{2+} (Fe[III]-reducers), and CO_2 to acetate (acetogens) or methane (methanogens; Sokolova et al. 2009; King and Weber 2007). Moreover, CODHs can fix the generated CO_2

autotrophically using reductive carbon fixation pathways, such as the Calvin-Benson cycle, reductive tricarboxylic acid cycle, Wood-Ljungdahl pathway, or 3-hydroxypropionate cycle (Ragsdale 2004).

Data search and phylogenetic analyses of CODHs have shown that CODHs belong to six distinct clades (Techtmann et al. 2012). CODHs do not coincide with the evolutionary lineage of the hosts and its function is not conserved in each clade (Techtmann et al. 2012). As it stands, functions of CODH have been predicted from the genomic context of the gene cluster including the CODH gene, although functions of most of CODH remains unknown (Techtmann et al. 2012).

With regard to genomic context, CODHs have been divided into four categories: within an acetyl-CoA synthase (ACS) gene cluster (category I), adjacent to an energy-converting hydrogenase (ECH) gene cluster (category II), adjacent to an electron transfer ferredoxin-like protein (CooF; Kerby et al. 1992) gene (except for the ECH gene cluster; category III), and not adjacent to *cooF* (category IV; Techtmann et al. 2012). Of the four categories, the function of category I and II CODHs is well characterized. The complex formed by category I CODH and ACS is thought to fix carbon via the Wood-Ljungdahl pathway (Ragsdale 2004). The category II CODH/ECH complex couples CO oxidation with H_2 production (hydrogenogenic CO metabolism). In this process, CooF transfers the intermediate electron between category II CODH/ECH complex, while a proton is translocated across the cytoplasmic membrane (Svetlitchnyi et al. 2001; Soboh et al. 2002). The generated electrochemical proton gradient then fuels ATP synthesis. These sequential reactions are considered to be responsible for energy conservation via hydrogenogenic CO metabolism (Svetlitchnyi et al. 2001; Soboh et al. 2002). Category III CODH is thought to be capable of using CooF

to transfer electrons to other functional proteins. A function of category IV CODH has not yet been predicted from its genomic context.

In hydrothermal environments, CO is supplied by volcanic gas, photochemical and thermochemical decomposition of organic matter, or byproduct of certain thermophiles (King and Weber 2007; Techtmann et al. 2009). So far, some anaerobic and thermophilic carboxydotrophs have been isolated (Sokolova et al. 2009) from hydrothermal environments where CO is supplied by volcanic gas, photochemical and thermochemical decomposition of organic matter, and as a byproduct of certain thermophiles (King and Weber 2007; Techtmann et al. 2009). Among them, 23 strains representing 13 genera (21 species) have been reported as hydrogenogenic carboxydotrophs (Sokolova and Lebedinsky 2013). Recently, in addition to these 23 strains, three thermophilic and hydrogenogenic carboxydotrophs, namely *Thermococcus balophilus* (Kozhevnikova et al. 2016), *Thermoanaerobacter kivui* (Weghoff and Müller 2016), and *Parageobacillus thermoglucosidasius* (Mohr et al. 2018) have been described. Although a number of strains of *Moorella thermoacetica* have been isolated, only one strain (*M. thermoacetica* strain AMP) exhibits hydrogenogenic carboxydotrophy whereas the other strains are acetogenic carboxydotrophs (Jiang et al. 2009).

Owing to ability of hydrogenogenic carboxydotroph to use potentially toxic CO and produce H_2 as the source of energy and carbon for other microbes, they are assumed to be important ‘CO scavengers’ or primary producers in the environment (Sokolova et al. 2009; Techtmann et al. 2009, Yoneda et al. 2013, Yoneda et al. 2015). Further, they are expected to be ‘microbial catalysis.’ Their hydrogenogenic CO-converting reaction has a potential application in the H_2 production as a renewable

fuel from CO in synthesis gas (syngas) (Henstra et al. 2007; Kim et al. 2013). In addition, CODH from *Carboxydothermus hydrogenoformans* is the model catalysis in CO_2 photo-reduction system (Woolerton et al. 2010). Despite their importance to microbial community or application, a few their genomes have been available.

The genus *Carboxydothermus* has been one of the most studied models of thermophilic carboxydotrophy. To date, five *Carboxydothermus* species have been described (Svetlichny et al. 1991; Slobodkin et al. 2006; Slepova et al. 2009; Novikov et al. 2011; Yoneda et al. 2012). All of them are strictly anaerobic, thermophilic, neutrophilic, and have been isolated as facultative carboxydotrophs from terrestrial hot springs. The genus is characterized by varied energy conservation strategies via CO oxidation. Of the five members in this genus, *C. hydrogenoformans*, *C. siderophilus*, *C. islandicus*, and *C. pertinax* can chemolithoautotrophically grow via hydrogenogenic CO metabolism (Yoneda et al. 2012; Sokolova and Lebedinsky 2013); whereas *C. ferrireducens* performs non-hydrogenogenic CO metabolism, in which CO oxidation is coupled with reduction of Fe (III) (Slobodkin et al. 2006). *C. pertinax* performs hydrogenogenic CO metabolism with iron reduction, in which H_2 production corresponds to approximately 69% of CO uptake (Yoneda et al. 2012), whereas in other hydrogenogenic members CO utilization results in nearly equimolar H_2 release. Among *Carboxydothermus* species, only *C. pertinax* can reduce thiosulfate or elemental sulfur as terminal electron acceptors during hydrogenogenic CO metabolism (Yoneda et al. 2012). These strongly suggest distinct strategies for energy conservation via CO metabolism in the genus *Carboxydothermus*.

The genus *Carboxydocella* also has been studied for hydrogenogenic CO metabolism. So far, three *Carboxydocella* species have been isolated from terrestrial hot

spring in Kamchatka Peninsula, Russia (Sokolova et al. 2002; Slepova et al. 2006; Slobodkina et al. 2012). They are thermophilic and strictly anaerobic, and form a deep phylogenetic branch of the class *Clostridia* (Sokolova et al. 2015). Except for *Carboxydocella manganica* (Slobodkina et al. 2012), *Carboxydocella thermautotrophica* and *Carboxydocella sporoproducens* were carboxydotroph (Sokolova et al. 2002; Slepova et al. 2006). Further, *C. thermautotrophica* is an obligate chemolithoautotroph growing exclusively by hydrogenogenic CO oxidation (Sokolova et al. 2002), whereas *C. sporoproducens* are facultative hydrogenogenic carboxydotroph (Slepova et al. 2006). These significant differences about hydrogenogenic carboxydotrophy in the genus *Carboxydocella* also raise distinct strategies for energy conservation via CO metabolism.

Previous studies on carboxydotrophs have provided clues on the mechanisms regulating CO metabolism. The CO-sensing transcriptional regulator CooA was first reported in the CODH−ECH gene cluster from *Rhodospirillum rubrum*, a hydrogenogenic carboxydotrophic bacterium (Shelver et al. 1995; Aono et al. 1996). The *cooA* gene is present in the genomic context of almost all category II CODH gene clusters and some category III CODH gene clusters, which are upregulated by activated CooA following CO sensing (Roberts et al. 2004). However, comprehensive gene expression pattern during hydrogenogenic CO metabolism is still limited.

For further investigation into hydrogenogenic CO metabolism, first, I determined draft genome sequences of anaerobic, thermophilic, and hydrogenogenic carboxydotrophs in the genus *Carboxydothermus* and *Carboxydocella*. Then, I investigated energy conservation via hydrogenogenic CO metabolism, using comparative genome in Chapter 2. Finally, I performed whole transcriptome analysis by

RNA sequencing (RNA-Seq) to understand a regulation mechanism of hydrogenogenic CO metabolism in *C. pertinax* in Chapter 3.

Chapter 2

Insight into carbon monoxide metabolism in hydrogenogenic carboxydotrophs revealed by genome analyses

Summary

Anaerobic hydrogenogenic carboxydotrophs are thought to scavenge potentially toxic CO and produce H_2 as available energy source for their surrounding microbial community. This hydrogenogenic carboxydotrophy relies on a complex of CO dehydrogenase−energy-converting hydrogenase (CODH−ECH) that is encoded in CODH−ECH gene cluster on hydrogenogenic carboxydotrophs. However, only a dozens of genomes of carboxydotrophs were available. To expand our knowledge on their hydrogenogenic CO metabolism, I determined draft genome sequences of hydrogenogenic carboxydotrophs in the genera *Carboxydothermus* and *Carboxydocella*. Draft genomes of two *Carboxydothermus* species showed that the CODH−ECH gene cluster was conserved only in *C. islandicus* as in other hydrogenogenic carboxydotrophs and the genome of *C. pertinax* lacked the gene for the CODH-I catalytic subunit encoded in the gene cluster, albeit their hydrogenogenic carboxydotrophy. Further, two hydrogenogenic *Carboxydocella* spp. strain JDF658 and ULO1, which were the first genomic reports in this genus, conserved a CODH−ECH gene cluster and possessed an additional unknown genomic context including a CODH gene on their genomes. The CODH with unknown genomic context in the genus *Carboxydocella* is potentially active and their physiological study could reveal its function in their hydrogenogenic CO metabolism.

The obtained genomic information strongly supported the hypothesis that the

CODH−ECH gene cluster would be considered to be responsible for hydrogenogenic carboxydotrophy. However, the lack of CODH gene in the CODH−ECH gene cluster in *C. pertinax* suggested that the organism performs alternative hydrogenogenic CO metabolism independent on the CODH gene product in the CODH−ECH gene cluster. To investigate energy conservation via hydrogenogenic CO metabolism in *C. pertinax*, I used comparative genome analysis of the genus *Carboxydothermus*. Comparative genome analysis revealed that *C. pertinax* could couple CODH-II alternatively to the distal ECH. Further, transcriptional analysis and cultivation study of *C. pertinax* showed that *C. pertinax* would couple CO oxidation by CODH-II with simultaneous reduction of not only H_2O by distantly encoded ECH but thiosulfate when grown under 100% CO.

2-1 Draft genome sequences of hydrogenogenic carboxydotrophs

Introduction

Hydrogenogenic CO metabolism has been partly explained by several whole genomes of carboxydotrophs. The genus *Carboxydothermus* has been one of the most studied models of thermophilic and hydrogenogenic carboxydotrophy. To date, five *Carboxydothermus* species have been described (Svetlichny et al. 1991; Slobodkin et al. 2006; Slepova et al. 2009; Novikov et al. 2011; Yoneda et al. 2012). Of these, only a complete genome of *Carboxydothermus hydrogenoformans* has been sequenced (Wu et al. 2005). The genus *Carboxydocella* also has been studied for hydrogenogenic CO metabolism. So far, three *Carboxydocella* species have been isolated from terrestrial hot spring in Kamchatka Peninsula, Russia (Sokolova et al. 2002; Slepova et al. 2006; Slobodkina et al. 2012). Their detail hydrogenogenic CO metabolism has still unveiled due to no available genome of *Carboxydocella* species.

C. hydrogenoformans possess five genes (*cooS-I* to *-V*) for catalytic subunit of CODHs (CODH-I to V) on their genome (Wu et al. 2005). According to genomic context of the gene clusters including each *cooS* and/or empirical evidence, their function is predicted as follows: CODH-I, energy conversion conjugated with ECH; CODH-II, NAD (P) H generation; CODH-III, carbon fixation in the Wood-Ljungdahl pathway conjugated with acetyl-CoA synthase (ACS); CODH-IV, oxidative stress response (Wu et al. 2005; Svetlitchnyi et al. 2001). CODH-V does not conserve sequences responsible for active center and its physiological function remains unknown (Inoue et al. 2013). *cooS-I* is arranged in gene cluster with an electron transfer protein (CooF; Kerby et al. 1992), Ni insertion protein (CooC; Kerby et al. 1997), a

transcriptional factor (CooA; Shelver et al. 1995), and ECH related genes. In most hydrogenogenic carboxydotrophs, *cooS-I* and ECH related genes form a gene cluster (CODH−ECH gene cluster) or are closely arranged (Sokolova et al. 2009). The hydrogenogenic CO utilization is performed by a complex of three enzymes; CODH-I, CooF and ECH complex (Svetlitchnyi et al. 2001; Soboh et al. 2002). However, limited genomic information prevents us from understanding their hydrogenogenic CO metabolism.

For further investigation into hydrogenogenic CO metabolism, I determined draft genome sequences of anaerobic, thermophilic, and hydrogenogenic carboxydotrophs in the genus *Carboxydothermus* and *Carboxydocella*. *Carboxydothermus pertinax* and *Carboxydothermus islandicus* are anaerobic, thermophilic and hydrogenogenic carboxydotrophs (Yoneda et al. 2012; Novikov et al. 2011). Two *Carboxydocella* spp. strain JDF658 and ULO1, isolated from distinct volcanic fronts in Japan in the process of the current study, also show hydrogenogenic carboxydotrophy. Draft genome sequences of these representative hydrogenogenic carboxydotrophs are expected to contribute to understand their hydrogenogenic CO metabolism.

Materials and methods

Strains, cultivation conditions, and genome extraction

C. pertinax Ug1 was isolated from acid hot spring (Yoneda et al. 2012) and cultivated routinely in our laboratory. *C. islandicus* SET IS-9 was obtained from Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany) as DSMZ 21278. They were grown at 65°C on modified DSM medium 507.

Modifications of DSM medium 507 were as follows: $MgCl_2.6H_2O$, $CaCl_2.2H_2O$, KH_2PO_4, and $NaHCO_3$ were decreased to 0.5 mM, 0.1 mM, 1.0 mM, and 5.0 mM, respectively (Zhao et al. 2011); resazurin and $Na_2S.9H_2O$ were omitted; and 2-(N-morpholino) ethanesulfonic acid (MES; final concentration, 10 mM) was supplemented as buffering agent. The pH of the medium was adjusted to 6.4 with 5 N NaOH solution. Genomic DNA of them was extracted using previously described NaOH methods (Mesbah et al. 1989).

Two strains of *Carboxydocella* spp., JDF658 and ULO1, were isolated from a deposit of an open-air stream from a hot spring well and the sediment of a maar lake, respectively. These spots are bordered on the volcanic fronts of Izu-Bonin Trench and Ryukyu Trench in Japan, respectively (Kamata and Kodama 1999). *Carboxydocella* spp., JDF658 and ULO1 were grown at 65 °C on the Bfc medium (Yoneda et al. 2012). Genomic DNA of two strains was extracted using a DNeasy blood and tissue kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions.

The vessels were sealed with a rubber stopper and a polypropylene screw cap. The strain was grown in 100 mL medium in a 250 mL Schott Duran bottle.

Genome sequences, Open Reading Flame (ORF) prediction, annotation

Genomic DNA of *C. pertinax* Ug1 and *C. islandicus* SET IS-9 were subjected to sequencing with the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA) using the 2×150-bp paired-end approach, generating 3,839,951 and 6,663,869 paired-end reads, respectively. High quality reads (Phred score above Q20 for 80% of the bases) were assembled into contigs using Velvet 1.2.07 or 1.2.10 software (Zerbino and Birney 2008). Genomic libraries were prepared from purified DNA of strains JDF658 and ULO1 using a Nextera® XT DNA sample prep kit (Illumina Inc.), followed

by sequencing with the Illumina MiSeq platform using a v2 reagent kit (2×150-bp paired-end reads). A total of 1,619,582 and 4,679,334 paired-end reads were generated for strains JDF658 and ULO1, respectively. High-quality reads (Phred score > Q30 for 80% of bases) were assembled into contigs using Velvet 1.2.10 (Zerbino and Birney 2008). The assembled contigs were subjected to the Microbial Genome Annotation Pipeline (MiGAP; http://www.migap.org/index.php/en) (Sugawara et al. 2009) to predict Open Reading Flames (ORFs), followed by manual curation. Subsequently, the protein sequences were annotated using BLASTP searches (Altschul et al. 1990; Altschul et al. 1997) against the Clusters of Orthologous Groups (COG) database (Tatusov 2000).

Phylogenetic analysis of CODH genes

For phylogenetic analysis of catalytic subunit genes of CODH, I retrieved three CODH amino acid sequences in clade D and 29 CODH amino acid sequences in clade F from the phylogenetic tree described previously (Techtmann et al. 2012) as references. The *C. hydrogenoformans* CooS-III gene containing a frameshift mutation (Wu et al. 2005) was translated to amino acid sequence with disregard of its frameshift mutation. The sequences were aligned using MAFFT 7.215 (Katoh et al. 2013) and gap positions were removed automatically using trimAL1.4 (Capella-Gutierrez et al. 2009). Phylogenetic reconstructions were performed and visualized as described above. Robustness of the topology in the phylogenetic trees was evaluated by bootstrap analysis based on 100 runs.

Results and discussion

General features of *C. pertinax* and *C. islandicus* genomes

The draft genomes of *C. pertinax* and *C. islandicus* were assembled into 96 (2.47 Mbp; G+C 40.7%) and 142 (2.39 Mbp; G+C 42.0%) contigs, respectively. The numbers of predicted ORFs were 2577 and 2480, respectively. Of these, 1,987 and 1,878 ORFs were assigned to the COG categories (Table 2-1-1), respectively.

Phylogenetic analysis of CODH genes from *C. pertinax* and *C. islandicus* genomes

I identified four CODH gene clusters (corresponding to CODH-II to CODH-V of *C. hydrogenoformans*) (Wu et al. 2005) in *C. pertinax* and five CODH gene clusters (corresponding to CODH-I to CODH-V of *C. hydrogenoformans*; Wu et al. 2005) in *C. islandicus* (Fig. 2-1-1). The CODH−ECH gene cluster was conserved in *C. islandicus*, like in other hydrogenogenic carboxydotrophs, whereas the genome of *C. pertinax* contained an ECH gene cluster but not the CODH−ECH gene cluster. In addition, *cooS*, encodes CODH-I, was not amplified when using a specific PCR primer set (forward primer, 5’-GCGGCGCGGGATTCCTTTAG-3’; reverse primer, 5’-AAGCCCGGCTGCCTTTCCTA-3’). These results suggested that in case of *C. pertinax*, complexes of gene products from the ECH gene cluster and its physically unlinked *cooS* were responsible for hydrogenogenic CO metabolism.

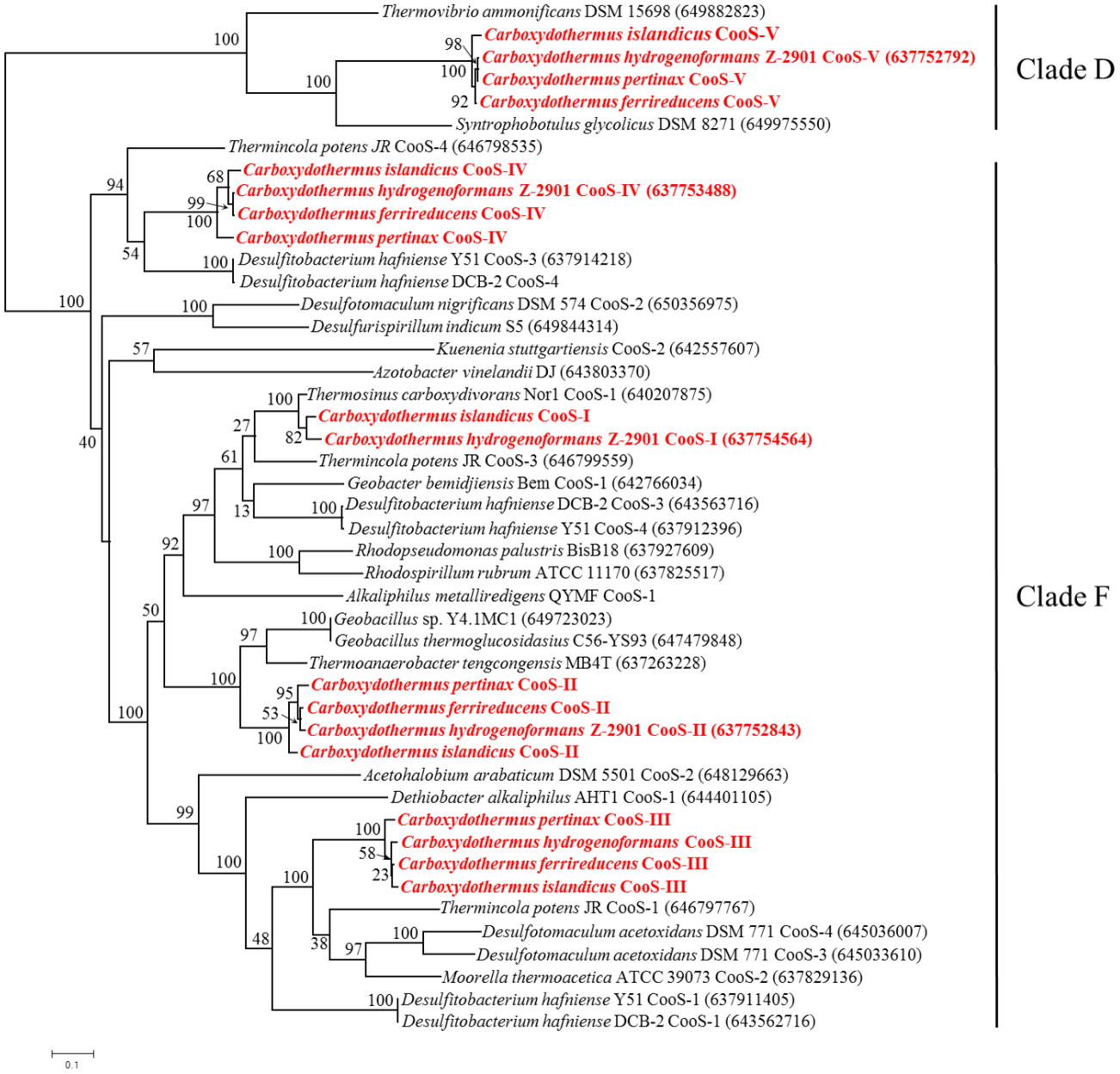


Fig. 2-1-1 Maximum likelihood phylogenetic tree based on the amino acid sequence of *cooS* genes

cooS genes from *Carboxydothermus* species are indicated in red and bold font. Each accession number is indicated in a parenthes.

Table 2-1-1 ORFs assigned to COG functional categories in *C. pertinax*

Gene ID	Length	E-value	Predicted function	COG
cpu_00000	1808	2.00E-38	HD-GYP domain	T
cpu_00020	332	2.00E-12	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	T
cpu_00030	1850	5.00E-40	Predicted polymerase	L
cpu_00040	317	1.00E-11	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	T
cpu_00050	230	4.00E-13	Serine phosphatase RsbU, regulator of sigma subunit	TK
cpu_00060	329	2.00E-16	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	T
cpu_00080	1037	6.00E-141	High-affinity nickel permease	P
cpu_00090	926	2.00E-65	Lysophospholipase	I
cpu_00100	191	2.00E-09	Uncharacterized protein, 4-oxalocrotonate tautomerase homolog	R
cpu_00110	470	7.00E-36	Predicted inhibitor of MCP methylation, homolog of CheC	N
cpu_00120	419	1.00E-14	Hemerythrin	P
cpu_00130	878	3.00E-63	FOG: GGDEF domain	T
cpu_00140	1151	2.00E-22	FOG: CheY-like receiver	T
cpu_00170	560	2.00E-41	Pyruvate-formate lyase-activating enzyme	O
cpu_00190	704	3.00E-13	UTP:GlnB (protein PII) uridylyltransferase	O
cpu_00200	773	3.00E-73	ATPases involved in chromosome partitioning	D
cpu_00210	611	1.00E-46	Uncharacterized metal-binding protein conserved in archaea	R
cpu_00220	749	6.00E-20	Multidrug resistance efflux pump	V
cpu_00230	785	3.00E-99	ABC-type multidrug transport system, ATPase component	V
cpu_00240	1121	7.00E-26	ABC-type multidrug transport system, permease component	V
cpu_00260	377	8.00E-12	MinD superfamily P-loop ATPase containing an inserted ferredoxin domain	C
cpu_00280	1394	2.00E-55	UDP-N-acetylmuramyl tripeptide synthase	M
cpu_00290	776	6.00E-106	Predicted glutamine amidotransferase	R
cpu_00300	1835	6.00E-28	Amino acid transporters	E
cpu_00310	1883	0	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	I
cpu_00340	566	1.00E-40	Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins	F
cpu_00350	458	3.00E-23	Putative translation initiation inhibitor, yjgF family	J
cpu_00370	488	2.00E-51	Sortase and related acyltransferases	M
cpu_00390	716	3.00E-99	NAD-dependent protein deacetylases, SIR2 family	K
cpu_00400	1991	0	Helicase subunit of the DNA excision repair complex	L
cpu_00410	2846	0	Excinuclease ATPase subunit	L
cpu_00430	1682	1.00E-68	N-methylhydantoinase A/acetone carboxylase, beta subunit	EQ
cpu_00440	1283	1.00E-92	Deacetylases, including yeast histone deacetylase and acetoin utilization protein	BQ
cpu_00450	293	2.00E-11	Acyl-CoA synthetase (NDP forming)	C
cpu_00470	380	5.00E-10	Prophage maintenance system killer protein	R
cpu_00480	1823	0	Nuclease subunit of the excinuclease complex	L
cpu_00490	803	9.00E-68	Predicted hydrolases of the HAD superfamily	R
cpu_00500	1022	6.00E-52	L-lactate dehydrogenase	C
cpu_00510	956	8.00E-79	Transcriptional regulator/sugar kinase	KG
cpu_00520	743	3.00E-44	Histidinol phosphatase and related hydrolases of the PHP family	ER
cpu_00530	878	3.00E-139	Predicted P-loop-containing kinase	R
cpu_00540	1241	7.00E-104	Uncharacterized conserved protein	S
cpu_00550	824	3.00E-32	Predicted xylanase/chitin deacetylase	G
cpu_00560	938	3.00E-95	Uncharacterized protein conserved in bacteria	S
cpu_00570	977	7.00E-10	Metal-dependent hydrolases of the beta-lactamase superfamily III	R
cpu_00580	488	9.00E-30	Acetyltransferase (isoleucine patch superfamily)	R
cpu_00590	587	6.00E-20	Predicted membrane protein	S
cpu_00600	1382	1.00E-152	DNA-directed RNA polymerase specialized sigma subunit, sigma54 homolog	K
cpu_00610	1007	0	Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase	G
cpu_00620	1187	0	3-phosphoglycerate kinase	G
cpu_00630	755	1.00E-126	Triosephosphate isomerase	G
cpu_00640	1520	0	Phosphoglyceromutase	G
cpu_00650	1286	0	Enolase	G
cpu_00660	233	3.00E-08	Preprotein translocase subunit SecG	U
cpu_00670	2060	0	Inorganic pyrophosphatase	C

Table 2-1-1 Continued

cpu_00680	407	6.00E-13	Uncharacterized protein, possibly involved in aromatic compounds catabolism	Q
cpu_00690	551	2.00E-72	Predicted hydrolase (HD superfamily)	R
cpu_00700	2177	0	Exoribonuclease R	K
cpu_00710	449	1.00E-79	tmRNA-binding protein	O
cpu_00730	1523	8.00E-23	Site-specific recombinases, DNA invertase Pin homologs	L
cpu_00800	563	2.00E-11	Uncharacterized conserved protein	S
cpu_00810	1124	1.00E-80	Uncharacterized conserved protein	S
cpu_00820	761	3.00E-26	SOS-response transcriptional repressors (RecA-mediated autopeptidases)	KT
cpu_00830	230	5.00E-14	Predicted transcriptional regulators	K
cpu_00910	938	6.00E-08	RecA/RadA recombinase	L
cpu_00920	950	2.00E-80	Bacteriophage protein gp37	S
cpu_00950	815	6.00E-39	DNA replication protein	L
cpu_00960	398	3.00E-35	Single-stranded DNA-binding protein	L
cpu_01020	1244	2.00E-09	Transposase and inactivated derivatives	L
cpu_01050	662	3.00E-10	Uncharacterized protein conserved in bacteria	S
cpu_01070	575	8.00E-08	Phage DNA packaging protein, Nu1 subunit of terminase	L
cpu_01080	941	2.00E-71	Bacteriophage tail assembly protein	R
cpu_01100	1553	1.00E-68	Bacteriophage capsid protein	R
cpu_01110	1085	1.00E-48	Protease subunit of ATP-dependent Clp proteases	OU
cpu_01190	1451	2.00E-24	Phage tail sheath protein FI	R
cpu_01200	521	3.00E-22	Phage tail tube protein FII	R
cpu_01230	2276	3.00E-28	Phage-related tail protein	S
cpu_01250	1010	4.00E-33	Phage protein D	R
cpu_01260	479	8.00E-28	Phage P2 baseplate assembly protein gpV	R
cpu_01280	296	1.00E-08	Phage baseplate assembly protein W	R
cpu_01290	1121	1.00E-46	Phage-related baseplate assembly protein	R
cpu_01300	563	3.00E-16	Bacteriophage P2-related tail formation protein	R
cpu_01340	410	4.00E-33	Phage-related holin (Lysis protein)	R
cpu_01350	725	2.00E-11	Putative peptidoglycan-binding domain-containing protein	M
cpu_01390	1955	2.00E-146	NADH:flavin oxidoreductases, Old Yellow Enzyme family	C
cpu_01410	566	2.00E-24	CBS-domain-containing membrane protein	T
cpu_01420	914	4.00E-44	Tetrahydromethanopterin S-methyltransferase, subunit H	H
cpu_01440	1847	5.00E-55	Predicted cobalamin binding protein	R
cpu_01460	440	6.00E-47	dUTPase	F
cpu_01470	1664	2.00E-37	Signal transduction histidine kinase	T
cpu_01490	647	4.00E-24	Uncharacterized protein conserved in bacteria	S
cpu_01500	575	5.00E-12	Spore germination protein	R
cpu_01520	797	6.00E-86	Glutamate racemase	M
cpu_01530	410	8.00E-38	Methylmalonyl-CoA mutase, C-terminal domain/subunit (cobalamin-binding)	I
cpu_01550	1448	0	Glutamate mutase epsilon subunit	E
cpu_01560	974	3.00E-41	Benzoyl-CoA reductase	E
cpu_01570	755	2.00E-68	Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)	I
cpu_01580	728	3.00E-42	Metal-dependent hydrolases of the beta-lactamase superfamily III	R
cpu_01600	764	4.00E-105	RNase PH	J
cpu_01610	608	2.00E-87	Xanthosine triphosphate pyrophosphatase	F
cpu_01620	470	1.00E-43	Predicted phosphoesterase	R
cpu_01630	1283	5.00E-118	FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor)	O
cpu_01640	587	1.00E-123	Protease subunit of ATP-dependent Clp proteases	OU
cpu_01650	1256	0	ATP-dependent protease Clp, ATPase subunit	O
cpu_01670	1643	3.00E-99	Predicted ATP-dependent protease	O
cpu_01680	2378	0	ATP-dependent Lon protease, bacterial type	O
cpu_01700	2630	0	Valyl-tRNA synthetase	J
cpu_01710	1298	3.00E-144	Folylpolyglutamate synthase	H
cpu_01750	626	7.00E-85	AT-rich DNA-binding protein	R
cpu_01760	563	1.00E-79	Nucleotide-binding protein implicated in inhibition of septum formation	D
cpu_01770	680	7.00E-91	DNA repair proteins	L
cpu_01780	1046	7.00E-178	Actin-like ATPase involved in cell morphogenesis	D

Table 2-1-1 Continued

cpu_01790	821	2.00E-58	Cell shape-determining protein	M
cpu_01810	1985	3.00E-115	Cell division protein FtsI/penicillin-binding protein 2	M
cpu_01820	377	8.00E-28	Septum formation inhibitor	D
cpu_01830	794	3.00E-156	Septum formation inhibitor-activating ATPase	D
cpu_01840	281	7.00E-17	Septum formation topological specificity factor	D
cpu_01900	1139	4.00E-106	Bacterial cell division membrane protein	D
cpu_01910	935	1.00E-19	Zn-dependent hydrolases, including glyoxylases	R
cpu_01920	653	9.00E-20	Membrane-bound metallopeptidase	D
cpu_01940	1862	1.00E-64	Fe-S oxidoreductase	C
cpu_01950	611	1.00E-30	Uncharacterized protein conserved in bacteria	S
cpu_01960	1166	3.00E-94	Ribonucleases G and E	J
cpu_01970	311	4.00E-45	Ribosomal protein L21	J
cpu_01980	299	2.00E-50	Ribosomal protein L27	J
cpu_02000	1286	0	Predicted GTPase	R
cpu_02010	377	4.00E-11	Transcriptional regulator	KT
cpu_02020	1223	1.00E-10	Permeases of the major facilitator superfamily	R
cpu_02030	854	2.00E-114	ABC-type tungstate transport system, permease component	H
cpu_02040	692	3.00E-57	ABC-type tungstate transport system, periplasmic component	H
cpu_02050	698	2.00E-49	ABC-type cobalt transport system, ATPase component	P
cpu_02060	353	5.00E-25	N-terminal domain of molybdenum-binding protein	R
cpu_02070	461	3.00E-33	Uncharacterized conserved protein	S
cpu_02080	1118	4.00E-180	Glutamate 5-kinase	E
cpu_02090	1247	0	Gamma-glutamyl phosphate reductase	E
cpu_02100	1256	2.00E-65	Mg/Co/Ni transporter MgtE (contains CBS domain)	P
cpu_02110	1250	2.00E-70	Mn2+ and Fe2+ transporters of the NRAMP family	P
cpu_02120	1037	6.00E-95	Uncharacterized conserved protein	S
cpu_02140	605	1.00E-64	Nicotinic acid mononucleotide adenylyltransferase	H
cpu_02150	251	8.00E-16	RNA-binding proteins (RRM domain)	R
cpu_02160	578	8.00E-50	Predicted HD superfamily hydrolase involved in NAD metabolism	H
cpu_02170	362	5.00E-41	Uncharacterized homolog of plant Iojap protein	S
cpu_02180	2465	0	Leucyl-tRNA synthetase	J
cpu_02190	641	4.00E-47	cAMP-binding proteins - catabolite gene activator	T
cpu_02210	761	0	ABC-type phosphate transport system, ATPase component	P
cpu_02230	1073	8.00E-163	Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase	J
cpu_02240	374	7.00E-52	NifU homolog involved in Fe-S cluster formation	C
cpu_02250	1178	0	Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes	E
cpu_02260	446	7.00E-57	Predicted transcriptional regulator	K
cpu_02270	1358	0	ATPase related to the helicase subunit of the Holliday junction resolvase	L
cpu_02280	320	3.00E-25	Thioredoxin domain-containing protein	O
cpu_02290	728	1.00E-97	Dinucleotide-utilizing enzymes	H
cpu_02300	1775	0	Aspartyl-tRNA synthetase	J
cpu_02310	1271	0	Histidyl-tRNA synthetase	J
cpu_02330	1298	3.00E-63	FOG: GGDEF domain	T
cpu_02340	731	7.00E-47	Predicted divalent heavy-metal cations transporter	P
cpu_02350	1379	5.00E-45	Predicted Fe-S oxidoreductases	R
cpu_02380	1955	1.00E-128	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_02390	941	4.00E-36	Predicted permeases	R
cpu_02400	1397	3.00E-65	Acetylornithine deacetylase	E
cpu_02410	1787	0	Carbon starvation protein, predicted membrane protein	T
cpu_02420	755	5.00E-63	Response regulator of the LytR/AlgR family	KT
cpu_02430	1658	0	Putative regulator of cell autolysis	T
cpu_02440	623	2.00E-44	Zn-dependent hydrolases, including glyoxylases	R
cpu_02450	437	9.00E-73	D-Tyr-tRNAtyr deacylase	J
cpu_02460	2150	0	Guanosine polyphosphate pyrophosphohydrolases/synthetases	TK
cpu_02470	2564	1.00E-120	Single-stranded DNA-specific exonuclease	L
cpu_02480	686	2.00E-61	Uncharacterized membrane protein, possible Na+ channel or pump	R
cpu_02490	926	2.00E-16	Permeases of the drug/metabolite transporter (DMT) superfamily	GER

Table 2-1-1 Continued

cpu_02500	830	4.00E-56	Predicted metal-dependent phosphoesterases (PHP family)	R
cpu_02510	719	5.00E-138	ABC-type branched-chain amino acid transport systems, ATPase component	E
cpu_02520	767	4.00E-131	ABC-type branched-chain amino acid transport systems, ATPase component	E
cpu_02530	977	3.00E-69	ABC-type branched-chain amino acid transport system, permease component	E
cpu_02540	884	2.00E-83	Branched-chain amino acid ABC-type transport system, permease components	E
cpu_02550	1130	9.00E-99	ABC-type branched-chain amino acid transport systems, periplasmic component	E
cpu_02560	1523	0	ABC-type uncharacterized transport systems, ATPase components	R
cpu_02570	1043	5.00E-106	ABC-type uncharacterized transport system, permease component	R
cpu_02580	935	3.00E-110	Uncharacterized ABC-type transport system, permease component	R
cpu_02590	992	3.00E-75	Uncharacterized ABC-type transport system	R
cpu_02600	3398	0	DNA-directed RNA polymerase, beta subunit/140 kD subunit	K
cpu_02610	3458	0	DNA-directed RNA polymerase, beta' subunit/160 kD subunit	K
cpu_02620	263	2.00E-10	Ribosomal protein HS6-type (S12/L30/L7a)	J
cpu_02630	380	4.00E-71	Ribosomal protein S12	J
cpu_02640	470	8.00E-86	Ribosomal protein S7	J
cpu_02650	2078	0	Translation elongation factors (GTPases)	J
cpu_02660	91	1.00E-18	GTPases - translation elongation factors	J
cpu_02670	845	8.00E-29	Predicted esterase of the alpha-beta hydrolase superfamily	R
cpu_02680	689	4.00E-70	DNA-binding protein, stimulates sugar fermentation	R
cpu_02710	797	3.00E-27	Histidinol phosphatase and related hydrolases of the PHP family	ER
cpu_02720	437	3.00E-31	CBS-domain-containing membrane protein	T
cpu_02730	2561	6.00E-73	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_02740	686	1.00E-64	Fe-S-cluster-containing hydrogenase components 1	C
cpu_02750	1247	2.00E-58	Polysulphide reductase	C
cpu_02760	431	7.00E-69	Uncharacterized protein conserved in bacteria	S
cpu_02770	935	3.00E-172	Predicted S-adenosylmethionine-dependent methyltransferase	M
cpu_02790	2219	7.00E-175	Cell division protein FtsI/penicillin-binding protein 2	M
cpu_02800	1475	0	UDP-N-acetylmuramyl tripeptide synthase	M
cpu_02810	1376	3.00E-157	UDP-N-acetylmuramyl pentapeptide synthase	M
cpu_02820	977	8.00E-64	UDP-N-acetylmuramyl pentapeptide phosphotransferase	M
cpu_02830	1358	4.00E-171	UDP-N-acetylmuramoylalanine-D-glutamate ligase	M
cpu_02840	1127	1.00E-96	Bacterial cell division membrane protein	D
cpu_02850	1109	1.00E-116	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase	M
cpu_02860	1337	0	UDP-N-acetylmuramate-alanine ligase	M
cpu_02870	908	7.00E-112	UDP-N-acetylmuramate dehydrogenase	M
cpu_02880	1262	0	UDP-N-acetylglucosamine enolpyruvyl transferase	M
cpu_02890	740	6.00E-19	Cell division septal protein	M
cpu_02900	716	4.00E-47	Uncharacterized protein conserved in bacteria	S
cpu_02910	695	9.00E-34	Uncharacterized protein conserved in bacteria	S
cpu_02920	329	9.00E-36	Uncharacterized conserved protein (small basic protein)	S
cpu_02930	1232	6.00E-135	Actin-like ATPase involved in cell division	D
cpu_02940	1058	2.00E-147	Cell division GTPase	D
cpu_02960	977	4.00E-97	Selenophosphate synthase	E
cpu_02980	713	1.00E-53	DNA-directed RNA polymerase specialized sigma subunit	K
cpu_02990	773	1.00E-84	DNA-directed RNA polymerase specialized sigma subunit	K
cpu_03010	257	4.00E-14	Uncharacterized conserved protein	S
cpu_03020	464	4.00E-75	Predicted transcriptional regulator, consists of a Zn-ribbon and ATP-cone domains	K
cpu_03030	2018	0	Oxygen-sensitive ribonucleoside-triphosphate reductase	F
cpu_03040	509	4.00E-27	Pyruvate-formate lyase-activating enzyme	O
cpu_03050	1514	1.00E-103	Predicted exonuclease of the beta-lactamase fold involved in RNA processing	J
cpu_03060	830	3.00E-84	Uncharacterized conserved protein	S
cpu_03070	695	5.00E-102	Response regulators	TK
cpu_03080	1304	2.00E-101	Signal transduction histidine kinase	T
cpu_03090	659	1.00E-63	Phosphate uptake regulator	P
cpu_03110	683	5.00E-94	Predicted enzyme with a TIM-barrel fold	R
cpu_03120	413	4.00E-28	Uncharacterized protein conserved in bacteria	S
cpu_03130	776	3.00E-89	Pyrroline-5-carboxylate reductase	E

Table 2-1-1 Continued

cpu_03140	263	1.00E-07	Predicted integral membrane protein	S
cpu_03150	740	5.00E-70	Uncharacterized conserved protein, contains S4-like domain	S
cpu_03160	506	3.00E-14	Cell division initiation protein	D
cpu_03170	3047	6.00E-75	Fe-S oxidoreductase	C
cpu_03180	668	2.00E-39	DNA polymerase III, epsilon subunit and related 3'-5' exonucleases	L
cpu_03190	1895	8.00E-117	Predicted signal-transduction protein containing cAMP-binding and CBS domains	T
cpu_03200	335	3.00E-23	Predicted membrane protein	S
cpu_03210	1529	5.00E-177	Predicted symporter	R
cpu_03220	1709	7.00E-63	Putative multicopper oxidases	Q
cpu_03230	2672	7.00E-67	Fe-S oxidoreductase	C
cpu_03240	581	1.00E-16	Transcriptional regulator	K
cpu_03250	1352	2.00E-96	tRNA and rRNA cytosine-C5-methylases	J
cpu_03260	1649	1.00E-63	Response regulator containing a CheY-like receiver domain and an HD-GYP domain	KT
cpu_03270	512	2.00E-68	Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins	F
cpu_03280	938	7.00E-72	Dioxygenases related to 2-nitropropane dioxygenase	R
cpu_03310	1940	1.00E-46	Signal transduction histidine kinase	T
cpu_03320	929	3.00E-43	ADP-ribosylglycohydrolase	O
cpu_03330	1271	5.00E-99	Na+/H+ antiporter NhaD and related arsenite permeases	P
cpu_03340	1277	7.00E-107	Na+/H+ antiporter NhaD and related arsenite permeases	P
cpu_03350	710	3.00E-29	Xanthine and CO dehydrogenases maturation factor, XdhC/CoxF family	O
cpu_03360	341	1.00E-21	Xanthine and CO dehydrogenases maturation factor, XdhC/CoxF family	O
cpu_03370	1427	0	Acetyl-CoA carboxylase, carboxyltransferase component (subunits alpha and beta)	I
cpu_03380	1316	4.00E-111	Xanthine/uracil permeases	F
cpu_03390	797	2.00E-82	Uncharacterized conserved protein	S
cpu_03410	1598	3.00E-142	Uncharacterized FAD-dependent dehydrogenases	R
cpu_03420	1331	6.00E-164	Phosphomannomutase	G
cpu_03430	1829	0	Glucosamine 6-phosphate synthetase	M
cpu_03500	290	5.00E-14	Predicted membrane protein	S
cpu_03510	191	6.00E-10	Predicted membrane protein	S
cpu_03520	422	7.00E-19	Transcriptional regulators	K
cpu_03530	578	2.00E-68	Putative NADPH-quinone reductase (modulator of drug activity B)	R
cpu_03540	581	4.00E-50	Predicted hydrolase (HD superfamily)	R
cpu_03570	572	1.00E-15	Transcriptional regulator	K
cpu_03590	227	2.00E-12	Predicted membrane protein	S
cpu_03600	683	3.00E-08	Cyclopropane fatty acid synthase and related methyltransferases	M
cpu_03610	770	1.00E-11	Predicted Zn peptidase	E
cpu_03620	344	2.00E-08	Predicted transcriptional regulators	K
cpu_03630	1544	1.00E-138	Type I restriction-modification system methyltransferase subunit	V
cpu_03640	1208	2.00E-18	Restriction endonuclease S subunits	V
cpu_03650	3101	0	Type I site-specific restriction-modification system	V
cpu_03670	797	4.00E-92	CO dehydrogenase maturation factor	D
cpu_03680	3824	2.00E-76	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	CP
cpu_03690	965	9.00E-57	Formate hydrogenlyase subunit 4	C
cpu_03700	431	4.00E-56	Ni,Fe-hydrogenase III small subunit	C
cpu_03710	497	2.00E-20	Formate hydrogenlyase subunit	C
cpu_03720	542	2.00E-16	Ni,Fe-hydrogenase III component G	C
cpu_03730	1082	2.00E-124	Ni,Fe-hydrogenase III large subunit	C
cpu_03740	344	9.00E-37	Zn finger protein HypA/HybF (possibly regulating hydrogenase expression)	R
cpu_03750	563	4.00E-38	Fe-S-cluster-containing hydrogenase components 2	C
cpu_03760	395	5.00E-09	Predicted transcriptional regulators	K
cpu_03790	926	5.00E-18	Permeases of the drug/metabolite transporter (DMT) superfamily	GER
cpu_03800	794	3.00E-32	Predicted transcriptional regulators	K
cpu_03810	1568	4.00E-16	Arabinose efflux permease	G
cpu_03820	1085	1.00E-60	Multidrug resistance efflux pump	V
cpu_03830	368	2.00E-39	Uncharacterized conserved protein	S
cpu_03840	761	3.00E-71	Predicted metal-dependent hydrolases with the TIM-barrel fold	R
cpu_03850	1136	8.00E-50	ABC-type uncharacterized transport system, periplasmic component	R

Table 2-1-1 Continued

cpu_03860	1808	3.00E-61	Signal transduction histidine kinase regulating C4-dicarboxylate transport system	T
cpu_03870	1331	0	Response regulator	T
cpu_03880	641	2.00E-58	Fe-S-cluster-containing hydrogenase components 1	C
cpu_03890	1232	2.00E-44	Polysulphide reductase	C
cpu_03900	3107	3.00E-57	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_03910	812	2.00E-59	FOG: GGDEF domain	T
cpu_03920	920	4.00E-18	Predicted xylanase/chitin deacetylase	G
cpu_03930	530	2.00E-09	Membrane-associated phospholipid phosphatase	I
cpu_03940	689	1.00E-37	Predicted xylanase/chitin deacetylase	G
cpu_03980	530	6.00E-30	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	K
cpu_03990	1148	1.00E-13	Uncharacterized proteins, LmbE homologs	S
cpu_04000	572	6.00E-51	Glycerol-3-phosphate responsive antiterminator (mRNA-binding)	K
cpu_04010	710	1.00E-78	Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)	G
cpu_04020	1502	0	Glycerol kinase	C
cpu_04030	1586	4.00E-115	Glycerol-3-phosphate dehydrogenase	C
cpu_04040	1178	5.00E-50	Anaerobic glycerol-3-phosphate dehydrogenase	E
cpu_04050	1211	1.00E-65	Fe-S oxidoreductase	C
cpu_04070	1343	1.00E-103	Putative regulator of cell autolysis	T
cpu_04080	755	8.00E-72	Response regulator of the LytR/AlgR family	KT
cpu_04090	1943	0	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	I
cpu_04100	371	6.00E-41	Phosphopantetheinyl transferase (holo-ACP synthase)	I
cpu_04110	1571	2.00E-73	Predicted sugar kinase	G
cpu_04120	1118	0	Alanine dehydrogenase	E
cpu_04130	1106	1.00E-151	Alanine racemase	M
cpu_04140	1013	1.00E-48	Glycosyltransferase	M
cpu_04150	2198	2.00E-113	Glycogen debranching enzyme	G
cpu_04160	1430	9.00E-162	Trehalose-6-phosphate synthase	G
cpu_04170	755	2.00E-23	Trehalose-6-phosphatase	G
cpu_04180	902	3.00E-15	Zn-dependent hydrolases, including glyoxylases	R
cpu_04190	269	2.00E-20	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_04200	815	2.00E-37	Cell wall-associated hydrolases (invasion-associated proteins)	M
cpu_04210	2270	0	Ribonucleotide reductase, alpha subunit	F
cpu_04220	824	1.00E-17	Predicted transcriptional regulators	K
cpu_04230	1415	3.00E-136	Dihydroorotase and related cyclic amidohydrolases	F
cpu_04240	575	3.00E-42	Amidases related to nicotinamidase	Q
cpu_04250	551	3.00E-98	3-polyprenyl-4-hydroxybenzoate decarboxylase	H
cpu_04260	1352	8.00E-135	3-polyprenyl-4-hydroxybenzoate decarboxylase and related decarboxylases	H
cpu_04270	272	6.00E-31	Carbon dioxide concentrating mechanism/carboxysome shell protein	QC
cpu_04280	977	5.00E-38	Ethanolamine utilization protein	E
cpu_04290	356	4.00E-24	Carbon dioxide concentrating mechanism/carboxysome shell protein	QC
cpu_04300	284	3.00E-33	Carbon dioxide concentrating mechanism/carboxysome shell protein	QC
cpu_04310	809	1.00E-91	Ethanolamine utilization protein, possible chaperonin	E
cpu_04330	272	5.00E-29	Carbon dioxide concentrating mechanism/carboxysome shell protein	QC
cpu_04340	1811	0	Adenine deaminase	F
cpu_04350	860	2.00E-73	Aerobic-type carbon monoxide dehydrogenase, middle subunit CoxM/CutM homologs	C
cpu_04360	473	4.00E-81	Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homologs	C
cpu_04370	1265	2.00E-109	Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL homologs	C
cpu_04380	1001	2.00E-86	Aerobic-type carbon monoxide dehydrogenase, large subunit CoxL/CutL homologs	C
cpu_04400	1265	1.00E-21	Uncharacterized MobA-related protein	R
cpu_04410	917	3.00E-112	Glutamate formiminotransferase	E
cpu_04420	623	3.00E-64	Methenyl tetrahydrofolate cyclohydrolase	E
cpu_04430	1328	1.00E-161	Permeases	R
cpu_04440	1334	5.00E-126	Cytosine deaminase and related metal-dependent hydrolases	FR
cpu_04450	1712	0	Adenine deaminase	F
cpu_04460	656	9.00E-59	L-serine deaminase	E
cpu_04470	899	2.00E-71	L-serine deaminase	E
cpu_04480	680	1.00E-45	Predicted metal-dependent hydrolase	R

Table 2-1-1 Continued

cpu_04490	587	2.00E-31	Response regulator with putative antiterminator output domain	T
cpu_04500	1328	0	Glutamine synthetase	E
cpu_04510	1985	3.00E-45	FOG: GGDEF domain	T
cpu_04520	1100	2.00E-72	Glutamate synthase domain 1	E
cpu_04530	1502	0	Glutamate synthase domain 2	E
cpu_04540	740	8.00E-57	Glutamate synthase domain 3	E
cpu_04550	551	7.00E-20	Response regulator with putative antiterminator output domain	T
cpu_04560	1064	1.00E-34	Glutamine synthetase	E
cpu_04570	1580	1.00E-36	Glutamate synthase domain 2	E
cpu_04580	1331	0	Glutamine synthetase	E
cpu_04590	1370	0	Pyruvate/2-oxoglutarate dehydrogenase complex	C
cpu_04600	1295	1.00E-75	Heterodisulfide reductase, subunit B	C
cpu_04610	254	1.00E-14	Predicted transcriptional regulators	K
cpu_04620	350	4.00E-40	Growth inhibitor	T
cpu_04630	665	5.00E-56	Predicted glutamine amidotransferases	R
cpu_04640	1130	8.00E-68	Imidazolonepropionase and related amidohydrolases	Q
cpu_04650	392	8.00E-24	FOG: CheY-like receiver	T
cpu_04670	527	2.00E-47	Predicted ATPase or kinase	R
cpu_04680	653	2.00E-25	Inactive homolog of metal-dependent proteases, putative molecular chaperone	O
cpu_04690	434	1.00E-31	Acetyltransferases	R
cpu_04700	1001	6.00E-156	Metal-dependent proteases with possible chaperone activity	O
cpu_04720	1298	4.00E-12	Predicted Fe-S oxidoreductases	R
cpu_04730	1283	0	Coenzyme F390 synthetase	H
cpu_04740	431	8.00E-57	ACT domain-containing protein	R
cpu_04750	578	5.00E-58	5-formyltetrahydrofolate cyclo-ligase	H
cpu_04760	485	5.00E-70	NADH:ubiquinone oxidoreductase 24 kD subunit	C
cpu_04770	1778	0	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	C
cpu_04780	1058	2.00E-131	Uncharacterized anaerobic dehydrogenase	R
cpu_04790	1610	0	Uncharacterized anaerobic dehydrogenase	R
cpu_04800	188	7.00E-17	Uncharacterized protein, 4-oxalocrotonate tautomerase homolog	R
cpu_04810	485	4.00E-33	Fe-S-cluster-containing hydrogenase components 2	C
cpu_04820	1901	1.00E-142	6Fe-6S prismane cluster-containing protein	C
cpu_04830	1184	4.00E-90	NAD(P)H-nitrite reductase	C
cpu_04840	497	1.00E-55	Rubrerythrin	C
cpu_04880	638	5.00E-50	Uncharacterized conserved protein	S
cpu_04890	368	1.00E-15	Uncharacterized conserved protein	S
cpu_04900	488	1.00E-41	Predicted nucleic-acid-binding protein containing a Zn-ribbon	R
cpu_04910	1253	2.00E-103	Acetyl-CoA acetyltransferase	I
cpu_04920	1145	5.00E-152	Acyl-CoA dehydrogenases	I
cpu_04930	1925	0	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	I
cpu_04940	1379	2.00E-151	Transcriptional regulator	KT
cpu_04950	1958	0	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	I
cpu_04960	1028	7.00E-72	Thiamine monophosphate synthase	H
cpu_04970	1289	0	Thiamine biosynthesis protein ThiC	H
cpu_04980	803	8.00E-94	Hydroxyethylthiazole kinase, sugar kinase family	H
cpu_04990	980	1.00E-102	Thiamine monophosphate kinase	H
cpu_05000	923	7.00E-38	Putative enzyme of poly-gamma-glutamate biosynthesis (capsule formation)	M
cpu_05010	1277	0	Predicted alternative tryptophan synthase beta-subunit (paralog of TrpB)	R
cpu_05020	524	2.00E-34	Nitroreductase	C
cpu_05030	902	9.00E-68	Transcriptional regulator	K
cpu_05040	332	4.00E-12	Predicted metal-binding protein	S
cpu_05060	614	5.00E-59	Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain)	H
cpu_05070	722	4.00E-54	ABC-type Co2+ transport system, permease component	P
cpu_05080	320	8.00E-36	ABC-type cobalt transport system, periplasmic component	P
cpu_05090	743	2.00E-17	ABC-type cobalt transport system	P
cpu_05100	1916	3.00E-104	Cobalamin biosynthesis protein CbiD	H
cpu_05110	605	2.00E-31	Precorrin-6B methylase 1	H

Table 2-1-1 Continued

cpu_05120	575	5.00E-48	Precorrin-6B methylase 2	H
cpu_05130	659	6.00E-49	Precorrin-2 methylase	H
cpu_05140	755	4.00E-126	Precorrin-4 methylase	H
cpu_05150	1010	2.00E-47	Cobalamin biosynthesis protein CbiG	H
cpu_05160	821	2.00E-116	Precorrin-3B methylase	H
cpu_05170	776	8.00E-56	Precorrin-6x reductase	H
cpu_05180	362	2.00E-19	Uncharacterized conserved protein	S
cpu_05190	614	3.00E-66	Precorrin isomerase	H
cpu_05200	506	2.00E-79	Adenosyl cobinamide kinase/adenosyl cobinamide phosphate guanylyltransferase	H
cpu_05210	1493	0	Cobyric acid synthase	H
cpu_05220	959	9.00E-105	Cobalamin biosynthesis protein CobD/CbiB	H
cpu_05230	1334	1.00E-151	Cobyrinic acid a,c-diamide synthase	H
cpu_05240	863	3.00E-39	Protein involved in propanediol utilization, and related proteins	Q
cpu_05250	1070	3.00E-68	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase	E
cpu_05280	758	4.00E-122	Uroporphyrinogen-III methylase	H
cpu_05290	767	4.00E-29	Cobalamin-5-phosphate synthase	H
cpu_05300	1016	1.00E-101	L-asparaginase II	E
cpu_05340	902	4.00E-57	ABC-type Fe3+-hydroxamate transport system, periplasmic component	P
cpu_05350	965	7.00E-72	ABC-type Fe3+-siderophore transport system, permease component	P
cpu_05360	1199	8.00E-104	ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components	PH
cpu_05370	509	2.00E-23	Predicted metal-binding protein	S
cpu_05380	428	4.00E-43	Predicted transcriptional regulator	K
cpu_05390	608	1.00E-36	DNA uptake protein and related DNA-binding proteins	L
cpu_05400	2180	1.00E-46	Predicted hydrolase (metallo-beta-lactamase superfamily)	R
cpu_05410	581	4.00E-30	Soluble lytic murein transglycosylase and related regulatory proteins	M
cpu_05420	989	9.00E-35	DNA polymerase III, delta subunit	L
cpu_05430	272	3.00E-20	Ribosomal protein S20	J
cpu_05460	1544	6.00E-98	Uncharacterized membrane protein, putative virulence factor	R
cpu_05470	1811	0	Membrane GTPase LepA	M
cpu_05480	1133	4.00E-121	Coproporphyrinogen III oxidase and related Fe-S oxidoreductases	H
cpu_05490	1025	6.00E-114	Transcriptional regulator of heat shock gene	K
cpu_05500	1565	2.00E-82	Chaperonin GroEL (HSP60 family)	O
cpu_05510	572	2.00E-37	Molecular chaperone GrpE (heat shock protein)	O
cpu_05520	1808	0	Molecular chaperone	O
cpu_05530	1148	6.00E-177	DnaJ-class molecular chaperone with C-terminal Zn finger domain	O
cpu_05540	917	3.00E-90	Ribosomal protein L11 methylase	J
cpu_05550	746	1.00E-79	Uncharacterized protein conserved in bacteria	S
cpu_05560	1304	0	2-methylthioadenine synthetase	J
cpu_05570	377	7.00E-40	Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases	FGR
cpu_05580	179	5.00E-08	Ribosomal protein S21	J
cpu_05590	446	1.00E-39	Uncharacterized conserved protein	S
cpu_05620	977	2.00E-145	Phosphate starvation-inducible protein PhoH, predicted ATPase	T
cpu_05630	2012	0	Predicted membrane-associated HD superfamily hydrolase	R
cpu_05640	428	4.00E-48	Predicted metal-dependent hydrolase	R
cpu_05650	482	4.00E-32	Diacylglycerol kinase	M
cpu_05660	632	2.00E-64	Uncharacterized conserved protein	S
cpu_05670	896	3.00E-141	GTPase	R
cpu_05680	1745	0	L-lactate permease	C
cpu_05690	1367	5.00E-137	FAD/FMN-containing dehydrogenases	C
cpu_05700	1244	1.00E-83	Fe-S oxidoreductase	C
cpu_05710	695	1.00E-63	Transcriptional regulators	K
cpu_05720	578	2.00E-27	Uncharacterized conserved protein	S
cpu_05730	2090	7.00E-165	Uncharacterized conserved protein containing a ferredoxin-like domain	C
cpu_05740	854	5.00E-51	Small-conductance mechanosensitive channel	M
cpu_05750	200	1.00E-17	Uncharacterized protein conserved in bacteria	S
cpu_05760	1088	0	Predicted GTPase, probable translation factor	J
cpu_05770	233	5.00E-17	Predicted redox protein, regulator of disulfide bond formation	O

Table 2-1-1 Continued

cpu_05780	1211	2.00E-12	Predicted transporter component	R
cpu_05790	785	2.00E-92	Uncharacterized conserved protein	S
cpu_05800	812	2.00E-39	ATP-dependent DNA ligase	L
cpu_05810	881	3.00E-91	Predicted eukaryotic-type DNA primase	L
cpu_05820	416	9.00E-62	Uncharacterized conserved protein	S
cpu_05830	1391	8.00E-138	Aspartyl aminopeptidase	E
cpu_05850	1001	1.00E-45	Uncharacterized membrane protein	S
cpu_05880	638	2.00E-60	Predicted phosphoribosyltransferases	R
cpu_05900	803	8.00E-96	ATP-utilizing enzymes of the PP-loop superfamily	R
cpu_05910	746	7.00E-98	NCAIR mutase (PurE)-related proteins	R
cpu_05920	1163	5.00E-144	Uncharacterized conserved protein	S
cpu_05930	893	4.00E-66	Transcriptional regulator	K
cpu_05940	1031	1.00E-76	Predicted membrane protein	S
cpu_05950	1151	6.00E-121	Selenocysteine lyase	E
cpu_05960	875	4.00E-54	Predicted transcriptional regulators	K
cpu_05970	764	1.00E-90	ATPases involved in chromosome partitioning	D
cpu_05980	707	9.00E-64	Predicted S-adenosylmethionine-dependent methyltransferase	M
cpu_05990	1895	0	NAD/FAD-utilizing enzyme apparently involved in cell division	D
cpu_06000	1385	0	Predicted GTPase	R
cpu_06010	620	2.00E-67	Predicted RNA-binding protein	R
cpu_06020	659	2.00E-60	Preprotein translocase subunit YidC	U
cpu_06030	227	8.00E-40	Uncharacterized conserved protein	S
cpu_06040	341	6.00E-23	RNase P protein component	J
cpu_06050	1364	0	ATPase involved in DNA replication initiation	L
cpu_06070	1115	2.00E-85	DNA polymerase sliding clamp subunit (PCNA homolog)	L
cpu_06080	260	1.00E-24	Uncharacterized conserved protein	S
cpu_06090	1061	8.00E-100	Recombinational DNA repair ATPase (RecF pathway)	L
cpu_06110	1907	0	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit	L
cpu_06120	2435	0	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit	L
cpu_06130	884	0	Pyridoxine biosynthesis enzyme	H
cpu_06140	575	3.00E-102	Predicted glutamine amidotransferase involved in pyridoxine biosynthesis	H
cpu_06150	1148	2.00E-172	Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase	E
cpu_06160	1577	1.00E-137	Phosphoglycerate dehydrogenase and related dehydrogenases	HE
cpu_06170	1286	0	Seryl-tRNA synthetase	J
cpu_06190	428	7.00E-20	Uncharacterized conserved protein	S
cpu_06200	593	2.00E-115	Adenylylsulfate kinase and related kinases	P
cpu_06210	746	2.00E-15	Predicted permeases	R
cpu_06220	1586	9.00E-39	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	C
cpu_06230	269	5.00E-08	Predicted ATPase, RNase L inhibitor (RLI) homolog	R
cpu_06240	1145	3.00E-157	ATP sulfurylase (sulfate adenylyltransferase)	P
cpu_06250	209	3.00E-10	Sulfur transfer protein involved in thiamine biosynthesis	H
cpu_06260	803	2.00E-97	Dinucleotide-utilizing enzymes	H
cpu_06270	1274	2.00E-73	Dissimilatory sulfite reductase (desulfoviridin), alpha and beta subunits	C
cpu_06280	230	6.00E-08	Predicted redox protein, regulator of disulfide bond formation	O
cpu_06290	1202	8.00E-142	O-acetylhomoserine sulfhydrylase	E
cpu_06300	461	1.00E-70	Cytosine/adenosine deaminases	FJ
cpu_06310	758	5.00E-30	Predicted xylanase/chitin deacetylase	G
cpu_06330	281	1.00E-22	Uncharacterized protein conserved in bacteria	S
cpu_06340	1550	3.00E-149	DNA polymerase III, gamma/tau subunits	L
cpu_06350	329	5.00E-33	Uncharacterized protein conserved in bacteria	S
cpu_06360	599	2.00E-106	Recombinational DNA repair protein (RecF pathway)	L
cpu_06380	1385	0	Uncharacterized FAD-dependent dehydrogenases	R
cpu_06390	1283	0	Adenylosuccinate synthase	F
cpu_06400	857	4.00E-70	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase	EH
cpu_06410	470	1.00E-26	Soluble P-type ATPase	R
cpu_06430	1373	1.00E-59	Arylsulfatase regulator (Fe-S oxidoreductase)	R
cpu_06450	650	1.00E-40	Membrane proteins related to metalloendopeptidases	M

Table 2-1-1 Continued

cpu_06480	254	4.00E-16	Regulators of stationary/sporulation gene expression	K
cpu_06490	1967	0	Methionyl-tRNA synthetase	J
cpu_06500	764	8.00E-116	Mg-dependent DNase	L
cpu_06510	524	2.00E-78	ATP:corrinoid adenosyltransferase	H
cpu_06520	1007	7.00E-34	Uncharacterized protein conserved in bacteria	S
cpu_06530	890	3.00E-101	Dimethyladenosine transferase (rRNA methylation)	J
cpu_06540	491	3.00E-38	Uncharacterized protein conserved in bacteria	S
cpu_06590	869	7.00E-21	Predicted glycosyltransferases	R
cpu_06600	866	3.00E-20	Predicted glycosyltransferases	R
cpu_06640	404	1.00E-09	Uncharacterized conserved protein	S
cpu_06650	290	7.00E-10	Uncharacterized protein conserved in bacteria	S
cpu_06670	551	6.00E-41	Uncharacterized protein with SCP/PR1 domains	S
cpu_06690	521	4.00E-58	Rubrerythrin	C
cpu_06700	332	2.00E-08	Rubrerythrin	C
cpu_06710	1124	2.00E-106	NAD(P)H-nitrite reductase	C
cpu_06720	158	4.00E-22	Rubredoxin	C
cpu_06730	968	8.00E-74	Mg2+ and Co2+ transporters	P
cpu_06750	1352	4.00E-71	Transcriptional regulator	K
cpu_06770	434	5.00E-58	3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydratases	I
cpu_06780	1046	2.00E-62	UDP-N-acetylmuramyl pentapeptide phosphotransferase	M
cpu_06810	731	1.00E-81	Teichoic acid biosynthesis proteins	M
cpu_06820	713	3.00E-37	Predicted sugar nucleotidyltransferases	M
cpu_06830	1049	1.00E-75	Glycerol dehydrogenase and related enzymes	C
cpu_06840	875	4.00E-179	UDP-glucose pyrophosphorylase	M
cpu_06850	737	4.00E-75	Predicted membrane protein	S
cpu_06860	1385	5.00E-122	Phosphomannomutase	G
cpu_06880	221	3.00E-10	Predicted redox protein, regulator of disulfide bond formation	O
cpu_06900	908	3.00E-58	Transcriptional regulator	K
cpu_06910	2171	4.00E-158	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_06920	539	8.00E-65	Fe-S-cluster-containing hydrogenase components 1	C
cpu_06930	857	1.00E-24	Formate-dependent nitrite reductase, membrane component	P
cpu_06940	578	3.00E-08	Uncharacterized component of anaerobic dehydrogenases	R
cpu_06960	833	7.00E-58	N-acetylmuramoyl-L-alanine amidase	M
cpu_06980	689	4.00E-97	Response regulators	TK
cpu_06990	1334	6.00E-65	Signal transduction histidine kinase	T
cpu_07000	197	3.00E-28	Ribosomal protein L31	J
cpu_07010	866	2.00E-112	Predicted metal-dependent enzyme	R
cpu_07020	1076	0	Protein chain release factor A	J
cpu_07030	857	5.00E-77	Methylase of polypeptide chain release factors	J
cpu_07040	1025	2.00E-93	Putative translation factor (SUA5)	J
cpu_07050	536	5.00E-28	Predicted membrane protein	S
cpu_07060	479	3.00E-40	Protein-tyrosine-phosphatase	T
cpu_07070	449	1.00E-75	Ribose 5-phosphate isomerase RpiB	G
cpu_07080	1265	0	Glycine/serine hydroxymethyltransferase	E
cpu_07090	446	3.00E-66	Deoxycytidylate deaminase	F
cpu_07110	1148	0	UDP-N-acetylglucosamine 2-epimerase	M
cpu_07140	686	5.00E-47	F0F1-type ATP synthase, subunit a	C
cpu_07150	236	3.00E-10	F0F1-type ATP synthase, subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K	C
cpu_07160	479	4.00E-27	F0F1-type ATP synthase, subunit b	C
cpu_07170	533	2.00E-41	F0F1-type ATP synthase, delta subunit	C
cpu_07180	1505	0	F0F1-type ATP synthase, alpha subunit	C
cpu_07190	848	8.00E-129	F0F1-type ATP synthase, gamma subunit	C
cpu_07200	1421	0	F0F1-type ATP synthase, beta subunit	C
cpu_07210	404	3.00E-42	F0F1-type ATP synthase, epsilon subunit (mitochondrial delta subunit)	C
cpu_07230	1259	0	UDP-N-acetylglucosamine enolpyruvyl transferase	M
cpu_07240	950	1.00E-68	Sporulation protein and related proteins	D
cpu_07260	416	4.00E-29	Predicted nucleic acid-binding protein, contains PIN domain	R

Table 2-1-1 Continued

cpu_07270	299	3.00E-27	Predicted nucleotidyltransferases	R
cpu_07280	257	1.00E-31	Uncharacterized conserved protein	S
cpu_07300	1016	2.00E-171	Actin-like ATPase involved in cell morphogenesis	D
cpu_07310	416	5.00E-18	Predicted nucleic acid-binding protein, contains PIN domain	R
cpu_07370	557	3.00E-26	Pyruvate:ferredoxin oxidoreductase gamma subunit	C
cpu_07380	740	3.00E-64	Pyruvate:ferredoxin oxidoreductase beta subunit	C
cpu_07390	1061	2.00E-81	Pyruvate:ferredoxin oxidoreductase alpha subunit	C
cpu_07400	206	2.00E-11	Dissimilatory sulfite reductase (desulfoviridin), alpha and beta subunits	C
cpu_07410	1034	3.00E-95	Glutamate dehydrogenase/leucine dehydrogenase	E
cpu_07420	2018	0	Transcriptional regulator	KT
cpu_07430	1247	0	Methylaspartate ammonia-lyase	E
cpu_07460	551	7.00E-60	Thymidine kinase	F
cpu_07470	392	4.00E-61	Predicted CoA-binding protein	R
cpu_07480	566	9.00E-28	Uncharacterized conserved protein	S
cpu_07490	1418	7.00E-20	Arabinose efflux permease	G
cpu_07500	926	1.00E-37	Leucyl aminopeptidase (aminopeptidase T)	E
cpu_07520	1004	5.00E-72	Predicted acyl-CoA transferases/carnitine dehydratase	C
cpu_07540	1592	3.00E-149	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_07550	545	1.00E-60	Predicted phosphatase homologous to the C-terminal domain	R
cpu_07560	674	9.00E-25	Multimeric flavodoxin WrbA	R
cpu_07570	1439	2.00E-46	FOG: GGDEF domain	T
cpu_07580	755	4.00E-24	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	T
cpu_07590	353	3.00E-15	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	T
cpu_07600	398	2.00E-24	Anti-sigma regulatory factor (Ser/Thr protein kinase)	T
cpu_07620	995	1.00E-91	Phosphate/sulphate permeases	P
cpu_07630	626	3.00E-31	Phosphate transport regulator (distant homolog of PhoU)	P
cpu_07640	806	2.00E-22	Zn-dependent hydrolases, including glyoxylases	R
cpu_07650	1595	2.00E-54	Signal transduction histidine kinase	T
cpu_07660	692	6.00E-87	Response regulators	TK
cpu_07670	383	2.00E-10	Lactoylglutathione lyase and related lyases	E
cpu_07680	1244	2.00E-84	Formate-dependent nitrite reductase, periplasmic cytochrome c552 subunit	P
cpu_07740	722	1.00E-74	ABC-type multidrug transport system, ATPase component	V
cpu_07750	1154	9.00E-175	Glycerate kinase	G
cpu_07760	1385	9.00E-98	H+/gluconate symporter and related permeases	GE
cpu_07770	1181	4.00E-76	Sugar diacid utilization regulator	KT
cpu_07780	218	2.00E-08	Transposase and inactivated derivatives	L
cpu_07790	1763	0	Aldehyde:ferredoxin oxidoreductase	C
cpu_07800	920	7.00E-63	Transcriptional regulator	K
cpu_07810	1400	2.00E-89	Di- and tricarboxylate transporters	P
cpu_07820	845	2.00E-127	Tartrate dehydratase alpha subunit/Fumarate hydratase class I, N-terminal domain	C
cpu_07830	554	2.00E-82	Tartrate dehydratase beta subunit/Fumarate hydratase class I, C-terminal domain	C
cpu_07840	1709	0	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	C
cpu_07850	542	2.00E-62	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	C
cpu_07860	1607	3.00E-78	Uncharacterized protein conserved in bacteria	S
cpu_07870	1139	6.00E-177	Homoserine acetyltransferase	E
cpu_07880	1247	0	O-acetylhomoserine sulfhydrylase	E
cpu_07900	1034	1.00E-17	HD-GYP domain	T
cpu_07910	2267	1.00E-99	Predicted signal transduction protein	T
cpu_07920	326	7.00E-09	Predicted transcriptional regulators	K
cpu_07930	1220	2.00E-165	Aspartokinases	E
cpu_07940	914	3.00E-111	Homoserine kinase	E
cpu_07950	1049	9.00E-142	Threonine synthase	E
cpu_07960	1292	2.00E-140	Homoserine dehydrogenase	E
cpu_07970	440	2.00E-54	ACT domain-containing protein	R
cpu_07990	1184	2.00E-38	ABC-type Na+ efflux pump, permease component	CP
cpu_08000	731	6.00E-108	ABC-type Na+ transport system, ATPase component	CP
cpu_08010	1259	2.00E-42	Methyl-accepting chemotaxis protein	NT

Table 2-1-1 Continued

cpu_08020	452	2.00E-15	Transposase and inactivated derivatives	L
cpu_08030	638	9.00E-39	Transposase and inactivated derivatives	L
cpu_08050	1010	3.00E-160	Glycerol-3-phosphate dehydrogenase	C
cpu_08060	590	3.00E-53	Predicted membrane protein	S
cpu_08070	1322	0	Predicted GTPases	R
cpu_08080	1295	2.00E-75	Fe-S oxidoreductase, related to NifB/MoaA family	C
cpu_08120	743	5.00E-12	Predicted membrane protein	S
cpu_08130	1964	2.00E-101	Penicillin tolerance protein	IM
cpu_08140	551	4.00E-40	1-acyl-sn-glycerol-3-phosphate acyltransferase	I
cpu_08150	668	3.00E-106	Cytidylate kinase	F
cpu_08160	1283	9.00E-164	5-enolpyruvylshikimate-3-phosphate synthase	E
cpu_08170	1088	2.00E-130	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase	E
cpu_08180	803	1.00E-140	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	E
cpu_08190	584	4.00E-106	Anthranilate/para-aminobenzoate synthases component II	EH
cpu_08200	353	7.00E-33	Chorismate mutase	E
cpu_08220	1184	6.00E-129	Predicted flavoproteins	R
cpu_08250	728	8.00E-97	16S rRNA uridine-516 pseudouridylate synthase	J
cpu_08260	530	2.00E-53	Uncharacterized membrane protein	S
cpu_08270	578	3.00E-77	Uncharacterized membrane protein, required for spore maturation in B.subtilis.	R
cpu_08280	425	2.00E-41	Uncharacterized conserved protein	S
cpu_08300	566	2.00E-48	Predicted transcriptional regulator containing the HTH domain	K
cpu_08310	740	5.00E-46	Uncharacterized conserved protein	S
cpu_08320	980	4.00E-134	Tryptophanyl-tRNA synthetase	J
cpu_08330	632	2.00E-18	Zn-dependent proteases	R
cpu_08340	2597	1.00E-63	tRNA nucleotidyltransferase/poly(A) polymerase	J
cpu_08360	1337	3.00E-143	Diaminopimelate decarboxylase	E
cpu_08370	986	4.00E-106	Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)	I
cpu_08380	1067	4.00E-39	Uncharacterized protein conserved in bacteria	S
cpu_08390	995	2.00E-52	Uncharacterized protein conserved in bacteria	S
cpu_08430	194	1.00E-07	Uncharacterized conserved protein	S
cpu_08440	752	3.00E-68	DNA-directed RNA polymerase specialized sigma subunit	K
cpu_08450	425	2.00E-25	Anti-sigma regulatory factor (Ser/Thr protein kinase)	T
cpu_08460	347	4.00E-18	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	T
cpu_08470	1133	3.00E-116	D-alanyl-D-alanine carboxypeptidase	M
cpu_08480	1166	0	Phosphopentomutase	G
cpu_08490	863	1.00E-129	Site-specific recombinase XerD	L
cpu_08500	614	3.00E-09	Uncharacterized membrane protein	S
cpu_08550	1541	5.00E-78	HD-GYP domain	T
cpu_08580	677	3.00E-23	Glycosyltransferases involved in cell wall biogenesis	M
cpu_08600	1121	4.00E-108	Uncharacterized membrane-anchored protein conserved in bacteria	S
cpu_08610	791	1.00E-25	Response regulator	TK
cpu_08630	1658	5.00E-178	ATPase involved in DNA repair	L
cpu_08640	452	5.00E-58	Arginine repressor	K
cpu_08650	575	5.00E-09	tRNA(1-methyladenosine) methyltransferase and related methyltransferases	J
cpu_08660	863	2.00E-78	Predicted sugar kinase	G
cpu_08670	800	5.00E-124	Predicted rRNA methylase	J
cpu_08680	1868	0	Deoxyxylulose-5-phosphate synthase	HI
cpu_08700	887	8.00E-92	Geranylgeranyl pyrophosphate synthase	H
cpu_08710	224	4.00E-10	Exonuclease VII small subunit	L
cpu_08720	1418	4.00E-158	Exonuclease VII, large subunit	L
cpu_08730	1394	2.00E-180	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	ER
cpu_08740	836	2.00E-62	2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases	HC
cpu_08750	941	1.00E-36	Metal-dependent proteases with possible chaperone activity	O
cpu_08760	434	7.00E-40	Transcription termination factor	K
cpu_08790	401	2.00E-35	Uncharacterized protein conserved in bacteria	S
cpu_08800	1349	0	Acetyl/propionyl-CoA carboxylase, alpha subunit	I
cpu_08810	1844	0	Pyruvate/oxaloacetate carboxyltransferase	C

Table 2-1-1 Continued

cpu_08890	947	3.00E-91	Uncharacterized protein conserved in bacteria	S
cpu_08920	557	2.00E-56	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)	J
cpu_08930	1064	5.00E-122	Xaa-Pro aminopeptidase	E
cpu_08940	455	2.00E-72	3-dehydroquinate dehydratase II	E
cpu_08950	1874	1.00E-68	HD-GYP domain	T
cpu_08960	1322	2.00E-114	Predicted Zn-dependent proteases and their inactivated homologs	R
cpu_08980	2216	1.00E-68	Predicted hydrolase of the HD superfamily (permuted catalytic motifs)	R
cpu_08990	476	3.00E-10	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_09000	812	1.00E-60	Uncharacterized protein predicted to be involved in DNA repair (RAMP superfamily)	L
cpu_09010	956	1.00E-43	Uncharacterized protein predicted to be involved in DNA repair (RAMP superfamily)	L
cpu_09020	1100	2.00E-16	Uncharacterized protein predicted to be involved in DNA repair (RAMP superfamily)	L
cpu_09030	488	2.00E-31	Uncharacterized protein predicted to be involved in DNA repair (RAMP superfamily)	L
cpu_09070	980	3.00E-12	Predicted transcriptional regulator	K
cpu_09080	1121	2.00E-11	Uncharacterized protein conserved in archaea	S
cpu_09100	1616	9.00E-67	FOG: EAL domain	T
cpu_09110	1232	3.00E-49	HD-GYP domain	T
cpu_09120	1421	2.00E-41	Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain	O
cpu_09130	1325	1.00E-41	tRNA nucleotidyltransferase/poly(A) polymerase	J
cpu_09140	1931	3.00E-67	Actin-like ATPase involved in cell division	D
cpu_09150	536	7.00E-09	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_09180	407	6.00E-56	Methylmalonyl-CoA mutase, C-terminal domain/subunit (cobalamin-binding)	I
cpu_09190	1652	0	Methylmalonyl-CoA mutase, N-terminal domain/subunit	I
cpu_09200	1634	0	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	I
cpu_09210	581	1.00E-41	ATP:corrinoid adenosyltransferase	H
cpu_09220	938	4.00E-125	Putative periplasmic protein kinase ArgK and related GTPases of G3E family	E
cpu_09240	425	2.00E-29	Rubrerythrin	C
cpu_09250	239	2.00E-20	Glutaredoxin and related proteins	O
cpu_09260	377	1.00E-23	Rhodanese-related sulfurtransferase	P
cpu_09270	719	4.00E-74	Prolipoprotein diacylglyceryltransferase	M
cpu_09280	464	2.00E-17	Uncharacterized protein involved in biosynthesis of c-type cytochromes	O
cpu_09300	1280	2.00E-114	Thioredoxin reductase	O
cpu_09330	1358	4.00E-32	Arylsulfatase regulator (Fe-S oxidoreductase)	R
cpu_09340	614	7.00E-28	Methylase involved in ubiquinone/menaquinone biosynthesis	H
cpu_09350	323	1.00E-34	Thioredoxin domain-containing protein	O
cpu_09360	734	1.00E-17	Methylase involved in ubiquinone/menaquinone biosynthesis	H
cpu_09380	770	2.00E-10	Predicted transcriptional regulators	K
cpu_09400	1235	6.00E-131	Collagenase and related proteases	O
cpu_09410	1013	2.00E-119	Predicted periplasmic solute-binding protein	R
cpu_09420	1535	5.00E-93	Uncharacterized membrane protein, putative virulence factor	R
cpu_09430	737	5.00E-138	UDP-glucose 4-epimerase	M
cpu_09440	224	4.00E-31	UDP-glucose 4-epimerase	M
cpu_09450	1394	2.00E-63	Uncharacterized vancomycin resistance protein	V
cpu_09470	416	4.00E-52	Predicted endonuclease involved in recombination	L
cpu_09480	953	2.00E-32	Predicted oxidoreductases of the aldo/keto reductase family	R
cpu_09490	257	7.00E-34	Uncharacterized protein conserved in bacteria	S
cpu_09510	2624	0	Alanyl-tRNA synthetase	J
cpu_09520	1034	8.00E-53	Predicted permease	R
cpu_09550	2774	0	Isoleucyl-tRNA synthetase	J
cpu_09560	728	6.00E-21	Predicted phosphoesterase	R
cpu_09590	1736	0	Phosphoenolpyruvate-protein kinase (PTS system EI component in bacteria)	G
cpu_09600	269	9.00E-28	Phosphotransferase system, HPr-related proteins	G
cpu_09610	452	3.00E-48	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)	GT
cpu_09620	1364	4.00E-139	Phosphotransferase system, fructose-specific IIC component	G
cpu_09630	953	2.00E-103	Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)	G
cpu_09640	767	5.00E-97	Transcriptional regulators of sugar metabolism	KG
cpu_09650	1571	2.00E-122	Isopropylmalate/homocitrate/citramalate synthases	E
cpu_09660	1124	6.00E-177	Isocitrate/isopropylmalate dehydrogenase	CE

Table 2-1-1 Continued

cpu_09670	488	7.00E-72	3-isopropylmalate dehydratase small subunit	E
cpu_09680	1265	0	3-isopropylmalate dehydratase large subunit	E
cpu_09690	1529	2.00E-180	Isopropylmalate/homocitrate/citramalate synthases	E
cpu_09700	1664	0	Thiamine pyrophosphate-requiring enzymes	EH
cpu_09710	992	0	Ketol-acid reductoisomerase	EH
cpu_09720	512	7.00E-73	Acetolactate synthase, small (regulatory) subunit	E
cpu_09730	1658	0	Thiamine pyrophosphate-requiring enzymes	EH
cpu_09740	1661	0	Dihydroxyacid dehydratase/phosphogluconate dehydratase	EG
cpu_09750	881	3.00E-91	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase	EH
cpu_09770	1895	0	Putative Ser protein kinase	T
cpu_09780	1136	3.00E-132	Uncharacterized conserved protein	S
cpu_09790	1334	7.00E-164	Uncharacterized conserved protein	S
cpu_09810	1067	2.00E-94	Geranylgeranyl pyrophosphate synthase	H
cpu_09820	1349	7.00E-170	Glutamyl-tRNA reductase	H
cpu_09830	935	7.00E-148	Porphobilinogen deaminase	H
cpu_09840	1034	2.00E-43	Uroporphyrinogen-III synthase	H
cpu_09850	983	0	Delta-aminolevulinic acid dehydratase	H
cpu_09860	995	2.00E-87	Predicted Fe-S oxidoreductases	R
cpu_09870	1298	0	Glutamate-1-semialdehyde aminotransferase	H
cpu_09880	305	2.00E-10	Uncharacterized conserved protein	S
cpu_09890	1625	6.00E-126	Predicted unusual protein kinase	R
cpu_09910	764	5.00E-57	CO dehydrogenase maturation factor	D
cpu_09920	2012	8.00E-107	6Fe-6S prismane cluster-containing protein	C
cpu_09930	2198	4.00E-174	CO dehydrogenase/acetyl-CoA synthase beta subunit	C
cpu_09940	1337	4.00E-129	CO dehydrogenase/acetyl-CoA synthase gamma subunit (corrinoid Fe-S protein)	C
cpu_09950	1895	4.00E-136	Uncharacterized metal-binding protein	R
cpu_09960	758	6.00E-64	CO dehydrogenase maturation factor	D
cpu_09970	932	4.00E-77	CO dehydrogenase/acetyl-CoA synthase delta subunit (corrinoid Fe-S protein)	C
cpu_09980	791	2.00E-43	Methionine synthase I, cobalamin-binding domain	E
cpu_09990	533	4.00E-19	Heterodisulfide reductase, subunit C	C
cpu_10000	863	1.00E-61	Heterodisulfide reductase, subunit B	C
cpu_10010	2681	2.00E-48	Heterodisulfide reductase, subunit A and related polyferredoxins	C
cpu_10020	359	7.00E-19	Coenzyme F420-reducing hydrogenase, delta subunit	C
cpu_10040	917	1.00E-66	5,10-methylenetetrahydrofolate reductase	E
cpu_10050	851	1.00E-24	Histidinol phosphatase and related hydrolases of the PHP family	ER
cpu_10060	734	6.00E-29	Putative primosome component and related proteins	L
cpu_10070	713	2.00E-30	DNA replication protein	L
cpu_10080	1592	3.00E-97	Predicted unusual protein kinase	R
cpu_10090	1106	2.00E-145	Uncharacterized Fe-S center protein	R
cpu_10110	833	2.00E-109	Spermidine synthase	E
cpu_10170	266	3.00E-29	Uncharacterized protein conserved in bacteria	S
cpu_10180	251	2.00E-19	Copper chaperone	P
cpu_10190	2504	0	Cation transport ATPase	P
cpu_10220	1853	7.00E-30	Predicted ATPase	R
cpu_10230	995	1.00E-18	DNA repair exonuclease	L
cpu_10240	1427	6.00E-12	ATPase involved in DNA repair	L
cpu_10270	482	5.00E-59	FOG: GGDEF domain	T
cpu_10280	362	3.00E-43	6-pyruvoyl-tetrahydropterin synthase	H
cpu_10290	647	8.00E-70	Predicted PP-loop superfamily ATPase	R
cpu_10300	557	2.00E-51	Uncharacterized conserved protein	S
cpu_10310	734	1.00E-38	Organic radical activating enzymes	O
cpu_10320	578	7.00E-114	GTP cyclohydrolase I	H
cpu_10340	803	5.00E-105	Predicted acid phosphatase	R
cpu_10360	2891	0	DNA polymerase III, alpha subunit	L
cpu_10370	2018	0	Translation elongation factors (GTPases)	J
cpu_10380	803	3.00E-90	Flagellar motor component	N
cpu_10390	800	1.00E-75	Flagellar motor protein	N

Table 2-1-1 Continued

cpu_10400	1571	1.00E-58	Methyl-accepting chemotaxis protein	NT
cpu_10410	470	6.00E-49	Chemotaxis signal transduction protein	NT
cpu_10420	2099	0	Chemotaxis protein histidine kinase and related kinases	NT
cpu_10430	1022	9.00E-121	Chemotaxis response regulator	NT
cpu_10460	1424	1.00E-77	Flagellar hook-associated protein	N
cpu_10470	863	2.00E-33	Flagellin and related hook-associated proteins	N
cpu_10490	455	7.00E-30	Uncharacterized protein conserved in bacteria	S
cpu_10500	221	2.00E-16	Carbon storage regulator (could also regulate swarming and quorum sensing)	T
cpu_10510	1772	1.00E-77	Uncharacterized protein conserved in bacteria	S
cpu_10530	1241	5.00E-57	Uncharacterized protein conserved in bacteria	S
cpu_10540	995	4.00E-139	Predicted nucleoside-diphosphate sugar epimerases	MG
cpu_10550	1142	1.00E-155	Predicted pyridoxal phosphate-dependent enzyme	M
cpu_10560	854	1.00E-73	Spore coat polysaccharide biosynthesis protein F, CMP-KDO synthetase homolog	M
cpu_10570	1022	7.00E-40	Spore coat polysaccharide biosynthesis protein, predicted glycosyltransferase	M
cpu_10580	539	1.00E-14	Acetyltransferases, including N-acetylases of ribosomal proteins	J
cpu_10590	1052	2.00E-165	Sialic acid synthase	M
cpu_10600	818	3.00E-48	Flagellin and related hook-associated proteins	N
cpu_10610	356	4.00E-17	Uncharacterized flagellar protein FlaG	N
cpu_10620	1457	2.00E-63	Flagellar capping protein	N
cpu_10630	377	1.00E-34	Flagellin-specific chaperone FliS	NUO
cpu_10660	176	6.00E-09	Flagellar basal body protein	N
cpu_10670	416	4.00E-56	Flagellar basal body rod protein	N
cpu_10680	287	5.00E-24	Flagellar hook-basal body protein	NU
cpu_10690	1586	9.00E-125	Flagellar biosynthesis/type III secretory pathway lipoprotein	NU
cpu_10700	575	1.00E-53	Flagellar motor switch protein	N
cpu_10710	449	3.00E-57	Flagellar motor switch protein	N
cpu_10730	1301	0	Flagellar biosynthesis/type III secretory pathway ATPase	NU
cpu_10760	401	2.00E-26	Flagellar hook capping protein	N
cpu_10780	785	6.00E-81	Flagellar basal body rod protein	N
cpu_10790	470	2.00E-18	Flagellar basal body-associated protein	N
cpu_10800	374	1.00E-19	Flagellar motor switch/type III secretory pathway protein	NU
cpu_10820	767	4.00E-108	Flagellar biosynthesis pathway, component FliP	NU
cpu_10830	269	1.00E-25	Flagellar biosynthesis pathway, component FliQ	NU
cpu_10840	767	4.00E-30	Flagellar biosynthesis pathway, component FliR	NU
cpu_10850	1070	2.00E-130	Flagellar biosynthesis pathway, component FlhB	NU
cpu_10860	2051	0	Flagellar biosynthesis pathway, component FlhA	NU
cpu_10870	1052	3.00E-75	Flagellar GTP-binding protein	N
cpu_10880	863	4.00E-64	ATPases involved in chromosome partitioning	D
cpu_10890	647	7.00E-12	Predicted glycosyltransferase	M
cpu_10900	752	3.00E-53	DNA-directed RNA polymerase specialized sigma subunit	K
cpu_10910	734	9.00E-62	Flagellar basal body rod protein	N
cpu_10920	491	3.00E-53	Chemotaxis protein; stimulates methylation of MCP proteins	NT
cpu_10930	764	3.00E-83	Methylase of chemotaxis methyl-accepting proteins	NT
cpu_10940	602	3.00E-40	Chemotaxis protein CheC, inhibitor of MCP methylation	NT
cpu_10950	365	9.00E-32	FOG: CheY-like receiver	T
cpu_10960	1007	1.00E-93	Flagellar motor switch protein	N
cpu_10970	1046	6.00E-40	Chemotaxis protein CheC, inhibitor of MCP methylation	NT
cpu_10980	1526	1.00E-35	Predicted polymerase	L
cpu_10990	668	1.00E-22	Uncharacterized protein conserved in bacteria	S
cpu_11000	1859	1.00E-67	Methyl-accepting chemotaxis protein	NT
cpu_11010	431	1.00E-55	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	E
cpu_11020	536	1.00E-28	Prephenate dehydratase	E
cpu_11030	1085	0	Archaeal fructose 1,6-bisphosphatase	G
cpu_11050	1130	4.00E-56	Glycosyltransferase	M
cpu_11060	716	7.00E-68	Phosphosulfolactate phosphohydrolase and related enzymes	HR
cpu_11070	1103	1.00E-42	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase	M
cpu_11090	818	4.00E-25	Methyl-accepting chemotaxis protein	NT

Table 2-1-1 Continued

cpu_11100	1361	5.00E-180	Anthranilate/para-aminobenzoate synthases component I	EH
cpu_11120	773	5.00E-70	Nucleoside-diphosphate-sugar pyrophosphorylase	MJ
cpu_11130	1088	5.00E-47	Nucleoside-diphosphate-sugar epimerases	MG
cpu_11140	1334	1.00E-117	Predicted pyridoxal phosphate-dependent enzyme apparently	M
cpu_11150	1796	4.00E-156	Thiamine pyrophosphate-requiring enzymes	EH
cpu_11170	878	2.00E-34	Nucleoside-diphosphate-sugar epimerases	MG
cpu_11180	782	7.00E-42	Predicted sugar phosphatases of the HAD superfamily	G
cpu_11190	980	1.00E-12	Glycosyltransferases involved in cell wall biogenesis	M
cpu_11200	491	4.00E-57	Molybdopterin biosynthesis enzymes	H
cpu_11210	437	5.00E-30	Uncharacterized protein conserved in bacteria	S
cpu_11220	473	4.00E-85	Molybdenum cofactor biosynthesis enzyme	H
cpu_11230	965	4.00E-122	Molybdenum cofactor biosynthesis enzyme	H
cpu_11240	779	4.00E-89	Uncharacterized protein required for formate dehydrogenase activity	C
cpu_11250	1904	3.00E-110	Molybdopterin biosynthesis enzyme	H
cpu_11260	1229	7.00E-151	Molybdopterin biosynthesis enzyme	H
cpu_11280	734	2.00E-16	Uncharacterized protein involved in formate dehydrogenase formation	O
cpu_11290	698	6.00E-31	Cytochrome b subunit of formate dehydrogenase	C
cpu_11300	782	2.00E-51	Fe-S-cluster-containing hydrogenase components 1	C
cpu_11310	2372	2.00E-72	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_11320	587	7.00E-31	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_11330	821	1.00E-42	Dinucleotide-utilizing enzymes	H
cpu_11350	1097	2.00E-30	Predicted Fe-S oxidoreductases	R
cpu_11360	725	1.00E-84	Aldehyde:ferredoxin oxidoreductase	C
cpu_11370	191	5.00E-20	Aldehyde:ferredoxin oxidoreductase	C
cpu_11380	983	0	Fructose-1,6-bisphosphatase/sedoheptulose 1,7-bisphosphatase and related proteins	G
cpu_11390	668	2.00E-82	Mn-containing catalase	P
cpu_11400	863	5.00E-68	ABC-type phosphate transport system, periplasmic component	P
cpu_11410	875	2.00E-87	ABC-type phosphate transport system, permease component	P
cpu_11420	857	2.00E-86	ABC-type phosphate transport system, permease component	P
cpu_11450	1250	3.00E-151	Tyrosyl-tRNA synthetase	J
cpu_11460	2420	0	Membrane carboxypeptidase (penicillin-binding protein)	M
cpu_11470	2354	5.00E-164	Mismatch repair ATPase (MutS family)	L
cpu_11480	2489	1.00E-119	Collagenase and related proteases	O
cpu_11500	575	6.00E-81	3-methyladenine DNA glycosylase	L
cpu_11510	1649	9.00E-63	Histidinol phosphatase and related hydrolases of the PHP family	ER
cpu_11520	254	3.00E-08	Uncharacterized protein conserved in bacteria	S
cpu_11530	2396	0	Phenylalanyl-tRNA synthetase beta subunit	J
cpu_11540	1016	2.00E-167	Phenylalanyl-tRNA synthetase alpha subunit	J
cpu_11550	785	8.00E-55	rRNA methylases	J
cpu_11560	662	7.00E-69	K+ transport systems, NAD-binding component	P
cpu_11570	1364	8.00E-87	Trk-type K+ transport systems, membrane components	P
cpu_11580	356	9.00E-63	Ribosomal protein L20	J
cpu_11590	194	2.00E-15	Ribosomal protein L35	J
cpu_11600	464	8.00E-81	Translation initiation factor 3 (IF-3)	J
cpu_11610	1904	0	Threonyl-tRNA synthetase	J
cpu_11620	599	3.00E-20	Uncharacterized protein conserved in bacteria	S
cpu_11630	794	8.00E-106	Tryptophan synthase alpha chain	E
cpu_11640	1208	0	Tryptophan synthase beta chain	E
cpu_11650	656	2.00E-82	Phosphoribosylanthranilate isomerase	E
cpu_11660	794	1.00E-104	Indole-3-glycerol phosphate synthase	E
cpu_11670	1019	7.00E-164	Anthranilate phosphoribosyltransferase	E
cpu_11680	563	1.00E-118	Anthranilate/para-aminobenzoate synthases component II	EH
cpu_11690	1478	0	Anthranilate/para-aminobenzoate synthases component I	EH
cpu_11700	767	1.00E-131	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	E
cpu_11730	242	7.00E-11	Predicted transcriptional regulators	K
cpu_11740	419	7.00E-08	Predicted transcriptional regulators	K
cpu_11770	590	1.00E-33	Predicted endonuclease	V

Table 2-1-1 Continued

cpu_11780	626	2.00E-21	Site-specific recombinase XerD	L
cpu_11790	746	3.00E-52	Predicted xylanase/chitin deacetylase	G
cpu_11800	1349	1.00E-85	Na+-driven multidrug efflux pump	V
cpu_11820	494	3.00E-61	Acetolactate synthase, small (regulatory) subunit	E
cpu_11830	590	3.00E-31	Uncharacterized conserved protein	S
cpu_11850	1001	2.00E-25	Zn-dependent hydrolases, including glyoxylases	R
cpu_11860	2045	7.00E-92	Fe-S oxidoreductase	C
cpu_11870	216	3.00E-41	Electron transfer flavoprotein, alpha subunit	C
cpu_11890	191	6.00E-17	Sec-independent protein secretion pathway components	U
cpu_11900	1046	2.00E-08	Ferredoxin	C
cpu_11910	491	2.00E-43	Molybdopterin-guanine dinucleotide biosynthesis protein	H
cpu_11920	1259	2.00E-130	Molybdopterin biosynthesis enzyme	H
cpu_11930	593	9.00E-46	Molybdopterin-guanine dinucleotide biosynthesis protein A	H
cpu_11940	200	9.00E-19	Sec-independent protein secretion pathway components	U
cpu_11950	821	9.00E-27	DMSO reductase anchor subunit	R
cpu_11960	2183	2.00E-173	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_11970	683	2.00E-24	Uncharacterized component of anaerobic dehydrogenases	R
cpu_11990	1232	6.00E-41	Sugar diacid utilization regulator	KT
cpu_12000	815	2.00E-39	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_12010	974	2.00E-57	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_12020	185	1.00E-18	Sec-independent protein secretion pathway components	U
cpu_12030	1067	3.00E-23	Imidazolonepropionase and related amidohydrolases	Q
cpu_12040	728	5.00E-59	Sec-independent protein secretion pathway component TatC	U
cpu_12050	431	8.00E-32	Predicted redox protein, regulator of disulfide bond formation	O
cpu_12060	575	9.00E-23	Membrane proteins related to metalloendopeptidases	M
cpu_12070	896	2.00E-91	Co/Zn/Cd efflux system component	P
cpu_12080	359	3.00E-17	Predicted transcriptional regulators	K
cpu_12100	758	4.00E-57	Membrane protease subunits, stomatin/prohibitin homologs	O
cpu_12110	359	2.00E-22	Membrane-bound serine protease (ClpP class)	O
cpu_12120	374	2.00E-14	Uncharacterized homolog of gamma-carboxymuconolactone decarboxylase subunit	S
cpu_12140	467	2.00E-21	Ribonuclease HI	L
cpu_12180	1190	3.00E-155	Uncharacterized flavoproteins	C
cpu_12190	209	1.00E-17	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_12210	1235	8.00E-106	Acyl-CoA dehydrogenases	I
cpu_12220	764	9.00E-87	Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)	I
cpu_12230	1217	2.00E-62	Benzoyl-CoA reductase	E
cpu_12240	1004	3.00E-24	Benzoyl-CoA reductase	E
cpu_12250	1640	0	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_12260	1115	2.00E-37	Methyl-accepting chemotaxis protein	NT
cpu_12280	671	3.00E-43	Predicted metal-dependent hydrolase	R
cpu_12290	3203	7.00E-134	Type I site-specific restriction-modification system	V
cpu_12300	1205	1.00E-32	Restriction endonuclease S subunits	V
cpu_12310	2444	2.00E-112	Type I restriction-modification system methyltransferase subunit	V
cpu_12340	2075	5.00E-16	Uncharacterized conserved protein	S
cpu_12360	416	5.00E-29	Molecular chaperone (small heat shock protein)	O
cpu_12370	710	8.00E-33	Uncharacterized membrane protein (homolog of Drosophila rhomboid)	R
cpu_12390	758	6.00E-48	Acyl-ACP thioesterase	I
cpu_12400	1226	6.00E-14	Sugar phosphate permease	G
cpu_12410	1601	3.00E-15	Arabinose efflux permease	G
cpu_12420	416	1.00E-15	Transcriptional regulators	K
cpu_12440	395	3.00E-25	Methylated DNA-protein cysteine methyltransferase	L
cpu_12450	920	1.00E-170	Cysteine synthase	E
cpu_12460	1622	0	Chaperonin GroEL (HSP60 family)	O
cpu_12470	284	3.00E-50	Co-chaperonin GroES (HSP10)	O
cpu_12490	269	2.00E-14	6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase	G
cpu_12500	1448	7.00E-17	Transglutaminase-like enzymes, putative cysteine proteases	E
cpu_12510	1766	1.00E-62	N-acetylmuramoyl-L-alanine amidase	M

Table 2-1-1 Continued

cpu_12530	407	2.00E-16	Uncharacterized conserved protein	S
cpu_12540	740	8.00E-45	Capsular polysaccharide biosynthesis protein	GM
cpu_12550	1235	2.00E-154	Lysine 2,3-aminomutase	E
cpu_12560	638	5.00E-15	Ferredoxin-thioredoxin reductase, catalytic subunit	C
cpu_12570	815	4.00E-87	Electron transfer flavoprotein, beta subunit	C
cpu_12580	1010	4.00E-136	Electron transfer flavoprotein, alpha subunit	C
cpu_12590	1295	5.00E-86	Dehydrogenases (flavoproteins)	C
cpu_12600	281	7.00E-31	Ferredoxin-like protein	C
cpu_12650	434	2.00E-17	Uncharacterized protein conserved in bacteria	S
cpu_12670	818	1.00E-28	Predicted permeases	R
cpu_12680	1070	4.00E-38	ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components	P
cpu_12690	770	6.00E-66	ABC-type nitrate/sulfonate/bicarbonate transport system, permease component	P
cpu_12700	740	2.00E-112	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	P
cpu_12710	566	8.00E-72	Uncharacterized conserved protein	S
cpu_12720	782	1.00E-54	Beta-propeller domains of methanol dehydrogenase type	R
cpu_12730	1796	4.00E-37	FOG: GGDEF domain	T
cpu_12740	833	2.00E-69	Periplasmic serine proteases (ClpP class)	OU
cpu_12770	920	3.00E-24	Zn-dependent hydrolases, including glyoxylases	R
cpu_12790	977	6.00E-133	Biotin synthase and related enzymes	H
cpu_12810	578	1.00E-30	Mn2+ and Fe2+ transporters of the NRAMP family	P
cpu_12820	680	3.00E-91	Allophanate hydrolase subunit 1	E
cpu_12830	764	2.00E-116	Uncharacterized proteins, homologs of lactam utilization protein B	R
cpu_12840	929	2.00E-103	Allophanate hydrolase subunit 2	E
cpu_12850	926	1.00E-76	Biotin synthase and related enzymes	H
cpu_12860	1256	5.00E-85	Uncharacterized protein conserved in bacteria	S
cpu_12870	419	2.00E-39	Predicted transcriptional regulator	K
cpu_12880	1334	0	Cytochrome bd-type quinol oxidase, subunit 1	C
cpu_12890	1016	7.00E-98	Cytochrome bd-type quinol oxidase, subunit 2	C
cpu_12900	1724	0	ABC-type transport system	CO
cpu_12910	1658	3.00E-155	ABC-type transport system	CO
cpu_12920	1598	1.00E-54	Response regulator containing a CheY-like receiver domain and an HD-GYP domain	KT
cpu_12930	605	9.00E-58	Formylmethanofuran dehydrogenase subunit E	C
cpu_12940	1106	1.00E-51	ABC-type Fe3+-hydroxamate transport system, periplasmic component	P
cpu_12950	1043	1.00E-77	ABC-type Fe3+-siderophore transport system, permease component	P
cpu_12960	773	1.00E-108	ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components	PH
cpu_12970	671	8.00E-89	ABC-type multidrug transport system, ATPase component	V
cpu_12990	560	5.00E-16	Regulators of stationary/sporulation gene expression	K
cpu_13000	1007	7.00E-13	Parvulin-like peptidyl-prolyl isomerase	O
cpu_13010	3485	0	Transcription-repair coupling factor (superfamily II helicase)	LK
cpu_13040	563	3.00E-91	Peptidyl-tRNA hydrolase	J
cpu_13050	590	2.00E-20	Ribosomal protein L25 (general stress protein Ctc)	J
cpu_13060	560	7.00E-10	Uncharacterized protein conserved in bacteria	S
cpu_13070	941	5.00E-170	Phosphoribosylpyrophosphate synthetase	FE
cpu_13080	1340	0	N-acetylglucosamine-1-phosphate uridyltransferase	M
cpu_13090	1205	4.00E-149	Threonine dehydratase	E
cpu_13100	746	2.00E-08	Molybdopterin-guanine dinucleotide biosynthesis protein A	H
cpu_13110	689	8.00E-55	Transcriptional regulators	K
cpu_13120	866	7.00E-92	4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase	I
cpu_13130	356	2.00E-19	Uncharacterized protein conserved in bacteria	S
cpu_13140	467	6.00E-38	Conserved protein	R
cpu_13150	1046	5.00E-14	Uncharacterized conserved protein	S
cpu_13190	596	1.00E-61	Uracil-DNA glycosylase	L
cpu_13210	1403	2.00E-161	Uncharacterized conserved protein	S
cpu_13220	2054	3.00E-55	FOG: GGDEF domain	T
cpu_13230	2048	2.00E-56	FOG: GGDEF domain	T
cpu_13240	1151	9.00E-110	Deacetylases, including yeast histone deacetylase and acetoin utilization protein	BQ
cpu_13250	629	1.00E-25	FOG: CBS domain	R

Table 2-1-1 Continued

cpu_13280	1133	3.00E-126	Periplasmic protease	M
cpu_13290	1109	2.00E-47	Membrane proteins related to metalloendopeptidases	M
cpu_13300	887	1.00E-65	Cell division protein	D
cpu_13310	680	3.00E-121	Predicted ATPase involved in cell division	D
cpu_13320	950	6.00E-156	Transketolase, C-terminal subunit	G
cpu_13330	842	6.00E-134	Transketolase, N-terminal subunit	G
cpu_13340	503	3.00E-52	Uncharacterized conserved protein	S
cpu_13350	1010	6.00E-134	Protein chain release factor B	J
cpu_13360	2624	0	Preprotein translocase subunit SecA (ATPase, RNA helicase)	U
cpu_13370	1136	2.00E-70	HD-GYP domain	T
cpu_13400	197	2.00E-31	Cold shock proteins	K
cpu_13410	533	5.00E-34	Ribosome-associated protein Y (PSrp-1)	J
cpu_13420	197	5.00E-33	Cold shock proteins	K
cpu_13430	689	2.00E-50	Predicted amidophosphoribosyltransferases	R
cpu_13440	416	8.00E-20	Uncharacterized conserved protein	S
cpu_13450	821	5.00E-53	Uncharacterized conserved protein	S
cpu_13460	962	3.00E-36	Fe-S oxidoreductase	C
cpu_13470	1208	3.00E-149	Uncharacterized conserved protein containing a ferredoxin-like domain	C
cpu_13510	527	3.00E-10	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	K
cpu_13520	743	9.00E-81	Pleiotropic transcriptional repressor	K
cpu_13540	1190	0	S-adenosylmethionine synthetase	H
cpu_13550	821	5.00E-24	FOG: CBS domain	R
cpu_13570	1055	1.00E-58	HD-GYP domain	T
cpu_13580	1214	1.00E-45	HD-GYP domain	T
cpu_13590	755	2.00E-23	Methyl-accepting chemotaxis protein	NT
cpu_13600	707	1.00E-59	Predicted branched-chain amino acid permease (azaleucine resistance)	E
cpu_13610	320	1.00E-17	Predicted membrane protein	S
cpu_13620	1241	4.00E-13	Pyruvate-formate lyase-activating enzyme	O
cpu_13630	1277	0	Transcription termination factor	K
cpu_13640	644	1.00E-96	Transaldolase	G
cpu_13650	854	5.00E-154	Fructose/tagatose bisphosphate aldolase	G
cpu_13670	410	1.00E-21	Response regulator	T
cpu_13680	1607	0	CTP synthase (UTP-ammonia lyase)	F
cpu_13690	1673	0	Arginyl-tRNA synthetase	J
cpu_13700	422	1.00E-16	Uncharacterized protein conserved in bacteria	S
cpu_13710	563	7.00E-41	Pyruvate:ferredoxin oxidoreductase gamma subunit	C
cpu_13720	1745	0	Indolepyruvate ferredoxin oxidoreductase, alpha and beta subunits	C
cpu_13730	1316	0	Coenzyme F390 synthetase	H
cpu_13740	1472	1.00E-65	Na+/proline symporter	ER
cpu_13750	746	1.00E-78	ABC-type uncharacterized transport system, permease component	R
cpu_13760	659	2.00E-56	ABC-type antimicrobial peptide transport system, ATPase component	V
cpu_13770	440	3.00E-19	Transcriptional regulators	K
cpu_13780	1178	1.00E-155	Aspartate/tyrosine/aromatic aminotransferase	E
cpu_13790	482	4.00E-27	Transcriptional regulators	K
cpu_13810	1217	2.00E-151	GTPases	R
cpu_13820	224	8.00E-08	Indolepyruvate ferredoxin oxidoreductase, alpha and beta subunits	C
cpu_13830	452	3.00E-09	Uncharacterized conserved protein	S
cpu_13860	845	3.00E-56	Zn-dependent protease with chaperone function	O
cpu_13880	869	4.00E-72	ABC-type phosphate/phosphonate transport system, periplasmic component	P
cpu_13890	737	3.00E-48	ABC-type phosphate/phosphonate transport system, permease component	P
cpu_13900	1838	1.00E-158	Asparagine synthase (glutamine-hydrolyzing)	E
cpu_13920	857	3.00E-10	Predicted permease, DMT superfamily	R
cpu_13940	1241	2.00E-12	Arabinose efflux permease	G
cpu_13950	1127	5.00E-09	Ferredoxin	C
cpu_13960	896	8.00E-22	Membrane protease subunits, stomatin/prohibitin homologs	O
cpu_13970	167	1.00E-16	Uncharacterized protein conserved in bacteria	S
cpu_13980	713	9.00E-88	Response regulators consisting of a CheY-like receiver domain	TK

Table 2-1-1 Continued

cpu_13990	1457	2.00E-65	Signal transduction histidine kinase	T
cpu_14000	1202	2.00E-37	ABC-type antimicrobial peptide transport system, permease component	V
cpu_14010	692	5.00E-118	ABC-type antimicrobial peptide transport system, ATPase component	V
cpu_14020	1184	3.00E-25	Membrane-fusion protein	M
cpu_14030	435	8.00E-25	Fe-S-cluster-containing hydrogenase components 2	C
cpu_14040	1910	1.00E-121	6Fe-6S prismane cluster-containing protein	C
cpu_14050	671	3.00E-28	cAMP-binding proteins - catabolite gene activator	T
cpu_14060	158	3.00E-12	Pyridoxine biosynthesis enzyme	H
cpu_14070	449	2.00E-26	Universal stress protein UspA and related nucleotide-binding proteins	T
cpu_14080	332	1.00E-19	Predicted flavin-nucleotide-binding protein	R
cpu_14090	812	2.00E-17	Predicted metal-dependent hydrolase of the TIM-barrel fold	R
cpu_14100	458	2.00E-21	Uncharacterized protein conserved in bacteria	S
cpu_14110	914	3.00E-163	Uncharacterized enzyme involved in pigment biosynthesis	Q
cpu_14120	1097	2.00E-45	Sugar kinases, ribokinase family	G
cpu_14130	1403	3.00E-54	Amino acid transporters	E
cpu_14160	1316	2.00E-174	Ammonia permease	P
cpu_14170	362	1.00E-49	Nitrogen regulatory protein PII	E
cpu_14180	1217	1.00E-17	Uroporphyrinogen-III decarboxylase	H
cpu_14190	632	1.00E-59	Predicted cobalamin binding protein	R
cpu_14200	1385	2.00E-52	Na+/melibiose symporter and related transporters	G
cpu_14210	1202	2.00E-18	Arabinose efflux permease	G
cpu_14230	1769	2.00E-120	Uncharacterized metal-binding protein	R
cpu_14240	788	1.00E-44	Methionine synthase I, cobalamin-binding domain	E
cpu_14250	632	5.00E-55	Predicted cobalamin binding protein	R
cpu_14270	1178	8.00E-21	Uroporphyrinogen-III decarboxylase	H
cpu_14280	632	9.00E-54	Predicted cobalamin binding protein	R
cpu_14290	779	8.00E-67	Transcriptional regulators	K
cpu_14310	425	1.00E-28	Uncharacterized conserved protein	S
cpu_14330	761	9.00E-35	Uncharacterized conserved protein	S
cpu_14340	1361	3.00E-135	Na+-dependent transporters of the SNF family	R
cpu_14350	932	1.00E-159	D-serine dehydratase	E
cpu_14370	1703	4.00E-99	Aldehyde:ferredoxin oxidoreductase	C
cpu_14380	1154	3.00E-129	Alcohol dehydrogenase, class IV	C
cpu_14390	1508	2.00E-18	Sugar diacid utilization regulator	KT
cpu_14400	479	2.00E-72	Uncharacterized conserved protein	S
cpu_14420	1112	2.00E-94	Trypsin-like serine proteases	O
cpu_14430	872	9.00E-21	Predicted periplasmic or secreted lipoprotein	R
cpu_14440	809	1.00E-134	Predicted methyltransferases	R
cpu_14450	749	3.00E-74	Predicted O-methyltransferase	R
cpu_14460	356	7.00E-18	Uncharacterized protein conserved in bacteria	S
cpu_14470	848	7.00E-107	Uncharacterized homolog of PSP1	S
cpu_14480	803	5.00E-19	DNA polymerase III, gamma/tau subunits	L
cpu_14490	611	1.00E-81	Thymidylate kinase	F
cpu_14500	1328	5.00E-71	Arginine/lysine/ornithine decarboxylases	E
cpu_14510	1505	2.00E-21	Subtilisin-like serine proteases	O
cpu_14530	684	5.00E-14	Transposase and inactivated derivatives	L
cpu_14540	257	4.00E-08	Transposase and inactivated derivatives	L
cpu_14550	1130	3.00E-96	Trypsin-like serine proteases	O
cpu_14560	794	2.00E-45	ABC-type Mn2+/Zn2+ transport systems, permease components	P
cpu_14570	689	1.00E-72	ABC-type Mn/Zn transport systems, ATPase component	P
cpu_14580	875	6.00E-76	ABC-type metal ion transport system	P
cpu_14590	422	6.00E-36	Fe2+/Zn2+ uptake regulation proteins	P
cpu_14630	1388	9.00E-145	Predicted Zn-dependent proteases and their inactivated homologs	R
cpu_14640	779	5.00E-12	Tfp pilus assembly protein PilO	NU
cpu_14650	530	2.00E-09	Tfp pilus assembly protein PilN	NU
cpu_14660	1055	5.00E-46	Tfp pilus assembly protein, ATPase PilM	NU
cpu_14680	545	3.00E-09	Tfp pilus assembly protein PilW	NU

Table 2-1-1 Continued

cpu_14690	356	4.00E-09	Tfp pilus assembly protein PilV	NU
cpu_14720	461	2.00E-16	Type II secretory pathway, pseudopilin PulG	NU
cpu_14730	434	4.00E-18	Type II secretory pathway, pseudopilin PulG	NU
cpu_14740	1208	9.00E-129	Type II secretory pathway, component PulF	NU
cpu_14750	1088	0	Tfp pilus assembly protein, pilus retraction ATPase PilT	NU
cpu_14760	1718	0	ATPase PulE/	NU
cpu_14770	1607	9.00E-135	3-dehydroquinate synthetase	E
cpu_14780	905	1.00E-58	Transcriptional regulator	K
cpu_14810	218	5.00E-08	Predicted redox protein, regulator of disulfide bond formation	O
cpu_14820	1154	7.00E-159	Chorismate synthase	E
cpu_14830	848	1.00E-116	Shikimate 5-dehydrogenase	E
cpu_14840	548	3.00E-50	Predicted hydrolase of the HAD superfamily	R
cpu_14850	980	1.00E-23	Molybdopterin biosynthesis enzyme	H
cpu_14860	1127	2.00E-105	ABC-type spermidine/putrescine transport systems, ATPase components	E
cpu_14870	683	5.00E-86	ABC-type molybdate transport system, permease component	P
cpu_14880	794	1.00E-74	ABC-type molybdate transport system, periplasmic component	P
cpu_14890	341	2.00E-19	N-terminal domain of molybdenum-binding protein	R
cpu_14900	704	6.00E-44	DNA-directed RNA polymerase specialized sigma subunit	K
cpu_14910	1793	0	Aldehyde:ferredoxin oxidoreductase	C
cpu_14920	545	3.00E-47	Carbonic anhydrases/acetyltransferases, isoleucine patch superfamily	R
cpu_14930	1211	1.00E-147	Uncharacterized conserved protein	S
cpu_14940	761	1.00E-111	NCAIR mutase (PurE)-related proteins	R
cpu_14950	824	6.00E-109	ATP-utilizing enzymes of the PP-loop superfamily	R
cpu_14960	1010	4.00E-63	Isopropylmalate/homocitrate/citramalate synthases	E
cpu_14970	899	1.00E-116	Acetaldehyde dehydrogenase (acetylating)	Q
cpu_14980	788	1.00E-89	2-keto-4-pentenoate hydratase	Q
cpu_14990	890	6.00E-12	Predicted nucleotide-utilizing enzyme	R
cpu_15000	767	2.00E-28	Nicotinate-nucleotide pyrophosphorylase	H
cpu_15020	977	9.00E-111	ABC-type multidrug transport system, ATPase component	V
cpu_15030	740	1.00E-11	ABC-type multidrug transport system, permease component	V
cpu_15050	578	1.00E-62	Predicted Fe-S oxidoreductase	R
cpu_15060	383	4.00E-36	Predicted Fe-S oxidoreductase	R
cpu_15070	1034	2.00E-123	Threonine aldolase	E
cpu_15080	1544	1.00E-54	Response regulator	KT
cpu_15090	917	3.00E-123	Thioredoxin reductase	O
cpu_15100	383	5.00E-47	Uncharacterized conserved protein	S
cpu_15120	311	2.00E-11	Response regulators consisting of a CheY-like receiver domain	TK
cpu_15140	1121	9.00E-143	Succinyl-CoA synthetase, beta subunit	C
cpu_15150	869	3.00E-138	Succinyl-CoA synthetase, alpha subunit	C
cpu_15160	1586	0	Methylmalonyl-CoA mutase, N-terminal domain/subunit	I
cpu_15170	398	3.00E-52	Methylmalonyl-CoA mutase, C-terminal domain/subunit	I
cpu_15180	410	5.00E-09	4-hydroxyphenylpyruvate dioxygenase and related hemolysins	ER
cpu_15190	1541	0	Acetyl-CoA carboxylase, carboxyltransferase component	I
cpu_15200	1547	0	Acetyl/propionyl-CoA carboxylase, alpha subunit	I
cpu_15210	392	2.00E-26	Biotin carboxyl carrier protein	I
cpu_15220	680	5.00E-56	Predicted Zn-dependent hydrolases of the beta-lactamase fold	R
cpu_15230	641	7.00E-51	Phosphatidylserine decarboxylase	I
cpu_15240	512	1.00E-37	Phosphatidylserine synthase	I
cpu_15250	833	3.00E-28	AraC-type DNA-binding domain-containing proteins	K
cpu_15260	1151	0	Alcohol dehydrogenase, class IV	C
cpu_15270	566	2.00E-26	Heterodisulfide reductase, subunit C	C
cpu_15280	863	5.00E-92	Heterodisulfide reductase, subunit B	C
cpu_15290	584	5.00E-81	Heterodisulfide reductase, subunit A and related polyferredoxins	C
cpu_15300	1382	0	Heterodisulfide reductase, subunit A and related polyferredoxins	C
cpu_15310	209	3.00E-38	Coenzyme F420-reducing hydrogenase, delta subunit	C
cpu_15320	953	7.00E-08	Uncharacterized conserved protein containing a ferredoxin-like domain	C
cpu_15340	842	4.00E-49	2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases	HC

Table 2-1-1 Continued

cpu_15350	1073	8.00E-50	Signal transduction histidine kinase	T
cpu_15370	596	3.00E-14	Predicted Fe-S oxidoreductases	R
cpu_15380	1184	5.00E-39	ABC-type Na+ efflux pump, permease component	CP
cpu_15390	731	2.00E-104	ABC-type Na+ transport system, ATPase component	CP
cpu_15410	1445	2.00E-45	Dinucleotide-utilizing enzymes	H
cpu_15440	659	4.00E-86	Predicted ATPase with chaperone activity	O
cpu_15460	656	5.00E-69	ABC-type multidrug transport system, ATPase component	V
cpu_15500	3272	8.00E-139	ATP-dependent nuclease, subunit B	L
cpu_15510	3503	3.00E-167	ATP-dependent exoDNAse beta subunit	L
cpu_15520	746	4.00E-101	ABC-type antimicrobial peptide transport system, ATPase component	V
cpu_15540	578	6.00E-15	Uncharacterized protein conserved in cyanobacteria	S
cpu_15560	263	2.00E-19	Uncharacterized conserved protein	S
cpu_15580	692	4.00E-41	Uncharacterized protein conserved in bacteria	S
cpu_15620	1070	3.00E-13	Exopolysaccharide biosynthesis protein	G
cpu_15670	746	4.00E-39	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_15690	941	7.00E-40	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_15710	2444	3.00E-47	Predicted helicases	R
cpu_15720	431	2.00E-19	RecB family exonuclease	L
cpu_15730	992	2.00E-106	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_15740	263	1.00E-25	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_15760	863	7.00E-106	Heterodisulfide reductase, subunit B	C
cpu_15770	1253	4.00E-68	Uncharacterized protein involved in benzoate metabolism	Q
cpu_15780	1121	1.00E-80	Xaa-Pro aminopeptidase	E
cpu_15790	788	1.00E-103	Transcriptional regulator	K
cpu_15800	617	2.00E-12	MinD superfamily P-loop ATPase containing an inserted ferredoxin domain	C
cpu_15830	1031	0	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	EP
cpu_15840	968	5.00E-152	ABC-type oligopeptide transport system, ATPase component	E
cpu_15850	1598	4.00E-153	ABC-type dipeptide transport system, periplasmic component	E
cpu_15860	1004	3.00E-144	ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	EP
cpu_15870	839	1.00E-114	ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	EP
cpu_15880	878	6.00E-42	MoxR-like ATPases	R
cpu_15890	1235	8.00E-33	Protein containing von Willebrand factor type A (vWA) domain	R
cpu_15900	704	2.00E-39	Uncharacterized proteins, LmbE homologs	S
cpu_15910	1139	2.00E-48	Glycosyltransferase	M
cpu_15920	392	5.00E-43	Fe2+/Zn2+ uptake regulation proteins	P
cpu_15930	1307	3.00E-141	Predicted transcriptional regulator containing CBS domains	K
cpu_15940	3353	0	DNA polymerase III, alpha subunit	L
cpu_15950	677	5.00E-11	K+ transport systems, NAD-binding component	P
cpu_15980	854	7.00E-153	Acetyl-CoA carboxylase beta subunit	I
cpu_15990	968	8.00E-169	Acetyl-CoA carboxylase alpha subunit	I
cpu_16000	962	2.00E-141	6-phosphofructokinase	G
cpu_16010	1751	0	Pyruvate kinase	G
cpu_16030	692	6.00E-42	Predicted xylanase/chitin deacetylase	G
cpu_16040	986	8.00E-53	D-alanyl-D-alanine carboxypeptidase	M
cpu_16050	1229	8.00E-130	Predicted Zn-dependent peptidases	R
cpu_16070	245	1.00E-10	Uncharacterized conserved protein	S
cpu_16080	791	7.00E-86	Dihydrodipicolinate reductase	E
cpu_16090	911	3.00E-09	Lactate dehydrogenase and related dehydrogenases	CHR
cpu_16100	569	1.00E-08	Archaeal flavoproteins	C
cpu_16110	1019	5.00E-160	Aspartate-semialdehyde dehydrogenase	E
cpu_16120	1232	3.00E-130	Aspartokinases	E
cpu_16130	881	7.00E-118	Dihydrodipicolinate synthase/N-acetylneuraminate lyase	EM
cpu_16140	1706	0	Predicted hydrolase of the metallo-beta-lactamase superfamily	R
cpu_16150	785	8.00E-73	Uncharacterized bacitracin resistance protein	V
cpu_16160	2222	3.00E-169	DNA segregation ATPase FtsK/SpoIIIE and related proteins	D
cpu_16170	476	9.00E-40	Cell wall hydrolyses involved in spore germination	M
cpu_16190	1040	5.00E-42	Predicted ATPase	R

Table 2-1-1 Continued

cpu_16200	689	6.00E-22	Uncharacterized protein conserved in bacteria	S
cpu_16210	1316	8.00E-163	2-methylthioadenine synthetase	J
cpu_16220	491	1.00E-57	Uncharacterized protein (competence- and mitomycin-induced)	R
cpu_16230	1031	8.00E-121	RecA/RadA recombinase	L
cpu_16240	440	3.00E-29	Uncharacterized protein conserved in bacteria	R
cpu_16250	1541	7.00E-66	Predicted HD superfamily hydrolase	R
cpu_16270	776	2.00E-133	Uncharacterized protein conserved in bacteria	S
cpu_16280	260	3.00E-45	Uncharacterized protein conserved in bacteria	S
cpu_16290	890	1.00E-81	Zn-dependent dipeptidase, microsomal dipeptidase homolog	E
cpu_16300	1151	2.00E-120	Aspartate/tyrosine/aromatic aminotransferase	E
cpu_16330	1451	2.00E-48	IMP dehydrogenase/GMP reductase	F
cpu_16340	860	3.00E-49	MoxR-like ATPases	R
cpu_16350	1286	2.00E-29	Protein containing von Willebrand factor type A (vWA) domain	R
cpu_16370	91	1.00E-18	GTPases - translation elongation factors	J
cpu_16380	647	1.00E-20	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	K
cpu_16390	506	2.00E-47	Predicted RNA-binding protein containing a PIN domain	R
cpu_16400	713	2.00E-72	rRNA methylases	J
cpu_16410	671	9.00E-53	Predicted alternative thymidylate synthase	F
cpu_16420	389	3.00E-40	Uncharacterized protein conserved in bacteria	S
cpu_16430	1397	0	Cysteinyl-tRNA synthetase	J
cpu_16440	671	6.00E-103	Serine acetyltransferase	E
cpu_16450	1436	0	Glutamyl- and glutaminyl-tRNA synthetases	J
cpu_16460	488	4.00E-86	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	I
cpu_16470	701	4.00E-74	4-diphosphocytidyl-2-methyl-D-erithritol synthase	I
cpu_16480	1082	2.00E-132	Integral membrane protein (PIN domain superfamily)	R
cpu_16490	482	1.00E-48	Transcriptional regulators, similar to M. xanthus CarD	K
cpu_16510	1058	9.00E-168	Predicted nucleic-acid-binding protein (contains the HHH domain)	R
cpu_16520	1337	0	Predicted ATP-dependent serine protease	O
cpu_16530	2441	0	ATPases with chaperone activity, ATP-binding subunit	O
cpu_16540	1055	1.00E-128	Arginine kinase	E
cpu_16550	521	2.00E-48	Uncharacterized protein with conserved CXXC pairs	S
cpu_16560	473	3.00E-55	Transcriptional repressor of class III stress genes	K
cpu_16570	1046	1.00E-52	HD-GYP domain	T
cpu_16580	227	1.00E-29	HD-GYP domain	T
cpu_16590	251	2.00E-15	Serine phosphatase RsbU, regulator of sigma subunit	TK
cpu_16600	554	7.00E-34	Serine phosphatase RsbU, regulator of sigma subunit	TK
cpu_16610	314	2.00E-08	Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	T
cpu_16630	1889	0	6Fe-6S prismane cluster-containing protein	C
cpu_16640	290	1.00E-35	Ribosomal protein S6	J
cpu_16650	428	4.00E-40	Single-stranded DNA-binding protein	L
cpu_16660	230	9.00E-28	Ribosomal protein S18	J
cpu_16670	986	4.00E-18	Predicted membrane protein	S
cpu_16680	458	1.00E-47	Ribosomal protein L9	J
cpu_16690	1919	2.00E-100	Predicted ATP-dependent protease	O
cpu_16700	1331	0	Replicative DNA helicase	L
cpu_16720	1088	3.00E-66	Pyruvate:ferredoxin oxidoreductase	C
cpu_16730	743	2.00E-64	Pyruvate:ferredoxin oxidoreductase	C
cpu_16740	563	1.00E-26	Pyruvate:ferredoxin oxidoreductase	C
cpu_16750	191	1.00E-11	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_16760	722	2.00E-23	Recombinational DNA repair protein (RecF pathway)	L
cpu_16770	872	0	Glycyl-tRNA synthetase, alpha subunit	J
cpu_16780	2072	0	Glycyl-tRNA synthetase, beta subunit	J
cpu_16790	641	4.00E-12	FOG: CBS domain	R
cpu_16800	830	8.00E-105	Uncharacterized protein conserved in bacteria	S
cpu_16810	2660	0	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase	G
cpu_16850	371	8.00E-20	Predicted transcriptional regulators	K
cpu_16870	830	3.00E-52	ABC-type multidrug transport system, ATPase component	V

Table 2-1-1 Continued

cpu_16920	338	5.00E-26	Predicted transcriptional regulators	K
cpu_16930	677	8.00E-08	Uncharacterized protein conserved in bacteria	S
cpu_16940	974	8.00E-80	Periplasmic serine proteases (ClpP class)	OU
cpu_16980	209	4.00E-19	Predicted ATPase with chaperone activity	O
cpu_17010	327	7.00E-15	Transposase and inactivated derivatives	L
cpu_17020	378	7.00E-12	Predicted dehydrogenase	R
cpu_17040	962	5.00E-128	tRNA-dihydrouridine synthase	J
cpu_17050	770	4.00E-97	Putative transcriptional regulator, homolog of Bvg accessory factor	K
cpu_17060	980	2.00E-80	Biotin-(acetyl-CoA carboxylase) ligase	H
cpu_17070	851	9.00E-141	Nicotinate-nucleotide pyrophosphorylase	H
cpu_17080	1559	0	Aspartate oxidase	H
cpu_17090	929	2.00E-157	Quinolinate synthase	H
cpu_17100	383	9.00E-75	Aspartate 1-decarboxylase	H
cpu_17110	851	4.00E-154	Panthothenate synthetase	H
cpu_17120	830	2.00E-153	Ketopantoate hydroxymethyltransferase	H
cpu_17130	875	2.00E-42	Uncharacterized conserved protein	S
cpu_17140	485	3.00E-62	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase	H
cpu_17150	359	2.00E-47	Dihydroneopterin aldolase	H
cpu_17160	1181	4.00E-110	Dihydropteroate synthase and related enzymes	H
cpu_17170	482	4.00E-10	Predicted HD superfamily hydrolase	R
cpu_17180	1670	0	Formyltetrahydrofolate synthetase	F
cpu_17190	395	9.00E-31	Predicted thioesterase	R
cpu_17210	923	1.00E-128	Thioredoxin reductase	O
cpu_17220	329	6.00E-26	Thioredoxin domain-containing protein	O
cpu_17290	1145	1.00E-12	Fatty acid/phospholipid biosynthesis enzyme	I
cpu_17310	1118	4.00E-87	Uncharacterized conserved protein	S
cpu_17320	941	3.00E-62	Transcriptional regulator	K
cpu_17330	1397	2.00E-177	Cobyrinic acid a,c-diamide synthase	H
cpu_17340	1082	9.00E-66	Predicted membrane protein	S
cpu_17360	317	2.00E-40	Dissimilatory sulfite reductase (desulfoviridin), gamma subunit	P
cpu_17370	1328	1.00E-49	Fe-S oxidoreductase	C
cpu_17380	734	1.00E-17	Nitrate reductase gamma subunit	C
cpu_17390	1292	7.00E-53	Fe-S oxidoreductase	C
cpu_17400	776	5.00E-123	Uroporphyrinogen-III methylase	H
cpu_17420	1058	5.00E-84	Dissimilatory sulfite reductase (desulfoviridin), alpha and beta subunits	C
cpu_17430	1193	2.00E-80	Dissimilatory sulfite reductase (desulfoviridin), alpha and beta subunits	C
cpu_17440	1610	0	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_17450	392	2.00E-27	Uncharacterized protein, possibly involved in aromatic compounds catabolism	Q
cpu_17460	1127	1.00E-32	High-affinity Fe2+/Pb2+ permease	P
cpu_17470	413	2.00E-31	Molecular chaperone (small heat shock protein)	O
cpu_17480	1163	1.00E-11	Arabinose efflux permease	G
cpu_17490	848	8.00E-87	Uncharacterized protein conserved in bacteria	S
cpu_17500	437	1.00E-30	Molecular chaperone (small heat shock protein)	O
cpu_17510	797	2.00E-120	ABC-type metal ion transport system, periplasmic component/surface antigen	P
cpu_17520	647	1.00E-97	ABC-type metal ion transport system, permease component	P
cpu_17530	1013	3.00E-171	ABC-type metal ion transport system, ATPase component	P
cpu_17550	905	2.00E-89	ABC-type multidrug transport system, ATPase component	V
cpu_17570	467	7.00E-21	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	K
cpu_17580	272	2.00E-10	Predicted membrane protein	S
cpu_17590	188	4.00E-09	Predicted membrane protein	S
cpu_17600	935	4.00E-15	Predicted membrane-bound metal-dependent hydrolases	R
cpu_17610	986	4.00E-36	Ca2+/Na+ antiporter	P
cpu_17630	413	2.00E-30	cAMP-binding proteins - catabolite gene activator	T
cpu_17640	239	4.00E-11	cAMP-binding proteins - catabolite gene activator	T
cpu_17650	539	3.00E-22	Uncharacterized conserved protein	S
cpu_17670	704	4.00E-33	Uncharacterized Fe-S protein	C
cpu_17680	746	3.00E-72	Dehydrogenases with different specificities	IQR

Table 2-1-1 Continued

cpu_17690	1652	7.00E-86	Activator of 2-hydroxyglutaryl-CoA dehydratase	I
cpu_17700	1256	3.00E-20	Benzoyl-CoA reductase	E
cpu_17720	1946	1.00E-164	NADH:flavin oxidoreductases, Old Yellow Enzyme family	C
cpu_17730	1205	1.00E-08	Transposase and inactivated derivatives	L
cpu_17750	569	2.00E-32	Multimeric flavodoxin WrbA	R
cpu_17760	668	1.00E-16	Uncharacterized conserved protein	S
cpu_17770	1550	9.00E-169	Transcriptional regulator	KT
cpu_17800	689	1.00E-102	ABC-type antimicrobial peptide transport system, ATPase component	V
cpu_17810	1139	3.00E-08	ABC-type transport system, involved in lipoprotein release, permease component	M
cpu_17820	692	3.00E-09	Antirepressor regulating drug resistance	KT
cpu_17840	239	7.00E-17	Predicted transcriptional regulator	K
cpu_17850	458	2.00E-46	Fe2+/Zn2+ uptake regulation proteins	P
cpu_17870	1175	3.00E-105	D-alanyl-D-alanine carboxypeptidase	M
cpu_17880	329	6.00E-36	Uncharacterized conserved protein	S
cpu_17890	371	1.00E-24	Integral membrane protein possibly involved in chromosome condensation	D
cpu_17900	404	6.00E-22	Integral membrane protein possibly involved in chromosome condensation	D
cpu_17920	1205	2.00E-114	Transposase and inactivated derivatives	L
cpu_17930	263	1.00E-27	DNA replication protein	L
cpu_17960	725	3.00E-08	Predicted hydrolase (metallo-beta-lactamase superfamily)	R
cpu_17980	1847	2.00E-17	Predicted ATP-dependent endonuclease of the OLD family	L
cpu_18000	374	2.00E-18	Uncharacterized conserved protein	S
cpu_18040	569	5.00E-15	Uncharacterized conserved protein	S
cpu_18100	542	6.00E-32	RecB family exonuclease	L
cpu_18120	995	9.00E-46	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_18130	641	3.00E-22	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_18140	2168	1.00E-83	Predicted helicases	R
cpu_18150	983	1.00E-111	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_18160	299	5.00E-28	Uncharacterized protein predicted to be involved in DNA repair	L
cpu_18240	1532	9.00E-59	HD-GYP domain	T
cpu_18250	1241	3.00E-44	Methyl-accepting chemotaxis protein	NT
cpu_18260	1208	4.00E-26	HD-GYP domain	T
cpu_18270	1175	2.00E-64	Transcriptional regulator/sugar kinase	KG
cpu_18280	1241	5.00E-71	ABC-type sugar transport system, periplasmic component	G
cpu_18290	881	3.00E-91	ABC-type sugar transport systems, permease components	G
cpu_18300	866	8.00E-81	ABC-type sugar transport system, permease component	G
cpu_18310	1058	3.00E-167	ABC-type sugar transport systems, ATPase components	G
cpu_18320	434	4.00E-22	6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase	G
cpu_18330	254	1.00E-07	Fe2+ transport system protein A	P
cpu_18340	1442	8.00E-120	Fe2+ transport system protein B	P
cpu_18350	599	1.00E-79	Fe2+ transport system protein B	P
cpu_18390	482	3.00E-13	Predicted membrane protein/domain	S
cpu_18400	533	3.00E-41	2'-5' RNA ligase	J
cpu_18410	617	1.00E-35	Multimeric flavodoxin WrbA	R
cpu_18420	998	3.00E-94	dGTP triphosphohydrolase	F
cpu_18430	1727	6.00E-134	DNA primase (bacterial type)	L
cpu_18440	1115	4.00E-138	DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32)	K
cpu_18450	698	1.00E-56	Predicted SAM-dependent methyltransferase	R
cpu_18460	1106	3.00E-46	Uncharacterized protein conserved in bacteria	S
cpu_18470	713	3.00E-12	Zn-ribbon protein, possibly nucleic acid-binding	R
cpu_18480	326	6.00E-38	Uncharacterized conserved protein	S
cpu_18490	872	3.00E-14	Predicted permease, DMT superfamily	R
cpu_18500	362	1.00E-68	Methylglyoxal synthase	G
cpu_18530	356	2.00E-41	Chorismate mutase	E
cpu_18540	1085	1.00E-80	Prephenate dehydrogenase	E
cpu_18550	1007	2.00E-141	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	E
cpu_18560	2675	5.00E-54	Anaerobic dehydrogenases, typically selenocysteine-containing	C
cpu_18570	596	3.00E-63	Fe-S-cluster-containing hydrogenase components 1	C

Table 2-1-1 Continued

cpu_18580	875	5.00E-09	Formate-dependent nitrite reductase, membrane component	P
cpu_18590	1070	2.00E-95	NaMN:DMB phosphoribosyltransferase	H
cpu_18600	1391	1.00E-103	Protoporphyrinogen oxidase	H
cpu_18610	878	2.00E-76	Protoheme ferro-lyase (ferrochelatase)	H
cpu_18620	1052	4.00E-154	Uroporphyrinogen-III decarboxylase	H
cpu_18630	1310	2.00E-173	Methylaspartate ammonia-lyase	E
cpu_18640	1265	5.00E-117	Benzoyl-CoA reductase	E
cpu_18650	734	5.00E-72	Activator of 2-hydroxyglutaryl-CoA dehydratase	I
cpu_18660	1082	1.00E-160	Glycine cleavage system T protein (aminomethyltransferase)	E
cpu_18670	392	4.00E-58	Glycine cleavage system H protein (lipoate-binding)	E
cpu_18680	1295	0	Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain	E
cpu_18690	1460	0	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	E
cpu_18700	803	8.00E-71	Uncharacterized homolog of biotin synthetase	S
cpu_18710	917	2.00E-53	Biotin synthase-related enzyme	R
cpu_18720	962	2.00E-148	ABC-type oligopeptide transport system, ATPase component	E
cpu_18730	1004	0	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	EP
cpu_18740	914	1.00E-111	ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	EP
cpu_18750	929	2.00E-114	ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	EP
cpu_18760	1625	1.00E-158	ABC-type oligopeptide transport system, periplasmic component	E
cpu_18770	1466	0	Lysyl-tRNA synthetase (class II)	J
cpu_18780	473	3.00E-51	Transcription elongation factor	K
cpu_18790	691	9.00E-43	Predicted dehydrogenase	R
cpu_18820	422	7.00E-28	Type II secretory pathway, component ExeA (predicted ATPase)	U
cpu_18830	497	2.00E-63	Uncharacterized protein conserved in bacteria	S
cpu_18870	1211	2.00E-135	Uncharacterized conserved protein	S
cpu_18880	1319	1.00E-11	Cytochrome c peroxidase	P
cpu_18890	392	1.00E-39	Uncharacterized conserved protein	S
cpu_18900	545	4.00E-15	Lysophospholipase L1 and related esterases	E
cpu_18910	845	5.00E-86	Predicted Co/Zn/Cd cation transporters	P
cpu_18920	890	7.00E-13	Permeases of the drug/metabolite transporter (DMT) superfamily	GER
cpu_18930	818	4.00E-58	ABC-type amino acid transport	ET
cpu_18940	221	9.00E-17	ABC-type amino acid transport system, permease component	E
cpu_18950	458	1.00E-44	HD superfamily phosphohydrolases	R
cpu_18960	686	1.00E-15	HD superfamily phosphohydrolases	R
cpu_18980	575	2.00E-18	Uncharacterized protein conserved in cyanobacteria	S
cpu_19030	725	3.00E-106	Bacteriophage protein gp37	S
cpu_19050	1700	6.00E-09	Uncharacterized conserved protein	S
cpu_19090	272	7.00E-16	Uncharacterized conserved protein	S
cpu_19100	1069	3.00E-44	Transposase and inactivated derivatives	L
cpu_19110	193	2.00E-22	DNA replication protein	L
cpu_19150	644	7.00E-79	Predicted ATPase with chaperone activity	O
cpu_19200	179	3.00E-30	Predicted ATPase with chaperone activity	O
cpu_19210	239	5.00E-13	Predicted ATPase with chaperone activity	O
cpu_19240	585	2.00E-70	ABC-type amino acid transport system, permease component	E
cpu_19250	728	3.00E-167	ABC-type polar amino acid transport system, ATPase component	E
cpu_19260	530	8.00E-18	Acetyltransferases, including N-acetylases of ribosomal proteins	J
cpu_19290	671	3.00E-28	ABC-type multidrug transport system, ATPase component	V
cpu_19300	2453	6.00E-98	Predicted membrane protein	S
cpu_19310	566	9.00E-19	Transcriptional regulator	K
cpu_19340	363	8.00E-25	DNA replication protein	L
cpu_19380	422	9.00E-30	Uncharacterized conserved protein	S
cpu_19400	632	2.00E-52	Methenyl tetrahydrofolate cyclohydrolase	E
cpu_19410	863	6.00E-111	5,10-methylene-tetrahydrofolate dehydrogenase	H
cpu_19420	620	3.00E-43	Uncharacterized membrane-associated protein	S
cpu_19430	746	2.00E-28	Uncharacterized Fe-S protein	C
cpu_19440	1112	1.00E-115	Leucyl aminopeptidase (aminopeptidase T)	E
cpu_19450	1292	3.00E-175	6Fe-6S prismane cluster-containing protein	C

Table 2-1-1 Continued

cpu_19460	2228	1.00E-132	ATP-dependent exoDNAse alpha subunit	L
cpu_19470	1094	2.00E-48	Glycosyltransferase	M
cpu_19490	758	7.00E-108	2-keto-4-pentenoate hydratase	Q
cpu_19500	383	8.00E-59	S-adenosylmethionine decarboxylase	E
cpu_19520	722	6.00E-95	Methylase involved in ubiquinone/menaquinone biosynthesis	H
cpu_19530	1448	0	3-polyprenyl-4-hydroxybenzoate decarboxylase and related decarboxylases	H
cpu_19540	866	5.00E-50	4-hydroxybenzoate polyprenyltransferase and related prenyltransferases	H
cpu_19550	1100	1.00E-152	Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes	HR
cpu_19560	845	3.00E-47	Predicted periplasmic solute-binding protein	R
cpu_19570	1076	5.00E-162	Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes	HR
cpu_19580	1400	4.00E-118	Selenocysteine synthase [seryl-tRNASer selenium transferase]	E
cpu_19590	1907	1.00E-105	Selenocysteine-specific translation elongation factor	J
cpu_19600	578	6.00E-21	Multimeric flavodoxin WrbA	R
cpu_19620	611	5.00E-31	Nitroreductase	C
cpu_19630	839	5.00E-63	Predicted eukaryotic-type DNA primase	L
cpu_19640	1067	3.00E-112	Predicted Rossmann fold nucleotide-binding protein involved in DNA uptake	LU
cpu_19650	2078	0	Topoisomerase IA	L
cpu_19660	1301	0	NAD(FAD)-utilizing enzyme possibly involved in translation	J
cpu_19670	890	2.00E-126	Site-specific recombinase XerD	L
cpu_19680	542	3.00E-100	ATP-dependent protease HslVU (ClpYQ), peptidase subunit	O
cpu_19690	1385	0	ATP-dependent protease HslVU (ClpYQ), ATPase subunit	O
cpu_19700	785	1.00E-119	Pleiotropic transcriptional repressor	K
cpu_19710	698	9.00E-146	Ribosomal protein S2	J
cpu_19720	653	3.00E-62	Translation elongation factor Ts	J
cpu_19730	728	6.00E-137	Uridylate kinase	F
cpu_19740	557	5.00E-96	Ribosome recycling factor	J
cpu_19760	164	1.00E-09	MinD superfamily P-loop ATPase containing an inserted ferredoxin domain	C
cpu_19770	767	1.00E-125	Undecaprenyl pyrophosphate synthase	I
cpu_19780	770	4.00E-39	CDP-diglyceride synthetase	I
cpu_19790	1052	3.00E-32	Predicted permease	R
cpu_19800	1157	0	1-deoxy-D-xylulose 5-phosphate reductoisomerase	I
cpu_19810	1031	2.00E-64	Predicted membrane-associated Zn-dependent proteases 1	M
cpu_19820	1064	0	Enzyme involved in the deoxyxylulose pathway of isoprenoid biosynthesis	I
cpu_19830	1709	0	Prolyl-tRNA synthetase	J
cpu_19840	653	2.00E-17	Glycosyltransferases involved in cell wall biogenesis	M
cpu_19850	659	7.00E-61	Uncharacterized membrane protein	S
cpu_19860	4247	0	DNA polymerase III, alpha subunit (gram-positive type)	L
cpu_19870	1394	0	Ammonia permease	P
cpu_19880	827	3.00E-172	Nitrogenase subunit NifH (ATPase)	P
cpu_19890	317	6.00E-38	Nitrogen regulatory protein PII	E
cpu_19900	362	3.00E-34	Nitrogen regulatory protein PII	E
cpu_19920	1298	3.00E-104	Nitrogenase molybdenum-iron protein, alpha and beta chains	C
cpu_19930	1478	6.00E-111	Nitrogenase molybdenum-iron protein, alpha and beta chains	C
cpu_19940	1823	6.00E-103	Nitrogenase molybdenum-iron protein, alpha and beta chains	C
cpu_19950	815	7.00E-25	Predicted Fe-S oxidoreductases	R
cpu_19960	806	1.00E-94	Isopropylmalate/homocitrate/citramalate synthases	E
cpu_19970	464	6.00E-55	Uncharacterized protein conserved in bacteria	S
cpu_19980	1040	9.00E-72	Transcription elongation factor	K
cpu_19990	278	2.00E-27	Predicted nucleic-acid-binding protein implicated in transcription termination	K
cpu_20010	2510	0	Translation initiation factor 2 (IF-2; GTPase)	J
cpu_20020	389	2.00E-36	Ribosome-binding factor A	J
cpu_20030	950	3.00E-64	Exopolyphosphatase-related proteins	R
cpu_20040	869	1.00E-69	Pseudouridine synthase	J
cpu_20050	926	9.00E-100	FAD synthase	H
cpu_20070	269	1.00E-41	Ribosomal protein S15P/S13E	J
cpu_20080	2198	0	Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase)	J
cpu_20090	866	3.00E-77	Endonuclease IV	L

Table 2-1-1 Continued

cpu_20100	680	2.00E-32	Cell wall hydrolyses involved in spore germination	M
cpu_20140	785	3.00E-94	ATPase component/photorepair protein PhrA	P
cpu_20160	1238	1.00E-123	Alcohol dehydrogenase, class IV	C
cpu_20170	1793	0	Aldehyde:ferredoxin oxidoreductase	C
cpu_20190	800	1.00E-66	Transcriptional regulator	K
cpu_20200	1154	2.00E-143	Acyl-CoA dehydrogenases	I
cpu_20210	1175	1.00E-115	Acetyl-CoA acetyltransferase	I
cpu_20220	1007	2.00E-69	3-hydroxy-3-methylglutaryl CoA synthase	I
cpu_20230	428	6.00E-32	Predicted nucleic-acid-binding protein containing a Zn-ribbon	R
cpu_20240	869	4.00E-114	3-hydroxyacyl-CoA dehydrogenase	I
cpu_20250	773	7.00E-82	Enoyl-CoA hydratase/carnithine racemase	I
cpu_20260	1187	1.00E-124	Acetyl-CoA acetyltransferase	I
cpu_20270	1160	6.00E-131	Acyl-CoA dehydrogenases	I
cpu_20280	800	8.00E-81	Enoyl-CoA hydratase/carnithine racemase	I
cpu_20290	1646	0	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_20310	1142	2.00E-141	Acyl-CoA dehydrogenases	I
cpu_20320	1652	0	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_20330	1292	8.00E-41	Uncharacterized conserved protein	S
cpu_20350	443	2.00E-17	Acyl dehydratase	I
cpu_20360	1427	1.00E-176	Aromatic ring hydroxylase	Q
cpu_20370	776	4.00E-45	Predicted hydrolases or acyltransferases	R
cpu_20380	635	3.00E-52	Multimeric flavodoxin WrbA	R
cpu_20420	722	6.00E-46	DNA replication protein	L
cpu_20430	1484	9.00E-25	Transposase and inactivated derivatives	L
cpu_20450	1112	5.00E-29	Retron-type reverse transcriptase	L
cpu_20470	302	1.00E-08	Transposase and inactivated derivatives	L
cpu_20490	2075	0	Highly conserved protein containing a thioredoxin domain	O
cpu_20540	3305	3.00E-33	Uncharacterized protein conserved in bacteria	S
cpu_20560	857	3.00E-10	Uncharacterized conserved	S
cpu_20570	461	6.00E-29	Fe-S-cluster-containing hydrogenase components 2	C
cpu_20580	1889	0	Aldehyde:ferredoxin oxidoreductase	C
cpu_20610	929	3.00E-98	ABC-type multidrug transport system, ATPase component	V
cpu_20620	1082	1.00E-19	ABC-type multidrug transport system, permease component	V
cpu_20630	242	2.00E-20	Superfamily I DNA and RNA helicases	L
cpu_20670	1172	1.00E-127	Nucleotidyltransferase/DNA polymerase involved in DNA repair	L
cpu_20690	2495	0	DNA polymerase I - 3'-5' exonuclease and polymerase domains	L
cpu_20700	791	1.00E-115	Formamidopyrimidine-DNA glycosylase	L
cpu_20710	623	2.00E-17	Predicted membrane protein	S
cpu_20720	590	8.00E-72	Dephospho-CoA kinase	H
cpu_20730	530	1.00E-29	Soluble lytic murein transglycosylase	M
cpu_20750	533	3.00E-29	Micrococcal nuclease (thermonuclease) homologs	L
cpu_20760	398	5.00E-20	Uncharacterized conserved protein	S
cpu_20770	539	5.00E-27	Uncharacterized conserved protein	R
cpu_20790	1487	7.00E-67	HD-GYP domain	T
cpu_20800	674	6.00E-76	Uncharacterized conserved protein	S
cpu_20820	2291	0	Membrane carboxypeptidase (penicillin-binding protein)	M
cpu_20830	1175	4.00E-137	Acetyl-CoA acetyltransferase	I
cpu_20840	656	1.00E-36	ABC-type Co2+ transport system, permease component	P
cpu_20860	809	3.00E-25	ABC-type cobalt transport system, permease component CbiQ	P
cpu_20870	812	6.00E-96	ABC-type cobalt transport system, ATPase component	P
cpu_20880	1607	4.00E-164	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_20890	1199	1.00E-70	Mn2+ and Fe2+ transporters of the NRAMP family	P
cpu_20900	1163	2.00E-54	Uncharacterized conserved protein	S
cpu_20930	1373	7.00E-59	Na+/H+ antiporter	C
cpu_20940	1718	9.00E-85	Glucoamylase and related glycosyl hydrolases	G
cpu_20960	461	2.00E-42	Uncharacterized conserved protein	S
cpu_20970	833	1.00E-109	Spermidine synthase	E

Table 2-1-1 Continued

cpu_20980	863	9.00E-85	Arginase/agmatinase/formimionoglutamate hydrolase, arginase family	E
cpu_21000	650	1.00E-17	Putative threonine efflux protein	E
cpu_21010	1475	9.00E-162	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	IQ
cpu_21040	2417	5.00E-84	3-hydroxyacyl-CoA dehydrogenase	I
cpu_21050	1172	1.00E-132	Acetyl-CoA acetyltransferase	I
cpu_21060	1742	8.00E-115	Acyl-CoA dehydrogenases	I
cpu_21070	596	2.00E-17	Transcriptional regulator	K
cpu_21080	1181	6.00E-118	Acetyl-CoA acetyltransferase	I
cpu_21090	851	2.00E-150	3-hydroxyacyl-CoA dehydrogenase	I
cpu_21100	1142	1.00E-157	Acyl-CoA dehydrogenases	I
cpu_21110	782	9.00E-93	Enoyl-CoA hydratase/carnithine racemase	I
cpu_21120	572	7.00E-86	Electron transfer flavoprotein, alpha subunit	C
cpu_21130	1181	8.00E-150	Acetyl-CoA acetyltransferase	I
cpu_21140	878	4.00E-28	Fe-S oxidoreductase	C
cpu_21150	638	3.00E-52	Predicted phosphatases	R
cpu_21160	1721	0	Transcriptional regulator	KT
cpu_21170	560	1.00E-16	Signal peptidase I	U
cpu_21180	470	5.00E-16	Uncharacterized protein involved in biosynthesis of c-type cytochromes	O
cpu_21190	1100	2.00E-60	Coenzyme F420-reducing hydrogenase, beta subunit	C
cpu_21200	1016	6.00E-104	TRAP-type C4-dicarboxylate transport system, periplasmic component	G
cpu_21210	1298	7.00E-146	Na+/H+-dicarboxylate symporters	C
cpu_21220	665	5.00E-82	Response regulator of citrate/malate metabolism	KT
cpu_21230	1556	1.00E-156	Signal transduction histidine kinase regulating citrate/malate metabolism	T
cpu_21240	1238	0	Malic enzyme	C
cpu_21250	791	3.00E-108	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	C
cpu_21260	344	1.00E-07	Succinate dehydrogenase, hydrophobic anchor subunit	C
cpu_21270	419	7.00E-14	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	C
cpu_21280	1712	0	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	C
cpu_21290	548	2.00E-80	Tartrate dehydratase beta subunit	C
cpu_21300	842	9.00E-135	Tartrate dehydratase alpha subunit	C
cpu_21310	1514	8.00E-164	Transcriptional regulator	KT
cpu_21320	611	5.00E-67	SOS-response transcriptional repressors (RecA-mediated autopeptidases)	KT
cpu_21350	695	4.00E-21	ABC-type transport system	O
cpu_21360	677	2.00E-37	ABC-type transport system	O
cpu_21370	701	1.00E-72	ABC-type multidrug transport system, ATPase component	V
cpu_21400	1994	8.00E-148	Cytochrome c biogenesis factor	O
cpu_21410	413	1.00E-12	Cytochrome c-type biogenesis protein CcmE	O
cpu_21430	1214	2.00E-170	Cystathionine beta-lyase family protein involved in aluminum resistance	P
cpu_21440	908	1.00E-18	ATPases of the AAA+ class	O
cpu_21450	1340	5.00E-53	Uncharacterized protein conserved in bacteria	S
cpu_21460	257	8.00E-33	Uncharacterized host factor I protein	R
cpu_21470	935	2.00E-138	tRNA delta(2)-isopentenylpyrophosphate transferase	J
cpu_21490	1724	2.00E-170	DNA mismatch repair enzyme (predicted ATPase)	L
cpu_21500	2525	0	Mismatch repair ATPase (MutS family)	L
cpu_21510	1319	0	2-methylthioadenine synthetase	J
cpu_21550	1067	4.00E-16	Dehydrogenases (flavoproteins)	C
cpu_21580	629	1.00E-34	Septum formation inhibitor	D
cpu_21590	530	1.00E-11	NTP pyrophosphohydrolases including oxidative damage repair enzymes	LR
cpu_21620	992	3.00E-62	Uncharacterized protein conserved in bacteria	S
cpu_21630	1139	2.00E-39	Uncharacterized protein conserved in bacteria	S
cpu_21640	950	3.00E-103	Activator of 2-hydroxyglutaryl-CoA dehydratase	I
cpu_21650	1466	2.00E-70	Signal transduction histidine kinase	T
cpu_21660	686	6.00E-96	Response regulators c	TK
cpu_21670	356	4.00E-31	Predicted endonuclease	L
cpu_21680	1430	2.00E-115	NADH:ubiquinone oxidoreductase subunit 2 (chain N)	C
cpu_21690	1520	4.00E-145	NADH:ubiquinone oxidoreductase subunit 4 (chain M)	C
cpu_21700	1931	0	NADH:ubiquinone oxidoreductase subunit 5 (chain L)	CP

Table 2-1-1 Continued

cpu_21710	308	3.00E-31	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K)	C
cpu_21720	509	6.00E-23	NADH:ubiquinone oxidoreductase subunit 6 (chain J)	C
cpu_21730	413	3.00E-40	Formate hydrogenlyase subunit 6	C
cpu_21740	1052	2.00E-111	NADH:ubiquinone oxidoreductase subunit 1 (chain H)	C
cpu_21750	1097	0	NADH:ubiquinone oxidoreductase 49 kD subunit	C
cpu_21760	428	4.00E-50	NADH:ubiquinone oxidoreductase 27 kD subunit	C
cpu_21770	500	2.00E-87	NADH:ubiquinone oxidoreductase 20 kD subunit	C
cpu_21780	356	5.00E-31	NADH:ubiquinone oxidoreductase subunit 3 (chain A)	C
cpu_21790	653	7.00E-73	Ribonuclease HII	L
cpu_21800	836	2.00E-70	Predicted GTPases	R
cpu_21810	341	2.00E-58	Ribosomal protein L19	J
cpu_21820	572	1.00E-92	Uncharacterized protein conserved in bacteria	S
cpu_21830	761	7.00E-143	tRNA-(guanine-N1)-methyltransferase	J
cpu_21840	512	4.00E-43	RimM protein, required for 16S rRNA processing	J
cpu_21850	227	2.00E-24	Predicted RNA-binding protein (contains KH domain)	R
cpu_21860	260	1.00E-37	Ribosomal protein S16	J
cpu_21870	1340	0	Signal recognition particle GTPase	U
cpu_21880	341	5.00E-19	Uncharacterized protein conserved in bacteria	S
cpu_21890	1373	4.00E-154	Ornithine/acetylornithine aminotransferase	E
cpu_21900	1172	7.00E-150	Predicted SAM-dependent methyltransferases	R
cpu_21910	1292	1.00E-127	Cytosine deaminase and related metal-dependent hydrolases	FR
cpu_21920	632	4.00E-42	Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases	G
cpu_21930	1235	0	S-adenosylhomocysteine hydrolase	H
cpu_21940	797	2.00E-106	Purine nucleoside phosphorylase	F
cpu_21950	908	3.00E-151	Signal recognition particle GTPase	U
cpu_21960	3557	0	Chromosome segregation ATPases	D
cpu_21970	701	4.00E-89	dsRNA-specific ribonuclease	K
cpu_21980	1238	0	3-oxoacyl-(acyl-carrier-protein) synthase	IQ
cpu_21990	233	2.00E-20	Acyl carrier protein	IQ
cpu_22000	743	3.00E-81	Dehydrogenases with different specificities	IQR
cpu_22010	938	2.00E-124	(acyl-carrier-protein) S-malonyltransferase	I
cpu_22020	944	2.00E-106	Dioxygenases related to 2-nitropropane dioxygenase	R
cpu_22030	992	1.00E-153	3-oxoacyl-[acyl-carrier-protein] synthase III	I
cpu_22040	992	3.00E-133	Fatty acid/phospholipid biosynthesis enzyme	I
cpu_22060	176	1.00E-20	Ribosomal protein L32	J
cpu_22070	497	2.00E-12	Predicted metal-binding, possibly nucleic acid-binding protein	R
cpu_22080	1196	0	Acetate kinase	C
cpu_22090	1004	2.00E-140	Phosphotransacetylase	C
cpu_22100	1082	1.00E-08	Uncharacterized protein conserved in bacteria	S
cpu_22120	488	4.00E-90	Phosphopantetheine adenylyltransferase	H
cpu_22130	542	3.00E-65	N6-adenine-specific methylase	L
cpu_22200	2033	0	RecG-like helicase	LK
cpu_22210	188	2.00E-21	Ribosomal protein L28	J
cpu_22220	860	3.00E-92	Disulfide bond chaperones of the HSP33 family	O
cpu_22230	536	5.00E-27	Predicted phosphoesterase	R
cpu_22240	458	3.00E-80	Riboflavin synthase beta-chain	H
cpu_22250	1208	5.00E-124	3,4-dihydroxy-2-butanone 4-phosphate synthase	H
cpu_22260	641	3.00E-94	Riboflavin synthase alpha chain	H
cpu_22270	1076	2.00E-81	Pyrimidine deaminase	H
cpu_22280	665	3.00E-108	Pentose-5-phosphate-3-epimerase	G
cpu_22290	872	1.00E-116	Predicted GTPases	R
cpu_22300	1832	6.00E-55	Serine/threonine protein kinase	RTKL
cpu_22310	743	1.00E-69	Serine/threonine protein phosphatase	T
cpu_22320	1028	3.00E-148	Predicted Fe-S-cluster redox enzyme	R
cpu_22330	1322	3.00E-119	tRNA and rRNA cytosine-C5-methylases	J
cpu_22340	491	4.00E-24	Uncharacterized conserved protein	S
cpu_22350	941	5.00E-137	Methionyl-tRNA formyltransferase	J

Table 2-1-1 Continued

cpu_22360	458	9.00E-68	N-formylmethionyl-tRNA deformylase	J
cpu_22370	2156	0	Primosomal protein N' (replication factor Y) - superfamily II helicase	L
cpu_22380	1199	7.00E-144	Phosphopantothenoylcysteine synthetase/decarboxylase	H
cpu_22390	206	5.00E-11	DNA-directed RNA polymerase, subunit K/omega	K
cpu_22400	614	2.00E-94	Guanylate kinase	F
cpu_22410	263	8.00E-48	Uncharacterized protein conserved in bacteria	S
cpu_22420	866	7.00E-93	Uncharacterized stress-induced protein	S
cpu_22430	1178	1.00E-147	Aspartate/tyrosine/aromatic aminotransferase	E
cpu_22440	1172	2.00E-150	Aspartate/tyrosine/aromatic aminotransferase	E
cpu_22450	824	4.00E-117	Diaminopimelate epimerase	E
cpu_22460	1721	3.00E-111	Predicted RNA-binding protein homologous to eukaryotic snRNP	K
cpu_22470	566	9.00E-51	Orotate phosphoribosyltransferase	F
cpu_22480	722	1.00E-82	Orotidine-5'-phosphate decarboxylase	F
cpu_22490	920	1.00E-93	Dihydroorotate dehydrogenase	F
cpu_22500	770	1.00E-39	2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases	HC
cpu_22510	3194	0	Carbamoylphosphate synthase large subunit (split gene in MJ)	EF
cpu_22520	1067	0	Carbamoylphosphate synthase small subunit	EF
cpu_22530	1277	2.00E-162	Dihydroorotase and related cyclic amidohydrolases	F
cpu_22540	935	1.00E-135	Aspartate carbamoyltransferase, catalytic chain	F
cpu_22550	551	2.00E-84	Pyrimidine operon attenuation protein/uracil phosphoribosyltransferase	F
cpu_22560	461	5.00E-13	Uncharacterized conserved protein	S
cpu_22570	920	9.00E-128	Pseudouridylate synthases, 23S RNA-specific	J
cpu_22580	434	8.00E-35	Lipoprotein signal peptidase	MU
cpu_22590	650	2.00E-22	DnaK suppressor protein	T
cpu_22610	1358	0	Uncharacterized conserved protein	S
cpu_22620	275	3.00E-31	ACT domain-containing protein	T
cpu_22630	875	2.00E-94	Preprotein translocase subunit SecF	U
cpu_22640	1226	2.00E-132	Preprotein translocase subunit SecD	U
cpu_22650	659	1.00E-13	Uncharacterized conserved protein	S
cpu_22660	278	3.00E-24	Preprotein translocase subunit YajC	U
cpu_22670	1115	0	Queuine/archaeosine tRNA-ribosyltransferase	J
cpu_22680	1016	0	S-adenosylmethionine:tRNA-ribosyltransferase-isomerase	J
cpu_22690	1415	4.00E-58	Sporulation protein and related proteins	D
cpu_22710	641	3.00E-13	DNA-directed RNA polymerase specialized sigma subunit	K
cpu_22730	536	6.00E-50	Uncharacterized protein conserved in bacteria	S
cpu_22740	1019	0	Holliday junction resolvasome, helicase subunit	L
cpu_22750	590	4.00E-52	Holliday junction resolvasome, DNA-binding subunit	L
cpu_22760	476	1.00E-53	Holliday junction resolvasome, endonuclease subunit	L
cpu_22770	734	4.00E-126	Uncharacterized conserved protein	S
cpu_22780	725	6.00E-90	NAD synthase	H
cpu_22790	935	1.00E-60	Transcriptional regulator	K
cpu_22800	1586	3.00E-178	Predicted kinase related to dihydroxyacetone kinase	R
cpu_22810	836	2.00E-69	Uncharacterized protein conserved in bacteria	S
cpu_22830	695	1.00E-87	16S rRNA uridine-516 pseudouridylate synthase	J
cpu_22840	941	5.00E-26	Soluble lytic murein transglycosylase and related regulatory proteins	M
cpu_22850	917	2.00E-99	Uncharacterized ABC-type transport system, permease component	R
cpu_22860	1109	2.00E-92	ABC-type uncharacterized transport system, permease component	R
cpu_22870	1526	0	ABC-type uncharacterized transport systems, ATPase components	R
cpu_22880	1067	6.00E-107	Uncharacterized ABC-type transport system	R
cpu_22890	995	1.00E-125	Hydrogenase maturation factor	O
cpu_22900	1076	1.00E-148	Hydrogenase maturation factor	O
cpu_22910	215	2.00E-18	Hydrogenase maturation factor	O
cpu_22920	2264	0	Hydrogenase maturation factor	O
cpu_22930	662	6.00E-65	Ni2+-binding GTPase	OK
cpu_22940	341	5.00E-30	Zn finger protein HypA/HybF (possibly regulating hydrogenase expression)	R
cpu_22950	461	5.00E-29	Ni,Fe-hydrogenase maturation factor	C
cpu_22960	632	9.00E-20	Thiosulfate reductase cytochrome B subunit (membrane anchoring protein)	C

Table 2-1-1 Continued

cpu_22970	1427	0	Ni,Fe-hydrogenase I large subunit	C
cpu_22980	1064	2.00E-147	Ni,Fe-hydrogenase I small subunit	C
cpu_22990	2189	2.00E-149	Response regulator	T
cpu_23000	1217	5.00E-30	Predicted GTPases	R
cpu_23010	1376	2.00E-67	Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes	HR
cpu_23030	686	2.00E-109	Deoxyribose-phosphate aldolase	F
cpu_23040	398	1.00E-58	Cytidine deaminase	F
cpu_23050	1328	0	Thymidine phosphorylase	F
cpu_23060	632	5.00E-62	Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases	G
cpu_23070	1034	0	Predicted translation initiation factor 2B subunit, eIF-2B alpha/beta/delta family	J
cpu_23080	256	4.00E-23	Electron transfer flavoprotein, alpha subunit	C
cpu_23090	1298	3.00E-118	Acetyl-CoA hydrolase	C
cpu_23100	1145	9.00E-160	Acyl-CoA dehydrogenases	I
cpu_23110	1085	5.00E-116	6-phosphofructokinase	G
cpu_23130	1643	3.00E-99	Predicted ATPase of the ABC class	R
cpu_23140	938	3.00E-110	D-alanine-D-alanine ligase and related ATP-grasp enzymes	M
cpu_23160	821	1.00E-23	Multimeric flavodoxin WrbA	R
cpu_23170	2300	5.00E-47	Serine/threonine protein kinase	RTKL
cpu_23180	812	3.00E-39	Purine nucleoside phosphorylase	F
cpu_23190	707	5.00E-11	Trans-aconitate methyltransferase	R
cpu_23200	623	3.00E-74	Adenylate kinase and related kinases	F
cpu_23220	1364	0	Gamma-aminobutyrate permease and related permeases	E
cpu_23230	893	3.00E-69	L-serine deaminase	E
cpu_23240	656	2.00E-60	L-serine deaminase	E
cpu_23250	1229	7.00E-155	Threonine dehydratase	E
cpu_23260	377	2.00E-51	Putative translation initiation inhibitor, yjgF family	J
cpu_23270	674	9.00E-70	Uncharacterized protein conserved in bacteria	S
cpu_23280	809	1.00E-65	Lipoate-protein ligase A	H
cpu_23300	167	6.00E-12	Threonine dehydrogenase and related Zn-dependent dehydrogenases	ER
cpu_23310	732	6.00E-12	Transposase and inactivated derivatives	L
cpu_23320	308	1.00E-56	Ribosomal protein S10	J
cpu_23330	629	2.00E-100	Ribosomal protein L3	J
cpu_23340	620	8.00E-83	Ribosomal protein L4	J
cpu_23350	287	6.00E-36	Ribosomal protein L23	J
cpu_23360	284	5.00E-53	Ribosomal protein S19	J
cpu_23370	341	5.00E-52	Ribosomal protein L22	J
cpu_23380	668	3.00E-108	Ribosomal protein S3	J
cpu_23390	431	3.00E-75	Ribosomal protein L16/L10E	J
cpu_23400	209	3.00E-20	Ribosomal protein L29	J
cpu_23410	296	9.00E-33	Ribosomal protein S17	J
cpu_23420	368	2.00E-76	Ribosomal protein L14	J
cpu_23430	323	4.00E-34	Ribosomal protein L24	J
cpu_23440	539	8.00E-102	Ribosomal protein L5	J
cpu_23450	398	6.00E-78	Ribosomal protein S8	J
cpu_23460	548	6.00E-87	Ribosomal protein L6P/L9E	J
cpu_23470	368	8.00E-55	Ribosomal protein L18	J
cpu_23480	503	5.00E-83	Ribosomal protein S5	J
cpu_23490	182	2.00E-18	Ribosomal protein L30/L7E	J
cpu_23500	440	4.00E-50	Ribosomal protein L15	J
cpu_23510	1259	6.00E-159	Preprotein translocase subunit SecY	U
cpu_23520	746	7.00E-123	Methionine aminopeptidase	J
cpu_23540	218	1.00E-38	Translation initiation factor 1 (IF-1)	J
cpu_23550	371	1.00E-64	Ribosomal protein S13	J
cpu_23560	386	2.00E-68	Ribosomal protein S11	J
cpu_23570	626	3.00E-81	Ribosomal protein S4 and related proteins	J
cpu_23580	941	1.00E-137	DNA-directed RNA polymerase, alpha subunit/40 kD subunit	K
cpu_23590	338	8.00E-52	Ribosomal protein L17	J

Table 2-1-1 Continued

cpu_23600	767	5.00E-88	ABC-type cobalt transport system, ATPase component	P
cpu_23610	839	2.00E-80	ABC-type cobalt transport system, ATPase component	P
cpu_23620	812	4.00E-44	ABC-type cobalt transport system, permease component CbiQ	P
cpu_23630	722	4.00E-95	Pseudouridylate synthase	J
cpu_23640	1004	9.00E-35	Ca2+/Na+ antiporter	P
cpu_23650	431	8.00E-73	Ribosomal protein L13	J
cpu_23660	392	7.00E-70	Ribosomal protein S9	J
cpu_23670	515	4.00E-15	Uncharacterized conserved protein	S
cpu_23680	725	1.00E-53	N-acetylmuramoyl-L-alanine amidase	M
cpu_23690	1922	1.00E-118	Predicted membrane protein	S
cpu_23700	1400	2.00E-53	Uncharacterized conserved protein	S
cpu_23710	992	4.00E-85	Pyruvate-formate lyase-activating enzyme	O
cpu_23720	1070	4.00E-08	Arabinose efflux permease	G
cpu_23730	1016	1.00E-138	Acetylglutamate semialdehyde dehydrogenase	E
cpu_23740	1199	0	N-acetylglutamate synthase (N-acetylornithine aminotransferase)	E
cpu_23750	884	1.00E-119	Acetylglutamate kinase	E
cpu_23760	1196	0	Ornithine/acetylornithine aminotransferase	E
cpu_23770	977	0	Ornithine carbamoyltransferase	E
cpu_23780	1205	0	Argininosuccinate synthase	E
cpu_23790	1373	0	Argininosuccinate lyase	E
cpu_23820	791	4.00E-77	Enoyl-CoA hydratase/carnithine racemase	I
cpu_23830	908	1.00E-28	Predicted membrane protein	S
cpu_23850	473	1.00E-11	Uncharacterized conserved protein	S
cpu_23920	629	1.00E-28	Response regulator of citrate/malate metabolism	KT
cpu_23930	1355	3.00E-33	Signal transduction histidine kinase	T
cpu_23950	461	1.00E-87	Predicted rRNA methylase (SpoU class)	J
cpu_23960	701	5.00E-99	Predicted EndoIII-related endonuclease	L
cpu_23970	188	8.00E-14	Ferredoxin	C
cpu_23980	917	3.00E-76	Pseudouridylate synthases, 23S RNA-specific	J
cpu_23990	1025	1.00E-64	Predicted oxidoreductases of the aldo/keto reductase family	R
cpu_24000	599	4.00E-46	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	K
cpu_24010	440	3.00E-14	Predicted membrane protein	S
cpu_24020	218	3.00E-16	Predicted heme/steroid binding protein	R
cpu_24030	728	4.00E-50	Hydantoin racemase	E
cpu_24040	1550	1.00E-113	N-methylhydantoinase A/acetone carboxylase, beta subunit	EQ
cpu_24050	1082	8.00E-175	Uncharacterized conserved protein	S
cpu_24060	1325	1.00E-50	Purine-cytosine permease and related proteins	F
cpu_24070	1568	4.00E-149	Transcriptional regulator	KT
cpu_24080	1109	5.00E-73	Uncharacterized conserved protein	S
cpu_24100	1328	5.00E-121	SAM-dependent methyltransferases related to tRNA (uracil-5-)-methyltransferase	J
cpu_24110	296	8.00E-08	Predicted nucleotidyltransferases	R
cpu_24130	692	9.00E-08	Predicted metal-dependent membrane protease	R
cpu_24140	998	2.00E-165	Isocitrate/isopropylmalate dehydrogenase	CE
cpu_24150	497	5.00E-61	3-isopropylmalate dehydratase small subunit	E
cpu_24160	1253	0	3-isopropylmalate dehydratase large subunit	E
cpu_24170	1151	2.00E-139	Isopropylmalate/homocitrate/citramalate synthases	E
cpu_24180	1445	0	Asp-tRNAAsn/Glu-tRNAGln amidotransferase B subunit (PET112 homolog)	J
cpu_24190	1463	0	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases	J
cpu_24200	284	5.00E-33	Asp-tRNAAsn/Glu-tRNAGln amidotransferase C subunit	J
cpu_24210	1994	0	NAD-dependent DNA ligase (contains BRCT domain type II)	L
cpu_24220	2156	0	Superfamily I DNA and RNA helicases	L
cpu_24230	329	2.00E-16	Predicted transcriptional regulators	K
cpu_24240	953	1.00E-75	Predicted permeases	R
cpu_24310	560	3.00E-18	Cytochrome c biogenesis protein	O
cpu_24340	638	2.00E-25	Predicted membrane protein	S
cpu_24360	236	2.00E-18	Arsenite efflux pump ACR3 and related permeases	P
cpu_24370	302	5.00E-26	Arsenite efflux pump ACR3 and related permeases	P

Table 2-1-1 Continued

cpu_24380	377	2.00E-12	Predicted transcriptional regulators	K
cpu_24390	611	2.00E-63	Phosphoribosyl-AMP cyclohydrolase	E
cpu_24400	779	2.00E-161	Imidazoleglycerol-phosphate synthase	E
cpu_24410	731	3.00E-96	ProFAR isomerase	E
cpu_24420	599	2.00E-99	Glutamine amidotransferase	E
cpu_24430	578	2.00E-100	Imidazoleglycerol-phosphate dehydratase	E
cpu_24440	1049	1.00E-113	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase	E
cpu_24450	1277	0	Histidinol dehydrogenase	E
cpu_24460	665	2.00E-63	ATP phosphoribosyltransferase	E
cpu_24470	977	6.00E-68	Histidyl-tRNA synthetase	J
cpu_24480	308	7.00E-44	Uncharacterized protein conserved in bacteria	S
cpu_24490	1016	2.00E-34	Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase	J
cpu_24500	998	2.00E-25	Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase	J
cpu_24510	1268	0	Phosphoribosylamine-glycine ligase	F
cpu_24520	1514	0	AICAR transformylase/IMP cyclohydrolase PurH	F
cpu_24530	629	2.00E-107	Folate-dependent phosphoribosylglycinamide formyltransferase PurN	F
cpu_24540	1031	9.00E-178	Phosphoribosylaminoimidazole (AIR) synthetase	F
cpu_24550	1358	0	Glutamine phosphoribosylpyrophosphate amidotransferase	F
cpu_24560	2183	0	FGAM synthase, synthetase domain	F
cpu_24570	704	3.00E-126	FGAM synthase, glutamine amidotransferase domain	F
cpu_24580	245	4.00E-27	FGAM synthase, PurS component	F
cpu_24590	707	2.00E-113	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase	F
cpu_24600	1292	0	Adenylosuccinate lyase	F
cpu_24610	488	5.00E-76	Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase	F
cpu_24620	1529	0	GMP synthase, PP-ATPase domain/subunit	F
cpu_24630	539	1.00E-93	Hypoxanthine-guanine phosphoribosyltransferase	F
cpu_24650	554	3.00E-65	Uncharacterized conserved protein	S
cpu_24660	320	1.00E-13	Predicted pyrophosphatase	R
cpu_24670	1262	2.00E-42	Uncharacterized conserved protein	S
cpu_24700	1004	6.00E-12	Nitrate/TMAO reductases, membrane-bound tetraheme cytochrome c subunit	C
cpu_24710	1328	6.00E-85	Sugar transferases involved in lipopolysaccharide synthesis	M
cpu_24720	947	3.00E-84	Nucleoside-diphosphate-sugar epimerases	MG
cpu_24730	980	8.00E-12	Predicted glycosyltransferases	R
cpu_24750	1094	2.00E-20	Uncharacterized protein conserved in bacteria	S
cpu_24760	1397	4.00E-153	Mannose-1-phosphate guanylyltransferase	M
cpu_24770	1382	1.00E-132	Phosphomannomutase	G
cpu_24780	944	8.00E-69	Nucleoside-diphosphate-sugar epimerases	MG
cpu_24790	1052	0	GDP-D-mannose dehydratase	M
cpu_24800	920	7.00E-08	Predicted glycosyltransferases	R
cpu_24820	140	1.00E-08	DNA replication protein	L
cpu_24870	1265	4.00E-96	Formate-dependent nitrite reductase, periplasmic cytochrome c552 subunit	P
cpu_24890	695	8.00E-18	Predicted hydrolase (HAD superfamily)	R
cpu_24900	1394	9.00E-94	Glucose-6-phosphate isomerase	G
cpu_24910	1220	3.00E-12	Arabinose efflux permease	G
cpu_24920	641	4.00E-69	Undecaprenyl pyrophosphate synthase	I
cpu_24930	692	4.00E-34	Predicted DNA-binding proteins	R
cpu_24970	617	2.00E-44	Thiamine pyrophosphokinase	H
cpu_24980	515	2.00E-35	Predicted membrane protein	S
cpu_24990	1292	4.00E-160	Superfamily II DNA and RNA helicases	LKJ
cpu_25000	707	2.00E-11	2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase	H
cpu_25030	1007	6.00E-12	Uncharacterized conserved protein	S
cpu_25040	656	3.00E-14	Tfp pilus assembly protein PilF	NU
cpu_25070	1103	8.00E-42	Methyl-accepting chemotaxis protein	NT
cpu_25080	1106	1.00E-44	Methyl-accepting chemotaxis protein	NT
cpu_25090	449	3.00E-88	Nucleoside diphosphate kinase	F
cpu_25100	1820	0	ATP-dependent Zn proteases	O
cpu_25110	1304	4.00E-49	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control	D

Table 2-1-1 Continued

cpu_25120	2429	1.00E-22	Serine phosphatase RsbU, regulator of sigma subunit	TK
cpu_25130	878	6.00E-69	Exopolyphosphatase	FP
cpu_25170	254	3.00E-15	Sporulation protein and related proteins	D
cpu_25180	269	3.00E-39	Bacterial nucleoid DNA-binding protein	L
cpu_25190	1412	8.00E-152	Protein containing tetrapyrrole methyltransferase	R
cpu_25240	800	6.00E-35	Type II secretory pathway, component ExeA (predicted ATPase)	U
cpu_25320	380	4.00E-38	Ribosomal protein L7/L12	J
cpu_25330	530	3.00E-49	Ribosomal protein L10	J
cpu_25340	695	7.00E-117	Ribosomal protein L1	J
cpu_25350	425	3.00E-79	Ribosomal protein L11	J
cpu_25360	521	9.00E-70	Transcription antiterminator	K
cpu_25370	263	6.00E-10	Preprotein translocase subunit SecE	U
cpu_25400	266	3.00E-21	Electron transfer flavoprotein, beta subunit	C
cpu_25420	503	1.00E-40	Predicted phosphoesterase	R
cpu_25430	1189	8.00E-51	Transposase and inactivated derivatives	L
cpu_25450	223	3.00E-08	Transposase and inactivated derivatives	L
cpu_25490	1202	0	GTPases - translation elongation factors	J
cpu_25530	800	4.00E-45	Type II secretory pathway, component ExeA (predicted ATPase)	U
cpu_25540	305	1.00E-32	DNA replication protein	L
cpu_25560	332	5.00E-29	Predicted ATPase with chaperone activity	O
cpu_25640	668	3.00E-26	Predicted integral membrane protein	S
cpu_25650	803	5.00E-34	Signal transduction histidine kinase regulating citrate/malate metabolism	T
cpu_25660	761	6.00E-51	Response regulator of the LytR/AlgR family	KT
cpu_25670	332	7.00E-30	Dehydrogenases with different specificities	IQR
cpu_25680	1406	1.00E-84	ABC-type multidrug transport system, ATPase component	V
cpu_25710	1037	1.00E-29	Membrane-fusion protein	M
cpu_25720	698	2.00E-120	ABC-type antimicrobial peptide transport system, ATPase component	V
cpu_25730	1190	3.00E-44	ABC-type antimicrobial peptide transport system, permease component	V
cpu_25740	1184	2.00E-92	ABC-type branched-chain amino acid transport systems, periplasmic component	E

General features of two *Carboxydocella* strains

The draft genomes of strains JDF658 and ULO1 were assembled into 270 and 180 contigs with a total length of approximately 2.60 Mbp and 2.70 Mbp, respectively. These draft genomes showed average G+C contents of 49.2% and 48.8%, containing 2,731 and 2,804 predicted ORFs, respectively.

Phylogenetic and comparative analysis of CODHs

I identified CODH genes in the two strains and compared their genomic context with their relatives and other carboxydotrophs to predict CODH function (Techtmann et al. 2012). In both strains, I identified putative six *cooS* genes. Of these, three distinct *cooS* that encoded conserved amino acid sequences in active centers (Fig.

2-1-2). The genomic contexts of the three *cooS* differed from each other, but well conserved to counterparts among our strains and *Carboxydocella sporoproducens* (IMG Genome ID; 2568526009) that were isolated from geologically distinct volcanic regions. Two of the three *cooS* were identified in a well-characterized genomic context: an ECH gene cluster and an ACS gene cluster. Complexes of these gene products are considered to be responsible for energy conversion by hydrogen production via hydrogenogenic CO metabolism and carbon fixation via the Wood–Ljungdahl pathway, respectively (Techtmann et al. 2012; Wu et al. 2005). The final *cooS* gene was identified in a genomic context of unknown function. This CODH was phylogenetically distinct from well-characterized CODHs, branching deeply from a clade including still less-understood CODHs that lack some conserved amino acids in active centers (Inoue et al. 2013). The third CODH is potentially active, and physiological study could reveal its function in the CO metabolism of *Carboxydocella* spp..

Collectively, four draft genome sequences (two species in the genus *Carboxydothermus* and two strains in the genus *Carboxydocella*) provided genomic insights into hydrogenogenic CO metabolism. In particular, the CODH−ECH gene cluster was found in the genomes of *C. islandicus* and two *Carboxydocella* strains as in other hydrogenogenic carboxydotrophs. However, the genome of *C. pertinax* lacked the CODH-I catalytic subunit gene in the gene cluster, albeit their hydrogenogenic carboxydotrophy. These results suggested that *C. pertinax* performed alternative hydrogenogenic CO metabolism independent of the CODH-I gene product. To understand hydrogenogenic CO metabolism in *C. pertinax*, I performed comparative genome analysis in the genus *Carboxydothermus* in next section.

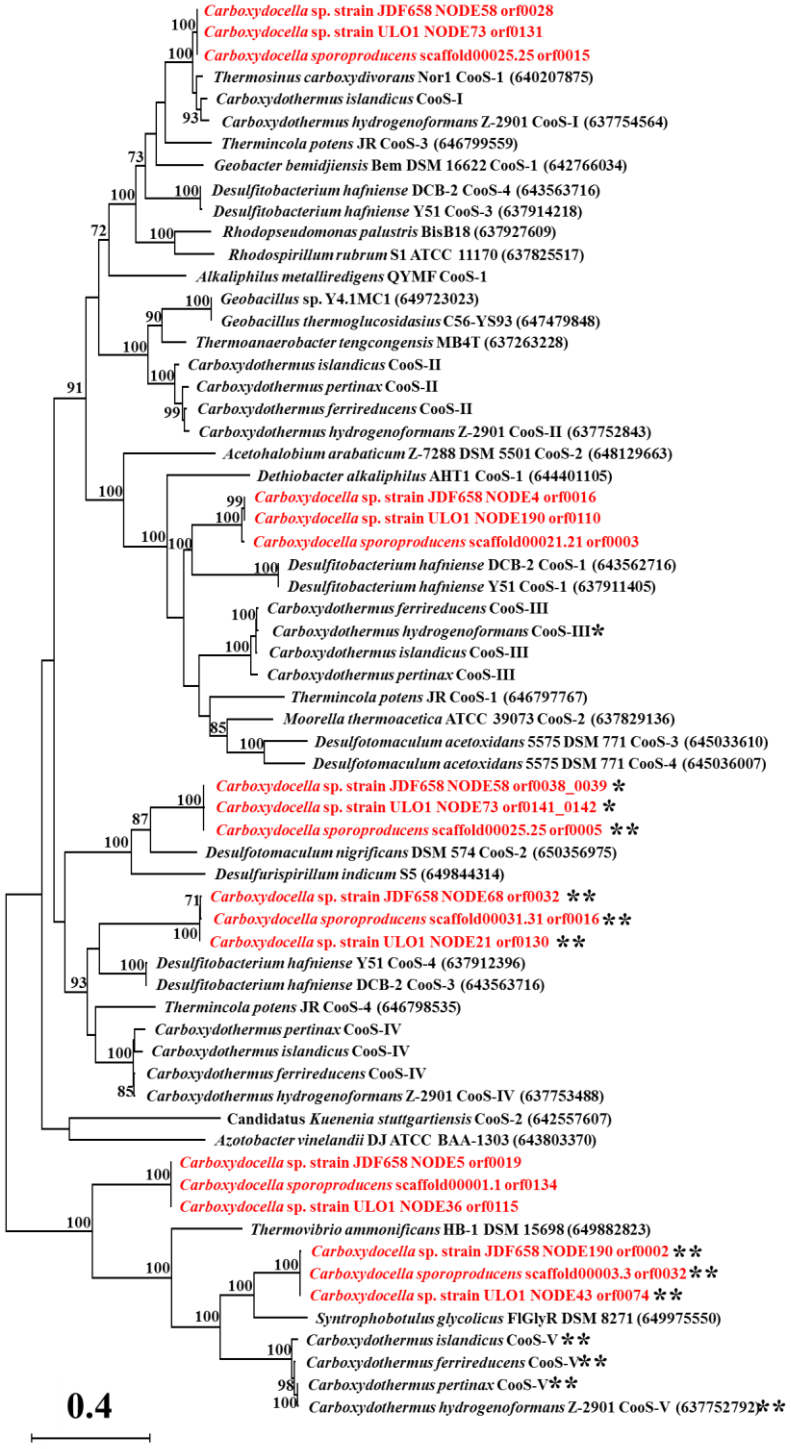


Fig. 2-1-2 Maximum likelihood phylogenetic tree based on the amino acid sequence of *cooS* genes

cooS genes from *Carboxydocella* species are indicated in red and bold font. Each accession number is indicated in parentheses. Only bootstrap support values (out of 100 runs) equal to or greater than 70 are shown.

*, frameshifted; **, amino acid sequences ware not conserved in its active site.

2-2 Comparative genome analysis of carboxydotrophs in the genus *Carboxydothermus*

Introduction

All members in the genus *Carboxydothermus* are strictly anaerobic, thermophilic, and carboxydotrophic bacteria (Svetlichny et al. 1991; Slobodkin et al. 2006; Slepova et al. 2009; Novikov et al. 2011; Yoneda et al. 2012). The genus is characterized by varied energy conservation strategies via CO oxidation. Except for *C. ferrireducens*, the other *Carboxydothermus* species are hydrogenogenic carboxydotrophs. Further, *C. pertinax* performs hydrogenogenic CO metabolism with iron reduction, in which H_2 production corresponds to approximately 69% of CO uptake (Yoneda et al. 2012), whereas in other hydrogenogenic members CO utilization results in nearly equimolar H_2 release. These physiological differences about CO metabolism strongly suggest distinct strategies for energy conservation via CO metabolism.

In addition to *C. hydrogenoformans* who possesses five *cooS* genes (*cooS-I* to -*V*) in the genome (Wu et al. 2005), I have sequenced the genomes of *C. islandicus* and *C. pertinax* (Chapter2 section1), and found that whereas *C. islandicus* also possessed five *cooS* genes (corresponding to *cooS-I* to -*V* of *C. hydrogenoformans*), *C. pertinax* contained only four *cooS* genes (corresponding to *cooS-II* to -*V* of *C. hydrogenoformans*). This suggests an alternative hydrogenogenic CO metabolism independent of the *cooS-I* gene product. In the present study, I investigated energy conservation via alternative hydrogenogenic CO metabolism in *C. pertinax*, using comparative genome analysis of available *Carboxydothermus* genomes, transcriptional analysis, and a cultivation study with thiosulfate as electron donor under 100% CO.

Materials and methods

Available genome databases for comparative genome analysis

For comparative genome analysis of four members of the *Carboxydothermus* genus, I used *C. islandicus* and *C. pertinax* draft genomes. In addition, I downloaded the *C. hydrogenoformans* Z-2911 complete genome (NC_007503.1) (Wu et al. 2005) and the *C. ferrireducens* DSM 11255 draft genome (NZ_ ATYG00000000) from the Reference Sequence Database (RefSeq) of the NCBI (Pruitt et al. 2007).

Comparative analysis of the available genomes in the genus *Carboxydothermus*

To assign Clusters of Orthologous Groups (COG) categories, the protein sequences from four *Carboxydothermus* species were annotated using BLASTp search (Altschul et al. 1990; Altschul et al. 1997) with an e-value of $1e^{-5}$ at an effective database size of 10^7 against the COG database (Tatusov 2000). Orthologous genes among the four *Carboxydothermus* species were identified using all-versus-all BLASTp search with an e-value of $1e^{-5}$ at an effective database size of 10^7 (Alcaraz et al. 2010). The orthologous genes shared by all analyzed genomes were defined as “core genes”, whereas the genes present in only one analyzed genome were defined as “unique genes”.

For prediction of metabolic pathways in the four *Carboxydothermus* species, BLAST KEGG Orthology And Links Annotation (KOALA) (Kanehisa et al. 2016) were used with manual curation. The manual curation process included manual verification of each gene’s annotation using Rapid Annotations using Subsystem Technology (RAST) (Aziz et al. 2008).

For comprehensive genome comparison, I generated synteny plots between the complete genome of *C. hydrogenoformans* and the other three draft genomes using

MUMmer 3.0 (Delcher et al. 2003). EasyFig was used to compare the genome comparison of sequences around the inversion region in *C. pertinax* and *C. hydrogenoformans* (Sullivan et al. 2011). I performed in silico subtractive hybridization analysis of the reference *C. hydrogenoformans* genome against *C. pertinax*, *C. islandicus*, and *C. ferrireducens* genomes, using BLASTp search with default parameters (Shao et al. 2010; Fukiya et al. 2004). The homology score (H-value) was calculated as follows: $H=i \times (lm/lq)$, where *i* was the level of identity of the region with the highest Bit score expressed at a frequency between 0 and 1, *lm* the length of the highest scoring matching sequence (including gaps), and *lq* the query length. H-values were expressed between 0 and 1. If the BLASTp e-value of the best-hit was ≥0.01, its H-value was defined as zero.

Phylogenies based on housekeeping genes and genomic similarity score (GSS).

In addition to phylogenetic analysis based on the 16S rRNA gene (Yoneda et al. 2012), phylogenetic analysis of the *Thermoanaerobacteraceae* members based on concatenated housekeeping genes and GSS was performed. Besides the four genomes in the genus *Carboxydothermus*, I collected 13 publicly available *Thermoanaerobacteraceae* genomes from NCBI and Integrated Microbial Genomes (Markowitz et al. 2006; Table 2-2-1).

As housekeeping genes, I used the amino acid sequences corresponding to the genes for ribosome recycling factor (*frr*), transcription elongation factor (*nusA*), 50S ribosomal protein L27 (*rpmA*), elongation factor Ts (*tsf*), and cytidine triphosphate synthetase (*pyrG*) from the genomes of the *Thermoanaerobacteraceae* species listed in Table 2-2-1. The concatenated sequences were aligned using MAFFT 7.215 (Katoh et al. 2013) and gap positions were removed automatically using trimAL1.4

(Capella-Gutierrez et al. 2009). Phylogenetic reconstructions were performed by the ML method using RAxML 8.2.4 (Stamatakis 2014) and visualized with MEGA 5.2 (Tamura et al. 2011). Robustness of the topology of the phylogenetic trees was evaluated by bootstrap analysis based on 100 runs.

To compute the similarity between genomes of *Thermoanaerobacteraceae* species (Table 2-2-1), I calculated GSS as previously described (Omae et al. 2017). The neighbor-joining tree was built using a GSS distance matrix (Alcaraz et al. 2010).

Table 2-2-1 Genome accession numbers for phylogenetic analysis

Organism	NCBI genome accesion	IMG genome accesion
Ammonifex degensii KC4	NC_013385	
Caldanaerobacter subterraneus subsp. *tengcongensis* MB4	NC_003869	
Carboxydothermus ferrireducens DSM 11255	NZ_ ATYG00000000	
Carboxydothermus hydrogenoformans Z-2911	NC_007503.1	
Carboxydothermus islandicus SET IS-9	BDJL01000000	
Carboxydothermus pertinax Ug1	BDJK01000000	
Moorella thermoacetica ATCC 39073	NC_007644	
Tepidanaerobacter acetatoxydans Re1	NC_019954	
Thermoacetogenium phaeum DSM 12270	NC_018870	
Thermoanaerobacter brockii subsp. *finnii* Ako-1	NC_014964	
Thermoanaerobacter italicus Ab9	NC_013921	
Thermoanaerobacter mathranii subsp. *mathranii* str. A3	NC_014209	
Thermoanaerobacter pseudethanolicus ATCC 33223	NC_010321	
Thermoanaerobacter sp. X513	NC_014538	
Thermoanaerobacter sp. X514	NC_010320	
Thermoanaerobacter wiegelii Rt8. B1	NC_015958	
Desulfovirgula thermocuniculi DSM 16036		2524023160

Comparative analysis of CODHs

For comparison of CODH gene cluster from *Carboxydothermus* species, EasyFig was used to generate the structural comparison of CODH-I–ECH gene clusters in the genus *Carboxydothermus* (Sullivan et al. 2011).

Cultivation study of *C. pertinax*

For RNA extraction, *C. pertinax* was grown on modified DSM medium 507 under 100% CO and 100% N_2 gas with 100 rotations per minutes. The gas phases of inoculated vessels were changed to 100% CO gas or 100% N_2 gas as control. Sodium thiosulfate (final concentration, 1 g/L) as terminal electron acceptor and pyruvate Na (final concentration, 2 g/L) as electron donor and carbon source were added to the medium in both gas phase conditions to collect enough cells for analysis. Growth was assessed by counting cell-derived fluorescent signals in samples using the S3e^{TM} Cell Sorter (Bio-Rad, Hercules, CA, USA). Cells were stained with SYBR Green I. CO consumption and H_2 production in the headspace were measured using a GC-2014 gas chromatographer (Shimadzu, Kyoto, Japan) equipped with a thermal conductivity detector and Shincarbon ST packed column (Shinwa Chemical Industries, Kyoto, Japan). Argon was used as the carrier gas. The *C. pertinax* frozen stock was kept in modified DSM medium 507 with thiosulfate.

To assess growth profiles of *C. pertinax* under 100% CO with thiosulfate as terminal electron acceptor, sodium thiosulfate was added to the medium (final concentration, 1 g /L). To clarify an effect of CO as electron donor and carbon source, pyruvate Na was not added in this case. Carboxydotrophic growth was assessed as described above. CO consumption and H_2 production in the headspace were measured as described above. Concentration of dissolved hydrogen sulfide was determined using

a colorimetric method (Cord-Ruwisch et al. 1985) with the following modifications: sample volume was decreased from 1.95 mL to 1 mL and acidic copper regent volume was increased from 0.05 mL to 1 mL. Absorbance of the resulting mixture was measured at 480 nm using an Ultraspec 2100 pro spectrophotometer (Amersham Biosciences, Little Chalfont, UK). The detection limit for dissolved hydrogen sulfide in culture was 50 μM. All cultivation study data are representative of mean values derived from at least triplicate biological determinations.

Reverse Transcription quantitative PCR (RT-qPCR)

Once the cells reached exponential phase (18 h after inoculation) (Fig. 2-2-1), they were collected on a 0.22-μm Durapore membrane filter (Millipore, Billerica, MA, USA). The filter was transferred to a tube containing 1 mL of RNA later (Qiagen, GmbH, Hilden, Germany) directly and the cell suspension was stored at -80 °C until use. Total RNA was isolated from 10 mL of the stored cell suspension using mirVana mRNA Isolation Kits (Invitrogen, Carlsbad, CA, USA) and contaminating DNA was removed using TURBO DNase (Invitrogen) according to the manufacturer's instructions. Then, total RNA was reverse transcribed to single strand cDNA (ss cDNA) using SuperScript III First-Strand Synthesis System (Invitrogen).

RT-qPCR primers were designed based on the gene sequences of *cooS-II* to *V* (encoding CODH-II to -V catalytic subunits), *cooA2F1* (encoding CooA and CooF in the CODH-II gene cluster), *cooLH* (encoding ECH large and small catalytic subunits in the CODH-I−ECH gene cluster), and *hyaAB* (encoding H_2-uptake Ni-Fe-hydrogenase large and small catalytic subunits) from the C. pertinax draft genome using Primer-BLAST (Ye et al. 2012; Table 2-2-2). The reaction mixture contained 12 μL of ss cDNA template, which was obtained as described above with 1.5 μL of SYBR®

Premix EX TaqTM II (Tli RNaseH Plus; TaKaRa, Kyoto, Japan), according to the manufacturer's instructions. PCR amplification was performed using the Thermal Cycler Dice Real Time System Single (TaKaRa). The cycling program consisted of an initial denaturing step of 2 min at 95 °C, followed by 35 cycles of 15 s at 95 °C, 15 s at 64 °C, 30 s at 72 °C, a final thermal denaturing step to generate melting curves to verify amplification specificity. The qPCR standard curve of each targeted gene showed a log linear relationship when a 10-fold dilution series of PCR products (from 102 to 108 copies/μL) from *C. pertinax* genomic DNA was used as template. Efficiency and R^2 value of each standard curve ranged from 94.7% (*cooL*) to 155.2% (*cooS-IV*) and from 0.983 (*cooS-II*) to 1.000 (*cooA2L* and *hyaB*), respectively (Table 2-2-2). represent the mean value of at least triplicate biological determinations.

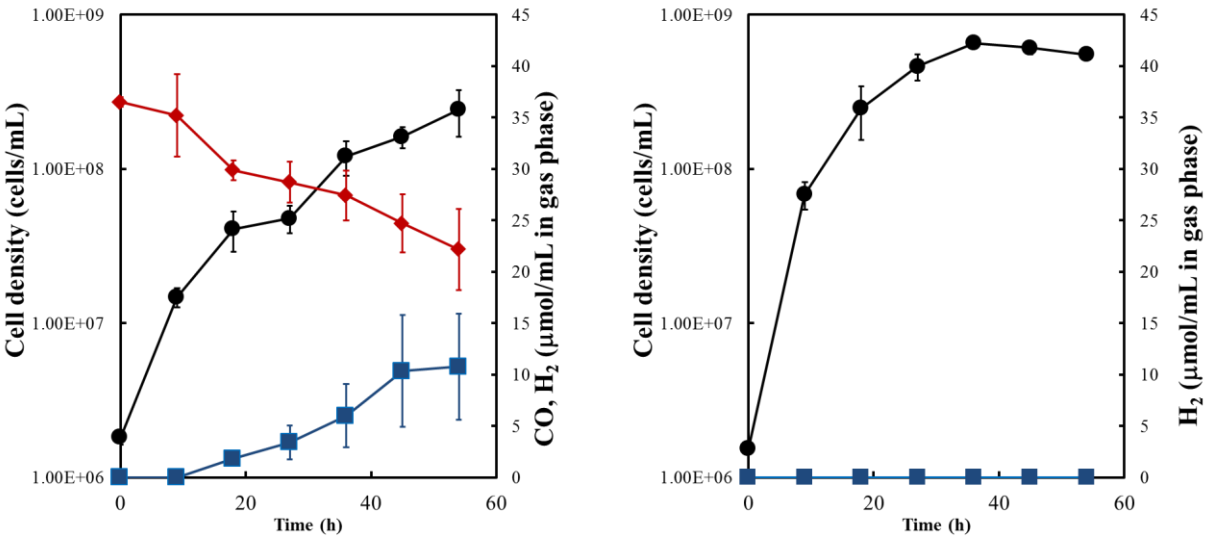


Fig. 2-2-1 Growth of *C. pertinax* under 100% CO (left) and 100% N_2 (right)
Different parameters are shown: black disks, cell density; red diamonds, CO concentration; blue squares, H_2 concentration. Plots represent means of at least three biological replicates. Error bars represent standard deviations.

All qPCR data for each standard curve represent the mean value of at least triplicate technical determinations. The detection limit for target genes transcript levels of mRNA was 1.00 × 102 copies/μL. The melting curve of all qPCR products had a single peak at approximately 82 °C to 84 °C (data not shown). Relative transcript levels were calculated using 16S rRNA gene transcripts as internal standard. All RT-qPCR data

Table 2-2-2 Primer sequences and efficiency of amplification for RT-qPCR

Target gene	Locus tag	Primer direction	Primer Sequence (5'-3')	Amplification efficiency (%)	R^2
cooS-II	cpu_14040	Forward	TCCGGTTGTAGCCTCAGCCC	133.2	0.983
		Reverse	AGGCCTATGGTTACGGCCCAGG		
cooS-III	cpu_09920	Forward	CCCTTCAATGTAGAAACTGCCCACG	111.4	0.993
		Reverse	CGGGGATGTGAACAGGTTTGCCA		
cooS-IV	cpu _04820	Forward	ACAGCGAAACTGCCGATACGGC	155.2	0.989
		Reverse	TGCTTCCACGCTAAAACCAGCCA		
cooS-V	cpu _16630	Forward	ACCGTTTGTCGGGGTGGCTTTA	103.8	0.998
		Reverse	TGCCGTAAATTCTAAGGCCTTTGGC		
cooA2	cpu _14050	Forward	CGGGTTGGTCCGGGTATTGGG	114.5	1.000
		Reverse	TGTACGCCCCGGTTGAGTGC		
cooF1	cpu _14030	Forward	AGCCTGCCCGTTTCACGCTA	125.4	0.996
		Reverse	GGGGCAGGAGGTAACGCAGA		
cooH	cpu _03730	Forward	GGCGGTAGCCAAGTCCGAGG	108.9	0.993
		Reverse	CCCGCCACTTAATGCGTTCCG		
cooL	cpu _03700	Forward	GGAACTGGCGACCACAGCCT	94.7	1.000
		Reverse	CCGGGCAGTAAGAGGCCCAG		
hyaA	cpu _22980	Forward	ATTGTCCGCGGCGCCAGTAT	107.4	0.998
		Reverse	CAGTTCACCCCGCCGTTCCA		
hyaB	cpu _22970	Forward	GCCGGCTGCGGTTAACCAAAAA	114.4	1.000
		Reverse	TCCGGGCACTATCCCGACCA		
16S rRNA	cpu _r00020	Forward	ACAGATGGGCCCGCGTCCTA	113.6	0.998
		Reverse	AGTCCCAGTGTGGCCGTCCA		

Results

Comparative genome analysis of four *Carboxydothermus* species

The genomes of all four *Carboxydothermus* species were similar in size, ranging from 2,386,596 bp for *C. islandicus* to 2,465,639 bp for *C. pertinax*, and in G+C content, ranging from 40.7% in *C. pertinax* to 42.0% in *C. islandicus* and *C. hydrogenoformans* (Table 2-2-3). The numbers of ORFs ranged from 2457 (*C. ferrireducens*) to 2620 (*C. hydrogenoformans*). Of these, 77.1% (*C. pertinax*), 75.7% (*C. islandicus*), 72.6% (*C. hydrogenoformans*), and 79.5% (*C. ferrireducens*) were assigned to COG categories (Fig. 2-2-2 and Table 2-2-3). The content of each COG category was similar across the four genomes. The four genomes shared 1816 core genes (69.3% to 73.9% of their total ORFs). Unique genes corresponded to 11.9%, 11.9%, 14.5%, and 9.0% of total ORFs in *C. pertinax*, *C. islandicus*, *C. hydrogenoformans*, and *C. ferrireducens*, respectively (Table 2-2-3).

Carboxydothermus species have been reported as facultative carboxydotrophs (Svetlichny et al. 1991; Slobodkin et al. 2006; Slepova et al. 2009; Yoneda et al. 2012). They can grow heterotrophically on lactate or pyruvate with available electron acceptors. *C. pertinax* and *C. ferrireducens* can also grow on glucose with available electron acceptors. Metabolic pathway analysis revealed that *C. pertinax* and *C. ferrireducens* contained genes for the entire glycolysis pathway, whereas *C. islandicus* and *C. hydrogenoformans* lacked phosphoglucomutase (EC 5.4.2.2) (Fig. 2-2-3A). *C. pertinax* and *C. ferrireducens* possessed only fructose phosphotransferase system permease (EC 2.7.1.202) and phosphopyruvate-protein phosphotransferase (EC 2.7.3.9), whereas *C. islandicus* and *C. hydrogenoformans* lost the entire phosphotransferase system (Fig. 2-2-3B). Furthermore, all of them lacked malate dehydrogenase (EC 1.1.1.37), citrate

synthase (EC 2.3.3.1), and aconitase hydratase (EC 4.2.1.3) in their tricarboxylic acid cycle (Fig. 2-2-3C). Among other well-known carbon fixation pathways, the Wood-Ljungdahl pathway and the incomplete 3-hydroxypropionate cycle was found in all species (Fig. 2-2-3D). Of the two key enzymes in the 3-hydroxypropionate cycle, only acetyl-CoA carboxylase (EC 6.4.1.2) was conserved in all species, whereas propionyl-CoA carboxylase (EC 6.4.1.3) could not be detected. Genes encoding ribulose-1,5-bisphosphate carboxylase/oxygenase (EC 4.1.1.39), which is a key enzyme in the Calvin-Benson cycle, was not found in any of the four genomes.

Table 2-2-3 General features of the genomes from four *Carboxydothermus* species

	Value for:			
Parameter	*C. pertinax*	*C. islandicus*	*C. hydrogenoformans*	*C. ferrireducens*
Genome size (bp)	2,465,639	2,386,596	2,401,520	2,441,992
G + C content (%)	40.7	42	42	41.9
No. of:				
ORFs	2,577	2,480	2620	2457
rRNAs	4	11	12	6
rRNAs	46	48	49	47
No. (%) of genes in the COG	1987 (77.1)	1878 (75.7)	1901 (72.6)	1954 (79.5)
No. of contigs	96	142	1	49
No. of CODH genes	4	5	5	4
No. of unique genes	307	296	379	221
RefSeq ID	BDJK01000000	BDJL01000000	NC_007503.1	NZ_ATYG00000000

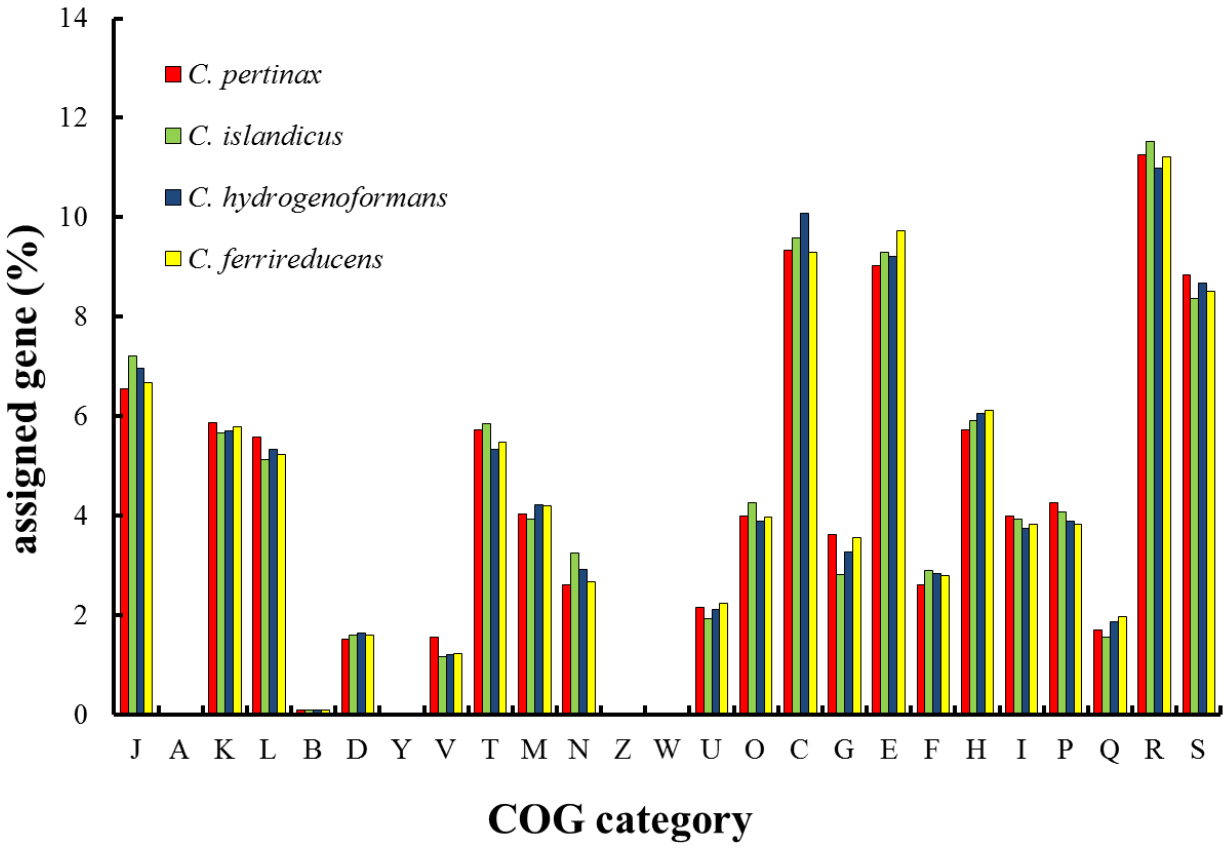
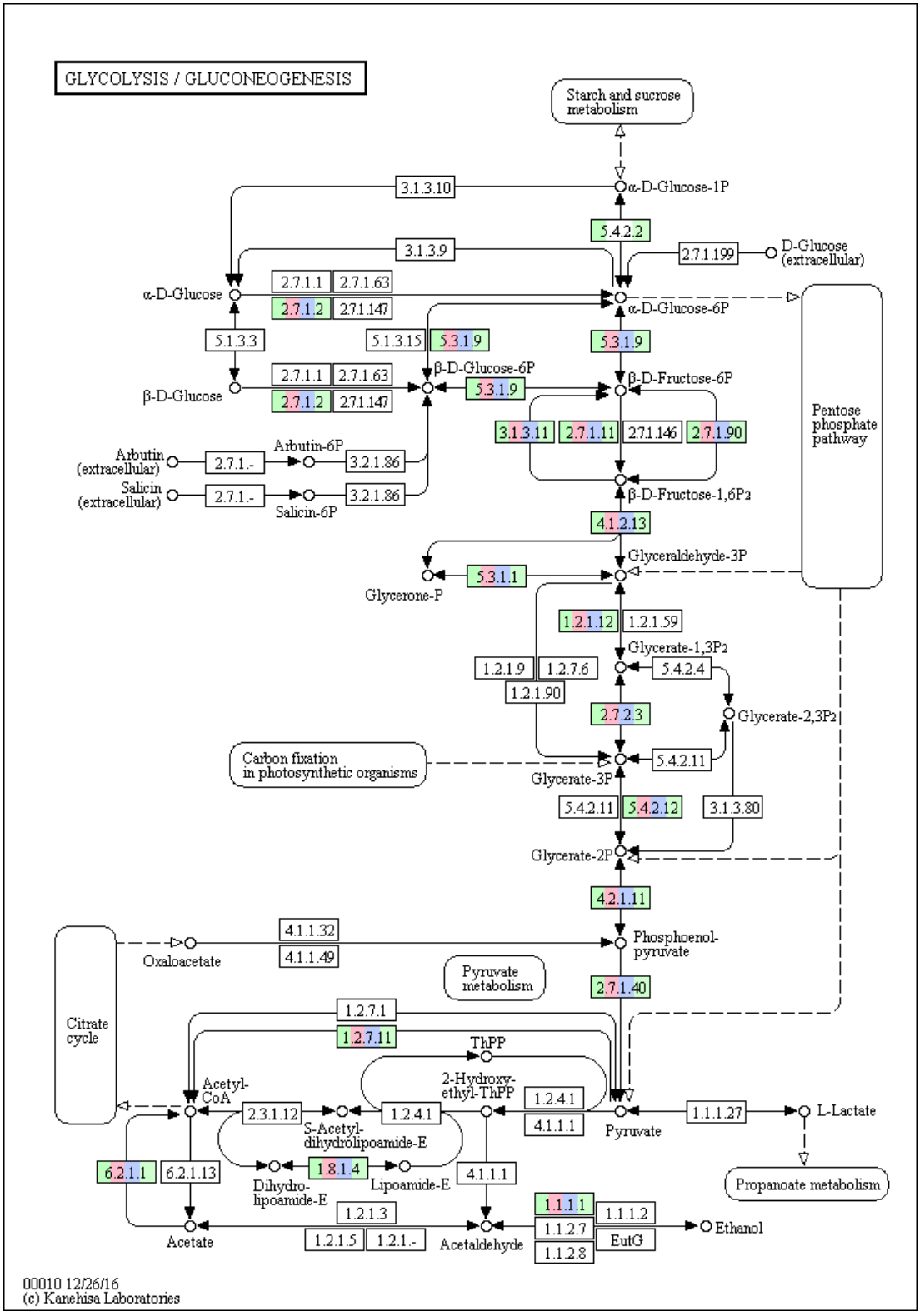


Fig. 2-2-2 Percentages of ORFs assigned to COG functional categories

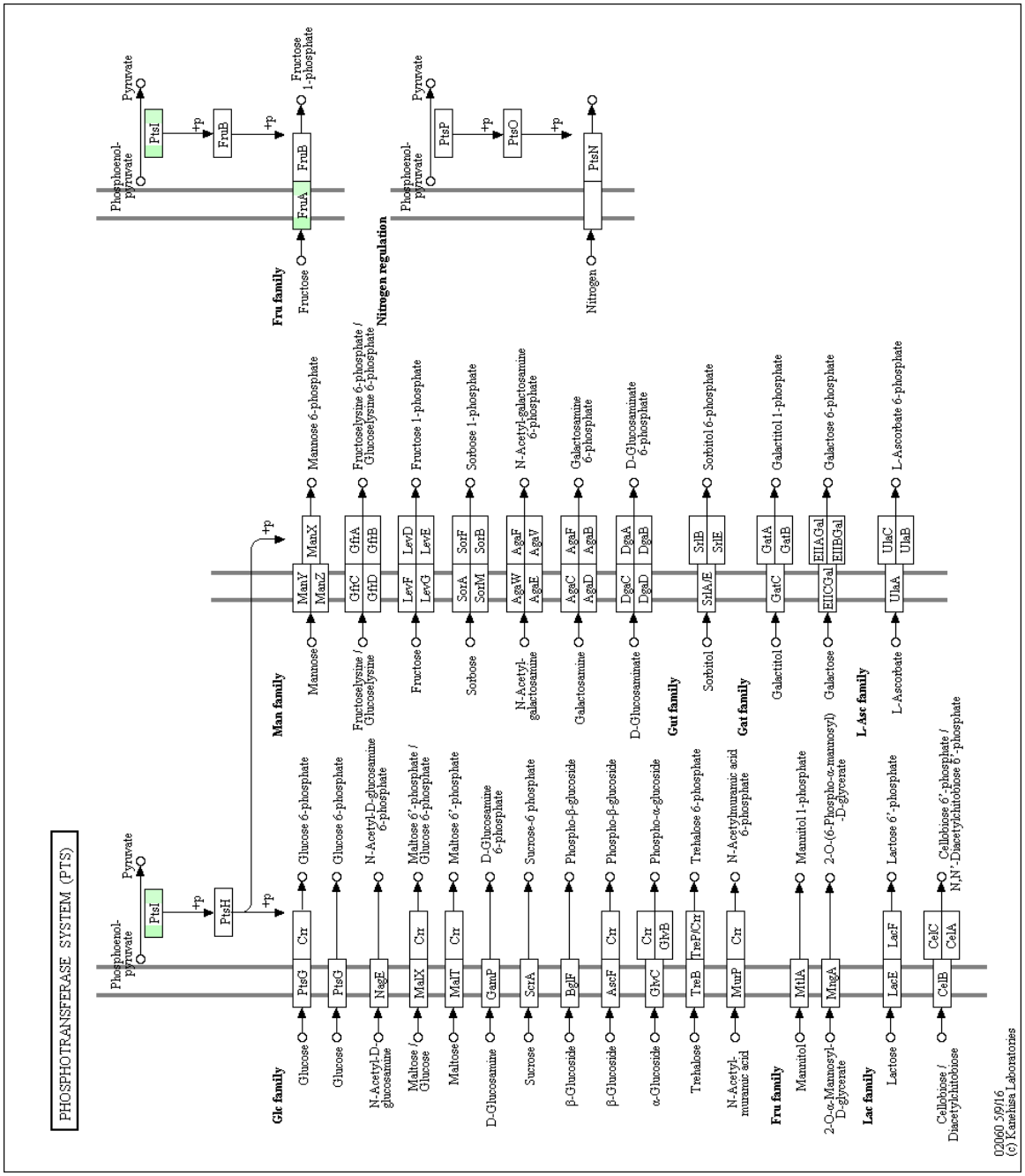
Red bar, *C. pertinax* (2,577 ORFs); light green bar, *C. islandicus* (2,480 ORFs); dark blue bar, *C. hydrogenoformans* (2,620 ORFs); yellow bar, *C. ferrireducens* (2,457 ORFs).

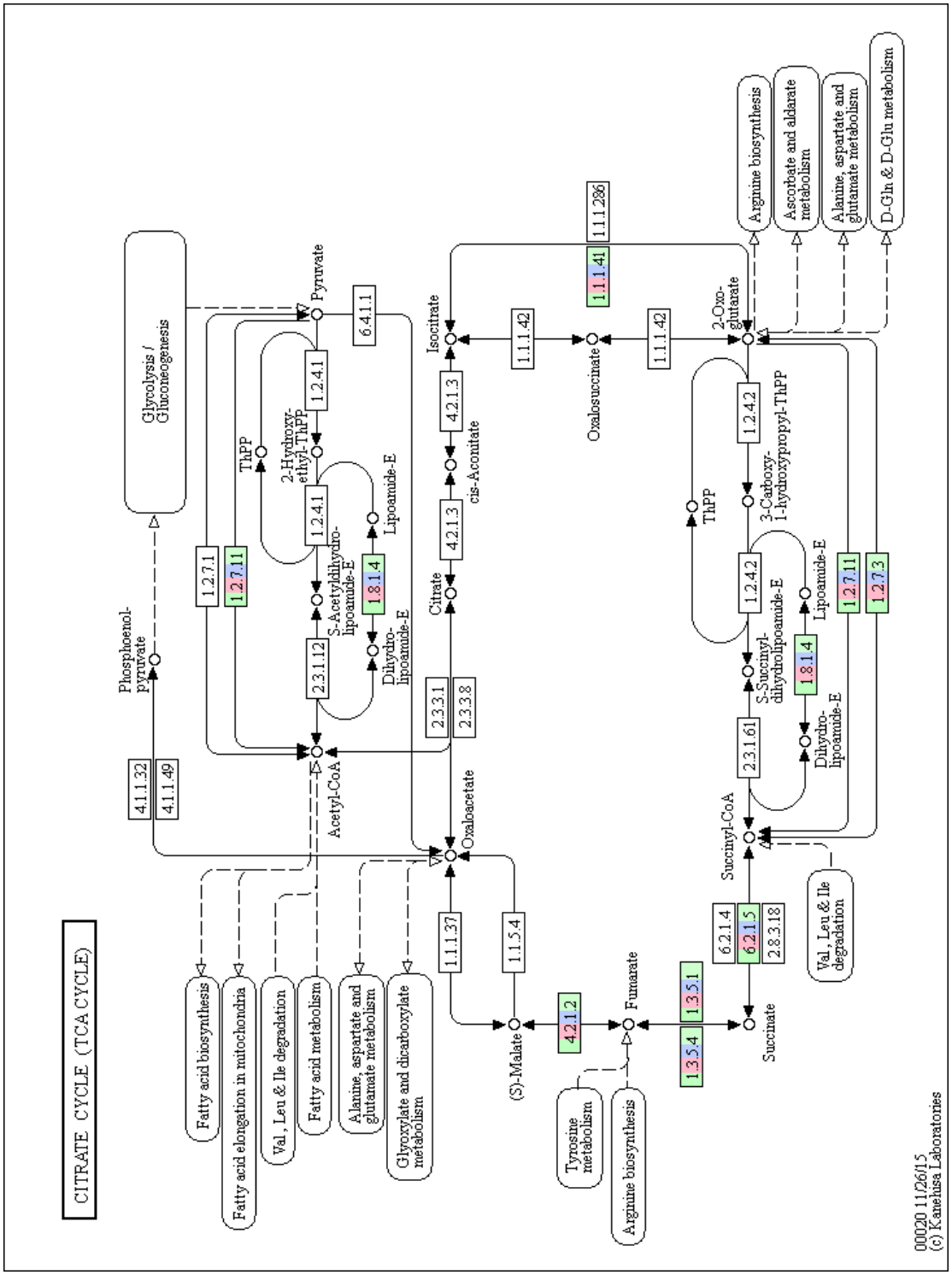
Functional categories: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control, cell division, and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination, and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; W, extracellular structures; Y, nuclear structure; and Z, cytoskeleton.

A


GLYCOLYSIS / GLUCONEOGENESIS
Starch and sucrose metabolism
α-D-Glucose-1P
3.1.3.10
5.4.2.2
3.1.3.9
2.7.1.199
D-Glucose (extracellular)
α-D-Glucose
2.7.1.1
2.7.1.63
2.7.1.2
2.7.1.147
α-D-Glucose-6P
5.1.3.3
5.1.3.15
5.3.1.9
β-D-Glucose-6P
β-D-Glucose
β-D-Fructose-6P
Pentose phosphate pathway
3.1.3.11
2.7.1.11
2.7.1.146
2.7.1.90
Arbutin (extracellular)
Arbutin-6P
3.2.1.86
Salicin (extracellular)
Salicin-6P
β-D-Fructose-1,6P2
4.1.2.13
Glyceraldehyde-3P
Glycerone-P
1.2.1.12
1.2.1.59
Glycerate-1,3P2
1.2.1.9
1.2.7.6
1.2.1.90
5.4.2.4
2.7.2.3
Glycerate-2,3P2
Carbon fixation in photosynthetic organisms
5.4.2.11
Glycerate-3P
5.4.2.12
3.1.3.80
Glycerate-2P
4.2.1.11
4.1.1.32
4.1.1.49
Oxaloacetate
Phosphoenol-pyruvate
Pyruvate metabolism
2.7.1.40
Citrate cycle
1.2.7.1
1.2.7.11
ThPP
Acetyl-CoA
2-Hydroxy-ethyl-ThPP
1.2.4.1
2.3.1.12
S-Acetyl-dihydrolipoamide-E
4.1.1.1
Pyruvate
1.1.1.27
L-Lactate
6.2.1.1
6.2.1.13
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
Propanoate metabolism
1.2.1.3
1.1.1.1
1.1.1.2
1.1.1.7
EutG
Ethanol
1.1.1.8
Acetate
1.2.1.5
1.2.1.-
Acetaldehyde
00010 12/26/16
(c) Kanehisa Laboratories

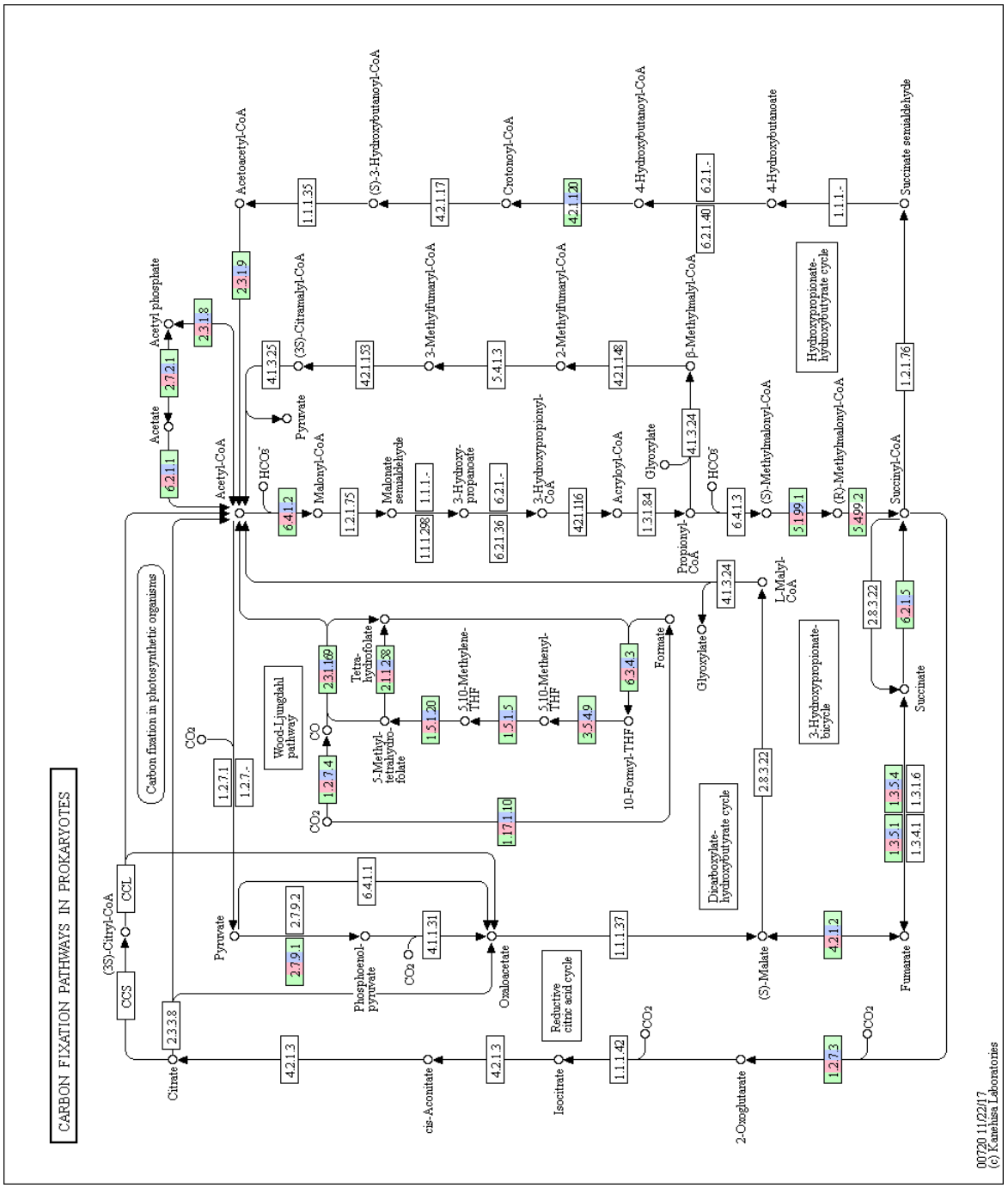
B


PHOSPHOTRANSFERASE SYSTEM (PTS)
Phosphoenol-pyruvate
Pyruvate
PtsI
PtsH
+p
Glc family
Glucose
Glucose 6-phosphate
PtsG
Crr
Glucose
Glucose 6-phosphate
PtsG
N-Acetyl-D-glucosamine
N-Acetyl-D-glucosamine 6-phosphate
NagE
Maltose / Glucose
Maltose 6'-phosphate / Glucose 6-phosphate
MalX
Maltose
Maltose 6'-phosphate
MalT
D-Glucosamine
D-Glucosamine 6-phosphate
GamP
Sucrose
Sucrose-6 phosphate
ScrA
β-Glucoside
Phospho-β-glucoside
BglF
β-Glucoside
Phospho-β-glucoside
AscF
α-Glucoside
Phospho-α-glucoside
GlvC
GlvB
Trehalose
Trehalose 6-phosphate
TreB
TreP/Crr
N-Acetyl-muramic acid
N-Acetylmuramic acid 6-phosphate
MurP
Fru family
Mannitol
Mannitol 1-phosphate
MtlA
2-O-α-Mannosyl-D-glycerate
2-O-(6-Phospho-α-mannosyl)-D-glycerate
MngA
Lac family
Lactose
Lactose 6'-phosphate
LacE
LacF
Cellobiose / Diacetylchitobiose
Cellobiose 6'-phosphate / N,N'-Diacetylchitobiose 6'-phosphate
CelB
CelC
CelA
Man family
Mannose
Mannose 6-phosphate
ManY
ManZ
ManX
Fructoselysine / Glucoselysine
Fructoselysine 6-phosphate / Glucoselysine 6-phosphate
GfrC
GfrD
GfrA
GfrB
Fructose
Fructose 1-phosphate
LevF
LevG
LevD
LevE
Sorbose
Sorbose 1-phosphate
SorA
SorM
SorF
SorB
N-Acetyl-galactosamine
N-Acetyl-galactosamine 6-phosphate
AgaW
AgaE
AgaF
AgaV
Galactosamine
Galactosamine 6-phosphate
AgaC
AgaD
AgaF
AgaB
D-Glucosaminate
D-Glucosaminate 6-phosphate
DgaC
DgaD
DgaA
DgaB
Gut family
Sorbitol
Sorbitol 6-phosphate
StlB
StlA/E
StlE
Gat family
Galactitol
Galactitol 1-phosphate
GatC
GatA
GatB
L-Asc family
Galactose
Galactose 6-phosphate
EIICGal
EIIAGal
EIIBGal
L-Ascorbate
L-Ascorbate 6-phosphate
UlaA
UlaC
UlaB
Fru family
Fructose
Fructose 1-phosphate
FruA
FruB
Nitrogen regulation
Nitrogen
PtsN
PtsO
PtsP
02060 5/9/16
(c) Kanehisa Laboratories


CITRATE CYCLE (TCA CYCLE)
Glycolysis / Gluconeogenesis
Phosphoenol-pyruvate
Pyruvate
Acetyl-CoA
Oxaloacetate
Citrate
cis-Aconitate
Isocitrate
Oxalosuccinate
2-Oxo-glutarate
Succinyl-CoA
Succinate
Fumarate
(S)-Malate
ThPP
2-Hydroxy-ethyl-ThPP
S-Acetyldihydro-lipoamide-E
Dihydro-lipoamide-E
Lipoamide-E
3-Carboxy-1-hydroxypropyl-ThPP
S-Succinyl-dihydrolipoamide-E
4.1.1.32
4.1.1.49
1.2.7.1
1.2.7.11
1.2.4.1
1.2.4.1
1.8.1.4
2.3.1.12
6.4.1.1
4.2.1.3
4.2.1.3
2.3.3.1
2.3.3.8
1.1.1.42
1.1.1.41
1.1.1.286
1.1.1.42
1.2.4.2
1.2.4.2
1.8.1.4
2.3.1.61
1.2.7.11
1.2.7.3
1.1.1.37
1.1.5.4
4.2.1.2
1.3.5.4
1.3.5.1
6.2.1.4
6.2.1.5
2.8.3.18
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Val, Leu & Ile degradation
Fatty acid metabolism
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
Tyrosine metabolism
Arginine biosynthesis
Val, Leu & Ile degradation
Arginine biosynthesis
Ascorbate and aldarate metabolism
Alanine, aspartate and glutamate metabolism
D-Gln & D-Glu metabolism
00020 11/26/15
(c) Kanehisa Laboratories

C

D


CARBON FIXATION PATHWAYS IN PROKARYOTES
Carbon fixation in photosynthetic organisms
Wood-Ljungdahl pathway
Reductive citric acid cycle
Dicarboxylate-hydroxybutyrate cycle
3-Hydroxypropionate-bicycle
Hydroxypropionate-hydroxybutyrate cycle
Acetyl phosphate
Acetate
Acetyl-CoA
Acetoacetyl-CoA
(S)-3-Hydroxybutanoyl-CoA
Crotonoyl-CoA
4-Hydroxybutanoyl-CoA
4-Hydroxybutanoate
Succinate semialdehyde
(3S)-Citramalyl-CoA
3-Methylfumaryl-CoA
2-Methylfumaryl-CoA
β-Methylmalyl-CoA
Pyruvate
Malonyl-CoA
Malonate semialdehyde
3-Hydroxy-propanoate
3-Hydroxypropionyl-CoA
Acryloyl-CoA
Glyoxylate
Propionyl-CoA
(S)-Methylmalonyl-CoA
(R)-Methylmalonyl-CoA
Succinyl-CoA
Succinate
Fumarate
(S)-Malate
L-Malyl-CoA
Glyoxylate
Formate
10-Formyl-THF
5,10-Methenyl-THF
5,10-Methylene-THF
5-Methyl-tetrahydro-folate
Tetra-hydrofolate
CO
CO2
HCO3-
(3S)-Citryl-CoA
CCL
CCS
Citrate
cis-Aconitate
Isocitrate
2-Oxoglutarate
Oxaloacetate
Phosphoenol-pyruvate
2.3.1.9
2.7.2.1
2.3.1.8
6.2.1.1
6.4.1.2
1.2.1.75
1.1.1.298
6.2.1.36
4.2.1.116
1.3.1.84
4.1.3.24
6.4.1.3
5.1.99.1
5.4.99.2
1.2.1.76
1.1.1.35
4.2.1.17
4.2.1.120
6.2.1.40
6.2.1.-
1.1.1.-
4.1.3.25
4.2.1.153
5.4.1.3
4.2.1.148
1.2.7.1
1.2.7.-
1.2.7.4
2.3.1.169
2.1.1.258
1.5.1.20
1.5.1.5
3.5.4.9
6.3.4.3
1.17.1.10
4.1.3.24
2.8.3.22
6.2.1.5
1.3.5.1
1.3.5.4
1.3.4.1
1.3.1.6
2.7.9.2
2.7.9.1
6.4.1.1
4.1.1.31
1.1.1.37
4.2.1.2
2.3.3.8
4.2.1.3
4.2.1.3
1.1.1.42
1.2.7.3
CO2
00720 11/22/17
(c) Kanehisa Laboratories

E

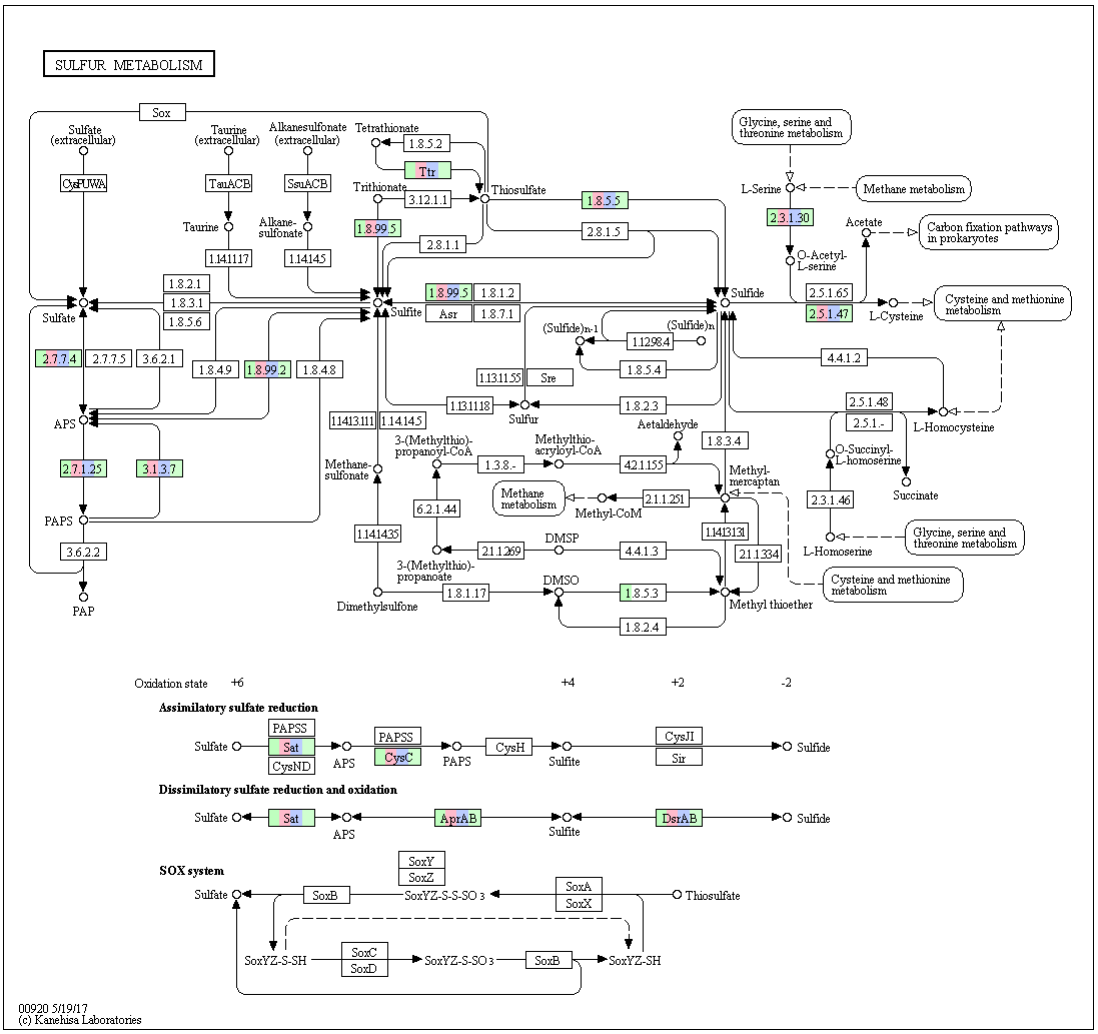


Fig. 2-2-3 Comparison of metabolic pathways in *C. pertinax*, *C. islandicus*, *C. hydrogenoformans*, and *C. ferrireducens*

A. Glycolysis and gluconeogenesis. B. Phosphotransferase system. C. TCA cycle. D. Carbon fixation pathways in prokaryotes. E. Sulfur metabolism

Green on left side, *C. pertinax*; red, *C. islandicus*; blue, *C. hydrogenoformans*; green on right side, *C. ferrireducens*. The empty box indicates that there are no ORFs assigned. Pathway maps were generated by KEGG.

To confirm phylogenetic relationships between the four *Carboxydothermus* species, I performed a phylogenetic analysis based on five housekeeping genes (Fig. 2-2-4A) and members of the *Thermoanaerobacteraceae*, whose genomes are available in databases. Accordingly, *C. pertinax* was the most deeply branched *Carboxydothermus* species, whereas *C. islandicus* was branched in the *Carboxydothermus* genus. Further, genome-wide phylogenetic analysis based on GSS confirmed the same topology (Fig. 2-2-4B). These results were supported by the homology scores (H-values) distribution between proteins of *C. hydrogenoformans* and those of other species (Fig. 2-2-5). Conserved proteins are defined by having H-value above 0.64 (Shao et al. 2010). Given the high H-value among *Carboxydothermus* species, most proteins appeared to be conserved. Interestingly, proteins of *C. pertinax* showed relatively low H-value relative to *C. hydrogenoformans*.

A comprehensive comparison between the four *Carboxydothermus* species was generated using the complete genome of *C. hydrogenoformans* as reference (Fig. 2-2-6). The genome of *C. hydrogenoformans* showed high co-linearity with the three other genomes. Comparisons among *Carboxydothermus* species showed disruptions of synteny and some prophage-containing regions. The genome of *C. pertinax* harbored a number of variable regions (>5000 bp) compared to *C. hydrogenoformans*, which were likely derived from insertion or inversion (Table 2-2-4). A large inversion (250 kbp in *C. hydrogenoformans*) was observed at the start of contig 6 in *C. pertinax* (Fig. 2-2-7). Multiple insertions and deletions divided this segment into three inversion regions (R6 to 8, Table 2-2-4). The notable variable region in *C. pertinax* (R7, Table 2-2-4) was the CODH-I−ECH gene cluster, which lacked *cooS* and *cooA* genes (Fig. 2-2-8). A finer scale analysis of the sequences upstream and downstream of the CODH-I−ECH gene

cluster revealed that sequences related to both *cooS* and *cooA* genes were absent in *C. pertinax*. In comparison, *C. ferrireducens* lacked the whole CODH-I−ECH gene cluster.

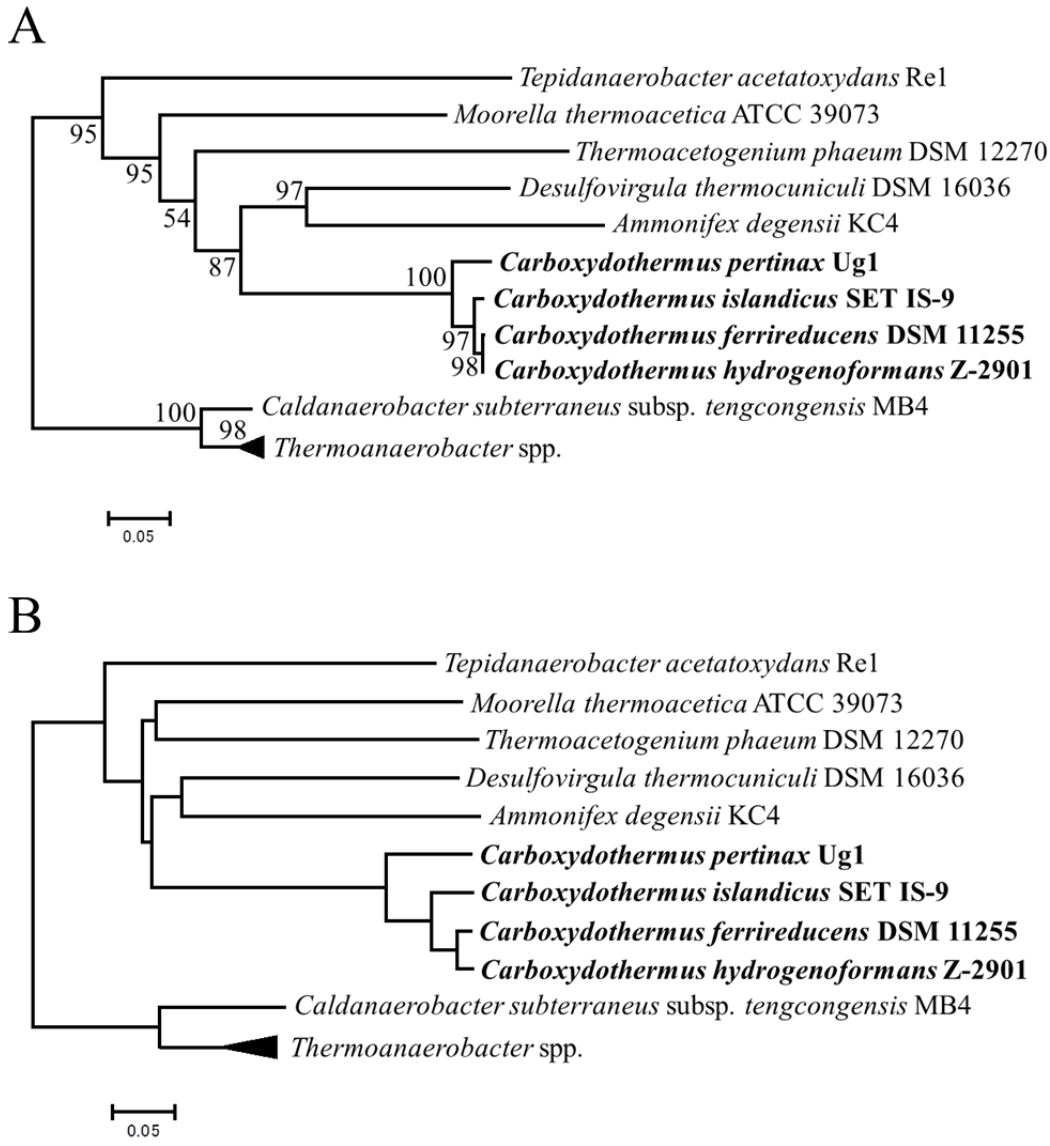


Fig. 2-2-4 Phylogenetic analyses among *Thermoanaerobacteraceae*
(A) Maximum likelihood analysis of five concatenated housekeeping genes (*frr*, *nusA*, *rpmA*, *tsf*, and *pyrG*). (B) GSS distance matrix is plotted as a neighbor-joining tree. *Carboxydothermus* species are indicated in bold font.

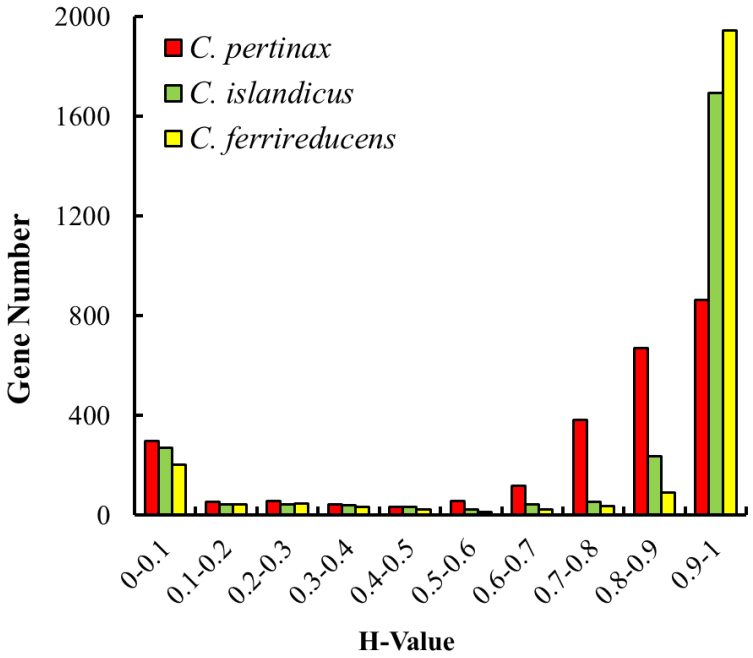


Fig. 2-2-5 Histograms of H-value (a homology measure score) for all proteins of *C. pertinax*, *C. islandicus*, and *C. ferrireducens* compared to *C. hydrogenoformans*
Red bar, *C. pertinax* (2,577 ORFs); light green bar, *C. islandicus* (2,480 ORFs); yellow bar, *C. ferrireducens* (2,457 ORFs).

Table 2-2-4 List of major insertion or inversion regions in *C. pertinax* compared to *C. hydrogenoformans*

major variable region	*C. pertinax*					
	status	contig number	length (bp)	position start	position end	number of ORFs
R1	insertion	10	6487	42951	49438	6
R2	insertion	23	26164	10895	37059	29
R3	inversion	55	177126	186423	363549	195
R4	insertion	17	43956	974	44930	44
R5	inversion	55	48194	136720	184914	45
R6	inversion	6	80991	318	81309	76
R7	inversion	6	8753	104851	113604	9
R8	inversion	6	25030	115963	140993	24

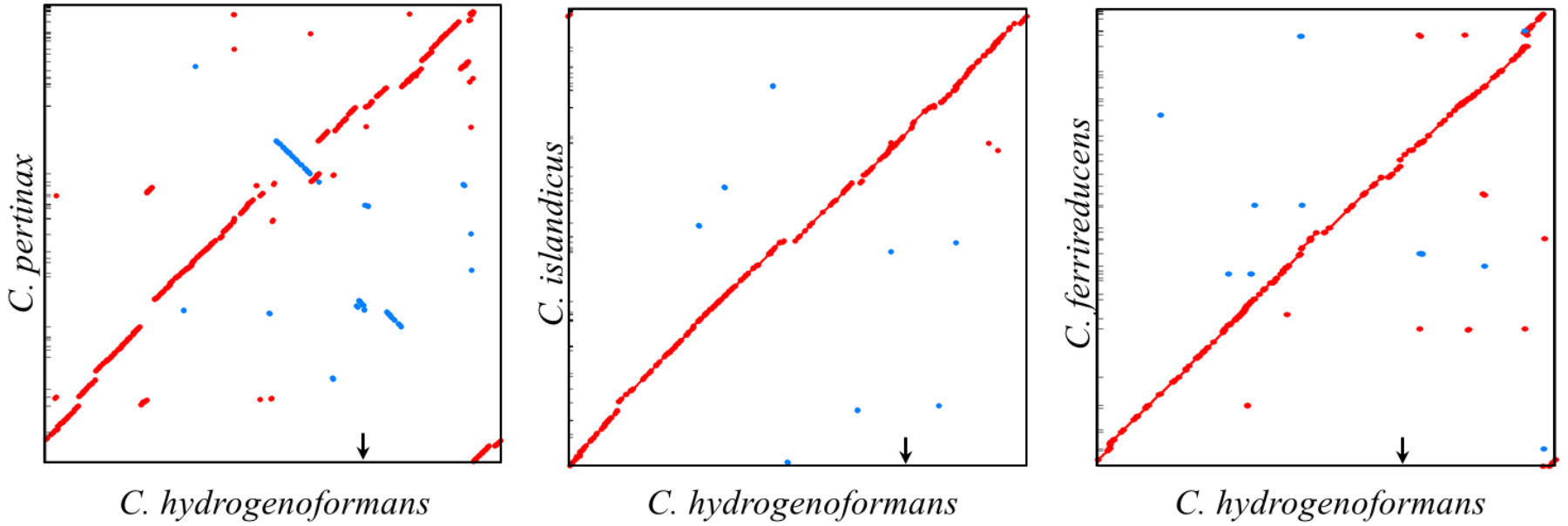


Fig. 2-2-6 Comparison between the genomes of *C. hydrogenoformans* and *C. pertinax*, and *C. islandicus* and *C. ferrireducens*

Synteny plots demonstrating the high co-linearity between the genomes of *C. hydrogenoformans* and *C. pertinax*, and *C. islandicus* and *C. ferrireducens*. Plots indicate a similar sequence in the same orientation (red) or reverse orientation (blue) shared by the two species. Scales on the y axis indicate contig boundaries. Arrows indicate the region encoding the CODH-I−ECH gene cluster in *C. hydrogenoformans*. A comparison of *C. hydrogenoformans* vs. *C. pertinax* (left), *C. islandicus* (middle), and *C. ferrireducens* (right) is presented.

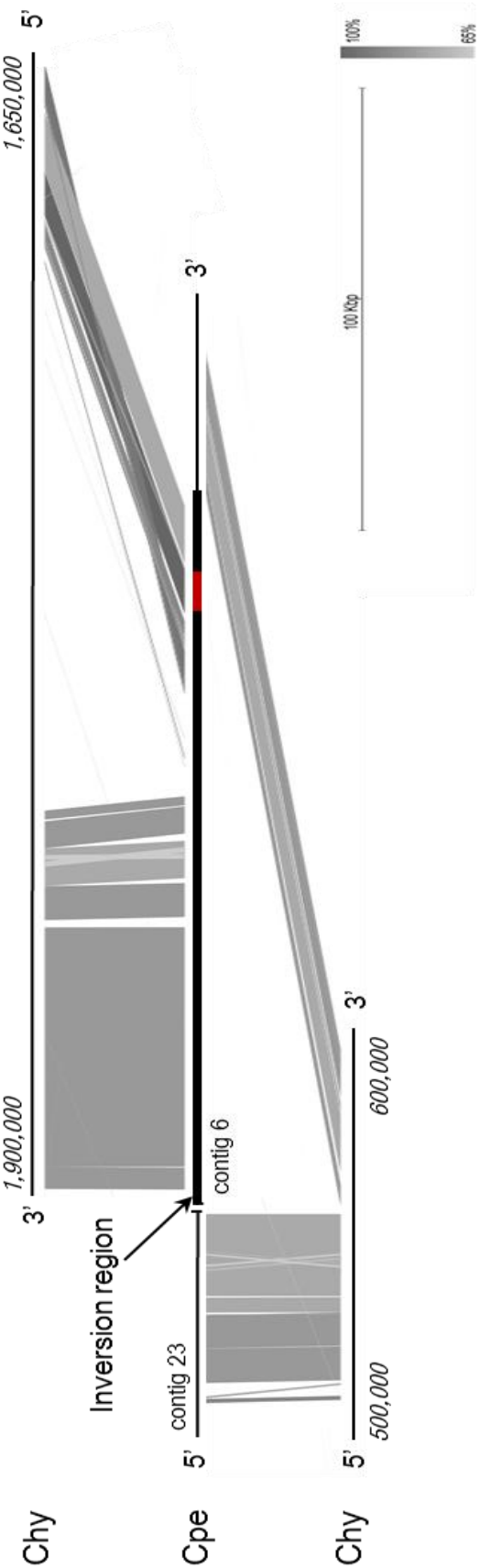


Fig. 2-2-7 Comparison of genome sequences around the inversion region in *C. pertinax* compared to *C. hydrogenoformans*
Vertical blocks between sequences indicate regions of shared similarity shaded according to BLASTn. Italic numbers indicate nucleotide position in *C. hydrogenoforans*. Cpe, *C. pertinax*; Chy, *C. hydrogenoformans*. Bold line, inversion region; red line, CODH-I–ECH gene cluster.

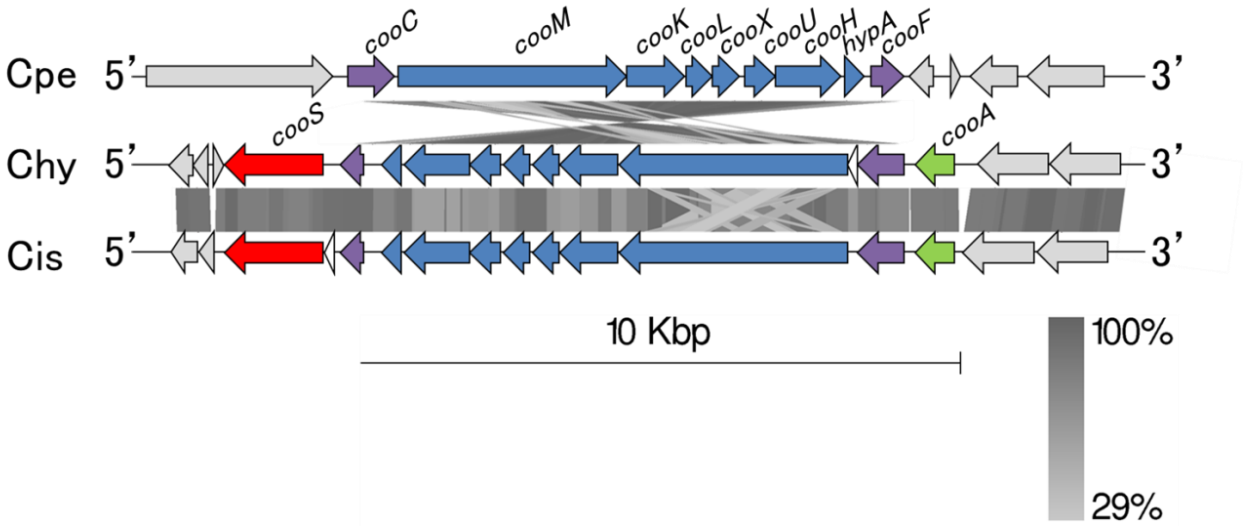


Fig. 2-2-8 Comparison of the CODH-I–ECH gene cluster in three *Carboxydothermus* species

Vertical blocks between sequences indicate regions of shared similarity shaded according to BLASTn. Cpe, *C. pertinax*; Chy, *C. hydrogenoformans*; Cis, *C. islandicus*. Red, *cooS*; green, *cooA*; purple, CODH-I related gene; blue, ECH-related gene; white, hypothetical gene; gray, unrelated neighboring gene.

Comparison of CODH-I gene clusters and identification of the CODH gene for CO oxidation in *C. pertinax*

C. pertinax, and *C. islandicus* possess four and five *cooS* genes, respectively. In addition, I report that *C. ferrireducens* possessed four *cooS* genes in its genome. Phylogenetic analysis of amino acid sequences of CooS confirmed that each *cooS* gene from the four *Carboxydothermus* species was closely related, with *C. islandicus* presenting all five *cooS* genes (corresponding to *cooS-I* to *cooS-V* in *C. hydrogenoformans*), whereas *C. pertinax* and *C. ferrireducens* lacked *cooS-I*.

In *C. hydrogenoformans*, the CODH-I−ECH gene cluster is composed of two functional sets of genes: the CODH-I gene cluster (*cooS-IFA*) and the ECH gene cluster (*cooMKLXUHhypA*) (Wu et al. 2005). *C. pertinax* and *C. islandicus* possessed the CODH-I−ECH gene cluster but the former lacked the genes encoding the CODH-I catalytic subunit (*cooS-I*) and its transcriptional regulator (*cooA*; Fig. 2-2-8). In contrast, *C. ferrireducens* completely lacked both CODH-I and ECH gene clusters. The genomic contexts of the other four CODH genes (*cooS-II* to *-V*) were conserved in the genus *Carboxydothermus*.

To identify which CODH is involved in CO oxidization in *C. pertinax*, I compared transcription of *C. pertinax* under 100% CO and 100% N_2 (Fig. 2-2-9; Table 2-2-2). Under both conditions, I added pyruvate Na (final concentration, 2 g/L) as electron donor and carbon source. Transcripts of *cooS-II* and *cooS-III* were detected (relative transcript levels to 16S rRNA gene were 2.44×10^{-2} copies/μL and 9.94×10^{-3} copies/μL, respectively) under 100% CO, but were below the detection limit (1.0×10^{2} copies/μL) under 100% N_2. Transcripts of *cooS-IV* and *cooS-V* were below the limit of detection (1.0×10^{2} copies/μL) in both conditions. Compared with relative transcript levels under 100% N_2, *cooS-II* was solely upregulated under 100% CO. In addition, *cooA2* and *cooF1* gene in the CODH-II gene cluster and *cooH* and *cool* gene in the CODH-I−ECH gene cluster were also upregulated under 100% CO. In contrast, transcripts of *hypAB* which encode H_2-uptake Ni-Fe-hydrogenase large and small catalytic subunit were below the limit of detection (1.0×10^{2} copies/μL) under both 100% CO and 100% N_2.

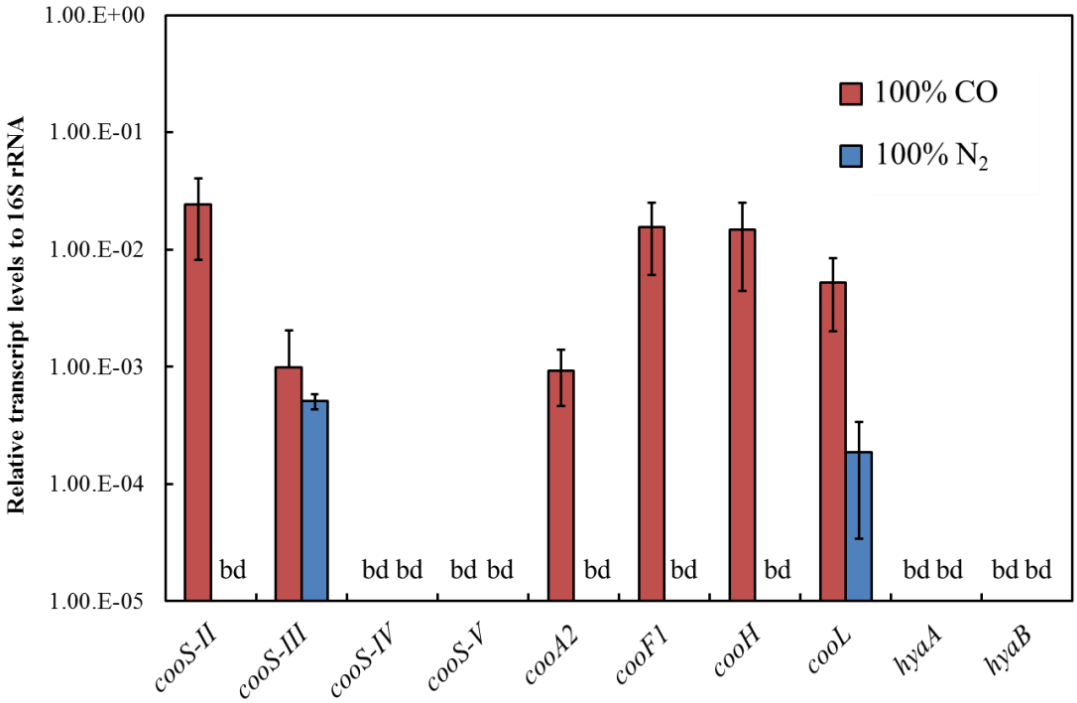


Fig. 2-2-9 Relative transcript levels in *C. pertinax* under 100% CO and 100% N_2
C. pertinax transcript levels relative to the 16S rRNA gene are shown. Red bar, cells grown under 100% CO; blue bar, cells grown under 100% N_2. bd, below the limit of detection. Bars represent the mean of at least three biological replicates. Error bars represent standard deviations.

Alternative CO metabolism with H_2 and hydrogen sulfide generation in *C. pertinax*

While other hydrogenogenic *Carboxydothermus* species cannot couple CO oxidation with hydrogen sulfide generation (Svetlichny et al. 1991; Slepov 2009, Yoneda et al. 2012), *C. pertinax* can couple CO oxidation with reduction of thiosulfate or elemental sulfur as terminal electron acceptors to produce hydrogen sulfide (Yoneda et al. 2012). To investigate differences in hydrogen sulfide generation with CO oxidation, I predicted related genes and compared those components. The dissimilatory sulfate reduction pathway was present in the four *Carboxydothermus* species (Fig. 2-2-3E). The reduction of thiosulfate or elemental sulfur was catalyzed by the PhsABC

(thiosulfate reductase) and PsrABC (elemental sulfur reductase) complexes, respectively. Both enzymes are highly similar to each other and belong to the dimethyl sulfoxide reductase family (Duval et al. 2008). PhsA and PsrA are catalytic subunits containing a molybdopterin guanine dinucleotide molybdenum cofactor (Berks and Hinsley 2002). PhsB and PsrB are electron transfer proteins possessing four iron-sulfur clusters (Hallenbeck et al. 1989). PhsC and PsrC are integral membrane proteins that anchor the remaining complex components to the membrane at a menaquinol oxidation site (Berks et al. 1995). All *Carboxydothermus* species possessed one gene cluster encoding the complex with conserved motifs for reduction of thiosulfate or elemental sulfur (Duval et al. 2008; Dietrich and Klimmek 2002) in each gene product (*C. pertinax*, cpu_06910 to 06930; *C. islandicus*, CISS_00980 to 01000; *C. hydrogenoformans*, CHY_2572 to 2574; and *C. ferrireducens*, WP_028052252.1 to WP_028052254.1). Moreover, the reduction of thiosulfate or elemental sulfur requires menaquinol, which is the most widespread microbial respiratory quinone (Nowicka et al. 2010), as the sole electron donor (Stoffels et al. 2012). The gene encoding Type II NADH dehydrogenase, which reduces menaquinone to menaquinol (Nantapong et al. 2005), was present only in *C. pertinax* (cpu_03530) as unique gene.

To reveal alternative hydrogenogenic CO metabolism in *C. pertinax*, I performed a cultivation study under 100% CO with or without thiosulfate. In these conditions, addition of thiosulfate as terminal electron acceptor affected growth and kinetic parameters as reported in Fig. 2-2-10 and Table 2-2-5. Measurement of dissolved hydrogen sulfide in the culture during hydrogenogenic carboxydotrophic growth with thiosulfate under 100% CO revealed that the concentration of dissolved hydrogen sulfide was below the limit of detection (50 μM) until after 27 h after inoculation (Fig.

2-2-11). Dissolved hydrogen sulfide was detected 36 h after inoculation (77.8 μM; standard deviation, 13.4 μM) and increased after that (186.3 μM; standard deviation, 14.5 μM at 102 h) (Fig. 2-2-11). The electrons from CO oxidation would be used in the reduction of H_2O to H_2 or thiosulfate to dissolved hydrogen sulfide in accordance with the equations: $CO + H_2O \rightarrow CO_2 + H_2$ or $CO + S_2O_3^{2-} + H_2O \rightarrow CO_2 + HS^- + SO_3^{2-} + H^+$. When CO was the electron donor, a mole of H_2 or dissolved hydrogen sulfide was generated. In culture, however, the total amount of dissolved hydrogen sulfide (14910 μmol; standard deviation, 1160 μmol) was in excess of the total amount of CO oxidized (2840 μmol; standard deviation, 240 μmol) (Fig. 2-2-11). Hydrogenogenic carboxydotrophs often require organic compounds (i.e., yeast extract) for their carboxydotrophic growth, even though they are considered as autotrophic carboxydotrophs (Omae et al. 2017). Hence, yeast extract, which was contained in the fresh medium (final concentration, 0.05 g/L), or organic compounds in the initial inoculum might have acted as indirect electron donor for the surplus reduction of dissolved hydrogen.

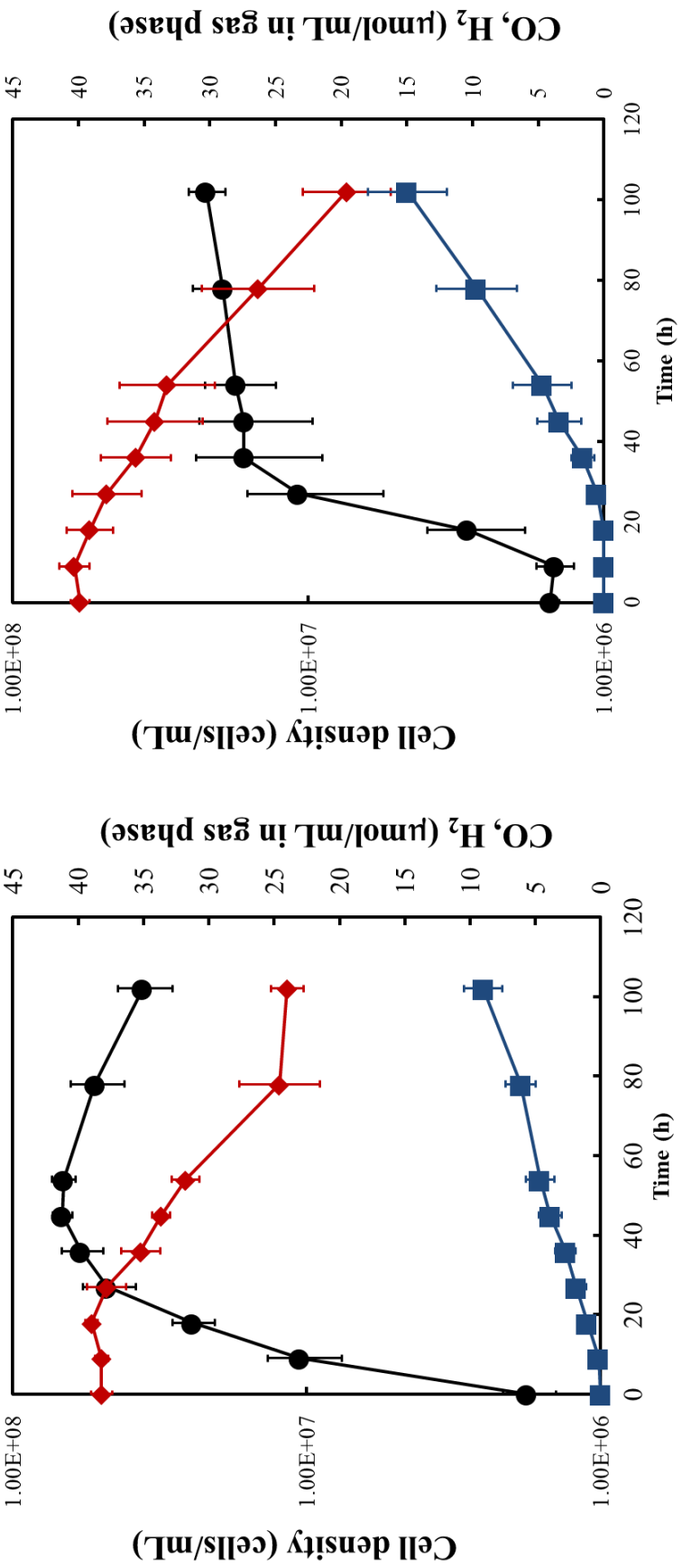


Fig. 2-2-10 Hydrogenogenic carboxydotrophic growth of *C. pertinax* with thiosulfate (left) and without thiosulfate (right)
Different parameters are shown: black disks, cell density; red diamonds, CO concentration; blue squares, H_2 concentration. Plots represent the mean of at least three biological replicates. Error bars represent standard deviations.

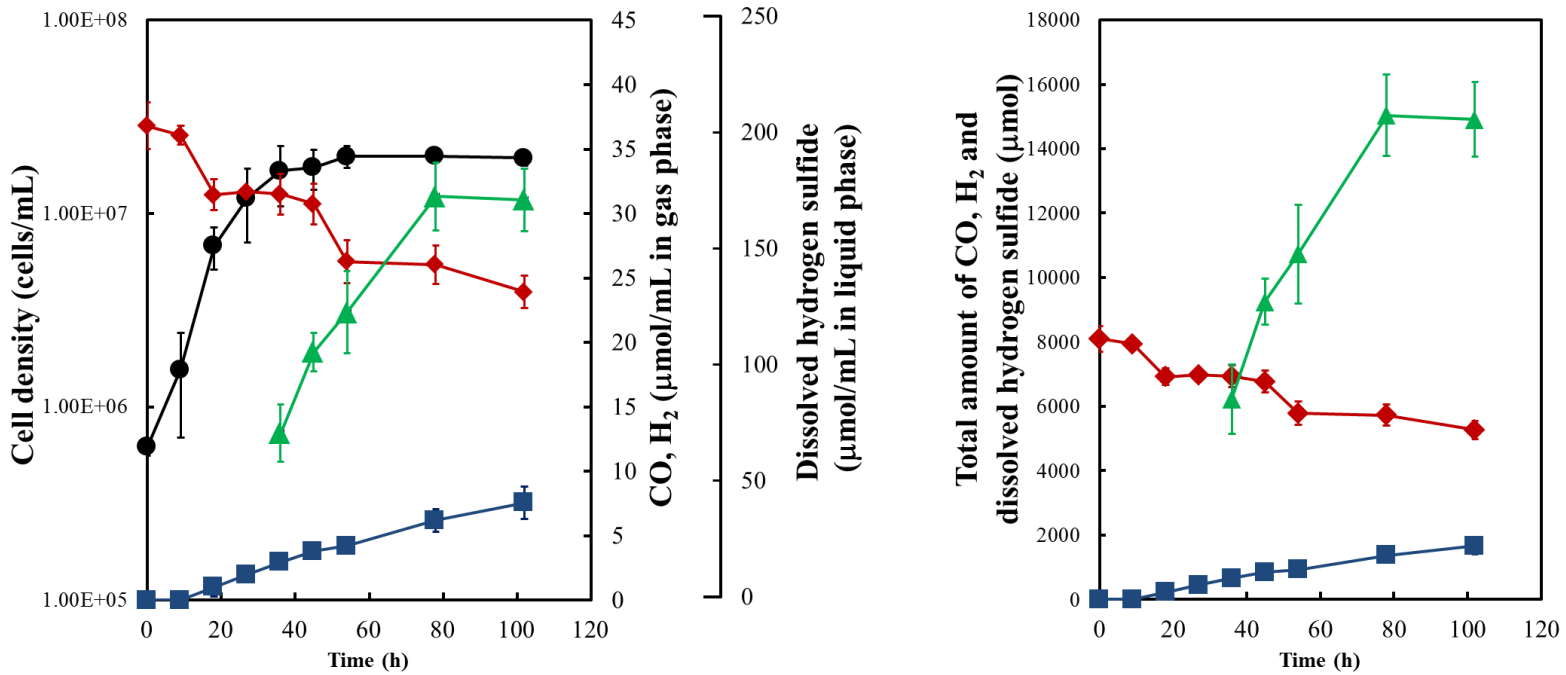


Fig. 2-2-11 Growth of *C. pertinax* under 100% CO with thiosulfate (left) and profiles of total amounts of CO, H_2 and dissolved hydrogen sulfide (right)

Different parameters are shown: black disks, cell density; red diamonds, CO concentration; blue squares, H_2 concentration; green triangles, dissolved hydrogen sulfide. Plots represent means of at least three biological replicates. Error bars represent standard deviations.

Table 2-2-5 Growth and kinetic parameters of *C. pertinax* during CO metabolism with thiosulfate and without thiosulfate

	With thiosulfate		Without thiosulfate		
Growth and kinetic parameter[a]	Value	SD[b]	Value	SD[b]	Fold difference
Maximum cell density (cells mL^{-1})	6.80×10^{7}	5.21×10^{6}	2.22×10^{7}	3.14×10^{6}	3.1 *
Maximum specific growth rate (h^{-1})	0.086	0.010	0.058	0.021	1.5
Final CO consumption (μmol mL^{-1})	14.6	1.72	20.3	3.55	0.7
Maximum specific CO consumption rate (μmol mL^{-1} h^{-1})	0.30	0.17	0.29	0.046	1.0
Final H_2 production (μmol mL^{-1} h^{-1})	9.00	1.46	14.6	3.00	0.6
Maximum specific H_2 production rate (μmol mL^{-1} h^{-1})	0.13	0.020	0.22	0.0025	0.6 *
CO:H_2 (mol:mol)	1:0.62		1:0.72		0.9

[a] Kinetic parameters were calculated with the data from Fig. 2-2-10

[b] SD, standard deviation

*, Significant difference (*p* value < 0.05) between parameters with thiosulfate and without thiosulfate

Discussion

Previous studies reported a range of available terminal electron acceptors in which *Carboxydothermus* members performed their CO metabolism (Svetlichny et al. 1991; Slobodkin et al. 2006; Slepova et al. 2009; Novikov et al. 2011; Yoneda et al. 2012). In the present study, I conducted a comparative genome analysis of four *Carboxydothermus* species to investigate the versatility of their CO metabolism. General features of their genomes and the percentage of ORFs assigned to each COG functional category were highly similar (Fig. 2-2-2 and Table 2-2-3), indicating that

their genomes were highly conserved. Consistent with previous phylogenetic analysis based on 16S rRNA, my phylogenetic analysis based on concatenated housekeeping genes and genomic similarity score confirmed the close relationship between *Carboxydothermus* species and the fact that *C. pertinax* was most deeply branched in this genus (Fig. 2-2-4). Close relationship was also confirmed by synteny plots (Fig. 2-2-6). A prominent feature was an inversion region in the CODH-I–ECH gene cluster of *C. pertinax* and the lack of *cooS* and *cooA* genes. To my knowledge, the CODH gene cluster is encoded within or adjacent to the ECH gene cluster in all hydrogenogenic carboxydotrophs (Sokolova and Lebedinsky 2013). The intact gene cluster expanded to phylogenetically diverse hydrogenogenic carboxydotrophs as a result of its vertical transmission and horizontal gene transfer (Techtmann et al. 2012; Sant'Anna et al. 2015). Thus, this result suggest that *cooS* and *cooA*, located at the ends of the CODH-I–ECH gene cluster in *C. pertinax*, were disrupted when inversion occurred, after which this species branched off from the other *Carboxydothermus* representatives.

The CODH/ECH complex is thought to play a critical role in energy conservation via hydrogenogenic CO metabolism (Ragsdale 2004). Further, expression of the CODH–ECH gene cluster is regulated by a transcription factor (Techtmann et al. 2012; Aono et al. 1996). In most cases, this is the CO-sensing activator CooA, whose gene is located within the CODH–ECH gene cluster (Roberts et al. 2004). Absence of both CODH-I and ECH gene clusters in the genome of the non-hydrogenogenic carboxydotroph *C. ferrireducens* is in line with this species' inability to produce H_2 during CO oxidation (Slobodkin et al. 2006). At the same time, absence of both *cooS-I* and *cooA* genes in the CODH-I–ECH gene cluster of *C. pertinax* seemed to be incompatible with *C. pertinax*'s hydrogenogenic carboxydotrophic growth.

Transcriptional analysis showed that genes encoding the CODH-II catalytic subunit and ECH catalytic large and small subunits were remarkably upregulated in *C. pertinax* under 100% CO. This result strongly suggests that *C. pertinax* couples CO oxidation by CODH-II with H_2 production by ECH for their hydrogenogenic carboxydotrophic growth. A previous study of *C. hydrogenoformans* showed that the estimated proportion of CODH-II to CODH-I in CODH−ECH complexes was estimated to be less than 10% (Soboh et al. 2002), however, it was not clear whether CODH-II interacted with the distal ECH (Soboh et al. 2002). On the other hand, *C. pertinax* might depend on CODH-II in terms of CO oxidation in their hydrogenogenic CO metabolism. In addition, although the *cooA* homologous gene (*cooA2*) in the CODH-II gene cluster was upregulated under 100% CO, the CooA-binding site (5′-TGTCA-N_6-CGACA-3′; He et al. 1996) was found only upstream (118 bp) of the CODH-I–ECH gene cluster but not upstream of the CODH-II gene cluster (data not shown). Thus, the mechanism regulating the CODH-II gene cluster during CO metabolism remains to be determined.

CO acts both as a substrate and an inhibitor for carboxydotrophs due to its high affinity for metalloproteins (Ragsdale 2004). Most carboxydotrophic sulfate-reducing bacteria in the genera *Desulfovibrio* and *Desulfotomaculum* are inhibited by a high CO concentration (Parshina et al. 2010). During carboxydotrophic growth, carboxydotrophic sulfate-reducing bacteria could couple CO oxidation to sulfate reduction with intermediate H_2 production under low CO (Zhou et al. 2011). *Carboxydothermus* species cannot reduce sulfate to hydrogen sulfide under 100% CO in the gas phase (Svetlichny et al. 1991; Slobodkin et al. 2006; Slepova et al. 2009; Novikov et al. 2011; Yoneda et al. 2012). My genomic approach showed that genes for the dissimilatory sulfate reduction pathway were conserved among *Carboxydothermus*

species (Fig. 2-2-3E). In sulfate-reducing mode, high CO may inhibit carboxydotrophic growth partly because CO is an inhibitor of hydrogenase or electron carriers involved in the reduction of sulfate (Mörsdorf et al. 1992; Parshina et al. 2010). *Carboxydothermus* species also contained genes with motifs essential for the reduction of thiosulfate or elemental sulfur (Duval et al. 2008; Dietrich et al. 2002). However, only *C. pertinax* possessed the gene encoding category II NADH dehydrogenase and could thus couple CO oxidation with reduction of thiosulfate or elemental sulfur as terminal electron acceptors to hydrogen sulfide.

Under 100% CO, the observed higher maximum cell density and maximum specific growth rate with thiosulfate (Fig. 2-2-10 and Table 2-2-5) suggested that *C. pertinax* could couple CO oxidation with H_2 and hydrogen sulfide production simultaneously. Such coupling may be more efficient than that with only H_2 and hydrogen sulfide production under 100% CO. Furthermore, final CO consumption with thiosulfate was lower than without thiosulfate, suggesting efficient CO oxidation. Based on thermodynamic considerations, CO oxidation can be coupled to the reduction of various terminal electron acceptors (Oelgeschläger et al. 2008). The reduction potential of thiosulfate ($E^{O'}$ [$S_2O_3^{2-}/HS^- + SO_3^{2-}$] = -402 mV) is higher than that of H_2O ($E^{O'}$ [$2H^+/H_2$] = -420 mV) under standard conditions (25 °C) (Thauer et al. 1977). Additionally, reduction of thiosulfate instead of H_2O is sufficiently exergonic to allow for the synthesis of 3 mol of ATP (Badziong and Thauer 1978). Therefore, *C. pertinax* could gain more proton motive force and grow more efficiently when thiosulfate was employed as the electron acceptor in CO metabolism. In this case, the actual energy calculated from the Nernst equation might be slightly lower at the temperature at which *C. pertinax* grows (65 °C). The lower values of final H_2 production and maximum

specific H_2 production rate with thiosulfate also suggest efficient CO consumption.

In general, hydrogenogenic carboxydotrophs can utilize CO with the production of nearly equimolar H_2, owing to the presence of the CODH/ECH complex (Ragsdale 2004). Under 100% CO, however, the amount of H_2 produced by *C. pertinax* was only 62% of the consumed CO with thiosulfate and 72% without thiosulfate (Fig. 2-2-10 and Table 2-2-5). This is less than expected from the equation of hydrogenogenic CO oxidation: $CO + H_2O \rightarrow CO_2 + H_2$. When thiosulfate was available as terminal electron acceptor, electron flow shifted to reduction of thiosulfate (Fig. 2-2-10). These results indicated that 72 % of the electrons from CO oxidation were consumed to reduce H_2O to H_2; the remaining electrons would provide reducing power to other reactions. In addition, the gap between the amount of H_2 produced and CO consumed has been observed also when *C. pertinax* coupled CO oxidation with Fe (III) reduction (Yoneda et al. 2012). From these findings, I suggest two possibilities: *C. pertinax* could utilize the intermediate H_2 as an electron donor for the reduction of thiosulfate by the H_2-uptake Ni-Fe-hydrogenase; alternatively, it could couple CO oxidation by its sole CODH-II with simultaneous reduction of mainly H_2O and thiosulfate under 100% CO. *C. pertinax* possessed an orthologous H_2-uptake Ni-Fe-hydrogenase (data not shown; cpu_22890 to 22980). Transcriptional analysis of *C. pertinax* which was performed with added both CO and pyruvate as electron donors showed that genes encoding the H_2-uptake Ni-Fe-hydrogenase large and small catalytic subunits were not transcribed (Fig. 2-2-10), suggesting that these genes were neither constitutive transcribed nor upregulated by addition of CO. Thus, the second option is more probable.

In this study, comparative genome analysis of four carboxydotrophic *Carboxydothermus* species revealed their distinct strategies for energy conservation

using the low redox potential of CO. So far, the functions of CODHs have been predicted by their genomic context (Techtmann et al. 2012). Although their genomic features were highly similar, the presence of the CODH-I/ECH complex influenced the ability to produce H_2 with CO oxidation. Interestingly, the hydrogenogenic carboxydotroph *C. pertinax* lacked both *cooS-I* and *cooA* genes in the CODH-I–ECH gene cluster, probably due to inversion. *C. pertinax* would thus couple simultaneous CO oxidation by its sole CODH-II with reduction of mainly H_2O and thiosulfate when grown under 100% CO. This makes *C. pertinax* a prominent candidate for understanding hydrogenogenic CO metabolism without an intact CODH–ECH gene cluster. Based on my findings that the function of CODH does not always correspond with the genomic context, it is expected that versatile CO metabolism will be inferred from available genomic databases.

Chapter 3

Whole transcriptome analysis of hydrogenogenic carboxydotrophic bacterium

Summary

To enhance our knowledge of its hydrogenogenic CO metabolism, I performed whole transcriptome analysis of *C. pertinax* grown under 100% CO or 100% N_2 using RNA sequencing. Of the 2,577 genes, 36 and 64 genes were differential expression genes (DEGs) with false discovery rate adjusted *P* value < 0.05 when grown under 100% CO or 100% N_2, respectively. Most of the DEGs (74% and 91%, respectively) were components of 23 gene clusters, suggesting switch between metabolisms via intensive expression changes in a relatively low number of gene clusters. Of the 9 significantly expressed gene clusters under 100% CO, CODH-II and ECH gene clusters were found. Only the ECH gene cluster was regulated by the CO-responsive transcriptional factor CooA, suggesting that others were separately regulated in the same transcriptional cascade as the ECH gene cluster. Of the 14 significantly expressed gene clusters under 100% N_2, ferrous iron transport gene cluster involved in anaerobic respiration and prophage region were found. Considering that the expression of the temperate phage was strictly repressed under 100% CO, hydrogenogenic CO metabolism might be stable for *C. pertinax*.

Introduction

Transcription of the genes in the CODH gene cluster in hydrogenogenic CO metabolism is activated by CO-responsive transcriptional factors CooA, RcoM (Kerby et al. 2008), and CorQR (Kim et al. 2015). Of these, CooA is well-characterized in hydrogenogenic carboxydotroph, *Rhodospirillum rubrum* (Aono et al. 1996; Roberts et al. 2004) and is present in most of the hydrogenogenic carboxydotrophs (Youn et al. 2004). CooA belongs to the cyclic AMP receptor protein family (Shelver et al. 1995). Homodimeric heme protein CooA is inactive in the absence of CO (Shelver et al. 1997). When CooA senses CO, ligand replacement in CooA leads to its conformational change making it active (Shelver et al. 1997). The active form of CooA can bind to the promoter region, thereby activating transcription through contacts with RNA polymerase (Leduc et al. 2001). In the *R. rubrum* genome, CODH gene cluster and ECH gene cluster are closely arranged and the CooA binding site is found in upstream of both *cooF* in the CODH gene cluster and *cooM* in the ECH gene cluster (Fox et al. 1996; Rajeev et al. 2012). Previous study shows that CooA homologs are divided into two phylogenetically distinct groups (CooA-1 and CooA-2). CooA-1 is found in the majority of CooA possessing carboxydotroph, whereas CooA-2 is found in some carboxydotrophs that possess multiple CODH gene clusters in their genomes (Techtmann et al. 2011). Further, CO binding assay of two CooA groups shows that both CooA-1 and CooA-2 are in their active forms in high CO concentration, whereas only CooA-2 is in active form even in high low CO concentration (Techtmann et al. 2011). This difference in CO activation in the two CooA groups enables the bacterium to regulate multiple CODH gene clusters across wide range of CO concentrations (Techtmann et al. 2011).

In addition to CooA, recent studies have reported other anaerobic CO responsive transcriptional factors RcoM and CO responsive regulatory system CorQR. Similar to CooA, RcoM whose gene is adjacent to CODH and ECH gene clusters in the hydrogenogenic carboxydotroph *Rubrivivax gelatinosus* (phylum Proteobacteria) (Wawrousek et al. 2014), possesses a potential CO-sensor domain containing heme (Kerby et al. 2008). CorQR pair is composed of CorQ with a DNA-binding domain of LysR-type transcriptional regulator (LTTR) family and CorR with 4-vinyl reductase domain instead of heme to sense CO (Kim et al. 2015). Genes encoding CorQR pair also flanked with CODH−ECH gene clusters in the hydrogenogenic carboxydotrophic archaeron, *Thermococcus onnurineus* (phylum Euryarchaeota) (Kim et al. 2015).

So far, only a few transcriptional studies about hydrogenogenic CO metabolism have been reported. A transcriptomic analysis of the sulfate-reducing carboxydotroph *Desulfovibrio vulgaris* using microarray reports that the expression of *cooS* is dependent on active CooA under low CO concentration (Rajeev et al. 2012). In addition, genome-wide primary transcriptome analysis of *T. onnurineus* shows that the expression of *cooS* gene is significantly upregulated when they utilize CO as the source of energy and carbon for H_2 production (Cho et al. 2017). However, there is limited data on comprehensive gene expression pattern of hydrogenogenic CO metabolism.

Despite of hydrogenogenic carboxydotrophy in *C. pertinax* (Yoneda et al. 2012), genome analysis in Chapter 2 shows that *C. pertinax* lacks genes encoding CODH-I catalytic subunit (CooS-I) and its transcriptional factor CooA-1 in their CODH-I−ECH gene cluster. Further, gene expression analysis in *C. pertinax* shows that CODH-II catalytic subunit gene (*cooS-II*) and distantly encoded ECH catalytic large and small subunit genes are remarkably upregulated under 100% CO, suggesting that *C.*

pertinax performs hydrogenogenic CO metabolism in which CODH-II couples with distal ECH. Since *C. pertinax* possesses one CooA homolog (CooA-2) unlike *C. hydrogenoformans* possessing two CooA homologs, their transcriptional regulation from CO response is expected to be simpler than *C. hydrogenoformans*. Therefore, the organism is helpful to understand the regulation mechanism of hydrogenogenic CO metabolism. In this Chapter, I performed whole transcriptome analysis by RNA sequencing (RNA-Seq) when they grow on pyruvate under 100% CO or100% N_2.

Materials and methods

Growth conditions for transcriptome analysis

To prepare RNA for transcriptome analysis, *C. pertinax* was grown at 65°C in modified DSM medium 507 under a headspace of 100% CO or 100% N_2 gas according to methods previously described in Chapter 2. I added sodium thiosulfate (final concentration, 1 g/L) as the terminal electron acceptor and pyruvate Na (final concentration, 2 g/L) as the electron donor and carbon source to the medium in both the gas phase conditions to collect enough cells for analysis. Growth was assessed by direct enumeration of SYBR Gold stained cells collected on 0.2 μm black polycarbonate membrane filters (Advantec, Tokyo, Japan) using a fluorescent microscope (Olympus, Tokyo, Japan). The cells exponentially grown in pre-culture were inoculated to fresh medium and cultivated routinely.

Transcriptome analysis of *C. pertinax*

When *C. pertinax* reached the late exponential phase as indicated by arrows (Fig. 3-1), 10 mL volume of the culture in both the conditions was collected and total RNA was extracted according to methods previously described in Chapter 2. After

removal of contaminating DNA using TURBO DNase (Invitrogen, Carlsbad, CA, USA), total RNA was purified using Agencourt RNAClean XP (Beckman Coulter, Brea, CA, USA) according to the manufacturer's instructions. Quantification was performed in the Agilent 2100 Bioanalyzer with Agilent RNA6000 pico kit (Agilent Technologies, Santa Clara, CA, USA). rRNA was depleted from the purified total RNA using Ribo-Zero™ Magnetic Kit (Bacteria) (Epicentre, Madison, WI, USA) according to the manufacturer's instructions. After rRNA removal, the purified total RNA, which was extracted from total 60 mL from the two replicates, was merged to obtain sufficient total RNA for RNA-Seq. Then, the merged total RNA was reverse transcribed to obtain double strand cDNA (ds cDNA) using PrimeScript Double Strand cDNA Synthesis Kit (TaKaRa Bio, Shiga, Japan) for RNA-Seq. The ds cDNA library was constructed with 75 bp paired-end libraries prepared by Nextera® XT DNA sample prep kit (Illumina, San Diego, CA, USA) and sequenced using Illumina MiSeq system (Illumina). For reverse transcription quantitative PCR (RT-qPCR), the purified total RNA was reverse transcribed to single strand cDNA (ss cDNA) using SuperScript III First-Strand Synthesis System (Invitrogen).

RNA-Seq data analysis

Generated high quality reads of *C. pertinax* with Q30 (sequence error rate lower than 0.1%) mapped to their draft genome (BDJK01000000) using Tophat version 2.0.13 (Trapnell et al. 2009) with Bowtie2 version 2.2.2 (mismatches $\leq$ 2 bp) (Langmead and Salzberg 2012). I manually removed contaminated rRNA reads which showed high similarity to *C. pertinax* 5S, 16S, and 23S rRNA. Read count per gene under both conditions was estimated using featureCounts (Liao et al. 2013). Rarefaction curves for reads under both conditions were constructed using PAST ver.3.17 (Hammer

et al. 2001). Gene expression levels were compared and normalized using the R statistical package edgeR (Robinson et al. 2010). Differential expression genes (DEGs) under 100% CO or 100% N_2were considered as significantly expressed when their log2 fold change was >1 or <-1, respectively and their false discovery rate (FDR) adjusted *P* value (*Q* value) was < 0.05.

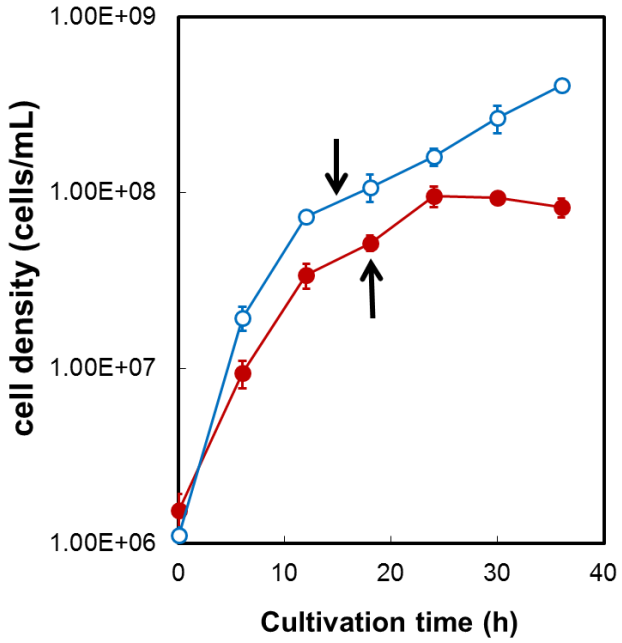


Fig. 3-1 Growth of *C. pertinax* under 100% CO and 100%N_2
Red closed circles, cell density under 100% CO; blue open circles, cell density under 100%N_2. Total RNA was extracted from *C. pertinax* when the cells reached late exponential phase as indicated by arrow. Plots represent the mean of three biological replicates. Error bars represent standard deviations.

Analysis to predict the function of DEGs

As a unit of polycistronic transcription, a gene cluster was predicted based on the following criteria: (i) intergenic distance between genes was less than 300 bp, (ii) genes were encoded in the same strand, and (iii) the visual pattern of read alignment to a

locus on *C. pertinax* genome. IGV software was used to visualize the patterns of read alignment (Thorvaldsdóttir et al. 2013). To predict the functions of DEGs, these protein sequences were annotated using BLASTp search (Altschul et al. 1990; Altschul et al. 1997) with an e-value of $1e^{-5}$ at an effective database size of 10^7 against the COG database (Tatusov 2000). The upstream regions (300 bp) of the predicted gene cluster containing DEGs and solely transcribed DEGs were collected from the *C. pertinax* draft genome. The primary sigma factor recognition sequences were predicted using BPROM with default parameters (Solovyev and Salamov 2011).

The promoter sequences were aligned separately for each motif (-10 motif and -35 motif). Logos were prepared using weblogo for visualization (Crooks et al. 2004). To discover motifs in the upstream regions, I used MEME ver5.0.2 program with default parameters (Bailey and Elkan 1994).

Prophage regions encoded in contigs (cutoff <1500 bp) from the genomes of the three *Carboxydothermus* species except for the already analyzed *C. hydrogenoformans* (Wu et al. 2005) were predicted using PHASTER (Arndt et al. 2016). To classify the predicted prophages in three *Carboxydothermus* species based on genome-wide similarities, a viral proteomic tree was generated by ViPTree (Nishimura et al. 2017b). The all-against-all distance matrix between viral reference genomes and the predicted phages was calculated on the basis of the normalized bit score of tBLASTx (S_G)(Nishimura et al. 2017a), and the proteomic tree was built with BIONJ using the distance matrix in the ViPTree.

RT-qPCR validation for the expression of CO metabolism related genes

To validate the RNA-Seq expression data, 10 predicted CO metabolism related genes were selected for RT-qPCR. RT-qPCR primers used have been described in

Chapter 2. PCR amplification was performed as Chapter 2. I served same cDNA samples as RNA-Seq to RT-qPCR. The detection limit of the transcript levels of mRNA was 1.00×10^2 copies/μL. Relative transcript amounts were calculated using *rrsD* (16S rRNA) transcripts as internal standard. All RT-qPCR data represent the mean value of at least triplicate biological determinations.

Accession number of the sequence

RNA-Seq data have been submitted to the DNA Data Bank of Japan Sequence Read Archive under the accession no. DRA007734.

Results

Overview of RNA-Seq

C. pertinax grew under both 100% CO and 100% N_2 conditions (Fig. 3-1). Under 100% CO (carboxydotrophic growth), doubling time was 2.3 h with a final cell density of $8.25\times10^7 \pm 1.05\times10^7$ cells/mL. Under 100% N_2 (heterotrophic growth), doubling time was 1.9 h with a final cell density of $4.07\times10^8 \pm 1.68\times10^7$ cells/mL. Total RNA was extracted from cells at late exponential phase in both conditions and subjected to RNA-Seq (Fig. 3-1). A total of approximately 3.54 M and 3.74 M high quality reads were obtained from 100% CO and 100% N_2 conditions, respectively. The total number of the mapped reads to the *C. pertinax* draft genome was 2.28 M and 1.95 M with 100% CO and 100% N_2, respectively (Table 3-1). Rarefaction curves of the mapped reads under both conditions showed that sequencing data from each condition was exhaustive to describe the transcriptional profile when read counts reached in approximately 1.0 M (Fig. 3-2), suggesting that the count of mapped genes was sufficient for RNA-Seq analysis under both conditions.

Table 3-1 Summary of sequencing data

	CO 100%	$N_2$100%
Raw reads	4,313,481	4,536,866
Q30 filtered reads	3,544,448	3,741,696
rRNA removed read	3,544,048	3,740,081
Mapped reads	2,279,615	1,947,491
Read count	2,150,439	1,839,810

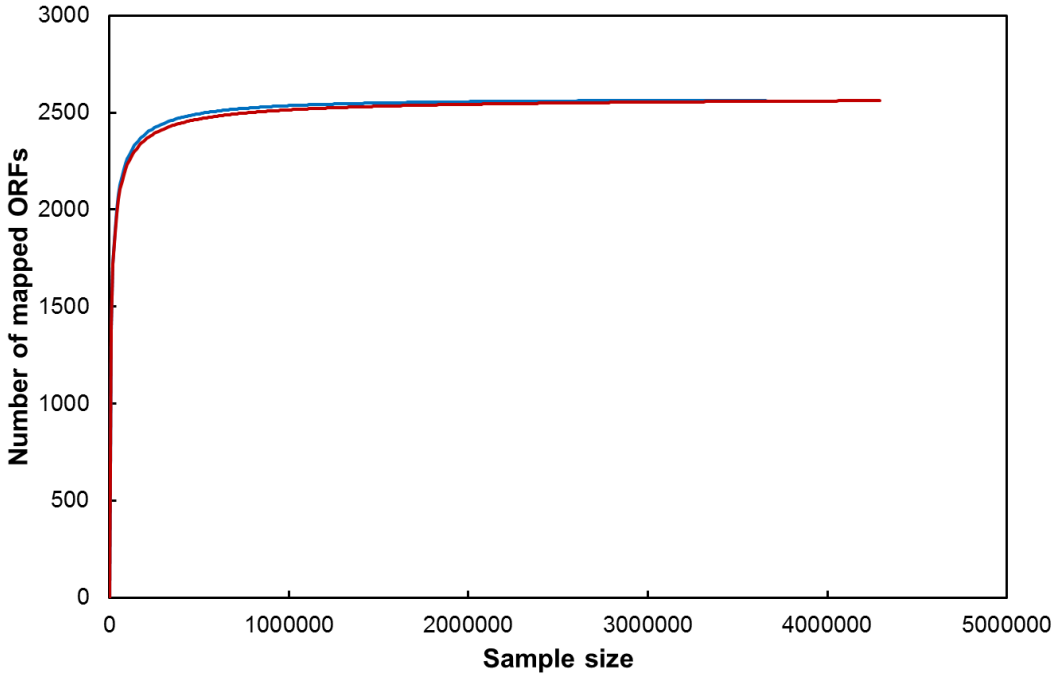


Fig. 3-2 Rarefaction curves of mapped reads from the cells grown under 100% CO and 100%N_2

Red line, 100%CO; blue line, 100%N_2.

Generally, central component of bacterial transcription is multi-subunit-DNA-dependent RNA polymerase. When I searched for sigma factors in *C. pertinax*, 13 sigma factors (one primary sigma factor, 11 alternative sigma factors as sigma^{70} family, and one sigma^{54}) were found (Table 3-2). Read count of the primary sigma factor was 6495 and 2855 under 100% CO and 100% N_2, respectively. These read counts were an order

of magnitude higher among the other sigma^{70} factors including sigma^{70} factors for sporulation under both conditions (Table 3-2). These results clearly indicated that the obtained RNA-Seq reads reflected expression patterns of exponentially growing cells.

Table 3-2 Read counts of sigma factors in *C. pertinax* under CO 100% and N_2 100%

		Read count	
sigma70 family	genes	CO 100%	N_2 100%
Primary sigma			
cpu_18440	*rpoD*	6495	2855
None essential primary like sigma			
N/A			
Alternative sigma			
Flagella sigma			
cpu_10900	*FliA*	922	474
cpu_13510	*FliA*	2	19
cpu_22710	*sigI*	123	106
Extracytoplasmic function sigma			
cpu_03980	*sigW*	203	258
cpu_17570	*rpoE*	19	14
cpu_24000	*rpoE*	127	80
Heat shock sigma			
N/A			
Sporulation sigma			
cpu_02980	*sigE*	36	38
cpu_02990	*sigG*	2	27
cpu_08440	*sigF*	40	48
cpu_14900	*sigK*	15	11
cpu_16380	*sigH*	850	637
sigma54 family			
cpu_00600	*rpoN*	2731	1276

N/A, not applicable

Under 100% CO or 100% N_2 conditions, 36 genes and 64 genes were identified as DEGs (*Q* value < 0.05), respectively (Fig. 3-3). A ratio of the 2,577 genes in *C. pertinax* draft genome to the DEGs was no more than 1.4% under 100% CO and 2.5% under 100% N_2 conditions, respectively. Seventy-four % of the DEGs under 100% CO and 91% of the DEGs under 100% N_2 were components of 9 and 14 gene clusters, respectively (Table 3-3). The remaining 8 DEGs under 100% CO and 6 DEGs under 100% N_2 were solely transcribed. As significantly expressed gene clusters are expected to play key roles in both the conditions, I classified them into three categories based on average read counts per kilobase of the genes in the gene clusters containing DEGs (Table 3-3): very high (average read counts ≥ 10,000), high (average read counts ≥ 1,000), and moderate (average read counts ≥ 100). Next, I described about these categorized gene clusters.

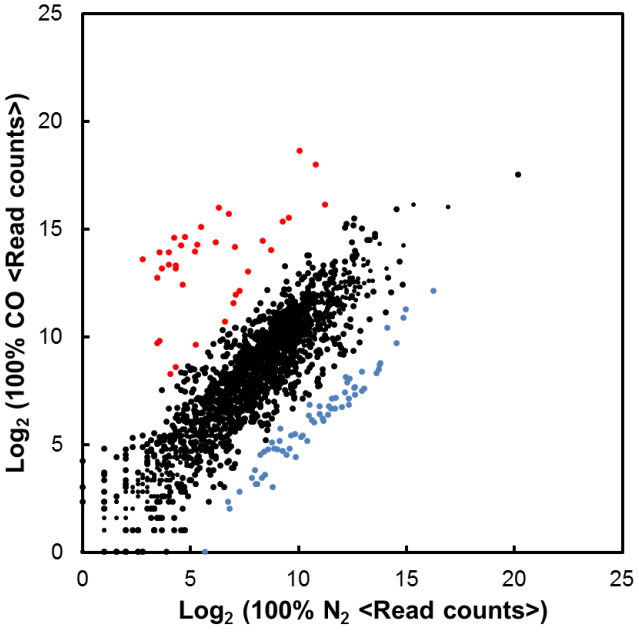


Fig. 3-3 Comparison of read counts between *C. pertina*x grown under 100% CO and 100% N_2

Each plot shows an open reading frame. Red plots, upregulated DEGs; blue plot, downregulated DEGs; black, unchanged genes.

Table 3-3 Overview of gene clusters containing DEGs under 100% CO and 100% N_2

metabolism category	locus tag	predicted function	fold change	No. of genes in gene cluster	No. of DEGs in gene cluster	average read count		expression category
						CO 100%	N_2 100%	
Significantlly expressed gene cluster under CO 100%								
Energy conservation	cpu_03670-03750	ECH	480-1760	9	9	35565	48	very high
	cpu_14030-14050	CODH-II	70-380	3	3	97573	405	very high
	cpu_22890-22980	H_2-uptake Ni-Fe-hydrogenase	20	1	10	471	118	moderate
Carbohydrate metabolism	cpu_07810-07850	anion/Na+ symporter, fumarate hydratase, and succinate dehydrogenase	460-960	5	5	12770	23	very high
	cpu_15750-15760	succinate dehydrogenase	420-640	2	2	21045	40	very high
Transcriptional regulation	cpu_03570-03580	TetR/AcrR family transcriptional regulator	20-30	1	2	593	46	moderate
	cpu_05930-05940	LysR-type transcriptional regulator	60-70	2	2	45670	710	very high
Unknown	cpu_03830-03840	unknown	40	2	2	22067	561	very high
	cpu_23290-23300	unknown	10	2	2	4326	58	high
Significantlly expressed gene cluster under N_2 100%								
Energy conservation	cpu_18560-18580	dimethyl sulfoxide reductase	30-40	3	3	205	6819	high
Prophage region	cpu_00830-00840			2	2	971	40766	very high
	cpu_00860-00940			9	9	465	15716	very high
	cpu_00950-01010			7	5	32	852	moderate
	cpu_01060-01070	myoviridae like prophage region	10-50	2	1	163	2382	high
	cpu_01080-01100			3	3	86	1504	high
	cpu_01110-01130			3	3	217	5618	high
	cpu_01140-01200			6	6	139	3903	high
	cpu_01220-01350			12	10	89	1428	high
Amino acid synthesis	cpu_11010-11030	aromatic amino acid biosynthesis	20-40	3	3	74	2021	high
	cpu_11630-11700	aromatic amino acid biosynthesis	20-30	8	4	26	292	moderate
Membrane transport	cpu_18330-18350	ferrous iron transport	10-20	3	2	1174	12640	very high
Unkown	cpu_13490-13500	unknown	20-50	2	2	35	1372	high
	cpu_24930-24940	unknown	20	2	2	1862	30424	very high

Significantly expressed gene clusters under 100% CO or 100% N_2

Of the six gene clusters in the 'very high' category under 100% CO (Table 3-3), the two gene clusters (cpu_14030-14050 and cpu_03670-03750; 70-380 and 480-1760 fold changes, respectively) were related to energy conservation and encoded CODH-II and ECH, which are essential proteins for hydrogenogenic CO metabolism as described in Chapter 2 (Fig. 3-4; Table 3-4). Further, the two gene clusters were related to carbohydrate metabolism (460-960 and 420-640 fold changes, respectively; Table 3-4). Of these, the gene cluster (cpu_07810-07850) composed of five genes encoding divalent anion/Na^+ symporter, fumarate hydratase α and β subunits, and succinate dehydrogenase flavoprotein, and iron sulfur subunit. The other gene cluster (cpu_15750-15760) composed of two genes encoding succinate dehydrogenase iron sulfur and cytochrome b subunit. These gene products catalyze the oxidoreductive reaction between succinate and malate in tricarboxylic acid (TCA) cycle. The gene cluster (cpu_05930-05940; 60-70 fold change) was composed of two genes encoding LTTR, involved in transcriptional regulation, and a hypothetical protein. In addition, a gene encoding another LTTR (cpu_07800) was also identified as a sole DEG (220 fold change). The function of rest of the gene cluster (cpu_03830-03840) was not predicted. Consequently, products of these six gene clusters were considered to play an important and specific role in their hydrogenogenic CO metabolism.

Next, we annotated three other categories (high or moderate; Table 3-3 and 3-4). These three gene clusters were composed of the following genes: two hypothetical genes (cpu_23290-23300; high), 10 H_2-uptake Ni-Fe-hydrogenase genes (cpu_22890-22980; moderate), and TetR/AcrR family transcriptional regulator gene and sodium/phosphate symporter (cpu_03570-03580; moderate).

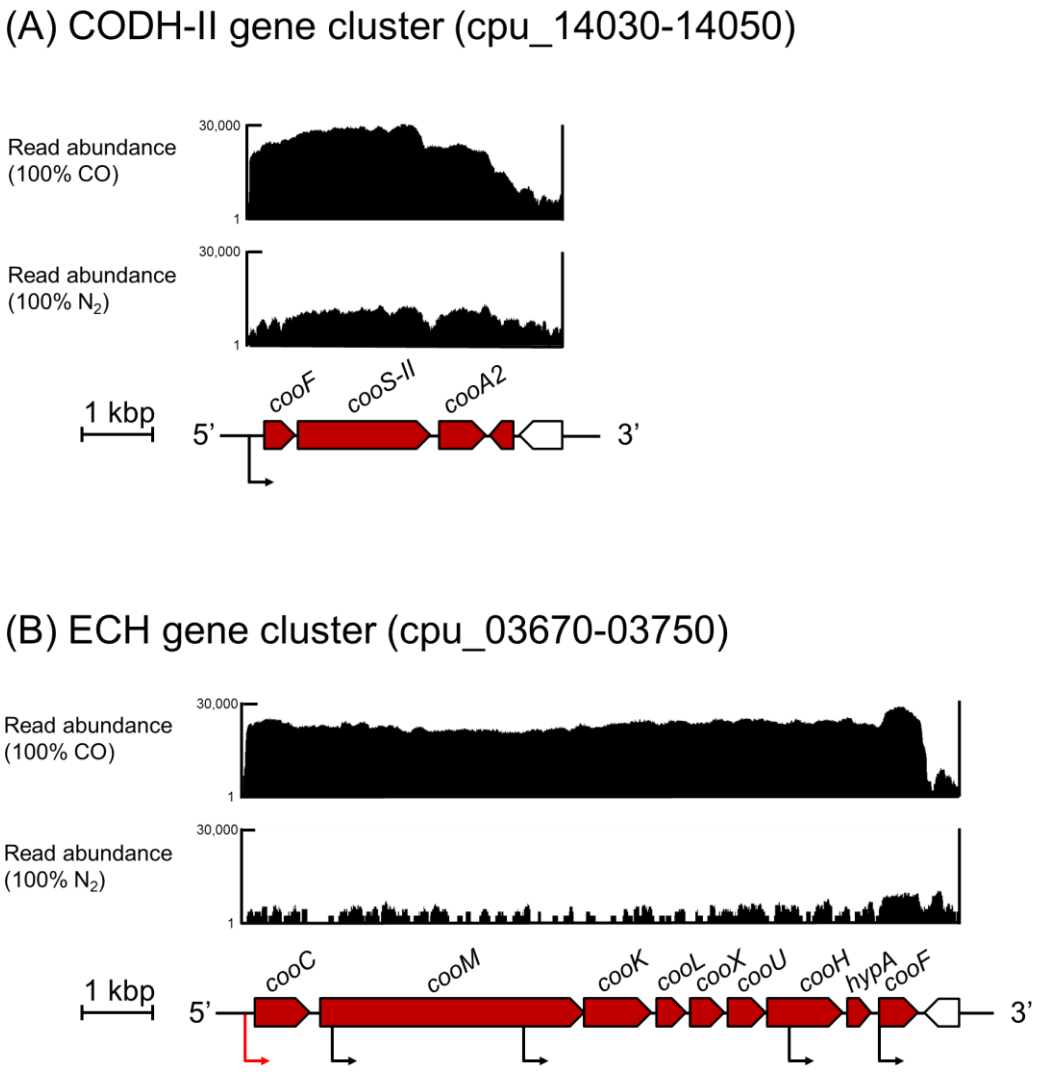


Fig. 3-4 Transcriptomic read mapping pattern of the CODH-II gene cluster (A) and the ECH gene cluster (B).

Part of genome-wide overview of reads mapped to the *C. pertinax* draft genome at samples in the CODH-II gene cluster (A) and the ECH gene cluster (B). Read abundance was displayed in a logarithmic scale of 1 to 30,000 by Integrative Genomics Viewer (Thorvaldsdóttir et al 2013). Red box, DEG under CO 100%; White box, not DEG under CO 100%. Positions of predicted sigma^{70} promoter sequence was predicted using BPROM with high linear discrimination function score (>5; Solovyev and Salamov 2011). Black arrows, the position of predicted sigma^{70} promoter sequence. Red arrow, the position of predicted sigma^{70} promoter sequence with already known CooA binding site.

Table 3-4 List of DEGs containing gene cluster under 100% CO

category	locus tag	annotation	read counts		fold change	*Q* value[a]
			100% CO	100% N_2		
Energy conservation	cpu_03670	CO dehydrogenase maturation factor CooC	24981	27	925	1.52E-07
	cpu_03680	Ni,Fe-hydrogenase subunit CooM	63907	79	809	1.59E-07
	cpu_03690	Ni,Fe-hydrogenase subunit CooK	24667	19	1298	9.67E-08
	cpu_03700	Ni,Fe-hydrogenase small catalytic subunit CooL	12330	7	1761	9.67E-08
	cpu_03710	Ni,Fe-hydrogenase subunit CooX	15301	12	1275	1.06E-07
	cpu_03720	Ni,Fe-hydrogenase component CooU	18965	24	790	1.74E-07
	cpu_03730	Ni,Fe-hydrogenase large catalytic subunit CooH	34400	45	764	1.74E-07
	cpu_03740	Zn finger protein HypA	9133	13	703	3.56E-07
	cpu_03750	Fe-S-cluster-containing hydrogenase component CooF	53076	110	483	6.11E-07
	cpu_14030	Fe-S-cluster-containing hydrogenase component CooF	21388	73	293	3.25E-06
	cpu_14040	CO dehydrogenase catalytic subunit CooS-II	402009	1065	377	1.08E-06
	cpu_14050	cAMP-binding protein CooA	22376	330	68	6.77.E-04
	cpu_22890	Hydrogenase maturation factor	601	188	3	1.00
	cpu_22900	Hydrogenase maturation factor	861	187	5	0.71
	cpu_22910	Hydrogenase maturation factor	97	36	3	1.00
	cpu_22920	Hydrogenase maturation factor	1641	741	2	1.00
	cpu_22930	Hydrogenase maturation factor	255	121	2	1.00
	cpu_22940	Zn finger protein HypA	39	5	8	0.502
	cpu_22950	Ni,Fe-hydrogenase maturation factor	306	17	18	3.96.E-02
	cpu_22960	Membrane anchoring protein	312	27	12	0.118
	cpu_22970	Ni,Fe-hydrogenase large catalytic subunit	364	34	11	0.139
	cpu_22980	Ni,Fe-hydrogenase small catalytic subunit	243	26	9	0.195
Carbohydrate metabolism	cpu_07810	Anion/Na+ symporter	15296	16	956	1.52E-07
	cpu_07820	Fumarate hydratase subunit alpha	9113	20	456	8.33E-07
	cpu_07830	Fumarate hydratase subunit beta	6738	11	613	4.62E-07
	cpu_07840	Succinate dehydrogenase flavoprotein subunit SdhA	19435	40	486	6.11E-07
	cpu_07850	Succinate dehydrogenase iron sulfur subunit SdhB	10128	20	506	6.11E-07
	cpu_15750	Succinate dehydrogenase iron sulfur subunit SdhB	10224	16	639	5.00E-06
	cpu_15760	Succinate dehydrogenase cytochrome b subunit SdhC	15629	37	422	7.00E-106
Transcriptional regulation	cpu_03570	TetR/AcrR family transcriptional regulator	387	20	19	2.81.E-02
	cpu_03580	Hypothetical protein	259	29	9	0.224588

Table 3-4 Continued

	cpu_05930	LysR-type transcriptional regulator	41169	616	67	6.77.E-04
	cpu_05940	Hypothetical protein	46738	755	62	7.86.E-04
Unknown	cpu_03830	Hypothetical protein	8223	206	40	2.62.E-03
	cpu_03840	Hypothetical protein	16649	429	39	2.70.E-03
	cpu_23290	Hypothetical protein	825	11	75	7.65.E-04
	cpu_23230	Hypothetical protein	883	12	74	8.27.E-04

[a] DEGs were considered when their fold change was >2 and *Q* value was < 0.05

Of the four gene clusters in 'very high' category under 100% N_2 (Table 3-3 and 3-5), the two gene clusters (cpu_00830-00840 and cpu_00860-00940; 40 and 30-50 fold changes, respectively) were a part of the prophage region in *C. pertinax* (Table 3-5). From viral proteomic tree using ViPTree, this prophage region (54 genes) was close to P2-like *Myoviridae* viruses, *Vibrio* virus X29 (NC_024369) and *Vibrio* phage phi 2 (KJ545483) (Fig.3-5). The gene cluster (cpu_18330-18350; 10-20 hold change) was composed of two genes encoding ferrous iron transport protein A and B (Lau et al. 2015). In addition, a gene (cpu_17460) encoding ferrous iron permease was identified as solely DEG (10 fold change). The function of the rest of the gene cluster (cpu_24930-24940) could not be predicted from their gene products. Products of these four gene products were also considered to play an important role in their heterotrophic growth.

Of the 10 gene clusters in high or moderate categories under 100% N_2 (Table 3-3 and 3-5), the gene cluster (cpu_18560-18580; high) was involved in energy conservation and encoded dimethyl sulfoxide reductase catalytic molybdopterin subunit, iron-sulfur cluster protein, and integral membrane protein. The six gene clusters (cpu_01060-01070, cpu_01080-01100, cpu_01110-01130, cpu_01140-01200,

and cpu_01220-01350; high, cpu_00950-01010; moderate) were found in the prophage region. Further, the two gene clusters (cpu_11010-11030, high; cpu_11630-11700, moderate) were involved in aromatic amino acid biosynthesis. The function of the rest of the gene cluster (cpu_13490-13500; high) was not predicted from their gene products.

Table 3-5 List of DEGs containing gene cluster under 100% N_2

category	locus tag	annotation	read counts		fold change	*Q* value[a]
			100% CO	100% N_2		
Energy conservation	cpu_18560	Dimethyl sulfoxide reductase catalytic molybdopterin subunit	437	14202	32	4.26.E-03
	cpu_18570	Dimethyl sulfoxide reductase iron-sulfur cluster protein	199	6235	31	0.0185
	cpu_18580	Dimethyl sulfoxide reductase integral membrane protein	105	4123	39	0.455
Prophage region	cpu_00830	Predicted transcriptional regulator	194	8618	44	6.77.E-04
	cpu_00840	Hypothetical protein	314	12604	40	7.86.E-04
	cpu_00860	Hypothetical protein	138	5249	38	8.62.E-04
	cpu_00870	Predicted transcriptional regulator	180	7856	44	6.77.E-04
	cpu_00880	Hypothetical protein	73	2227	31	1.71.E-03
	cpu_00890	Hypothetical protein	108	2752	25	2.70.E-03
	cpu_00900	Hypothetical protein	201	6354	32	1.46.E-03
	cpu_00910	Hypothetical protein	828	24265	29	1.74.E-03
	cpu_00920	Bacteriophage protein gp37	359	13483	38	8.62.E-04
	cpu_00930	FOG: LysM repeat	101	3313	33	1.36.E-03
	cpu_00940	Hypothetical protein	113	5213	46	6.73.E-04
	cpu_00950	DNA replication protein	21	942	45	7.91.E-04
	cpu_00960	Single-stranded DNA-binding protein	5	108	22	1.12.E-02
	cpu_00970	Hypothetical protein	3	12	4	0.97
	cpu_00980	Hypothetical protein	1	7	7	0.94
	cpu_00990	Hypothetical protein	1	51	51	1.40.E-02
	cpu_01000	Hypothetical protein	26	615	24	3.90.E-03
	cpu_01010	Holliday junction resolvase - archaeal type	28	767	27	2.70.E-03
	cpu_01060	Hypothetical protein	37	361	10	5.69.E-02
	cpu_01070	Phage DNA packaging protein	69	1589	23	3.67.E-03

Table 3-5 Continued

Prophage region	cpu_01080	Bacteriophage tail assembly protein	115	1466	13	2.11.E-02
	cpu_01090	Hypothetical protein	11	234	21	8.05.E-03
	cpu_01100	Bacteriophage capsid protein	138	3000	22	4.00.E-03
	cpu_01110	Protease subunit of ATP-dependent Clp proteases	168	4557	27	2.25.E-03
	cpu_01120	Hypothetical protein	83	2644	32	1.46.E-03
	cpu_01130	Hypothetical protein	268	5340	20	5.29.E-03
	cpu_01140	Hypothetical protein	4	112	28	6.13.E-03
	cpu_01150	Hypothetical protein	40	1126	28	2.39.E-03
	cpu_01160	Hypothetical protein	102	2971	29	1.89.E-03
	cpu_01170	Hypothetical protein	68	2316	34	1.33.E-03
	cpu_01180	Hypothetical protein	42	1187	28	2.24.E-03
	cpu_01190	Phage tail sheath protein FI	331	7902	24	3.09.E-03
	cpu_01200	Phage tail tube protein FII	65	1711	26	2.62.E-03
	cpu_01210	Hypothetical protein	11	321	29	3.17.E-03
	cpu_01220	Hypothetical protein	27	519	19	7.21.E-03
	cpu_01230	Phage-related tail protein	142	3437	24	3.09.E-03
	cpu_01240	P2-like prophage tail protein X	9	273	30	3.12.E-03
	cpu_01250	Phage protein D	85	2043	24	3.15.E-03
	cpu_01260	Phage P2 baseplate assembly protein gpV	43	821	19	6.66.E-03
	cpu_01270	Phage protein U	34	437	13	2.31.E-02
	cpu_01280	Phage baseplate assembly protein W	12	349	29	3.15.E-03
	cpu_01290	Phage-related baseplate assembly protein	81	1424	18	7.84.E-03
	cpu_01300	Bacteriophage P2-related tail formation protein	36	560	16	1.28.E-02
	cpu_01310	Phage-related tail fibre protein	211	2041	10	5.10.E-02
	cpu_01320	Hypothetical protein	140	1181	8	7.59.E-02
	cpu_01330	Hypothetical protein	25	344	14	2.04.E-02
	cpu_01340	Phage-related holin (Lysis protein)	27	385	14	1.85.E-02
	cpu_01350	Putative peptidoglycan-binding domain-containing protein	45	930	21	5.32.E-03
Amino acid synthesis	cpu_11010	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase AroA	23	700	30	1.99.E-03
	cpu_11020	Prephenate dehydratase PheA	36	1371	38	1.01.E-03
	cpu_11030	Fructose 1,6-bisphosphatase	109	2053	19	6.38.E-03
	cpu_11630	Tryptophan synthase alpha chain TrpA	58	320	6	0.225
	cpu_11640	Tryptophan synthase beta chain TrpB	53	577	11	3.77.E-02
	cpu_11650	Phosphoribosylanthranilate isomerase TrpF	14	258	18	9.46.E-03

Table 3-5 continued

Amino acid	cpu_11660	Indole-3-glycerol phosphate synthase TrpC	15	127	8	0.100
synthesis	cpu_11670	Anthranilate phosphoribosyltransferase TrpD	8	95	12	5.69.E-02
	cpu_11680	Anthranilate/para-aminobenzoate synthases component II PabA	9	83	9	0.108
	cpu_11690	Anthranilate/para-aminobenzoate synthases component I TrpE	28	472	17	1.08.E-02
	cpu_11700	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	9	264	29	3.32.E-03
Membrane	cpu_18330	Fe^{2+} transport system protein A	232	4930	21	4.26.E-03
transport	cpu_18340	Fe^{2+} transport system protein B	1375	18020	13	1.85.E-02
	cpu_18350	Fe^{2+} transport system protein B	996	3659	4	0.455
Unknown	cpu_13490	Hypothetical protein	7	154	22	8.69.E-03
	cpu_13500	Hypothetical protein	8	450	56	6.77.E-04
	cpu_24930	Hypothetical protein	1867	29942	16	9.46.E-03
	cpu_24940	Hypothetical protein	278	4763	17	8.03.E-03

[a] DEGs were considered when their fold change was >2 and *Q* value was < 0.05

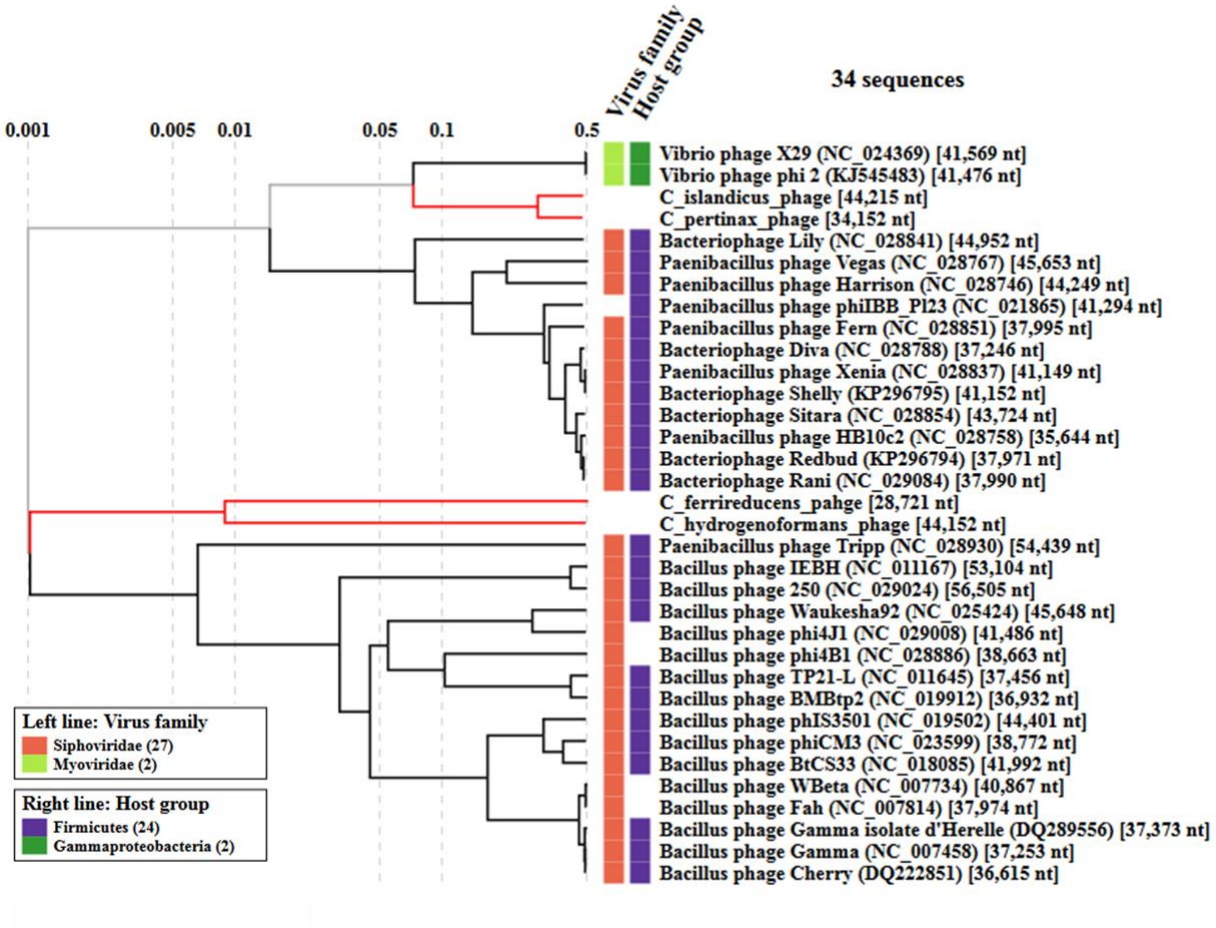


Fig. 3-5 A part of proteomic tree with four phages from four *Carboxydothermus* species

Branch lengths are logarithmically scaled from the root of the entire proteomic tree in VipTree database. Red branch represents phage in 4 *Carboxydothermus* species.

Expressions of genes involved in hydrogenogenic CO metabolism

Apart from largely up-regulated CODH-II gene cluster, RNA-Seq showed no significant expression changes of the CODH (CODH-III to V). In addition to ECH, *C. pertinax* possesses the H_2-uptake Ni-Fe-hydrogenase gene cluster. In this gene cluster, read count of all genes were relatively low in both the conditions and only one gene (cpu_22950) encoding maturation factor for H_2-uptake Ni-Fe-hydrogenase was identified as DEG.

To validate RNA-Seq, I performed gene expression analysis for 10 genes (*coos-IIA2F1* in CODH-II gene cluster, *cooS-III* to *V* in CODH-II to V gene cluster, *cooLH* in ECH gene cluster, and *hyaAB* in H_2-uptake Ni-Fe-hydrogenase gene cluster) using RT-qPCR (Fig. 3-6). The relative transcript levels of five genes (*cooS-II*, *cooA2*, *cooF1*, *cooH*, and *cooL*) under 100% CO were 20-640 folds higher than those under 100% N_2 (Fig. 3-6). Relative transcript levels of the three remaining genes (*cooS-III* to *V*) were not changed. Of them, relative transcript levels of *cooS-IV* were below the limit of detection (1.00×10^2 copies/μL). Further, relative transcript levels of the two genes (*hyaA* and *hyaB*) were relatively higher under 100% CO. Consequently, RT-qPCR analysis for these genes supported the RNA-Seq data (Fig. 3-6).

Analysis of regulation mechanism under hydrogenogenic CO metabolism

To predict the regulation mechanism under hydrogenogenic CO metabolism, I explored transcription start site (TSS) from 300 bp upstream of 36 transcriptional

regions (23 gene clusters containing DEGs and 13 solely DEGs), except for two transcriptional regions, which were significantly transcribed by the polar effect of their upstream. I assigned 36 TSSs from all promoter regions, regardless of their transcriptional categories (very high, high, and moderate). The 5'-untranslated region lengths of these transcripts were ranged from 19 to 239 bp. The -35 motif and -10 motif (Wösten et al. 1998) for primary sigma factor were strictly conserved among all the promoter regions in both the conditions (Fig. 3-7).

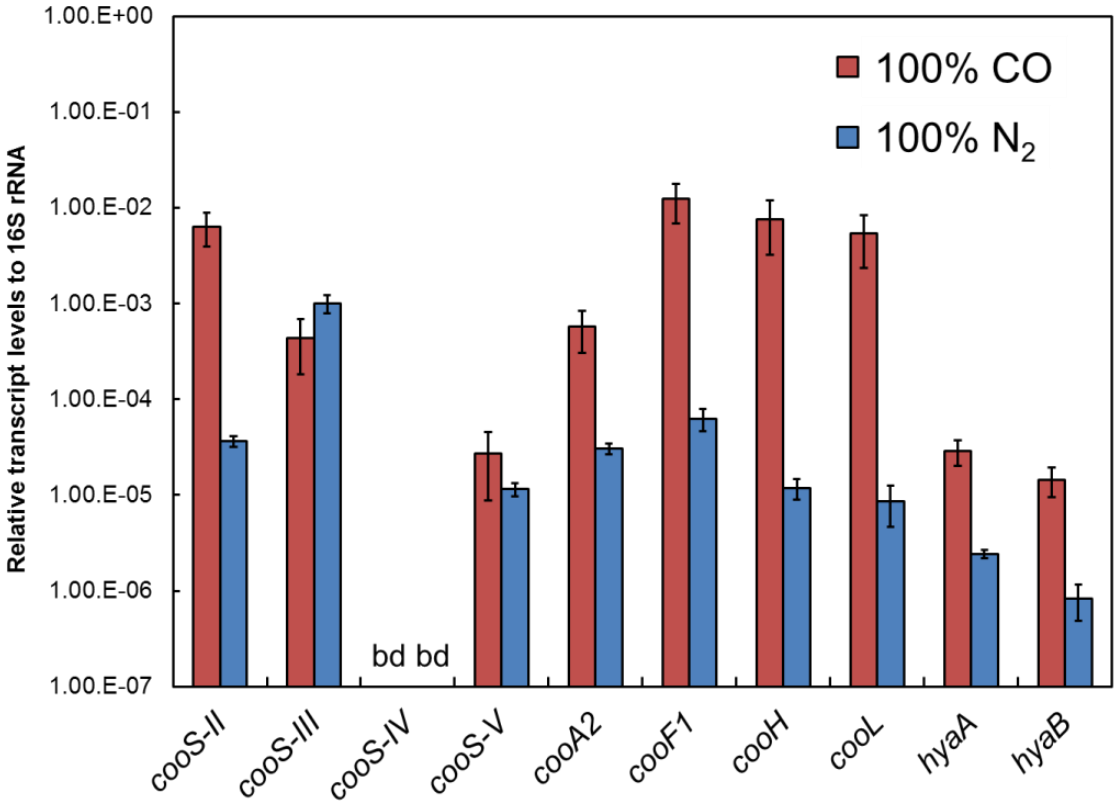


Fig. 3-6 Transcript levels relative to 16S rRNA using RT-qPCR method under 100% CO and 100% N_2

C. pertinax transcript levels relative to the 16S rRNA gene are shown. Red bar, cells grown under 100% CO; blue bar, cells grown under 100% N_2. bd, below the limit of detection. Bars represent the mean of at least three biological replicates. Error bars represent standard deviations.

Among the 16 promoter regions of the significantly expressed transcriptional regions under 100% CO, the CooA binding site was conserved only in the ECH gene cluster. Expect for the ECH gene cluster, I searched conserved binding motifs for transcriptional factor among other promoter regions. However, there is no consensus motif with a significantly low E-value in the three transcriptional categories. Among the 20 promoter regions of the significantly expressed transcriptional regions under 100% N_2, there is no consensus motif with a significantly low E-value in three transcriptional categories.

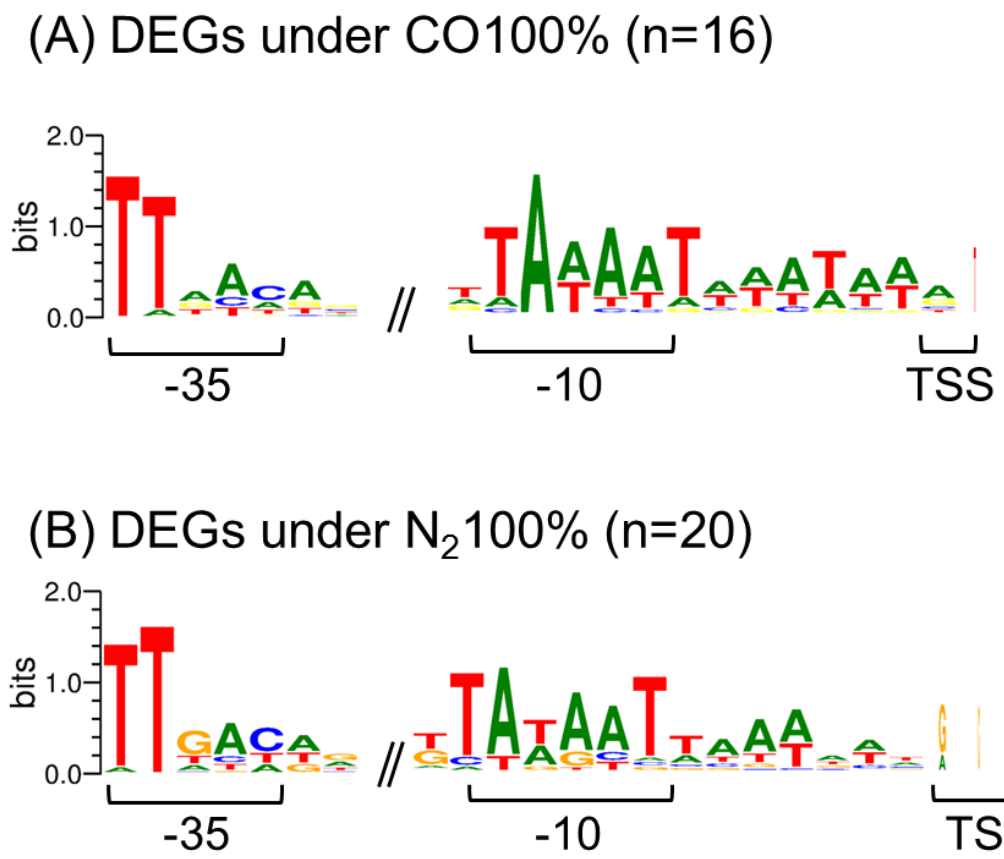


Fig. 3-7 Consensus sequence for primary sigma factors among upstream regions of the significantly expressed operons under 100% CO (A) and 100% N_2 (B)
The -35 and -10 motifs were identified from upstream regions of DEGs containing gene cluster and sole DEGs. The height of the letters is proportional to their frequency.

Discussion

In this study, I performed the whole transcriptome analysis of *C. pertinax* possessing only one CO-responsive transcriptional factor (CooA-2) as the first RNA-Seq report in hydrogenogenic carboxydotrophic bacterium. Total RNA from *C. pertinax* cells which grew under 100% CO or 100% N_2 were compared to understand the regulation mechanism between hydrogenogenic CO metabolism and heterotrophic metabolism. RNA-Seq data suggested that *C. pertinax* switched its metabolism by considerable expression changes in a relatively low number of gene clusters.

C. pertinax shows that electrons from CO oxidation are consumed to reduce H_2O to H_2; the remaining electrons are utilized in the reduction of thiosulfate by thiosulfate reductase (cpu_06910-06930) as described in Chapter 2. Genes involved in the transport of ferrous iron (very high category) were largely up-regulated under in 100% N_2, suggesting that the ferrous irons were positively transported into cells in heterotrophic growth. Considering that ferrous irons were incorporated as hemes or iron-sulfur clusters in various energy-generating and regulatory proteins (Braun and Hantk 2011), *C. pertinax* could conserve energy via anaerobic respiration in the modified DSM medium 507 containing sodium thiosulfate. In contrast, in hydrogenogenic CO metabolism, transport of ferrous iron (very high category) was significantly and incompletely reduced to low expression (Table 3-3). Instead, both CODH-II and ECH were largely upregulated. In addition, expressions of the genes encoding thiosulfate reductase were unchanged (data not shown). Therefore, owing to the sensitivity of heme protein to CO, *C. pertinax* would switch from anaerobic respiration to hydrogenogenic CO metabolism retaining a part of reducing power from CO oxidation to reduction of thiosulfate. In addition, owing to the high sensitivity of

CooA-2 to CO (Techtmann et al. 2011), *C. pertinax* would switch their metabolisms in response to low CO concentration.

Generally, carboxydotrophs can fix CO_2 or CO to acetyl-CoA via Wood-Ljungdahl pathway (Ragsdale 2004). In fact, transcriptome analysis of the acetogenic carboxydotroph, *Clostridium ljungdahlii* shows that a gene cluster containing most of the genes for Wood-Ljungdahl pathway is significantly upregulated in their CO metabolism (Tan et al. 2013). In RNA-Seq data, however, expressions of genes in CODH-III−ACS gene cluster were not up-regulated, regardless of CO addition. This result suggested that *C. pertinax* could fix carbon via other pathways in response to CO. Actually, a frameshift of CODH catalytic subunit gene in two hydrogenogenic carboxydotrophs suggested that carbon fixation via Wood-Ljungdahl pathway is not essential for hydrogenogenic CO metabolism (Wu et al. 2005; Omae et al. 2017).

In addition to Wood-Ljungdahl pathway, *C. pertinax* is predicted to possess an incomplete TCA cycle and an incomplete 3-hydroxypropionate cycle as described in Chapter 2. Among the incomplete TCA cycle related genes, genes for fumarate hydratase (EC: 4.2.1.2; cpu_07820-07830 and cpu_21290-21300) and succinate dehydrogenase (EC: 1.3.5.1; cpu_078240-07850, cpu_15750-15760, and cpu_21260-21280) are multi copies (Fukuyama et al. 2017). Of these gene sets, one encoding fumarate hydratase (cpu_07820-07830) and succinate dehydrogenase (cpu_07840-07850 and cpu_15750-15760; Table 3-4) were largely expressed under 100% CO (very high category), implying these gene clusters were up-regulated in response to CO and products of these genes enhanced carbon fixation via a part of reductive TCA cycle. Further, *C. pertinax* can conserve energy via their

hydrogenogenic CO metabolism in which CO oxidation is coupled with H_2 production and CO_2 is generated as a by-product of this reaction (Yoneda et al. 2012; Fukuyama et al. 2018). Therefore, I hypothesized that, in addition to Wood-Ljungdahl pathway, *C. pertinax* fixed CO_2 via reductive incomplete TCA cycle in response to CO.

In bacteria, self-cleaving activity of LexA repressor is stimulated by activated recombinase A (RecA) during response to DNA damage (SOS response) (Butala et al. 2008). As the temperate prophage also utilizes the bacterial SOS response system, host RecA promotes the entry of lytic phase when SOS response occurs in the host. In the prophage region, a gene encoding transcription repressor LexA (cpu_00820) was expressed under both conditions, whereas other genes such as DNA replication (cpu_00950) and many structure proteins were significantly downregulated under hydrogenogenic CO metabolism condition, suggesting that the lysis of the temperate phage was strictly repressed with 100% CO. In addition, these data might imply that hydrogenogenic CO metabolism is more stable than heterotrophic metabolism for *C. pertinax*.

Of the known anaerobic CO-responsive transcriptional factors, two transcriptional factors (CooA and RcoM for carboxydotrophic bacteria) possess heme domain (Kerby et al. 2008) and the other (CorQR for carboxydotrophic archaea) factor possesses 4-vinyl reductase domain to sense CO (Kim et al. 2015). Of the three transcriptional factors as DEGs under 100% CO, TetR/AcrR family transcriptional factor (cpu_03570) acts as the repressor when it senses cellular environmental dynamics (Deng et al. 2013). On the other hand, LTTR (cpu_05930 and cpu_07800) is a widespread transcriptional factor in bacteria and acts as either activators or repressors of single or operonic genes (Maddocks and Oyston 2008). However, the LTTRs were

not considered to be CO–responsive transcriptional factor, because it possessed neither heme nor 4-vinyl reductase domain. When I explored transcriptional factor with known CO-sensing domain from *C. pertinax* using BLASTp search, only CooA-2 with heme domain was found. The CooA binding site was conserved only upstream of the ECH gene cluster strongly indicating that the ECH gene cluster was regulated by CO-responsive transcriptional factor CooA regulation. Collectively, it was suggested that only the ECH gene cluster was regulated by active CooA and others were regulated secondarily in the same transcriptional cascade as the ECH gene cluster. Otherwise, the expression of gene clusters, expect for the ECH gene cluster, were regulated by an undiscovered CO-responsive transcriptional factor.

Chapter 4

General overview

In this study, I determined draft genome sequences of hydrogenogenic carboxydotrophs in the genera *Carboxydothermus* and *Carboxydocella*. The obtained genomic information from hydrogenogenic carboxydotrophs strongly supported the hypothesis that the CODH−ECH gene cluster would be considered to be responsible for hydrogenogenic carboxydotrophy. However, the lack of CODH gene in the CODH−ECH gene cluster in *C. pertinax* suggested that the organism performs alternative hydrogenogenic CO metabolism independent on the CODH gene product in the CODH−ECH gene cluster. To investigate energy conservation via hydrogenogenic CO metabolism in *C. pertinax*, I used comparative genome analysis of the genus *Carboxydothermus* (Chapter 2). Based on genomic information on *C. pertinax*, I provided insight into regulation mechanism of hydrogenogenic CO metabolism (Chapter 3).

General features of genomes of four members in the genus *Carboxydothermus* were highly similar. Their relationship was also confirmed by genome-wide phylogenetic analysis. From synteny plot against *C. hydrogenoformans* genome, *C. pertinax* lacked the CODH catalytic subunit and its transcriptional regulator encoding genes in this gene cluster probably due to inversion. Their hydrogenogenic carboxydotrophy resulted from coupling CODH-II with distally encoded ECH. In this process, *C. pertinax* would couple simultaneous CO oxidation with reduction of mainly H_2O and thiosulfate to gain more proton motive force and grow more efficiently by employing thiosulfate as the electron acceptor in CO metabolism.

Next, I performed whole transcriptome analysis for *C. pertinax* possessing only one CO-responsive transcriptional factor in order to understand the regulation mechanism of hydrogenogenic CO metabolism. RNA-Seq data suggested that *C. pertinax* switched its metabolism through intense expression changes of relatively low number of gene clusters. *C. pertinax* could conserve energy via anaerobic respiration with thiosulfate reduction in heterotrophic metabolism. Because of a high sensitivity of heme protein to CO, on the other hand, *C. pertinax* would switch anaerobic respiration to hydrogenogenic CO metabolism remaining a part of reducing power from CO oxidation to the reduction of thiosulfate.

Anaerobic hydrogenogenic carboxydotrophs are thought to fill a vital niche with scavenging potentially toxic CO and producing H_2 as available energy source for thermophilic microbes. This hydrogenogenic carboxydotrophy relies on a CODH−ECH gene cluster. This feature is thought to be as common to these organisms. In this study, however, genomic and transcriptional studies on hydrogenogenic carboxydotroph showed a versatility of CO metabolism owing to very low redox potential of CO.

Acknowledgements

This study could not have been completed without help of many people. First of all, I would like to express my sincere gratitude to Professor Yoshihiko Sako, who introduced me into the exciting world of thermophiles and encouraged me. I would like to express my deep appreciate to Associate Professor Takashi Yoshida, who always gave me helpful advices and comments. I am grateful to Professor Shigeki Sawayama for reviewing on my thesis.

I also want to thank my colleagues in the Laboratory of Marine Microbiology for their supports and encouragements. Especially, I would like to offer my special thanks to Dr. Takahiro Inoue, Mr. Kyosuke Takao, Dr. Yasuko Yoneda, Mr. Daichi Morimoto, Mr. Kimiho Omae, Mr. Hiroyasu Watai, Mr. Eitaro Ikeda, Mr Tatsuki Oguro, Ms. Hikaru Izumihara, Mr. Kento Tominaga, Mr. Tatsuhiro Isozaki, and all members belong to CO team for valuable discussion and technical help. I am deeply grateful to Ms. Ritsuko Mizuno, Ms. Mika Uehara, and Ms. Maho Masuda for giving me a lot of encouragement.

Finally, I would like to thank my family for giving me warm support throughout my student life.

References

Alcaraz LD, Moreno-Hagelsieb G, Eguiarte LE, Souza V, Herrera-Estrella L and Olmedo G (2010) Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* 11: 332.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

Aono S, Nakajima H, Saito K and Okada M (1996) A novel heme protein that acts as a carbon monoxide-dependent transcriptional activator in *Rhodospirillum rubrum*. *Biochem. Biophys. Res. Commun.* 228: 752–756.

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, and Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44: W16–W21.

Aziz RK et al. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9: 75.

Badziong W and Thauer RK (1978) Growth yields and growth rates of *Desulfovibrio vulgaris* (Marburg) growing on hydrogen plus sulfate and hydrogen plus thiosulfate as the sole energy sources. *Arch. Microbiol.* 117: 209–214.

Bailey TL and Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Altman, R, Brutlag, D, Karp, P, Lathro, R, and Searls, D, eds. Menlo Park, CA. 28-36

Berks BC and Hinsley AP (2002) Specificity of respiratory pathways involved in the reduction of sulfur compounds by *Salmonella enterica*. *Microbiology* 148: 3631–3638.

Berks BC, Page MD, Richardson DJ, Reilly A, Cavill A, Outen F and Ferguson SJ (1995) Sequence analysis of subunits of the membrane-bound nitrate reductase from a denitrifying bacterium: the integral membrane subunit provides a prototype for the dihaem electron-carrying arm of a redox loop. *Mol. Microbiol.* 15: 319–331.

Brady AL, Sharp CE, Grasby SE and Dunfield PF (2015) Anaerobic carboxydotrophic bacteria in geothermal springs identified using stable isotope probing. *Front. Microbiol.* 6: 897.

Braun V and Hantke K (2011) Recent insights into iron import by bacteria. *Curr. Opin. Chem. Biol.* 15: 328–334.

Butala M, Žgur-Bertok D and Busby SJW (2009) The bacterial LexA transcriptional repressor. *Cell. Mol. Life Sci.* 66: 82–93.

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

Cho S, Kim MS, Jeong Y, Lee BR, Lee JH, Kang SG and Cho BK (2017) Genome-wide primary transcriptome analysis of H_2-producing archaeon *Thermococcus onnurineus* NA1. *Sci. Rep.* 7: 43044.

Cord-Ruwisch R (1985) A quick method for the determination of dissolved and precipitated sulfides in cultures of sulfate-reducing bacteria. *J. Microbiol. Methods* 4: 33–36.

Crooks GE, Hon G, Chandonia J-M and Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res.* 14: 1188–1190.

Delcher AL, Salzberg SL, Phillippy AM, Delcher AL, Salzberg SL and Phillippy AM (2003) Using MUMmer to Identify Similar Regions in Large Sequence Sets. *Current Protocols in Bioinformatics*, p. 10.3.1-10.3.18. John Wiley and Sons, Inc., Hoboken, NJ, USA

Deng W, Li C and Xie J (2013) The underling mechanism of bacterial TetR/AcrR family transcriptional repressors. *Cell. Signal.* 25: 1608–1613.

Dietrich W and Klimmek O (2002) The function of methyl-menaquinone-6 and polysulfide reductase membrane anchor (PsrC) in polysulfide respiration of *Wolinella succinogenes*. *Eur. J. Biochem.* 269: 1086–1095.

Dobbek H, Gremer L, Meyer O and Huber R (1999) Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanylcysteine. *Proc. Natl. Acad. Sci. U. S. A.* 96: 8884–8889.

Dobbek H, Svetlitchnyi V, Gremer L, Huber R and Meyer O (2001) Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* 293: 1281–1285.

Duval S, Ducluzeau A-L, Nitschke W and Schoepp-Cothenet B (2008) Enzyme phylogenies as markers for the oxidation state of the environment: The case of respiratory arsenate reductase and related enzymes. *BMC Evol. Biol.* 8: 206.

Fox J, He Y, Shelver D, Roberts G and Ludden P (1996) Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum*. *J. Bacteriol.* 178: 6200–6208.

Fukiya S, Mizoguchi H, Tobe T and Mori H (2004) Extensive genomic diversity in

pathogenic *Escherichia coli* and *Shigella* Strains revealed by comparative genomic hybridization microarray. *J. Bacteriol.* 186: 3911–3921.

Grahame DA and DeMoll E (1995) Substrate and accessory protein requirements and thermodynamics of acetyl-CoA synthesis and cleavage in *Methanosarcina barkeri*. *Biochemistry* 34: 4617–4624.

Hallenbeck PC, Clark MA and Barrett EL (1989) Characterization of anaerobic sulfite reduction by *Salmonella typhimurium* and purification of the anaerobically induced sulfite reductase. *J. Bacteriol.* 171: 3008–3015.

Hammer Ø, Harper DAT and Ryan PD (2001) PAST-PAlaeontological STatistics.

He Y, Shelver D, Kerby RL and Roberts GP (1996) Characterization of a CO-responsive transcriptional activator from *Rhodospirillum rubrum*. *J. Biol. Chem.* 271: 120–123.

Henstra AM, Sipma J, Rinzema A and Stams AJM (2007) Microbiology of synthesis gas fermentation for biofuel production. *Curr. Opin. Biotechnol.* 18: 200–206.

Inoue T, Takao K, Yoshida T, Wada K, Daifuku T, Yoneda Y, Fukuyama K and Sako Y (2013) Cysteine 295 indirectly affects Ni coordination of carbon monoxide dehydrogenase-II C-cluster. *Biochem. Biophys. Res. Commun.* 441: 13–17.

Jiang B, Henstra A-M, Paulo PL, Balk M, van Doesburg W and Stams AJM (2009) Atypical one-carbon metabolism of an acetogenic and hydrogenogenic *Moorella thermoacetica* strain. *Arch. Microbiol.* 191: 123–131.

Kamata H and Kodama K (1999) Volcanic history and tectonics of the Southwest Japan Arc. *Isl. Arc* 8: 393–403.

Kanehisa M, Sato Y, Kawashima M, Furumichi M and Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44: D457–

D462.

Katoh K and Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30: 772–780.

Kerby RL, Hong SS, Ensign SA, Coppoc LJ, Ludden PW and Roberts GP (1992) Genetic and physiological characterization of the *Rhodospirillum rubrum* carbon monoxide dehydrogenase system. *J. Bacteriol.* 174: 5284–5294.

Kerby RL, Ludden PW and Roberts GP (1997) In vivo nickel insertion into the carbon monoxide dehydrogenase of *Rhodospirillum rubrum*: molecular and physiological characterization of *cooCTJ*. *J. Bacteriol.* 179: 2259–2566.

Kerby RL, Youn H and Roberts GP (2008) RcoM: a new single-component transcriptional regulator of CO metabolism in bacteria. *J. Bacteriol.* 190: 3336–3343.

Kim MS et al. (2015) A novel CO-responsive transcriptional regulator and enhanced H_2 production by an engineered *Thermococcus onnurineus* NA1 strain. *Appl. Environ. Microbiol.* 81: 1708–1714.

Kim MS et al. (2013) CO-dependent H_2 production by genetically engineered *Thermococcus onnurineus* NA1. *Appl. Environ. Microbiol.* 79: 2048–2053.

King GM and Weber CF (2007) Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5: 107–118.

Kochetkova TV., Rusanov II, Pimenov NV., Kolganova TV., Lebedinsky AV., Bonch-Osmolovskaya EA and Sokolova TG (2011) Anaerobic transformation of carbon monoxide by microbial communities of Kamchatka hot springs. *Extremophiles* 15: 319–325.

Kozhevnikova DA, Taranov EA, Lebedinsky AV., Bonch-Osmolovskaya EA and Sokolova TG (2016) Hydrogenogenic and sulfidogenic growth of *Thermococcus* archaea on carbon monoxide and formate. *Microbiology* 85: 400–410.

Langmead B and Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.

Leduc J, Thorsteinsson M V, Gaal T and Roberts GP (2001) Mapping CooA. RNA polymerase interactions. Identification of activating regions 2 and 3 in CooA, the co-sensing transcriptional activator. *J. Biol. Chem.* 276: 39968–39973.

Liao Y, Smyth GK and Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930.

Maddocks SE and Oyston PCF (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* 154: 3609–3623.

Markowitz VM et al. (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* 34: D344–D348.

Mesbah M, Premachandran U and Whiteman WB (1989) Precise measurement of the G+C content of deoxyribonucleic acid by high-performance liquid chromatography. *Int. J. Syst. Bacteriol.* 39: 159–167.

Mohr T, Aliyu H, Küchlin R, Polliack S, Zwick M, Neumann A, Cowan D and de Maayer P (2018) CO-dependent hydrogen production by the facultative anaerobe *Parageobacillus thermoglucosidasius*. *Microb. Cell Fact.* 17: 108.

Mörsdorf G, Frunzke K, Gadkari D and Meyer O (1992) Microbial growth on carbon monoxide. *Biodegradation* 3: 61–82.

Nantapong N, Otofuji A, T. Migita C, Adachi O, Toyama H and Matsushita K

(2005) Electron Transfer Ability from NADH to Menaquinone and from NADPH to Oxygen of Type II NADH Dehydrogenase of *Corynebacterium glutamicum*. *Biosci. Biotechnol. Biochem.* 69: 149–159.

Nishimura Y, Watai H, et al. (2017a) Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* 2: e00359-16.

Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H and Goto S (2017b) ViPTree: the viral proteomic tree server. *Bioinformatics* 106: 12826–12831.

Novikov AA, Sokolova TG, Lebedinsky AV, Kolganova TV and Bonch-Osmolovskaya EA (2011) *Carboxydothermus islandicus* sp. nov., a thermophilic, hydrogenogenic, carboxydotrophic bacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 61: 2532–2537.

Nowicka B and Kruk J (2010) Occurrence, biosynthesis and function of isoprenoid quinones. *Biochim. Biophys. Acta - Bioenerg.* 1797: 1587–1605.

Oelgeschläger E and Rother M (2008) Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Arch. Microbiol.* 190: 257–269.

Omae K, Yoneda Y, Fukuyama Y, Yoshida T and Sako Y (2017) Genomic analysis of *Calderihabitans maritimus* KKC1, a thermophilic, hydrogenogenic, carboxydotrophic bacterium isolated from marine sediment. *Appl. Environ. Microbiol.* 83: e00832-17.

Parshina SN, Sipma J, Henstra AM and Stams AJM (2010) Carbon monoxide as an electron donor for the biological reduction of sulphate. *Int. J. Microbiol.* 2010: 319527.

Pruitt KD, Tatusova T and Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.

Nucleic Acids Res. 35: D61-5.

Ragsdale SW (2004) Life with carbon monoxide. *Crit. Rev. Biochem. Mol. Biol.* 39: 165–195.

Rajeev L et al. (2012) Deletion of the *Desulfovibrio vulgaris* carbon monoxide sensor invokes global changes in transcription. *J. Bacteriol.* 194: 5783–5793.

Roberts GP, Youn H and Kerby RL (2004) CO-sensing mechanisms. *Microbiol. Mol. Biol. Rev.* 68: 453–473.

Robinson MD, McCarthy DJ and Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.

Sant'Anna FH, Lebedinsky AV, Sokolova TG, Robb FT and Gonzalez JM (2015) Analysis of three genomes within the thermophilic bacterial species *Caldanaerobacter subterraneus* with a focus on carbon monoxide dehydrogenase evolution and hydrolase diversity. *BMC Genomics* 16: 757.

Santiago B, Schübel U, Egelseer C and Meyer O (1999) Sequence analysis, characterization and CO-specific transcription of the cox gene cluster on the megaplasmid pHCG3 of *Oligotropha carboxidovorans*. *Gene* 236: 115–124.

Shao Y, He X, Harrison EM, Tai C, Ou H-Y, Rajakumar K and Deng Z (2010) mGenomeSubtractor: a web-based tool for parallel in silico subtractive hybridization analysis of multiple bacterial genomes. *Nucleic Acids Res.* 38: W194–W200.

Shelver D, Kerby RL, He Y and Roberts GP (1995) Carbon monoxide-induced activation of gene expression in *Rhodospirillum rubrum* requires the product of *cooA*, a member of the cyclic AMP receptor protein family of transcriptional

regulators. *J. Bacteriol.* 177: 2157–2163.

Shelver D, Kerby RL, He Y and Roberts GP (1997) CooA, a CO-sensing transcription factor from *Rhodospirillum rubrum*, is a CO-binding heme protein. *Proc. Natl. Acad. Sci.* 94: 11216–11220.

Slepova T V., Sokolova TG, Lysenko AM, Tourova TP, Kolganova TV., Kamzolkina OV., Karpov GA and Bonch-Osmolovskaya EA (2006) *Carboxydocella sporoproducens* sp. nov., a novel anaerobic CO-utilizing/H_2-producing thermophilic bacterium from a Kamchatka hot spring. *Int. J. Syst. Evol. Microbiol.* 56: 797–800.

Slepova TV, Sokolova TG, Kolganova TV, Tourova TP and Bonch-Osmolovskaya EA (2009) *Carboxydothermus siderophilus* sp. nov., a thermophilic, hydrogenogenic, carboxydotrophic, dissimilatory Fe(III)-reducing bacterium from a Kamchatka hot spring. *Int. J. Syst. Evol. Microbiol.* 59: 213–217.

Slobodkin AI, Sokolova TG, Lysenko AM and Wiegel J (2006) Reclassification of *Thermoterrabacterium ferrireducens* as *Carboxydothermus ferrireducens* comb. nov., and emended description of the genus *Carboxydothermus*. *Int. J. Syst. Evol. Microbiol.* 56: 2349–2351.

Slobodkina GB, Panteleeva AN, Sokolova TG, Bonch-Osmolovskaya EA and Slobodkin AI (2012) *Carboxydocella manganica* sp. nov., a thermophilic, dissimilatory Mn(IV)- and Fe(III)-reducing bacterium from a Kamchatka hot spring. *Int. J. Syst. Evol. Microbiol.* 62: 890–894.

Soboh B, Linder D and Hedderich R (2002) Purification and catalytic properties of a CO-oxidizing:H_2-evolving enzyme complex from *Carboxydothermus hydrogenoformans*. *Eur. J. Biochem.* 269: 5712–5721.

Sokolova T and Lebedinsky A (2013) CO-Oxidizing Anaerobic Thermophilic Prokaryotes. Thermophilic Microbes in Environmental and Industrial Biotechnology. 2nd ed, (Satyanarayana T, Littlechild J, K, ed), pp. 203–231.

Sokolova TG (2015) *Carboxydocella*. *Bergey's Manual of Systematics of Archaea and Bacteria*, pp. 1–4. John Wiley and Sons, Ltd, Chichester, UK

Sokolova TG, Henstra AM, Sipma J, Parshina SN, Stams AJM and Lebedinsky AV (2009) Diversity and ecophysiological features of thermophilic carboxydotrophic anaerobes. *FEMS Microbiol. Ecol.* 68: 131–141.

Sokolova TG, Kostrikina NA, Chernyh NA, Tourova TP, Kolganova TV and Bonch-Osmolovskaya EA (2002) *Carboxydocella thermautotrophica* gen. nov., sp. nov., a novel anaerobic, CO-utilizing thermophile from a Kamchatkan hot spring. *Int. J. Syst. Evol. Microbiol.* 52: 1961–1967.

Solovyev V and Salamov A (2011) *Metagenomics and its applications in agriculture, biomedicine, and environmental studies*. Li, RW, ed. Nova Science Publisher's

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stoffels L, Krehenbrink M, Berks BC and Unden G (2012) Thiosulfate reduction in *Salmonella enterica* is driven by the proton motive force. *J. Bacteriol.* 194: 475–485.

Sugawara H, Ohyama A, Mori H and Kurokawa K (2009) Microbial Genome Annotation Pipeline (MiGAP) for diverse users, abstr S001-1–2. *Twent. Int. Conf. Genome Informatics.* Yokohama, Japan.

Sullivan MJ, Petty NK and Beatson SA (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* 27: 1009–1010.

Svetlichny VA, Sokolova TG, Gerhardt M, Ringpfeil M, Kostrikina NA and Zavarzin GA (1991) *Carboxydothermus hydrogenoformans* gen. nov., sp. nov., a CO-utilizing thermophilic anaerobic bacterium from hydrothermal environments of Kunashir Island. *Syst. Appl. Microbiol.* 14: 254–260.

Svetlitchnyi V, Dobbek H, Meyer-Klaucke W, Meins T, Thiele B, Römer P, Huber R and Meyer O (2004) A functional Ni-Ni-[4Fe-4S] cluster in the monomeric acetyl-CoA synthase from *Carboxydothermus hydrogenoformans*. *Proc. Natl. Acad. Sci. U. S. A.* 101: 446–451.

Svetlitchnyi V, Peschel C, Acker G and Meyer O (2001) Two membrane-associated NiFeS-carbon monoxide dehydrogenases from the anaerobic carbon-monoxide-utilizing eubacterium *Carboxydothermus hydrogenoformans*. *J. Bacteriol.* 183: 5134–5144.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731–2739.

Tan Y, Liu J, Chen X, Zheng H and Li F (2013) RNA-seq-based comparative transcriptome analysis of the syngas-utilizing bacterium *Clostridium ljungdahlii* DSM 13528 grown autotrophically and heterotrophically. *Mol. Biosyst.* 9: 2775–2784.

Tatusov RL (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28: 33–36.

Techtmann SM, Colman AS, Murphy MB, Schackwitz WS, Goodwin LA and Robb FT (2011) Regulation of multiple carbon monoxide consumption pathways

in anaerobic bacteria. *Front. Microbiol.* 2: 147.

Techtmann SM, Colman AS and Robb FT (2009) "That which does not kill us only makes us stronger": the role of carbon monoxide in thermophilic microbial consortia. *Environ. Microbiol.* 11: 1027–1037.

Techtmann SM, Lebedinsky AV, Colman AS, Sokolova TG, Woyke T, Goodwin L and Robb FT (2012) Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. *Front. Microbiol.* 3: 132.

Thauer RK, Jungermann K and Decker K (1977) Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* 41: 100–180.

Thorvaldsdottir H, Robinson JT and Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14: 178–192.

Toshchakov SV. et al. (2018) Genomic insights into energy metabolism of *Carboxydocella thermautotrophica* coupling hydrogenogenic CO oxidation with the reduction of Fe(III) minerals. *Front. Microbiol.* 9: 1759.

Trapnell C, Pachter L and Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–1111.

Wawrousek K, Noble S, Korlach J, Chen J, Eckert C, Yu J and Maness PC (2014) Genome annotation provides insight into carbon monoxide and hydrogen metabolism in Rubrivivax gelatinosus. *PLoS One* 9: e114551.

Weghoff MC and Müller V (2016) CO metabolism in the thermophilic acetogen *Thermoanaerobacter kivui*. *Appl. Environ. Microbiol.* 82: 2312–2319.

Woolerton TW, Sheard S, Reisner E, Pierce E, Ragsdale SW and Armstrong FA (2010) Efficient and clean photoreduction of CO_2 to CO by enzyme-modified

TiO_2 nanoparticles using visible light. *J. Am. Chem. Soc.* 132: 2132–2133.

Wu M et al. (2005) Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet.* 1: e65.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S and Madden TL (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13: 134.

Yoneda Y, Kano S I., Yoshida T, Ikeda E, Fukuyama Y, Omae K, Kimura-Sakai S, Daifuku T, Watanabe T and Sako Y (2015) Detection of anaerobic carbon monoxide-oxidizing thermophiles in hydrothermal environments. *FEMS Microbiol. Ecol.* 91: fiv093.

Yoneda Y, Yoshida T, Daifuku T, Kitamura T, Inoue T, Kano S and Sako Y (2013) Quantitative detection of carboxydotrophic bacteria *Carboxydothermus* in a hot aquatic environment. *Fundam. Appl. Limnol.* 182: 161–170.

Yoneda Y, Yoshida T, Kawaichi S, Daifuku T, Takabe K and Sako Y (2012) *Carboxydothermus pertinax* sp. nov., a thermophilic, hydrogenogenic, Fe(III)-reducing, sulfur-reducing carboxydotrophic bacterium from an acidic hot spring. *Int. J. Syst. Evol. Microbiol.* 62: 1692–1697.

Youn H, Kerby RL, Conrad M and Roberts GP (2004) Functionally critical elements of CooA-related CO sensors. *J. Bacteriol.* 186: 1320–1329.

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

Zhao Y, Cimpoia R, Liu Z and Guiot SR (2011) Orthogonal optimization of *Carboxydothermus hydrogenoformans* culture medium for hydrogen production from carbon monoxide by biological water-gas shift reaction. *Int. J. Hydrogen*

Energy 36: 10655–10665.

Zhou J et al. (2011) How sulphate-reducing microorganisms cope with stress: lessons from systems biology. *Nat. Rev. Microbiol.* 9: 452–466.

Publication list

1. **Fukuyama Y, Omae K, Yoneda Y, Yoshida T and Sako Y.** (2017) Draft genome sequences of *Carboxydothermus pertinax* and *C. islandicus*, hydrogenogenic carboxydotrophic bacteria. *Genome announcements* 5: e01648-16.
2. **Fukuyama Y, Oguro T, Omae K, Yoneda Y, Yoshida T and Sako Y.** (2017) Draft genome sequences of two hydrogenogenic carboxydotrophic bacteria, *Carboxydocella* sp. strains JDF658 and ULO1, isolated from two distinct volcanic fronts in Japan. *Genome announcements* 5: e00242-17.
3. **Fukuyama Y, Omae K, Yoneda Y, Yoshida and, Sako Y.** (2018) Insight into energy conservation via alternative carbon monoxide metabolism in *Carboxydothermus pertinax* revealed by comparative genome analysis. (2018). *Appl. Environ. Microbiol.*, 84: e00458-18.
4. **Fukuyama Y, Omae K., Yoshida T and Sako Y.** Transcriptome analysis of *Carboxydothermus pertinax* provides insight into their alternative CO metabolism. (submitted to *Extremophiles*)
5. **Inoue T, Takao K, Fukuyama Y, Yoshida T and Sako Y.** (2014) Over-expression of carbon monoxide dehydrogenase-I with an accessory protein co-expression: a key enzyme for carbon dioxide reduction. *Biosci. Biotechnol. Biochem.*, 78: 582-587.
6. **Yoneda Y, Kano SI, Yoshida T, Ikeda E, Fukuyama Y., Omae K, Kimura-Sakai S, Daifuku T, Watanabe T and Sako Y.** (2015) Detection of anaerobic carbon monoxide-oxidizing thermophiles in hydrothermal environments. *FEMS Microbiol. Ecol.*, 91: 1-9.

7. **Omae K, Yoneda Y, Fukuyama Y, Yoshida T and Sako Y.** (2017) Genomic analysis of *Calderihabitans maritimus* KKC1, a thermophilic hydrogenogenic carboxydotrophic bacterium isolated from marine sediment. *Appl. Environ. Microbiol.*, 83:e00832-17.