*Acta Oceanol. Sin.*, 2015, Vol. 34, No. 4, P. 92–113
DOI: 10.1007/s13131-015-0650-7
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

# Ecological functions of uncultured microorganisms in the cobalt-rich ferromanganese crust of a seamount in the central Pacific are elucidated by fosmid sequencing

HUO Yingyi[1, 2], CHENG Hong[3], Anton F. Post[4], WANG Chunsheng[1, 2], JIANG Xiawei[5], PAN Jie[3], WU Min[3], XU Xuewei[1, 2*]

[1] Laboratory of Marine Ecosystem and Biogeochemistry, Second Institute of Oceanography, State Oceanic Administration, Hangzhou 310012, China

[2] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Hangzhou 310012, China

[3] College of Life Sciences, Zhejiang University, Hangzhou 310058, China

[4] The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biology Laboratory, Woods Hole, MA 02543, USA

[5] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, the First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, China

Received 3 January 2014; accepted 5 July 2014



**Abstract**

Cobalt-rich ferromanganese is an important seafloor mineral and is abundantly present in the seamount crusts. Such crusts form potential hotspots for biogeochemical activity and microbial diversity, yet our understanding of their microbial communities is lacking. In this study, a cultivation-independent approach was used to recover genomic information and derive ecological functions of the microbes in a sediment sample collected from the cobalt-rich ferromanganese crust of a seamount region in the central Pacific. A total of 78 distinct clones were obtained by fosmid library screening with a 16S rRNA based PCR method. Proteobacteria and MGI Thaumarchaeota dominated the bacterial and archaeal 16S rRNA gene sequence results in the microbial community. Nine fosmid clones were sequenced and annotated. Numerous genes encoding proteins involved in metabolic functions and heavy metal resistance were identified, suggesting alternative metabolic pathways and stress responses that are essential for microbial survival in the cobalt-rich ferromanganese crust. In addition, genes that participate in the synthesis of organic acids and exoploymers were discovered. Reconstruction of the metabolic pathways revealed that the nitrogen cycle is an important biogeochemical process in the cobalt-rich ferromanganese crust. In addition, horizontal gene transfer (HGT) events have been observed, and most of them came from bacteria, with some occurring in archaea and plants. Clone W4-93a, belonging to MGI Thaumarchaeota, contained a region of gene synteny. Comparative analyses suggested that a high frequency of HGT events as well as genomic divergence play important roles in the microbial adaption to the deep-sea environment.

**Key words:** seamount, cobalt-rich ferromanganese crust, metagenome, horizontal gene transfer

**Citation:** Huo Yingyi, Cheng Hong, Anton F. Post, Wang Chunsheng, Jiang Xiawei, Pan Jie, Wu Min, Xu Xuewei. 2015. Ecological functions of uncultured microorganisms in the cobalt-rich ferromanganese crust of a seamount in the central Pacific are elucidated by fosmid sequencing. Acta Oceanologica Sinica, 34(4): 92–113, doi: 10.1007/s13131-015-0650-7

## 1 Introduction

Seamounts are widespread and defined as topographic rises from the ocean floor with a limited area across the summit, which is below sea level or emerges above the sea surface only for short periods of time (Menard, 1964; Staudigel et al., 2010). Recently, 33 452 seamounts (elevation of > 1 000 m) were identified in global bathymetric datasets, most of them (57.2%) located in the Pacific Ocean (Yesson et al., 2011). However, the number of seamounts remains under debate due to the different definitions of what constitutes a seamount as well as the variation in techniques used to count them (Hillier and Watts, 2007; Iyer et al., 2012; Wessel et al., 2010). Although seamounts have a high degree of biodiversity, harbour unique biological communities, display high levels of endemism, represent hotspots of nutrient cycling and support commercial fisheries, fewer than 300 seamounts have been thoroughly sampled, and the majority of these studies have focused on hydrothermal vents (Clark et al., 2010; Duffy, 2008; Emerson and Moyer, 2010; Rowden et al., 2010; Schlacher et al., 2010).

Deposits of cobalt-rich, oxidised ferromanganese in crusts that cover seamounts were first discovered in 1980 (Craig et al., 1982; Ito et al., 2008; Muiños et al., 2013). These crusts usually

Foundation item: China Ocean Mineral Resources R & D Association COMRA Special Foundation under contract Nos DY125-15-R-03 and DY125-13-E-01; the National Natural Science Foundation of China under contract No. 41276173; the Zhejiang Provincial Natural Science Foundation of China under contract No. LQ13D060002; the Scientific Research Fund of the Second Institute of Oceanography, SOA under contract No. JT1305.

*Corresponding author, E-mail: xuxw@sio.org.cn

grow at very slow rates (1–10 μm per $10^3$ years) and exist only on a few seamounts (Fu et al., 2005; Koschinsky and Hein, 2003). Cobalt-rich ferromanganese crusts are a rich source of metals such as ferromanganese oxide, cobalt, copper, nickel, platinum and other rare earth elements (Fu et al., 2005; Zhang et al., 2008). Currently, our knowledge of biological interactions at the cobalt-rich ferromanganese crusts is very limited, and more research is needed into their resident microbial communities and their ecosystem functions to evaluate the environmental impacts of future crust exploration and mining.

The role of biogenesis in cobalt-rich ferromanganese crust formation on seamounts remains controversial and poorly understood. Previous results from a northern Pacific seamount indicated that crust accretion is not a purely physicochemical process as it also involves microbial processes (Verlaan, 1992). Microorganisms act as biological nuclei for the formation of cobalt-rich crusts, suggesting that biomineralization is indispensable in the mineral formation process (Wang and Müller, 2009). Scanning electron microscopy studies suggest that biological processes are involved in the formation of the ferromanganese crusts covering seamounts in the central Atlantic (Wang et al., 2011). Several studies of microbial diversity in sediments from cobalt-rich ferromanganese crusts of central Pacific seamounts have been performed using culture-independent approaches (Jiang et al., 2012; Liao et al., 2011). Phylogenetic analyses of bacterial and archaeal 16S rRNA clone libraries have revealed the predominance of *Proteobacteria* and marine archaeal group I (MGI), and it has been suggested that members of this community may be involved in sulphur, nitrogen and metal cycling in cobalt-rich ferromanganese crusts (Liao et al., 2011). Recently, nine novel lipolytic enzymes were identified, suggesting that the microbial populations participate in carbon degradation, calcium deposition and contribute to biomineralization (Jiang et al., 2012).

Recent work revealed the microbial community structure and diversity of the microbes in cobalt-rich ferromanganese crusts by using molecular and electron microscopy approaches. However, the functional aspects of the microbes should not be overlooked. In this study, a metagenomic library of deep-sea sediment collected from a cobalt-rich ferromanganese crust region from a seamount was screened to describe the genome content and biological properties of uncultivated microorganisms. In total, approximately 21 000 randomly selected fosmid clones were subjected to PCR-based screening, 35 archaeal and 43 bacterial 16S rRNA gene-containing clones were obtained, and nine fragments were sequenced and analysed. To our knowledge, this is the first report of a metagenomic approach to identify the role of microorganisms in the formation of cobalt-rich ferromanganese crusts of seamounts.

## 2 Materials and methods

### 2.1 *Sample collection, geochemical properties analysis and library construction*

A deep-sea sediment sample was collected from the skirt of a seamount located in the cobalt-rich crust deposit region in the central Pacific Ocean. The sample collection method, location, DNA extraction procedure and library construction information were described in detail in an earlier publication (Jiang et al., 2012). The elemental composition of the sample was determined by the Zhejiang Institute of Geology and Mineral Resources using various methods, including the gravimetric method (for $SiO_2$), inductively coupled plasma-atomic emission spectrometry (ICP-AES; for $Al_2O_3$, $Fe_2O_3$, CaO, MgO, $TiO_2$, MnO and $P_2O_5$), atomic absorption spectrometry (AAS; for $K_2O$ and $Na_2O$) and inductively coupled plasma-mass spectrometry (ICP-MS; for Ni, Co, V, Cr, Cu, Zn, Cd, Pb, Mo and Ba).

### 2.2 *16S rRNA gene screening*

Approximately 21 000 randomly selected fosmid clones were subjected to PCR-based screening. The DNA of pooled fosmid clones was extracted using the Axygen Plasmid Miniprep Kit (Axygen Biotechnology, Hangzhou, China). The fosmid DNA templates were treated with the plasmid-safe ATP-dependent DNase (Epicentre Biotechnologies, Madison, Wisconsin, USA) to remove the chromosomal DNA contamination of the host strain (*Escherichia coli* EPI300). Primers Ar20F (5′-TTCCGGTTGATCCYGCCTGA-3′) and Arch958R (5′-TCCGGCGTTGAMTCCAATT-3′) (DeLong, 1992) were used for archaeal 16S rRNA gene amplification. Primers 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 23S1R (5′-GGGTTTCCCCATTCGGAAATC-3′) were used for the identification of the bacterial 16S rRNA gene with the adjacent intergenic spacer region (ISR) (García-Martínez et al., 1996). Thirty-five cycles of amplification were carried out under the following conditions for the archaeal 16S rRNA gene: denaturation at 94°C for 45 s, annealing at 55°C for 45 s and elongation at 72°C for 1 min. Thirty-five cycles of amplification of bacterial 16S rRNA gene were performed under the following conditions: denaturation at 94°C for 15 s, annealing at 50°C for 30 s and elongation at 72°C for 2 min (Martín-Cuadrado et al., 2007). PCR fragments were extracted from the gel using the AXYGEN gel extraction kit (AXYGEN, Hangzhou, China) and then cloned into the pMD19-T vector (TAKARA, Dalian, China). The 16S rRNA gene fragments were sequenced with the primers M13F/M13R (archaea) and 27F/1492R (bacteria). The closest relatives of the 16S rRNA sequences were obtained from the NCBI GenBank database using blastn. Evolutionary distances were calculated according to Kimura's two-parameter correction method. Neighbour-joining trees were constructed with a bootstrap value of 500 using MEGA version 5.0 (Tamura et al., 2011). The accession numbers of the 16S rRNA gene sequences are JQ013299-JQ013333 (archaeal) and JQ013334-JQ013376 (bacterial).

### 2.3 *Fragments sequencing and analysis*

Two archaeal and seven bacterial fosmid clones were selected and their sequences were determined by sequencing using a Roche 454 GS-FLX and Illumina/Solexa Genome Analyzer II platforms (Tongji-SCBIT Biotechnology Co., Ltd). Gaps were closed with the help of targeted PCR. The PCR products were sequenced by primer walking. Open reading frames (ORFs) were predicted using MetaGeneMark (Zhu et al., 2010). Gene identification was obtained by submitting the deduced protein sequences for ortholog/homolog searches in the NCBI nr database, the Cluster of Orthologous Groups (COG) database (Tatusov et al., 2000) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) using blastp. A threshold e-value of 1e-5 was used for all analyses. For phylogenetic analysis of ORFs, a blastp search of protein sequences in the NCBI nr database was carried out with default parameters, and the sequences for the best blast hits were retrieved from the database. Neighbour-joining trees were constructed with a bootstrap value of 500 using the Poisson model option in the MEGA 5.0 phylogenetic software package (Tamura et al., 2011). The Genbank accessions of the fosmid sequences are JQ085817-JQ085825.

## 3 Results

### 3.1 *Elemental composition*

The elemental composition of the sediment sample SEAM02 is shown in Table 1. Compared with offshore sediment (Zhao, 1988) and continental crust (Wedepohl, 1995), the concentrations of $Al_2O_3$ and CaO were much lower in the sample, whereas those of $Na_2O$, MnO, Ni, Co, Cu, Zn, Pb and Ba were higher. Specifically, $Na_2O$, MnO, Co, Cu and Ba, reached concentrations of more than twice those found in the offshore sediment and continental crust. Taken together, the sample SEAM02 was mainly composed of $SiO_2$ (54.07%), $Al_2O_3$ (8.35%), $Na_2O$ (7.18%) and $Fe_2O_3$ (4.14%) and was rich in metals, including Mn, Ni, Co, Cu, Zn, Pb and Ba.

**Table 1.** Geochemical properties of the sediment sample SEAM02, offshore sediment and continental crust

| | SEAM02 | Offshore sediment[1)] | Continental crust[2)] |
|---|---|---|---|
| $SiO_2$/% | 54.07 | 54.43 | 61.50 |
| $Al_2O_3$/% | 8.35 | 12.03 | 15.10 |
| $Fe_2O_3$/% | 4.14 | 4.59 | 6.28 |
| CaO/% | 0.92 | 10.05 | 5.50 |
| MgO/% | 2.58 | 1.84 | 3.70 |
| $K_2O$/% | 2.10 | 1.98 | 2.40 |
| $Na_2O$/% | 7.18 | 2.24 | 3.20 |
| $TiO_2$/% | 0.48 | 0.57 | 0.68 |
| MnO/% | 0.33 | 0.12 | 0.10 |
| $P_2O_5$/% | 0.22 | 0.12 | 0.18 |
| Ni/μg·g$^{-1}$ | 74 | 27 | 56 |
| Co/μg·g$^{-1}$ | 52 | 14 | 24 |
| V/μg·g$^{-1}$ | 94 | – | 98 |
| Cr/μg·g$^{-1}$ | 57 | 43 | 126 |
| Cu/μg·g$^{-1}$ | 134 | 20 | 25 |
| Zn/μg·g$^{-1}$ | 83 | 72 | 65 |
| Cd/μg·g$^{-1}$ | 0.09 | – | – |
| Pb/μg·g$^{-1}$ | 35 | 28 | 14.8 |
| Mo/μg·g$^{-1}$ | 4.4 | – | – |
| Ba/μg·g$^{-1}$ | 1 821 | 431 | 584 |

Notes: [1)] Data of the East China Sea from Zhao (1988); [2)] data from Wedepohl (1995).

### 3.2 *Microbial community composition*

A total of 35 archaeal 16S rRNA-containing clones were obtained from the fosmid library. All of the archaeal clones belonged to Marine Group I (MGI) in the phylum Thaumarchaeota except clone W5-61a, which was grouped into Marine Benthic Group A (MBGA) (Fig. 1a). These sequences were most closely related to those from other deep-sea sediment environments, including sediments from the Pacific nodule province (Xu et al., 2005), the Weddell Sea of Antarctica (Gillan and Danis, 2007), the southern Mariana Trough (Kato, Kobayashi et al., 2009; Kato, Yanagawa et al., 2009), the east Pacific rise (Ehrhardt et al., 2007; Li et al., 2008), the Mediterranean cold seep (Heijs et al., 2007) and the Barents Sea cold seep (Lösekann et al., 2007). This result indicated that our clones represent members of a common and abundant archaeal community in deep-sea sediments. In accordance with findings from previous microbial diversity studies of deep-sea sediment with cobalt-rich crust deposits (Liao et al., 2011), MGI was the dominant archaeal group. The sequence of clone W5-61a, the only clone that affiliated with members of group MBGA, shared less than 90% identity with previously reported clones, with the exception of clones aEPR13S208 (97.6%) and YS16As04 (92.4%) retrieved from the East Pacific Rise (Li et al., 2008) and the Southern Mariana Trough (Kato, Kobayashi et al., 2009), respectively. There are no cultivated species reported for MBGA members, and only a single strain, *Nitrosopumilus maritimus* SCM1, within the MGI group has been isolated so far (Könneke et al., 2005), leaving us with little insight into the genetic make-up of their genomes and the physiological functions they encode.

A total of 43 bacterial fosmid clones were obtained that distributed over seven bacterial groups (Fig. 1b): Alphaproteobacteria (6 clones), Betaproteobacteria (1 clone), Gammaproteobacteria (14 clones), Deltaproteobacteria (3 clones), Actinobacteria (11 clones), Gemmatimonadetes (3 clones) and Chlorobi (1 clone). The other 4 clones (W4-21b, W4-50b, W5-15b and W5-77b) could not be assigned to any taxonomic division. A total of 38 bacterial clones showed high identities with uncultured clones from the deep-sea surface sediments of the south Atlantic Ocean (Schauer et al., 2009), the seafloor lavas of the east Pacific rise (EPR) and the Hawaiian basalts (Santelli et al., 2008), indicating these bacteria may be common to these deep-sea environments. In addition, the 16S rRNA gene sequences of Clones W4-21b, W5-15b and W5-102b did not have matches with greater than 90% identity to other sequences in the database, suggesting they might represent novel taxa for the deep-sea environment, and they may have unique adaptation to the cobalt-rich crust environment.

### 3.3 *Gene content of fosmid clones*

To obtain more genomic information on microbial adaptation to deep-sea sediments, we sequenced two archaeal and seven bacterial genome fragments. The fosmid insert sizes ranged from 23 to 45 kb, with G+C content ranging from 36% to 65%. A detailed description of DNA insert sizes, ORF positions, predicted functions, closest relatives, and COG classification are summarised in Appendix Table A1. In addition, inserts were subjected to gene annotation, revealing a range of gene functions.

#### 3.3.1 *Archaeal fosmid clones*

To the best of our knowledge, Clone W5-61a is the first genome fragment to be sequenced for a MBGA member of the Thaumarchaeota. The fragment has a G+C content of 48.6%, considerably higher than that of MGI Thaumarchaeota (approximately 34%) (Blainey et al., 2011; Walker et al., 2010). Clone W5-61a (32 142 bp) contains 25 ORFs, of which 23 of them could be assigned with the COG classification system. Of these, 17 encode functions in basic metabolic processes: leucine biosynthesis (ORFs 1-3), ion transport (ORFs 7-9), DNA protection and repair (ORFs 13-14) and purine metabolism (ORFs 15-22). Most of the predicted genes had the most significant blast hits to members of archaea, whereas ORF2, ORF3 and ORF6 showed the highest identity to bacterial genes. ORF2 and ORF3 encoded the large and small subunits of 3-isopropylmalate dehydratase, and phylogeny indicates that they may have been transferred from Firmicutes via HGT (Appendix Figs A1a and b). Both ORF6 and ORF11 were identified as encoding the pyrroloquinoline quinone biosynthesis protein C (PqqC); however, only 53.4% of the amino acid residues were similar between them. The phylogenetic tree was constructed based on bacterial and available archaeal original PqqC genes (Appendix Fig. A1c). The two genes formed dis-

tinct phyletic lines with high bootstrap values towards the periphery of the bacterial lineage, revealing that they may be of bacterial origin. ORF12 encodes a PQQ-dependent alcohol dehydrogenase, which is rarely found in archaea. The function of archaeal PQQ-dependent alcohol dehydrogenase is still unknown; however, bacterial alcohol dehydrogenase had been proved to oxidize various alcohols for bioenergy generation (Adachi et al., 2007). This gene might be important for the energy requirement of Archaea in deep sea sediments.

Clone W4-93a (34 190 bp) was most closely related to *Candidatus* Nitrosoarchaeum limnia within the marine group I Thaumarchaeota (>95.8% 16S sequence identity). The G+C content of this fragment was 36.3%, and a total of 49 ORFs were predicted. Deduced amino sequences indicated that only 21 proteins (44.7%) could be assigned a physiological function with the COG classification system, and nine proteins (19.1%) did not show significant similarity to any proteins in the NCBI nr database. Apart from these nine predicted proteins that had no significant relatives, the other ORFs were most closely related to known archaeal genes found in the members of Thaumarchaeota (17 ORFs to *Candidatus* Nitrosoarchaeum limnia, 16 ORFs to *Nitrosopumilus maritimus*, 4 ORFs to *Candidatus* Cenarchaeum symbiosum and 2 ORFs to *Candidatus* Nitrosoarchaeum koreensis). Interestingly, ORF25 was most closely related to the nitrogen regulatory protein P-II from *Thermococcus sibiricus*, a hyperthermophilic member of the Euryarchaeota (Miroshnichenko et al., 2001) (Appendix Table A1). Phylogenetic analysis also indicated that this gene may have been obtained from euryarchaeotal species by HGT (Appendix Fig. A1d).

Gene organisation and synteny was determined by comparing Clone W4-93a with two MGI Thaumarchaeota fragments (Fig. 2), *Nitrosopumilus maritimus* SCM1 and Clone 74A4 (Béjà et al.,

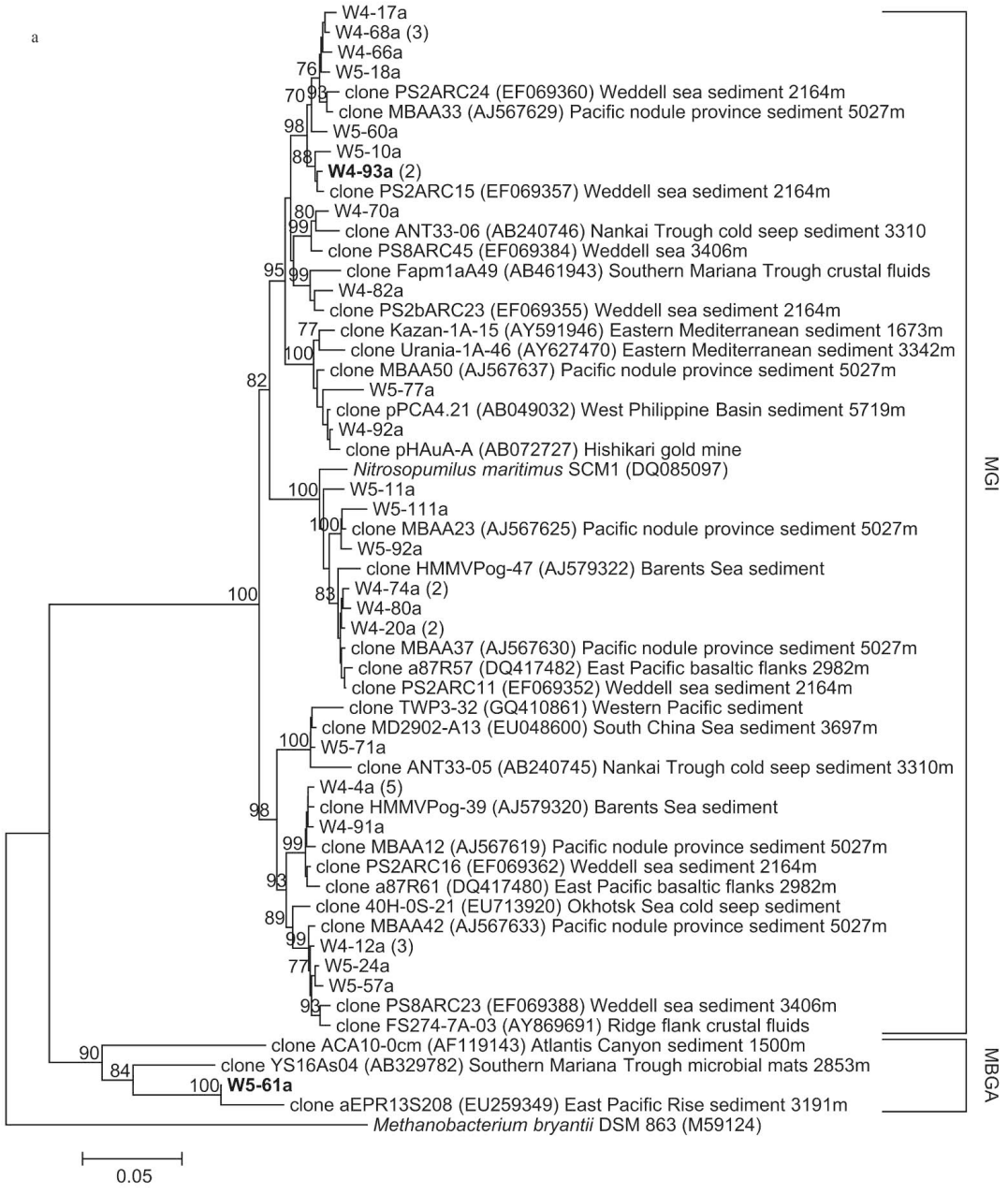


Fig. 1.

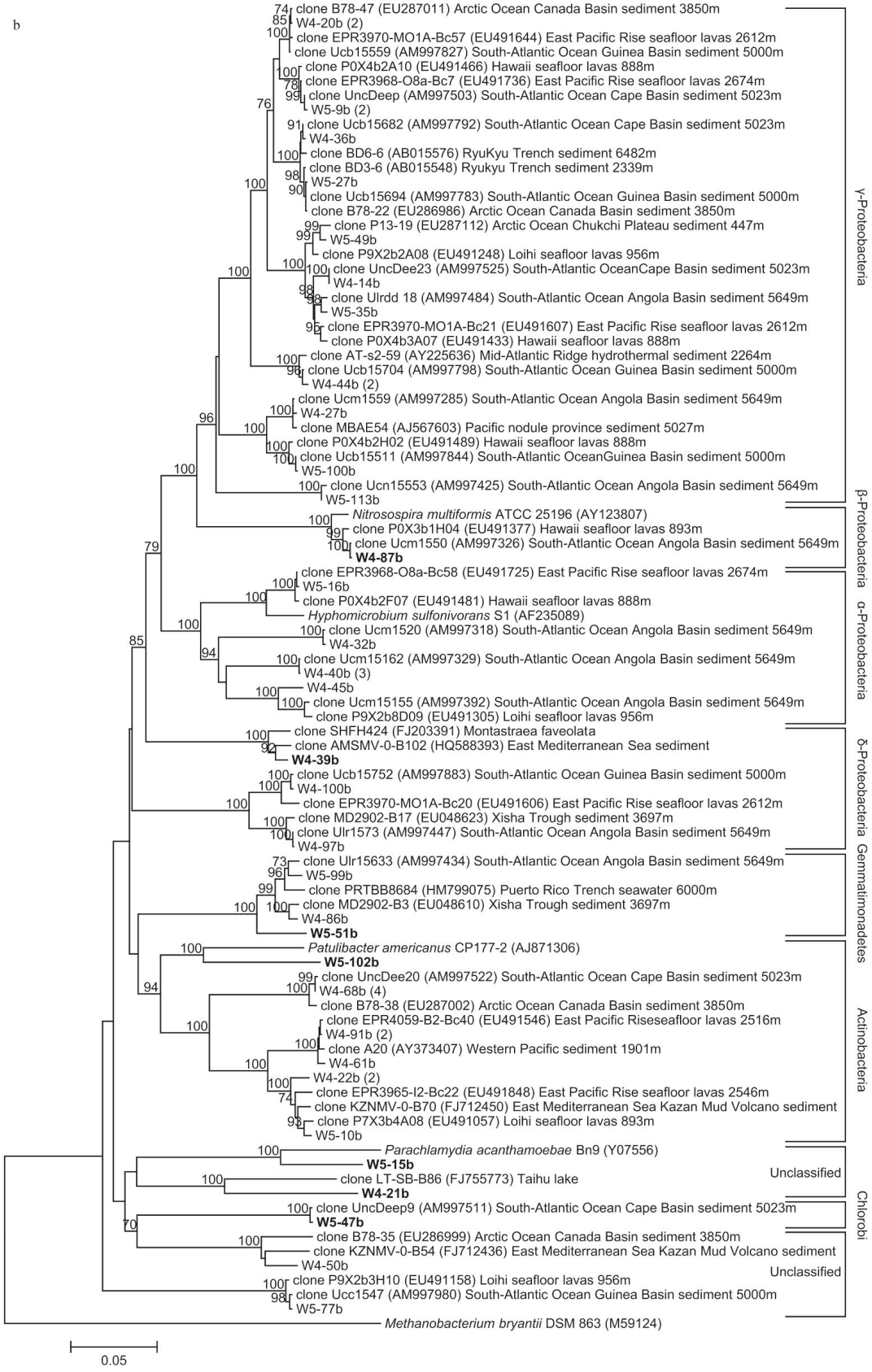


**Fig. 1.** Neighbour-joining trees of archaeal (a) and bacterial (b) 16S rRNA genes amplified from the metagenomic library SEAM02. Numbers in parentheses represent the total number of clones with the same 16S rRNA sequence. The environments where the relatives were obtained from are given after their NCBI accession numbers. Numbers at nodes correspond to bootstrap values based on 500 replicates, and the values less than 70% were omitted. Bar is 0.05 substitutions per nucleotide position.

2002). *N. maritimus* SCM1 isolated from seawater from the Seattle Aquarium was the first member of the Thaumarchaeota to have been cultured (Könneke et al., 2005), and this isolate was identified as a key chemolithoautotrophic ammonia-oxidiser in the marine environment. Fosmid Clone 74A4 was obtained from a surface sample in the Southern Ocean (Béjà et al., 2002). The 16S rRNA gene sequence of Clone W4-93a was greater than 95% identical to that of *N. maritimus* SCM1 as well as Clone 74A4. Comparative genomic analysis showed that the gene content and arrangement in the three related MGI Thaumarchaeota genome fragments were largely conserved at the 5′-end of the 16S-23S rRNA operon, but not at the 3′-end (Fig. 2). Upstream of the 16S-23S rRNA operon, a colinear region spanning 12 genes (approximately 13 kbp), was shared by the three fragments. Downstream of the 16S-23S rRNA genes, however, Strain SCM1 harboured an approximately 26.5 kbp fragment, whereas both Clones W4-93a and 74A4 did not contain this region. A single gene (ORF48, biotin—acetyl-CoA-carboxylase ligase) was common to the three fragments. Although Clone W4-93a lacked a genome fragment near the 16S-23S rRNA genes, it contained a nitrogen regulatory protein P-II gene (ORF25) and a nitroreductase gene (ORF31) that participates in nitrogen cycling.

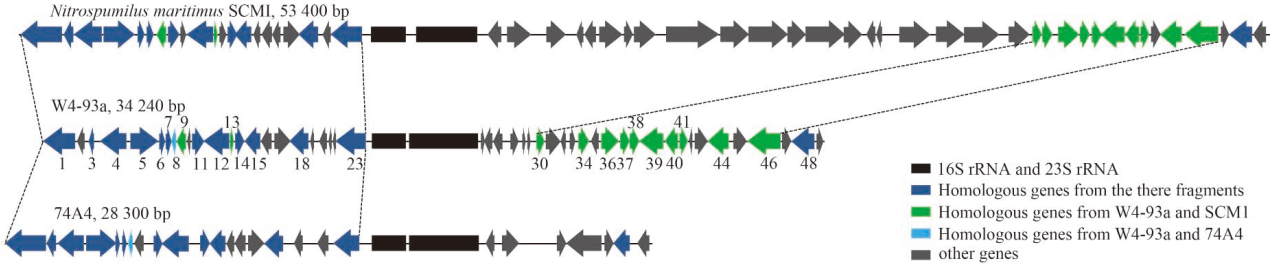


**Fig. 2.** Gene maps of fosmid Clone W4-93a, *Nitrosopumilus maritimus* SCM1 and fosmid 74A4. The rRNA genes were used as an alignment point. ORFs: 1, ATP-dependent DNA ligase; 3, hypothetical protein; 4, fructose-1,6-bisphosphatase; 5, translation elongation factor EF-1 alpha; 6, ribosomal protein S10; 7, RNA polymerase Rbp10; 8, C2H2 Zn finger protein; 9, hypothetical protein; 11, rossmann fold nucleotide-binding protein; 12, 3-hydroxybutyryl-CoA dehydrogenase; 13, hypothetical protein; 14, HIT superfamily hydrolase; 15, DnaJ class molecular chaperone; 18, TPR repeat-containing protein; 23, glutamate-1-semialdehyde aminotransferase; 30, hypothetical protein; 34, peptide methionine sulfoxide reductase; 36, alpha/beta hydrolase; 37, AbrB family transcription regulator; 38, hypothetical protein; 39, poly(R)-hydroxyalkanoic acid synthase subunit PhaC; 40, hypothetical protein; 41, hypothetical protein; 44, transcription factor TFIIB cyclin-related protein; 46, TPR repeat-containing protein; 48, biotin--acetyl-CoA-carboxylase ligase.

3.3.2 *Bacterial fosmid clones*

Clone W4-39b contained 28 ORFs and a 16S-23S-5S rRNA operon with two tRNA genes (tRNA-Ala and tRNA-Ile). The 16S rRNA and 23S rRNA sequences identified this clone as belonging to the Deltaproteobacteria. COG classification was successful in assigning functions to most of the predicted proteins, including genes responsible for L-glutamate synthesis (ORF7-8), glycerophospholipids metabolism (ORF10-11), *de novo* purine biosynthesis (ORF18), substrate transport (ORF13, ORF15-16 and ORF23) and transposition (ORF20-22). ORF20-22 encoded transposases that clustered with Alphaproteobacteria in a monophyletic tree (Appendix Fig. A1f). ORF24-26 had a small number of close relatives, including the genera *Erythrobacter, Novosphingobium, Sphingobium* and *Sphingomonas* of Alphaproteobacteria and the genera *Gallionella, Methylotenera* and *Limnobacter* of Betaproteobacteria (Appendix Figs A1g–i). Intracellular ammonium is incorporated into carbon skeletons via the glutamate/glutamine synthase pathway. Interestingly, ORF7 and ORF8, which encoded small and large subunits of glutamate synthase, were most similar to those found in Actinobacteria (Appendix Fig. A1e).

Clone W4-87b exhibited the highest 16S and 23S rRNA gene sequence identities to *Nitrosospira multiformis* ATCC 25196 (97.8% and 97.1%, respectively), which is an ammonia-oxidising bacterium (AOB) isolated from soil (Norton et al., 2008). Among the predicted 18 ORFs, only seven ORFs were related to the Betaproteobacteria class. ORF4 showed a clearly Bacteroidetes origin (Appendix Fig. A1j). Clone W4-87b encoded a urease accessory protein (UreD) involved in the activation of a urease that hydrolyses urea to ammonia. Considering the phylogenetic analyses of the rRNA genes, as well as that of *ureD*, we considered that Clone W4-87b might have derived from an AOB similar to N. *multiformis* which plays an important role in the nitrogen cycle. In addition, some oxidative stress resistance genes, including thioredoxin (ORF3 and ORF13), superoxide dismutase (ORF9) and universal stress protein A (ORF10) were found in Clone W4-87b.

A similar situation was observed for Clone W5-47b, which showed the highest 16S rRNA sequence identity to *Ignavibacterium album* Mat9-16 (92.6%) belonging to the Chlorobi. Chlorobi are obligate anaerobic photoautotrophic bacteria. Clone W5-47b contained several antibiotic-synthesising and metabolic genes from other bacteria, as well as archaea, to survive in the deep-sea environment. Examples of these were ORF1 (penicillin synthesis), ORF2-5 (pentose phosphate pathway), ORF10 (streptomycin biosynthesis), ORF17 and ORF24 (purine and pyridine metabolism) and ORF22 (poly-γ-glutamate synthesis).

The 16S and 23S rRNA genes found in Clone W5-102b identified this fragment as being derived from a member of the Actinobacteria. Most ORFs (54.8%) were predicted to be hypothetical proteins, and 13 ORFs had no orthologs in the NCBI nr database. Some ORFs might have been acquired by HGT, including ORF6 from Firmicutes (Appendix Fig. A1k), ORF10 from other bacteria (Appendix Fig. A1l), ORF18 from Euryarchaeota (Appendix Fig. A1m), ORF25 and ORF30 from Chloroflexi and ORF26 from Bacteroidia.

Tree topology with a high bootstrap value (100%) revealed that Clone W5-51b fell within a cluster composed of Gemmati-

monadetes members, and formed an independent clade (Fig. 1b). However, this clone was distinguishable from the known Gemmatimonadetes species based on the low (<85%) (Appendix Table A1) identities of its 16S rRNA gene and other functional genes. Upstream of the 5S-23S-16S rRNA genes, Clone W5-51b fragments were partially conserved, and most ORFs (10/13, 76.9%) exhibited the highest sequence identities with genes found in *Gemmatimonas aurantiaca*. Downstream of the rRNA gene cluster, ORFs were similar to those of other Gemmatimonadetes members, and most of them were more closely related to Proteobacteria. Phylogenetic analysis revealed that ORF17 and ORF18-19 might have been acquired by HGT from Alphaproteobacteria and Gammaproteobacteria, respectively (Appendix Figs A1n–o). Our data suggest that Clone W5-15b presented a new lineage in the phylum Gemmatimonadetes.

Clones W4-21b and W5-15b belonged to unknown taxonomic groups within the phylogenetic tree (Fig. 1) and showed very low 16S rRNA sequence identity with known bacterial species (<78%, Appendix Table A1). Some important functional genes were detected in these two clones. In Clone W4-21b, ORF22 encoded a SpoIID/LytB domain-containing protein that is involved in sporulation (Lopez-Diaz et al., 1986); ORF24 encodes a heavy metal resistance protein (CzcD) that is able to mediate metal efflux and so enhance the cell's resistance to cobalt, zinc, cadmium, etc. (Nies, 1992); and ORF15 (RadC), ORF16 (RecJ), ORF29 (DNA polymerase) and ORF30 (LexA repressor) were related to DNA repair. Sporulation, heavy metal resistance and DNA repair are important mechanisms of environmental stress resistance in microbes, indicating that Clone W4-21b belonged to a spore-forming bacterium adapted the deep-sea cobalt-rich ferromanganese crust environment. In Clone W5-15b, ORF1 (adenylate/guanylate cyclase), ORF3 (CspE), ORF6 (PspC), ORF9 (PhoH), ORF22 (glutathione S-transferase) and ORF26 (RecB) were recognized as the stress response regulating genes, and ORF32 encoded a poly-γ-glutamate synthesis protein. Poly-γ-glutamate is a natural polymer synthesised by gram-positive bacteria. It allows bacteria to survive at high salt concentrations and may also act as a virulence factor or a storage element for carbon and nitrogen precursors or as an energy source (Candela and Fouet, 2006). The above genes reveal that Clone W5-15b was derived from a gram-positive bacterium and was able to resist various stresses, including low temperature, high osmotic pressure and low nutrient availability. In addition, potential HGT of ORF22 from plant was identified (Appendix Fig. A1q).

## 4 Discussion

Recently, we published the first assessment of the bacterial and archaeal diversity in the sediment collected from a cobalt-rich ferromanganese crust (Liao et al., 2011). Proteobacteria and MGI Thaumarchaeota dominated the bacterial and archaeal communities, respectively. In addition, the microbial diversity inside nodules and in the surrounding sediments collected from a cobalt-rich ferromanganese crust were compared (Wu et al., 2013). Here, we focused on the ecological functions of these microbial communities with a special emphasis on their adaptive ability and survival in cobalt-rich ferromanganese crusts.

Manganese, cobalt, copper and nickel are the most important metal elements available in cobalt-rich ferromanganese crust. The content of Co + Cu + Ni is one of the significant evaluation indicators of the crust. Our study revealed that the concentrations of Mn, Ni, Co and Cu in the sediment from Sta. SEAM02 were more than twice those found in offshore sediments or in continental crust (Table 1), whereas they were at much lower concentrations than those in the cobalt-rich ferromanganese crust (data not shown). Some oxidative stress resistance genes, including thioredoxin, superoxide dismutase and PqqC, were detected in our fosmid library. A gene encoding a cobalt-zinc-cadmium resistance protein (ORF24 in clone W4-21b) was also found, suggesting that some microbes that inhabit in the sediment adapt to heavy metal toxicity (Appendix Table A1). Recent works demonstrate that cobalt-rich ferromanganese crusts are formed by biologically driven processes involving microbes (Wang et al., 2009; Wang and Müller, 2009), and that microorganisms are responsible for the bulk of Mn oxide formation (Tebo et al., 2004). Some microorganisms produce high amounts of organic acids or exopolymers to aggregate the metal granules (Gadd, 2007; Guibaud et al., 2009), and the surfaces of the budding and sheathed bacteria are surrounded by Mn oxide, which forms a shell and protects the microbes from heavy metal invasion (Ghiorse, 1984; Santelli et al., 2011). Many genes participating in the synthesis of organic acids and exoploymers were discovered, such as isopropylmalate synthase, poly-hydroxyalkanoic acid synthase and polyglutamate synthase.

The large abundance of phosphorus (Pi) and barium (Ba) in the sediment indicated that the seamount cobalt-rich ferromanganese crust region prossessed a high level of biological productivity, which may directly support a rich diversity of microbes as well as benthos on the seafloor. Oceanic Pi is usually enriched by marine organisms and settles into the sediment (Delaney, 1998). The concentration of $P_2O_5$ in the sediment from station SEAM02 was 0.22% higher than that from offshore sediment and continental crust (Table 1). Some genes involved in the P cycle were found in the fosmid library, and the ratio of those genes to the functional genes was 12.6% (22/174). Sedimentary Ba has been used as a proxy for the reconstruction of past oceanic productivity (Schenau and De Lange, 2001). The concentration of Ba in the sediment was 4.2 and 3.1 times than that in the offshore sediment and the continental crust, respectively.

In a deep-sea environment, the electron acceptors are reduced in a sequential order based on the free energy yield: $O_2$, $NO_3^-$ and $NO_2^-$, Mn and Fe, $SO_4^{2-}$, and $CO_2$ (Wright et al., 2012). The nitrogen cycle plays a key role not only in energy metabolism but also in the element cycle, which affects the Mn and Fe oxidation and even mineralisation process. Ammonia oxidation is a key process in marine nitrogen cycling and is executed by several microbial groups, including aerobic chemoautotrophic archaea (ammonia-oxidising archaea, AOA) and bacteria (ammonia-oxidising bacteria, AOB) (Beman et al., 2012; Li and Gu, 2013). *Nitrosopumilus maritimus* SCM1 was the first discovered isolate of MGI Thaumarchaeota and the first archaeal strain observed to undergo ammonia oxidation (Könneke et al., 2005). This strain grows chemoautotrophically by oxidising ammonia aerobically and assimilating carbon through the 3-hydroxypropionate/4-hydroxybutyrate pathway (Walker et al., 2010), indicating that marine MGI Thaumarchaeota may be important to the nitrogen and carbon cycles in ecosystem. In our study, MGI Thaumarchaeota was the most abundant archaea in the deep-sea sediment (Fig. 1). Ammonia monooxygenase (AMO) is a key enzyme in the ammonia oxidation process. An *amo* gene, having 98.0% amino acid sequence similarity with that from *N. maritimus* SCM1, was observed in fosmid-end sequences (data not shown). Clone W4-93a contained a 3-hydroxybutyryl-CoA dehydrogenase gene (ORF11), which is involved in the 3-hydroxypropionate/4-hydroxybutyrate cycle for autotrophic car-

bon fixation. Clone W4-87b, a potential chemoautotrophic AOB (97.8% 16S rRNA gene identity with *Nitrosospira multiformis*), was annotated. Although the *amo* gene was not observed in Clone W4-87b, an UreD gene (ORF4), which participates in the hydrolysis of urea to ammonia, was detected. The enzyme could provide a substrate for ammonia oxidation. In addition, Clone W4-93a contained putative genes for a nitrogen regulatory protein (ORF25) and a nitroreductase (ORF31). It is noteworthy that the combined microbial community and functional genes imply a nitrogen cycle was an important biogeochemical process in the deep-sea sediment from the seamount cobalt-rich ferromanganese crust region.

Our results not only support previous observations showing a relatively high abundance of Proteobacteria and MGI Thaumarchaeota in the microbial community (Liao et al., 2011) but also reveal the genomic and biological properties of the microbes in the deep-sea sediment from seamount cobalt-rich ferromanganese deposit region. In total, 78 clones containing archaeal and bacterial 16S rRNA genes were screened from 21 000 clones in the metagenome library; Proteobacteria (55.8%) and MGI Thaumarchaeota (97.1%) dominated in the bacterial and archaeal communities, respectively. The continued analysis of microbial genomic data indicated that numerous genes are involved in metabolism, which is the most abundant gene group (44.0% in genes assigned to the COG classification system) (Fig. 3). Interestingly, a significant proportion of genes (8.2%) were related to DNA transport and metabolism. The plankton and microbe from the upper layers of seawater sink to the seafloor and become an energy resource there. The total DNA sinking to the seafloor was estimated to $1.26\times10^7$ metric tons year$^{-1}$, and up to 0.45 gigatons of extracellular DNA is present in the top 10 cm of deep-sea sediments (Dell'Anno and Danovaro, 2005). Recent research has found the DNA-eating ability of *Escherichia coli* during long-term survival (Finkel and Kolter, 2001; Finkel, 2006). Considering most of the planktons and microbes from the upper layers of seawater are not able to survive in the deep-sea environment, the release of their DNA into the sediment might be used as an important nutrient for indigenous microbes. Several antibiotic-synthesising genes (penicillin and streptomycin) were also detected in the metagenome library. Antibiotics can inhibit or kill some microbes, as well as benthos, and help the microbes to occupy an ecological niche. Therefore, it is essential for microbes to have alternative metabolic pathways to survive in the deep-sea environment.

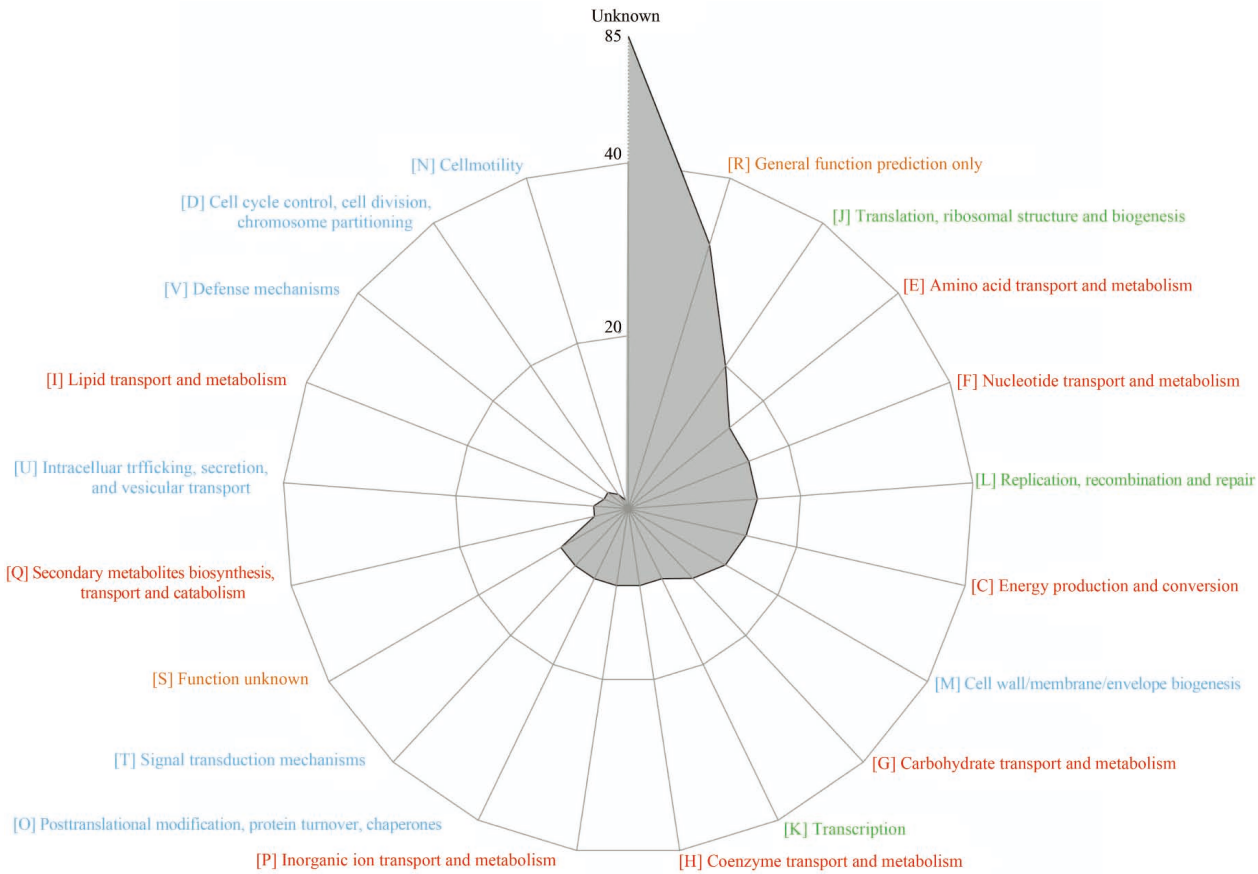


**Fig. 3.** Functional classification of genes of the nine fosmids according to COG classification system. Blue represents the COG corresponding to the "cellular processes and signalling", green "information storage and processing", red "metabolism", and orange "poorly characterised".

Genomic divergence and HGT played important roles in the microbial adaption to the heavy metal rich and cold deep-sea environment. Previous comparative genomic studies of uncultivated marine planktonic archaea from different oceanic regions revealed significant genomic divergences, regardless of the 16S rRNA gene sequence variation (Béjà et al., 2002; Martin-Cuadrado et al., 2008). In our study, the comparative genomic analysis of Clone W4-93a, *N. maritimus* SCM1 and fosmid Clone 74A4 suggested that considerable genome divergence exists at the genus level (95.2% 16S rRNA gene sequence identity) between

sedimentary and planktonic lineages (Fig. 2). Although the 16S rRNA gene sequences of fosmid W4-87b and *N. multiformis* ATCC 25196 showed a high identity (97.8%), their G+C content (44.7% and 53.9%) and genome synteny surrounding the rRNA operon (data not shown) were surprisingly different. CloneW5-51b also showed different genome synteny with *Gemmatimonas aurantiaca* T-27$^T$, which is the only isolate in class Gemmatimonadetes (Zhang et al., 2003). These differences may have been caused by genome evolution during adaption to different habitats. Many HGT events have been observed in genomes, and the HGT rate was 11.4% (23/201) among the seven fragments of known phylogenetic lineages (W5-61a, W4-93a, W4-39b, W4-87b, W5-47b, W5-102b and W5-51b). However, 17.9% (36/201) of the predicted genes have no significant relatives, and the origin of some cannot be determined due to a lack of sufficient information in the database. Taken together, the 11.4% HGT rate should be most likely an underestimation. Most HGTs were from bacteria to bacteria, with a few possibly were from bacteria to Thaumarchaeota, from Euryarchaeota to Thaumarchaeota, from Euryarchaeota to bacteria, and even from eukarya to bacteria (Appendix Fig. A1). Most of the genes acquired through HGT were involved in metabolism, including carbon and energy metabolism (isopropylmalate dehydratase, pyrroloquinoline quinone biosynthesis protein C, glutamate synthase, cytochrome *c* family protein, MIP family channel protein and alcohol dehydrogenase) and nitrogen metabolism (nitrogen regulatory protein and urease accessory protein UreD).

In conclusion, element concentrations in the sediment from the seamount cobalt-rich ferromanganese crust region are different from those in other marine or terrestrial environments. The large abundances of heavy metals (Mn, Ni, Co, Cu), P and Ba in the sediment from Sta. SEAM02 implied a unique microbial community with high biodiversity. Microbes inhabiting the cobalt-rich ferromanganese crust region not only adapt to high amounts of heavy metal but also might participate in the biomineralization process, as observed at the gene level. Alternative metabolic pathways and a variety of stress genes are essential for microbial survival in the deep-sea environment. Genomic divergence and HGT may played important roles in the microbial adaption to the deep-sea environment. Some microbes, which come from the upper seawater, might obtain a series of new features and adapt to this harsh environment via high frequency HGT events. The information gathered via the rRNA-gene based PCR screening method provided insight only into the genomic regions directly adjacent to rRNA operons. However, this is the first metagenomic study of deep-sea sediment from the cobalt-rich ferromanganese crust region, giving us some insights into the genetic and functional information about uncultured microorganisms in the cobalt-rich ferromanganese crust region.

**References**


Adachi O, Ano Y, Toyama H, et al. 2007. Biooxidation with PQQ- and FAD-dependent dehydrogenases. In: Schmid R D, Urlacher V B, eds. Modern Biooxidation: Enzymes, Reactions and Applications. Hoboken, NJ: John Wiley & Sons, Inc, 41

Béjà O, Koonin E V, Aravind L, et al. 2002. Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. Applied and Environmental Microbiology, 68(1): 335–345

Beman J M, Popp B N, Alford S E. 2012. Quantification of ammonia oxidation rates and ammonia-oxidizing archaea and bacteria at high resolution in the Gulf of California and eastern tropical North Pacific Ocean. Limnology and Oceanography, 57(3): 711–726

Blainey P C, Mosier A C, Potanina A, et al. 2011. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. PLoS One, 6(2): e16626

Candela T, Fouet A. 2006. Poly-gamma-glutamate in bacteria. Molecular Microbiology, 60(5): 1091–1098

Clark M R, Rowden A A, Schlacher T, et al. 2010. The ecology of seamounts: structure, function, and human impacts. Ann Rev Mar Sci, 2: 253–278

Craig J D, Andrews J E, Meylan M A. 1982. Ferromanganese deposits in the Hawaiian Archipelago. Marine Geology, 45(1–2): 127–157

Delaney M L. 1998. Phosphorus accumulation in marine sediments and the oceanic phosphorus cycle. Global Biogeochemical Cycles, 12(4): 563–572

Dell'Anno A, Danovaro R. 2005. Extracellular DNA plays a key role in deep-sea ecosystem functioning. Science, 309(5744): 2179

DeLong E F. 1992. Archaea in coastal marine environments. Proceedings of the National Academy of Sciences of the United States of America, 89(12): 5685–5689

Duffy E J. 2008. "Seamount". CenSeam: a global census of marine life on seamounts content partner and national oceanic and atmospheric administration content source. In: Cleveland C J, ed. Encycleopedia of Earth. Washington DC: Environmental Information Coalition, National Council for Science and the Environment

Ehrhardt C J, Haymon R M, Lamontagne M G, et al. 2007. Evidence for hydrothermal Archaea within the basaltic flanks of the East Pacific Rise. Environmental Microbiology, 9(4): 900–912

Emerson D, Moyer C L. 2010. Microbiology of seamounts: common patterns observed in community structure. Oceanography, 23: 148–163

Finkel S E. 2006. Long-term survival during stationary phase: evolution and the GASP phenotype. Nat Rev Microbiol, 4(2): 113–120

Finkel S E, Kolter R. 2001. DNA as a nutrient: novel role for bacterial competence gene homologs. Journal of Bacteriology, 183(21): 6288–6293

Fu Yazhou, Peng Jiantang, Qu Wenjun, et al. 2005. Os isotopic compositions of a cobalt-rich ferromanganese crust profile in Central Pacific. Chin Sci Bull, 50(18): 2106–2112

Gadd G M. 2007. Geomycology: biogeochemical transformations of rocks, minerals, metals and radionuclides by fungi, bioweathering and bioremediation. Mycological Research, 111(Pt 1): 3–49

García-Martínez J, Martínez-Murcia A, Antón A I, et al. 1996. Comparison of the small 16S to 23S intergenic spacer region (ISR) of the rRNA operons of some Escherichia coli strains of the ECOR collection and *E. coli* K-12. Journal of Bacteriology, 178(21): 6374–6377

Ghiorse W C. 1984. Biology of iron- and manganese-depositing bacteria. Annu Rev Microbiol, 38: 515–550

Gillan D C, Danis B. 2007. The archaebacterial communities in Antarctic bathypelagic sediments. Deep-Sea Research Part II: Topical Studies in Oceanography, 54(16–17): 1682–1690

Guibaud G, van Hullebusch E, Bordas F, et al. 2009. Sorption of Cd(II) and Pb(II) by exopolymeric substances (EPS) extracted from activated sludges and pure bacterial strains: Modeling of the metal/ligand ratio effect and role of the mineral fraction. Bioresource Technology, 100(12): 2959–2968

Heijs S K, Haese R R, van der Wielen P W, et al. 2007. Use of 16S rRNA gene based clone libraries to assess microbial communities potentially involved in anaerobic methane oxidation in a Mediterranean cold seep. Microbial Ecology, 53(3): 384–398

Hillier J K, Watts A B. 2007. Global distribution of seamounts from ship-track bathymetry data. Geophysical Research Letters, 34(13): L13304

Ito M, Tsunekawa M, Yamaguchi E, et al. 2008. Estimation of degree of liberation in a coarse crushed product of cobalt-rich ferromanganese crust/nodules and its gravity separation. International Journal of Mineral Processing, 87(3–4): 100–105

Iyer S D, Mehta M C, Das P, et al. 2012. Seamounts—characteristics, formation, mineral deposits and biodiversity. Geologica Acta, 10(3): 295–308

Jiang Xiawei, Xu Xuewei, Huo Yingyi, et al. 2012. Identification and characterization of novel esterases from a deep-sea sediment metagenome. Archives of Microbiology, 194(3): 207–214

Könneke M, Bernhard A E, de la Torre J R, et al. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature, 437(7058): 543–546

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research, 28(1): 27–30

Kato S, Kobayashi C, Kakegawa T, et al. 2009. Microbial communities in iron-silica-rich microbial mats at deep-sea hydrothermal fields of the Southern Mariana Trough. Environmental Microbiology, 11(8): 2094–2111

Kato S, Yanagawa K, Sunamura M, et al. 2009. Abundance of *Zetaproteobacteria* within crustal fluids in back-arc hydrothermal fields of the Southern Mariana Trough. Environmental Microbiology, 11(12): 3210–3222

Koschinsky A, Hein J R. 2003. Uptake of elements from seawater by ferromanganese crusts: solid-phase associations and seawater speciation. Marine Geology, 198(3–4): 331–351

Li Meng, Gu Jidong. 2013. Community structure and transcript responses of anammox bacteria, AOA, and AOB in mangrove sediment microcosms amended with ammonium and nitrite. Applied Microbiology and Biotechnology, 97(22): 9859–9874

Li Youxun, Li Fuchao, Zhang Xiaowen, et al. 2008. Vertical distribution of bacterial and archaeal communities along discrete layers of a deep-sea cold sediment sample at the East Pacific Rise (~13°N). Extremophiles, 12(4): 573–585

Liao L, Xu X W, Jiang X W, et al. 2011. Microbial diversity in deep-sea sediment from the cobalt-rich crust deposit region in the Pacific Ocean. FEMS Microbiology Ecology, 78(3): 565–585

Lopez-Diaz I, Clarke S, Mandelstam J. 1986. spoIID Operon of Bacillus subtilis: cloning and sequence. Journal of General Microbiology, 132(2): 341–354

Lösekann T, Knittel K, Nadalig T, et al. 2007. Diversity and abundance of aerobic and anaerobic methane oxidizers at the Haakon Mosby Mud Volcano, Barents Sea. Applied and Environmental Microbiology, 73(10): 3348–3362

Martín-Cuadrado A B, López-García P, Alba J C, et al. 2007. Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. PLoS One, 2(9): e914

Martin-Cuadrado A B, Rodriguez-Valera F, Moreira D, et al. 2008. Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. The ISME Journal, 2(8): 865–886

Menard H W. 1964. Marine Geology of the Pacific. New York: McGraw-Hill

Miroshnichenko M, Hippe H, Stackebrandt E, et al. 2001. Isolation and characterization of *Thermococcus sibiricus* sp. nov. from a Western Siberia high-temperature oil reservoir. Extremophiles, 5(2): 85–91

Muiños S B, Hein J R, Frank M, et al. 2013. Deep-sea Fe-Mn crusts from the Northeast Atlantic Ocean: composition and resource considerations. Marine Georesources & Geotechnology, 31(1): 40–70

Nies D H. 1992. Resistance to cadmium, cobalt, zinc, and nickel in microbes. Plasmid, 27(1): 17–28

Norton J M, Klotz M G, Stein L Y, et al. 2008. Complete genome sequence of *Nitrosospira multiformis*, an ammonia-oxidizing bacterium from the soil environment. Applied and Environmental Microbiology, 74(11): 3559–3572

Rowden A A, Dower J F, Schlacher T A, et al. 2010. Paradigms in seamount ecology: fact, fiction and future. Marine Ecology, 31(Supp): 226–241

Santelli C M, Orcutt B N, Banning E, et al. 2008. Abundance and diversity of microbial life in ocean crust. Nature, 453(7195): 653–656

Santelli C M, Webb S M, Dohnalkova A C, et al. 2011. Diversity of Mn oxides produced by Mn(II)-oxidizing fungi. Geochimica et Cosmochimica Acta, 75(10): 2762–2776

Schauer R, Bienhold C, Ramette A, et al. 2009. Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. The ISME Journal, 4(2): 159–170

Schenau S J, De Lange G J. 2001. Phosphorus regeneration vs. burial in sediments of the Arabian Sea. Marine Chemistry, 75(3): 201–217

Schlacher T A, Rowden A A, Dower J F, et al. 2010. Seamount science scales undersea mountains: new research and outlook. Marine Ecology, 31(Supp): 1–13

Staudigel H, Koppers A A P, Plank T A, et al. 2010. Seamounts in the subduction factory. Oceanography, 23(1): 176–181

Tamura K, Peterson D, Peterson N, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28(10): 2731–2739

Tatusov R L, Galperin M Y, Natale D A, et al. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research, 28(1): 33–36

Tebo B M, Bargar J R, Clement B G, et al. 2004. Biogenic manganese oxides: properties and mechanisms of formation. Annual Review of Earth and Planetary Sciences, 32: 287–328

Verlaan P A. 1992. Benthic recruitment and manganese crust formation on seamounts. Marine Biology, 113(1): 171–174

Walker C B, de la Torre J R, Klotz M G, et al. 2010. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. Proceedings of the National Academy of Sciences of the United States of America, 107(19): 8818–8823

Wang Xiaohong, Müller W E G. 2009. Marine biominerals: perspectives and challenges for polymetallic nodules and crusts. Trends in Biotechnology, 27(6): 375–383

Wang Xiaohong, Schlossmacher U, Natalio F, et al. 2009. Evidence for biogenic processes during formation of ferromanganese crusts from the Pacific Ocean: implications of biologically induced mineralization. Micron, 40(5–6): 526–535

Wang Xiaohong, Wiens M, Schröder H, et al. 2011. Molecular biomineralization: toward an understanding of the biogenic origin of polymetallic nodules, seamount crusts, and hydrothermal vents. In: Müller W E G, ed. Molecular Biomineralization. Berlin Heidelberg: Springer, 77–110

Wedepohl K H. 1995. The composition of the continental crust. Geochimica et Cosmochimica Acta, 59(7): 1217–1232

Wessel P, Sandwell D T, Kim S S. 2010. The global seamount census. Oceanography, 23(1): 24–33

Wright J J, Konwar K M, Hallam S J. 2012. Microbial ecology of expanding oxygen minimum zones. Nat Rev Microbiol, 10(6): 381–394

Wu Yuehong, Liao Li, Wang Chunsheng, et al. 2013. A comparison of microbial communities in deep-sea polymetallic nodules and the surrounding sediments in the Pacific Ocean. Deep-Sea Research Part I: Oceanographic Research Papers, 79: 40–49

Xu Meixiang, Wang Peng, Wang Fengping, et al. 2005. Microbial diversity at a deep-sea station of the Pacific nodule province. Biodiversity & Conservation, 14(14): 3363–3380

Yesson C, Clark M R, Taylor M L, et al. 2011. The global distribution of seamounts based on 30 arc seconds bathymetry data. Deep-Sea Research Part I: Oceanographic Research Papers, 58(4): 442–453

Zhang Hui, Sekiguchi Y, Hanada S, et al. 2003. *Gemmatimonas aurantiaca* gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. International Journal of Systematic and Evolutionary Microbiology, 53(4): 1155–1163

Zhang Fuyuan, Zhang Weiyan, Zhu Kechao, et al. 2008. Distribution characteristics of cobalt-rich ferromanganese crust resources on submarine seamounts in the Western Pacific. Acta Geologica Sinica, 82(4): 796–803

Zhao Qiyuan. 1988. Ocean Geochemistry. Beijing: The Geological Publishing House

Zhu Wenhan, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Research, 38(12): e132

# Appendix:

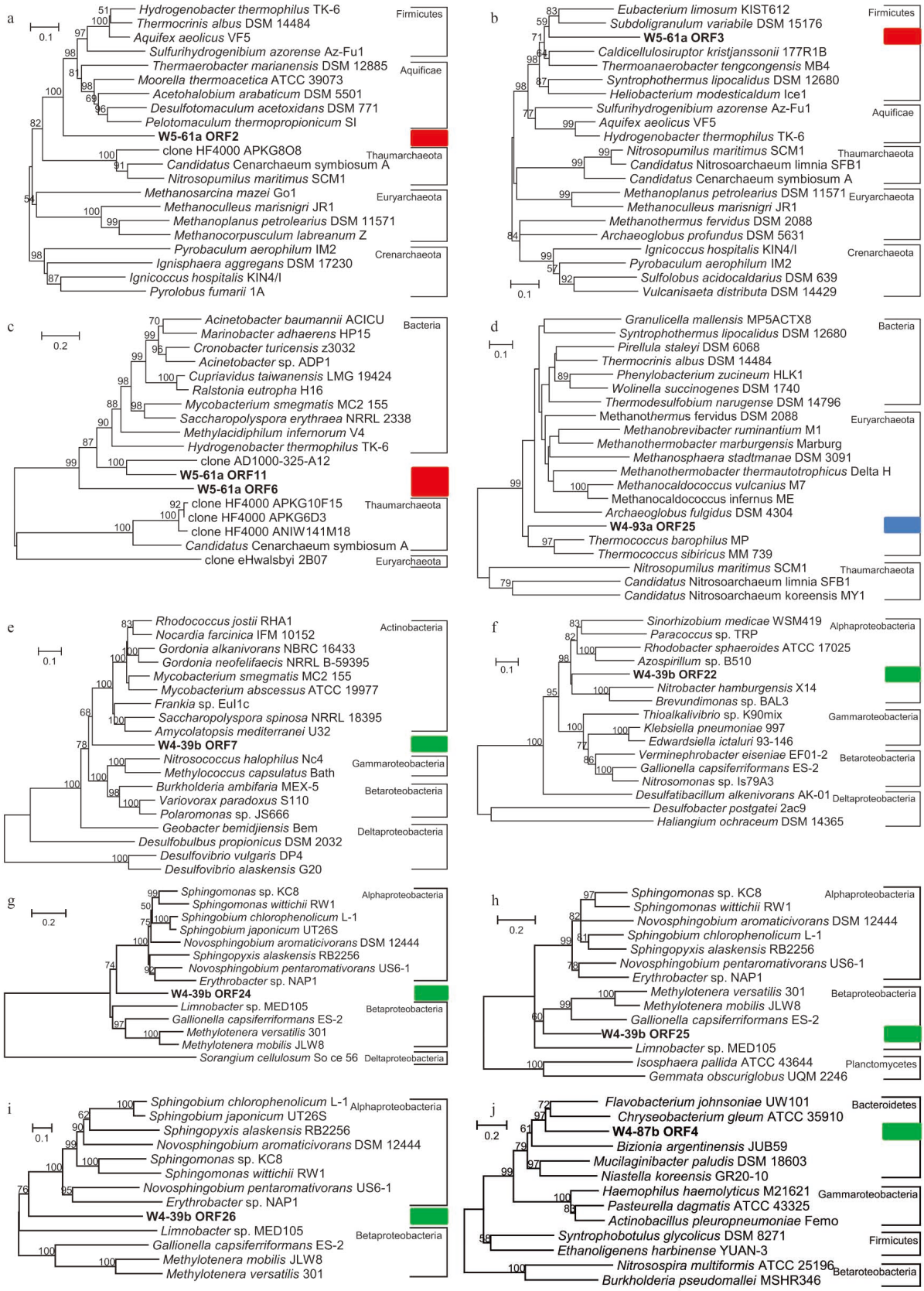


Fig. A1.

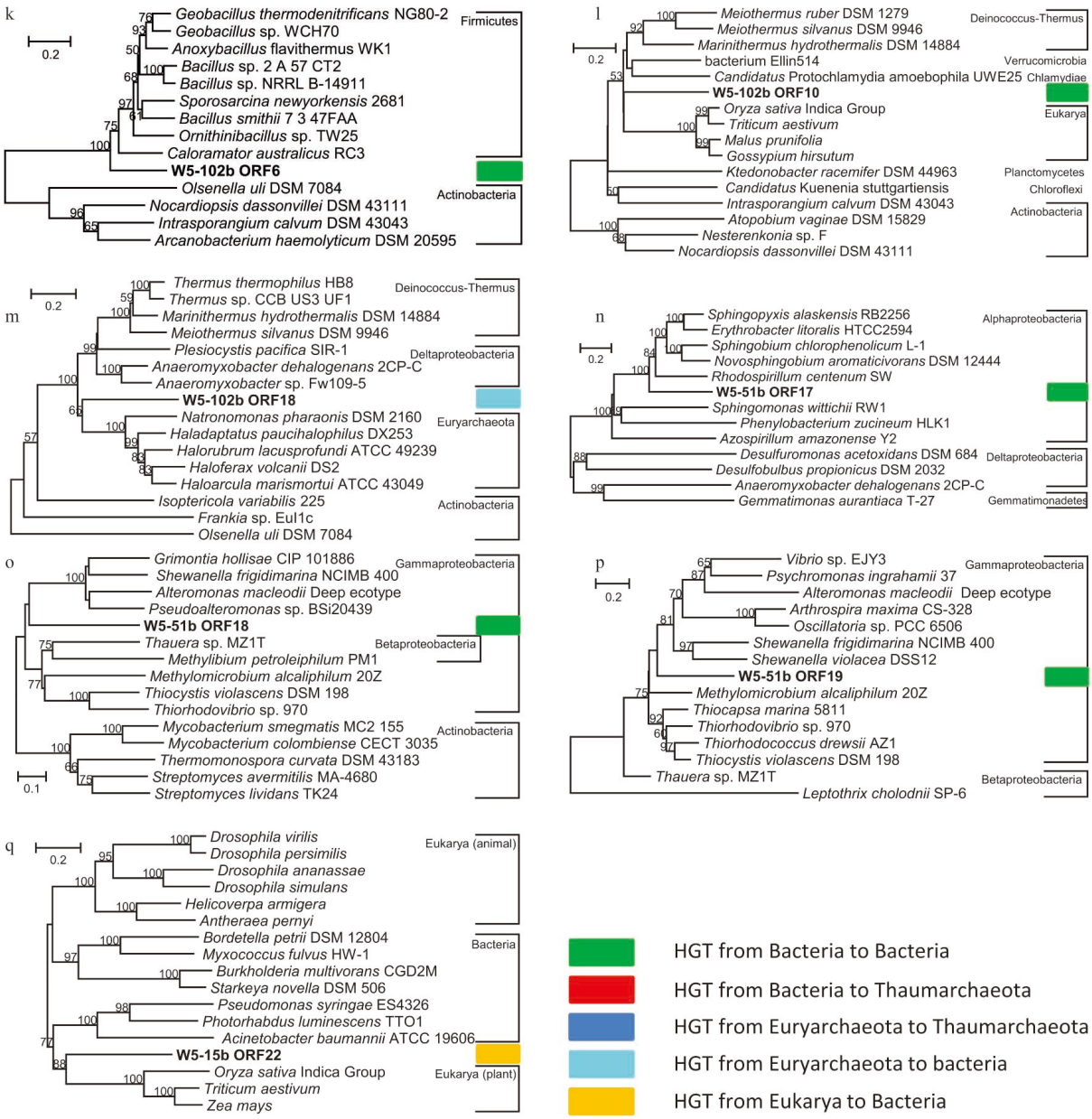


**Fig. A1.** Neighbor-joining phylogenetic trees illustrating HGT events. a. Large subunit of 3-isopropylmalate dehydratase (ORF2 OF w5-61a); b. small subunit of 3-isopropylmalate dehydratase (ORF3 of W5-61a); c. pyrroloquinoline quinone biosynthesis protein C (ORF6 and ORF11 of W5-61a); d. hypothetical protein (ORF25 of W4-93a); e. small subunit of glutamate synthase (ORF7 of W4-39b); f. transposase (ORF22 of W4-39b); g. cyclic nucleotide-binding potein (ORF24 of W4-39b); h. hypothetical protein (ORF25 of W4-39b); i. cytochrome c family protein (ORF26 of W4-39b); j. urease accessory protein UreD (ORF4 of W4-87b); k. thymidine kinase (ORF6 of W5-102b); l. MIP family channel pritein (ORF10 of W5-102b); m. alcohol dehydrogenase (ORF18 of W5-102b); n. TonB-dependent receptor (ORF17 of W5-15b); o. hypothetical protein (ORF18 of W5-51b); p. peptidase (ORF19 of W5-51b); and q. glutarthione S-transferase (ORF22 of W5-15b). Amino acid sequences obtained in this study is shown in bold. Numbers at nodes correspond to boostrap values based on 500 replicates, and the values less than 50% were omitted.

**Table A1.** Predicted proteins and RNA encoded by the fosmid clones

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| W5-61a, 32 237 bp, 48.6% G+C content | | | | | | | | | | | |
| 1 | 3–656 | 217 | isopropylmalate/citramalate/homocitrate synthase | *Methanopyrus kandleri* | *Methanopyri* | *Euryarchaeota* | 111/208(53%) | 156/208(75%) | 2E-66 | COG0119 | [E] |
| 2 | 670–1947 | 425 | 3-isopropylmalate dehydratase, large subunit | *Thermocrinis albus* | *Aquificae* | *Aquificae* | 240/419(57%) | 303/419(72%) | 7E-170 | COG0065 | [E] |
| 3 | 1951–2451 | 166 | 3-isopropylmalate dehydratase, small subunit | *Thermoanaerobacter italicus* | *Clostridia* | *Firmicutes* | 97/160(61%) | 114/160(71%) | 4E-64 | COG0066 | [E] |
| 4 | 2852–2448 | 134 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 5 | 3998–2904 | 364 | NAD-dependent epimerase/dehydratase | *Thermoproteus neutrophilus* | *Thermoprotei* | *Crenarchaeota* | 132/307(43%) | 193/307(63%) | 4E-72 | COG1089 | [M] |
| 6 | 4823–4125 | 232 | pyrroloquinoline quinone biosynthesis protein PqqC | *Burkholderia ambifaria* | *Betaproteobacteria* | *Proteobacteria* | 77/214(36%) | 117/214(55%) | 4E-38 | COG5424 | [H] |
| 7 | 6898–5177 | 573 | iron complex ABC transporter permease | *Candidatus* Caldiarchaeum subterraneum | – | *Thaumarchaeota* | 197/540(36%) | 329/540(61%) | 2E-112 | COG1178 | [P] |
| 8 | 8033–6984 | 349 | iron complex ABC transporter substrate-binding protein | *Candidatus* Caldiarchaeum subterraneum | – | *Thaumarchaeota* | 123/310(40%) | 180/310(58%) | 3E-74 | COG1840 | [P] |
| 9 | 9096–8101 | 331 | iron complex ABC transporter substrate-binding protein | *Candidatus* Caldiarchaeum subterraneum | – | *Thaumarchaeota* | 109/319(34%) | 179/319(56%) | 4E-53 | COG1840 | [P] |
| 10 | 9199–10014 | 271 | transcriptional regulator ArsR family | *Archaeoglobus veneficus* | *Archaeoglobi* | *Euryarchaeota* | 61/246(25%) | 107/246(44%) | 1E-09 | COG1719 | [R] |
| 11 | 10727–10011 | 238 | pyrroloquinoline quinone biosynthesis protein PqqC | clone AD1000-325-A12 | – | *Thaumarchaeota* | 123/228(54%) | 175/228(77%) | 4E-95 | COG5424 | [H] |
| 12 | 12728–10950 | 592 | PQQ-dependent alcohol dehydrogenase | *Candidatus* Caldiarchaeum subterraneum | – | *Thaumarchaeota* | 234/584(40%) | 347/584(59%) | 2E-143 | COG4993 | [G] |
| 13 | 12877–13668 | 263 | DNA methylase N-4/N-6 | clone KM3-153-F8 | – | *Thaumarchaeota* | 162/239(68%) | 191/239(80%) | 2E-122 | COG0863 | [L] |
| rRNA | 17721–14669 | | 23S rRNA gene | *Candidatus* Nitrososphaera gargensis | – | *Thaumarchaeota* | 2273/2921(78%) | – | 0 | – | – |
| rRNA | 19402–17904 | | 16S rRNA gene | *Candidatus* Nitrososphaera gargensis | – | *Thaumarchaeota* | 1253/1500(84%) | – | 0 | – | – |
| 14 | 19665–20507 | 280 | apurinic endonuclease Apn1 | *Methanoculleus marisnigri* | *Methanomicrobia* | *Euryarchaeota* | 150/279(54%) | 190/279(68%) | 3E-101 | COG0648 | [L] |
| 15 | 20580–20849 | 89 | phosphoribosylformylglycinamidine synthase, purS component | *Candidatus* Cenarchaeum symbiosum | – | *Thaumarchaeota* | 42/85(49%) | 57/85(67%) | 7E-22 | COG1828 | [F] |
| 16 | 20842–21564 | 240 | phosphoribosylformylglycinamidine synthase glutamine amidotransferase subunit | *Methanohalophilus mahii* | *Methanomicrobia* | *Euryarchaeota* | 119/230(52%) | 168/230(73%) | 6E-88 | COG0047 | [F] |
| 17 | 21569–23773 | 734 | phosphoribosylformylglycinamidine synthase synthetase subunit | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 347/729(48%) | 489/729(67%) | 0 | COG0046 | [F] |
| 18 | 23776–25179 | 467 | glutamine phosphoribosylpyrophosphate amidotransferase | *Candidatus* Cenarchaeum symbiosum | – | *Thaumarchaeota* | 218/457(48%) | 302/457(66%) | 2E-152 | COG0034 | [F] |
| 19 | 25453–26517 | 354 | glutamine phosphoribosylpyrophosphate amidotransferase | *Sulfolobus islandicus* | *Thermoprotei* | *Crenarchaeota* | 150/313(48%) | 214/313(68%) | 2E-101 | COG0034 | [F] |
| 20 | 26519–28045 | 508 | phosphoribosylamine/glycine ligase | *Vulcanisaeta distributa* | *Thermoprotei* | *Crenarchaeota* | 216/495(44%) | 301/495(61%) | 2E-125 | COG0151 | [F] |
| 21 | 28156–29007 | 283 | phosphoribosylaminoimidazole-succinocarboxamide synthase | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 146/274(53%) | 190/274(69%) | 5E-99 | COG0152 | [F] |
| 22 | 29092–30072 | 326 | phosphoribosylformylglycinamidine cyclo-ligase | *Thermococcus barophilus* | *Thermococci* | *Euryarchaeota* | 146/280(52%) | 198/280(71%) | 1E-87 | COG0150 | [F] |
| 23 | 30056–31075 | 339 | glyoxylate reductase | *Ferroplasma acidarmanus* | *Thermoplasmata* | *Euryarchaeota* | 147/282(52%) | 202/282(72%) | 3E-95 | COG1052 | [CHR] |
| 24 | 32121–31099 | 340 | cysteine synthase | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 160/309(52%) | 212/309(69%) | 5E-93 | COG0031 | [E] |
| 25 | 32237–32142 | 31 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| W4-93a, 34 240 bp, 36.3% G+C content | | | | | | | | | | | |
| 1 | 1402–2 | 466 | ATP-dependent DNA ligase | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 356/467(76%) | 419/467(90%) | 0 | COG1793 | [L] |
| 2 | 1821–1492 | 109 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| 3 | 2211–1990 | 73 | hypothetical protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 63/73(82%) | 68/73(93%) | 1E-39 | – | – |
| 4 | 3630–2482 | 382 | fructose-1,6-bisphosphatase | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 317/378(84%) | 352/378(93%) | 0 | COG1980 | [G] |
| 5 | 3831–5048 | 405 | translation elongation factor EF-1 alpha | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 352/405(87%) | 378/405(93%) | 0 | COG5256 | [J] |
| 6 | 5131–5367 | 78 | ribosomal protein S10 | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 71/77(92%) | 77/77(100%) | 6E-35 | COG0051 | [J] |
| 7 | 5409–5639 | 76 | RNA polymerase Rbp10 | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 50/76(66%) | 57/76(75%) | 2E-17 | – | – |
| 8 | 5836–5636 | 66 | C2H2 Zn finger protein | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 45/65(69%) | 53/65(82%) | 1E-17 | – | – |
| 9 | 6246–5836 | 136 | hypothetical protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 108/136(79%) | 122/136(90%) | 4E-58 | – | – |
| 10 | 6458–6333 | 41 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 11 | 6536–7045 | 169 | rossmann fold nucleotide-binding protein | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 120/169(71%) | 145/169(86%) | 4E-58 | COG1611 | [R] |
| 12 | 8135–7035 | 366 | 3-hydroxybutyryl-CoA dehydrogenase | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 310/366(85%) | 339/366(93%) | 0 | COG1250 | [I] |
| 13 | 8227–8397 | 56 | hypothetical protein | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 42/56(75%) | 48/56(86%) | 3E-14 | – | – |
| 14 | 8412–8819 | 135 | HIT superfamily hydrolase | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 83/133(62%) | 109/133(82%) | 2E-43 | COG0537 | [FGR] |
| 15 | 9495–8821 | 224 | DnaJ class molecular chaperone | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 130/227(57%) | 168/227(74%) | 1E-66 | COG0484 | [O] |
| 16 | 9993–9535 | 152 | transcriptional regulator | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 80/149(54%) | 103/149(69%) | 9E-34 | COG1917 | [S] |
| 17 | 10086–10790 | 234 | hypothetical protein | *Candidatus* Nitrosoarchaeum koreensis | – | *Thaumarchaeota* | 124/230(54%) | 180/230(78%) | 2E-68 | – | – |
| 18 | 11610–10792 | 272 | TPR repeat-containing protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 163/272(60%) | 215/272(79%) | 2E-94 | COG0457 | [R] |
| 19 | 11978–11802 | 58 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 20 | 12432–12103 | 109 | hypothetical protein | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 76/106(72%) | 95/106(90%) | 6E-40 | – | – |
| 21 | 12577–12443 | 44 | hypothetical protein | No significant similarities found | – | – | – | – | – | – | – |
| 22 | 12791–12648 | 47 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 23 | 14124–12817 | 435 | glutamate-1-semialdehyde aminotransferase | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 316/432(73%) | 368/432(85%) | 0 | COG0001 | [H] |
| rRNA | 14409–15878 | | 16S rRNA gene | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 1395/1466(95%) | – | 0 | – | – |
| rRNA | 16043–19040 | | 23S rRNA gene | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 2724/3013(91%) | – | 0 | – | – |
| 24 | 19163–19324 | 53 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 25 | 19680–19321 | 119 | nitrogen regulatory protein P-II | *Thermococcus sibiricus* | *Thermococci* | *Euryarchaeota* | 58/119(49%) | 71/119(60%) | 6E-19 | COG0347 | [E] |
| 26 | 20115–19723 | 130 | hypothetical protein | *Candidatus* Cenarchaeum symbiosum | – | *Thaumarchaeota* | 96/130(74%) | 111/130(85%) | 1E-50 | – | – |
| 27 | 20481–20828 | 115 | hypothetical protein | *Candidatus* Cenarchaeum symbiosum | – | *Thaumarchaeota* | 57/115(50%) | 89/115(77%) | 6E-29 | – | – |
| 28 | 20989–21123 | 44 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 29 | 21189–21317 | 42 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| 30 | 21617–21985 | 122 | hypothetical protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 93/118(79%) | 109/118(92%) | 2E-47 | – | – |
| 31 | 22025–22675 | 216 | nitroreductase | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 152/209(73%) | 176/209(84%) | 2E-86 | COG0778 | [C] |
| 32 | 22983–22777 | 68 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 33 | 23090–23371 | 93 | hypothetical protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 73/93(78%) | 85/93(91%) | 3E-36 | COG3432 | [K] |
| 34 | 23450–23917 | 155 | peptide methionine sulfoxide reductase | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 109/151(72%) | 128/151(85%) | 1E-61 | COG0225 | [O] |
| 35 | 24349–24002 | 115 | universal stress protein A domain-containing protein | *Candidatus* Cenarchaeum symbiosum | – | *Thaumarchaeota* | 62/113(55%) | 85/113(75%) | 1E-29 | – | – |
| 36 | 24445–25233 | 262 | alpha/beta hydrolase | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 172/260(66%) | 219/260(84%) | 9E-107 | COG0596 | [R] |
| 37 | 25278–25703 | 141 | AbrB family transcriptional regulator | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 96/141(68%) | 123/141(87%) | 8E-53 | – | – |
| 38 | 25700–26125 | 141 | hypothetical protein | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 98/140(70%) | 121/140(86%) | 9E-53 | – | – |
| 39 | 27242–26172 | 356 | poly(R)-hydroxyalkanoic acid synthase subunit PhaC | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 264/351(75%) | 315/351(90%) | 3E-165 | COG3243 | [I] |
| 40 | 27822–27226 | 198 | hypothetical protein | *Candidatus* Cenarchaeum symbiosum | – | *Thaumarchaeota* | 132/183(72%) | 158/183(86%) | 1E-75 | – | – |
| 41 | 27877–28269 | 130 | hypothetical protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 94/130(72%) | 114/130(88%) | 9E-49 | – | – |
| 42 | 28450–28307 | 47 | hypothetical protein | *Candidatus* Nitrosoarchaeum koreensis | – | *Thaumarchaeota* | 21/36(58%) | 31/36(86%) | 0.000007 | – | – |
| 43 | 28556–29110 | 184 | hypothetical protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 65/160(41%) | 99/160(62%) | 6E-26 | – | – |
| 44 | 30039–29116 | 307 | transcription factor TFIIB cyclin-related protein | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 250/303(83%) | 285/303(94%) | 6E-149 | COG1405 | [K] |
| 45 | 30257–30838 | 193 | DNA protection protein DPS | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 159/194(82%) | 180/194(93%) | 2E-92 | COG2406 | [R] |
| 46 | 32294–30858 | 478 | TPR repeat-containing protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 281/479(59%) | 358/479(75%) | 0 | COG0457 | [R] |
| 47 | 32386–32790 | 134 | hypothetical protein | *Candidatus* Nitrosoarchaeum limnia | – | *Thaumarchaeota* | 37/75(49%) | 50/75(67%) | 1E-08 | – | – |
| 48 | 33791–32787 | 334 | biotin—acetyl-CoA-carboxylase ligase | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 210/332(63%) | 272/332(82%) | 5E-122 | COG0340 | [H] |
| 49 | 33891–34190 | 99 | hypothetical protein | *Nitrosopumilus maritimus* | – | *Thaumarchaeota* | 47/96(49%) | 59/96(61%) | 2E-17 | – | – |
| W4-39b, 45 174 bp, 65.3% G+C content | | | | | | | | | | | |
| 1 | 2793–925 | 622 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 2 | 2932–4959 | 675 | hypothetical protein | *Methylobacterium nodulans* | *Alphaproteobacteria* | *Proteobacteria* | 57/173(33%) | 89/173(51%) | 2E-10 | – | – |
| rRNA | 5361–5245 | | 5S rRNA gene | *Ammonifex degensii* | *Clostridia* | *Firmicutes* | 87/102(87%) | – | 6E-25 | – | – |
| rRNA | 8460–5480 | | 23S rRNA gene | *Pelobacter carbinolicus* | *Deltaproteobacteria* | *Proteobacteria* | 2496/2998(83%) | – | 0 | – | – |
| tRNA | 8741–8669 | | tRNA-Ala | *Chondromyces crocatus* | *Deltaproteobacteria* | *Proteobacteria* | 71/73(97%) | – | 8E-26 | – | – |
| tRNA | 8898–8825 | | tRNA-Ile | *Acidobacterium capsulatu* | *Acidobacteria* | *Acidobacteria* | 69/74(93%) | – | 5E-22 | – | – |
| rRNA | 10670–9125 | | 16S rRNA gene | *Geobacter bemidjiensis* | *Deltaproteobacteria* | *Proteobacteria* | 1300/1507(86%) | – | 0 | – | – |
| 3 | 12273–11041 | 410 | tyrosyl-tRNA synthetase | *Halorhodospira halophila* | *Gammaproteobacteria* | *Proteobacteria* | 209/385(54%) | 271/385(70%) | 1E-139 | COG0162 | [J] |
| 4 | 12863–12270 | 197 | 5-formyltetrahydrofolate cyclo-ligase | *Haliangium ochraceum* | *Deltaproteobacteria* | *Proteobacteria* | 75/181(41%) | 104/181(57%) | 2E-26 | COG0212 | [H] |
| 5 | 13173–12877 | 98 | hypothetical protein | *Terriglobus saanensis* | *Acidobacteria* | *Acidobacteria* | 28/64(44%) | 41/64(64%) | 9E-6 | – | – |
| 6 | 13554–13186 | 122 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 7 | 15043–13574 | 489 | glutamate synthase (NADH) small subunit | *Frankia* sp. | *Actinobacteria* | *Actinobacteria* | 291/483(60%) | 347/483(72%) | 0 | COG0493 | [ER] |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| 8 | 19565–15036 | 1509 | glutamate synthase (NADH) large subunit | *Rubrobacter xylanophilus* | *Actinobacteria* | *Actinobacteria* | 875/1488(59%) | 1096/1488(74%) | 0 | COG0069 | [E] |
| 9 | 20101–19562 | 179 | AsnC family transcriptional regulator | *Myxococcus fulvus* | *Deltaproteobacteria* | *Proteobacteria* | 66/146(45%) | 96/146(66%) | 2E-35 | COG1522 | [K] |
| 10 | 20956–20207 | 249 | glycerophosphoryl diester phosphodiesterase | *Paenibacillus* sp. | *Bacilli* | *Firmicutes* | 98/236(42%) | 138/236(58%) | 4E-44 | COG0584 | [C] |
| 11 | 22431–20941 | 496 | glycerol kinase | *Desulfotomaculum kuznetsovii* | *Clostridia* | *Firmicutes* | 311/494(63%) | 394/494(80%) | 0 | COG0554 | [C] |
| 12 | 23637–22465 | 390 | signal recognition particle-docking protein FtsY | *Syntrophobacter fumaroxidans* | *Deltaproteobacteria* | *Proteobacteria* | 174/302(58%) | 219/302(73%) | 3E-104 | COG0552 | [U] |
| 13 | 23666–24892 | 408 | major facilitator superfamily permease | *Thauera* sp. | *Betaproteobacteria* | *Proteobacteria* | 139/332(42%) | 190/332(57%) | 9E-55 | COG0477 | [GEPR] |
| 14 | 28940–25221 | 1239 | chromosome segregation protein SMC | *Anaeromyxobacter dehalogenans* | *Deltaproteobacteria* | *Proteobacteria* | 127/212(60%) | 167/212(79%) | 9E-70 | COG1196 | [D] |
| 15 | 30973–29105 | 622 | ABC transporter related protein | *Geobacter metallireducens* | *Deltaproteobacteria* | *Proteobacteria* | 282/629(45%) | 394/629(63%) | 2E-165 | COG1132 | [V] |
| 16 | 32787–31009 | 592 | ABC transporter related protein | *Geobacter uraniireducens* | *Deltaproteobacteria* | *Proteobacteria* | 240/519(46%) | 333/519(64%) | 2E-154 | COG1132 | [V] |
| 17 | 33452–32784 | 222 | glycosyl transferase family 2 | *Dethiosulfovibrio peptidovorans* | *Synergistia* | *Synergistetes* | 105/225(47%) | 134/225(60%) | 2E-51 | COG0463 | [M] |
| 18 | 33522–34637 | 371 | phosphoribosylaminoimidazole synthetase | *Geobacter metallireducens* | *Deltaproteobacteria* | *Proteobacteria* | 183/332(55%) | 231/332(70%) | 4E-122 | COG0150 | [F] |
| 19 | 35158–34610 | 182 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 20 | 35505–35861 | 118 | transposase | *Octadecabacter antarcticus* | *Alphaproteobacteria* | *Proteobacteria* | 45/82(55%) | 62/82(76%) | 9E-24 | COG3547 | [L] |
| 21 | 35926–36054 | 42 | transposase | *Methylobacterium populi* | *Alphaproteobacteria* | *Proteobacteria* | 25/41(61%) | 31/41(76%) | 1E-08 | COG3547 | [L] |
| 22 | 37348–36329 | 339 | transposase | *Rhodobacter sphaeroides* | *Alphaproteobacteria* | *Proteobacteria* | 203/334(61%) | 237/334(71%) | 6E-128 | COG3547 | [L] |
| 23 | 37706–38221 | 171 | ABC transporter, extracellular solute-binding protein | *Grimontia hollisae* | *Gammaproteobacteria* | *Proteobacteria* | 95/153(62%) | 117/153(76%) | 4E-60 | COG1840 | [P] |
| 24 | 38270–40660 | 796 | cyclic nucleotide-binding protein | *Gallionella capsiferriformans* | *Betaproteobacteria* | *Proteobacteria* | 411/796(52%) | 558/796(70%) | 0 | COG0492 | [O] |
| 25 | 40709–41551 | 280 | hypothetical protein | *Novosphingobium* sp. | *Alphaproteobacteria* | *Proteobacteria* | 132/275(48%) | 177/275(64%) | 2E-79 | – | – |
| 26 | 41555–43252 | 565 | cytochrome c family protein | *Novosphingobium* sp. | *Alphaproteobacteria* | *Proteobacteria* | 191/584(33%) | 257/584(44%) | 1E-71 | – | – |
| 27 | 43808–44773 | 321 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 28 | 44816–45172 | 118 | hypothetical protein | *Limnobacter* sp. | *Betaproteobacteria* | *Proteobacteria* | 44/84(52%) | 57/84(68%) | 2E-14 | – | – |
| W4-87b, 23 526 bp, 44.7% G+C content | | | | | | | | | | | |
| 1 | 616–392 | 74 | 50S ribosomal protein L31 | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 64/74(86%) | 70/74(95%) | 2E-42 | COG0254 | [J] |
| 2 | 2115–856 | 419 | transcription termination factor Rho | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 395/419(94%) | 410/419(98%) | 0 | COG1158 | [K] |
| 3 | 2655–2329 | 108 | thioredoxin | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 87/103(84%) | 93/103(90%) | 3E-61 | COG0526 | [OC] |
| 4 | 3338–4189 | 283 | urease accessory protein UreD | *Flavobacterium johnsoniae* | *Flavobacteria* | *Bacteroidetes* | 126/263(48%) | 180/263(68%) | 2E-86 | COG0829 | [O] |
| 5 | 5138–4230 | 302 | esterase/lipase/thioesterase family protein | *Nitrosomonas* sp. | *Betaproteobacteria* | *Proteobacteria* | 173/292(59%) | 227/292(78%) | 3E-132 | COG0596 | [R] |
| 6 | 5893–5321 | 190 | YceI family protein | *Janthinobacterium Marseille* | *Betaproteobacteria* | *Proteobacteria* | 94/183(51%) | 130/183(71%) | 1E-60 | COG2353 | [S] |
| 7 | 6797–5976 | 273 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| rRNA | 7275–7163 | | 5S rRNA gene | *Nitrosomonas* sp. | *Betaproteobacteria* | *Proteobacteria* | 107/112(96%) | – | 2E-42 | – | – |
| rRNA | 10465–7481 | | 23S rRNA gene | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 2273/2341(97%) | – | 0 | – | – |
| tRNA | 11038–10965 | | tRNA-Ile | *Nitrosomonas* sp. | *Betaproteobacteria* | *Proteobacteria* | 72/72(100%) | – | 5E-28 | – | – |
| tRNA | 10954–10881 | | tRNA-Ala | *Acidiphilium multivorum* | *Alphaproteobacteria* | *Proteobacteria* | 73/74(99%) | – | 2E-26 | – | – |
| rRNA | 12687–11164 | | 16S rRNA gene | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 1491/1524(98%) | – | 0 | – | – |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| tRNA | 12091–12018 | | tRNA-Ile | *Pelodictyon phaeoclathratiforme* | *Chlorobea* | *Chlorobi* | 73/74(99%) | – | 2E-27 | – | – |
| 8 | 14209–13007 | 400 | tyrosyl-tRNA synthetase | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 314/398(79%) | 362/398(91%) | 0 | COG0162 | [J] |
| 9 | 14857–14228 | 209 | superoxide dismutase | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 153/201(76%) | 180/201(90%) | 9E-116 | COG0605 | [P] |
| 10 | 16111–15608 | 167 | universal stress protein A UspA | *Planctomyces maris* | *Planctomycetacia* or *Planctomycea* | *Planctomycetes* or *Planctobacteria* | 61/157(39%) | 86/157(55%) | 7E-22 | COG0589 | [T] |
| 11 | 16226–16522 | 98 | hypothetical protein | *Nakamurella multipartita* | *Actinobacteria* | *Actinobacteria* | 26/59(44%) | 35/59(59%) | 2E-8 | – | – |
| 12 | 16576–16779 | 67 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 13 | 17022–17525 | 167 | thioredoxin family protein | *Teredinibacter turnerae* | *Gammaproteobacteria* | *Proteobacteria* | 62/141(44%) | 89/141(63%) | 2E-38 | COG0526 | [OC] |
| 14 | 17536–17757 | 73 | lipoprotein | *Saccharophagus degradans* | *Gammaproteobacteria* | *Proteobacteria* | 27/50(54%) | 28/50(56%) | 8E-11 | – | – |
| 15 | 17842–18951 | 369 | hypothetical protein | *Hydrogenivirga* sp. | *Aquificae* | *Aquificae* | 92/280(33%) | 139/280(50%) | 5E-31 | – | – |
| 16 | 19074–21560 | 828 | hypothetical protein | *Candidatus* Cloacamonas acidaminovorans | – | candidate division WWE1 | 46/91(51%) | 62/91(68%) | 5E-18 | COG3391 | [S] |
| 17 | 22039–22407 | 122 | response regulator receiver protein | *Sphaerobacter thermophilus* | *Sphaerobacteridae* | *Chloroflexi* | 37/118(31%) | 63/118(53%) | 2E-12 | COG4566 | [T] |
| 18 | 23302–22643 | 219 | hypothetical protein | *Croceibacter atlanticus* | *Flavobacteria* | *Bacteroidetes* | 92/212(43%) | 135/212(64%) | 7E-42 | COG1738 | [S] |
| W5-47b, 34 525 bp, 36.8% G+C content | | | | | | | | | | | |
| 1 | 1237–2 | 411 | peptidase S45 penicillin amidase | *Truepera radiovictrix* | *Deinococci* | *Deinococcus-Thermus* | 142/362(39%) | 206/362(57%) | 6E-76 | COG2366 | [R] |
| 2 | 1621–2526 | 301 | 6-phosphogluconate dehydrogenase | *Dysgonomonas mossii* | *Bacteroidia* | *Bacteroidetes* | 190/296(64%) | 243/296(82%) | 2E-146 | COG1023 | [G] |
| 3 | 2550–4070 | 506 | glucose-6-phosphate dehydrogenase | *Dysgonomonas gadei* | *Bacteroidia* | *Bacteroidetes* | 309/505(61%) | 397/505(79%) | 0 | COG0364 | [G] |
| 4 | 4067–4831 | 254 | 6-phosphogluconolactonase | *Haliscomenobacter hydrossis* | *Sphingobacteria* | *Bacteroidetes* | 115/227(51%) | 161/227(71%) | 1E-82 | COG0363 | [G] |
| 5 | 4908–7742 | 944 | bifunctional transaldolase/phosoglucose isomerase | *Candidatus* Koribacter sp. | *Acidobacteria* | *Acidobacteria* | 428/948(45%) | 583/948(62%) | 0 | COG0176 | [G] |
| 6 | 7720–8454 | 244 | HAD-superfamily hydrolase | *Thermotoga neapolitana* | *Thermotogae* | *Thermotogae* | 72/209(34%) | 114/209(55%) | 2E-22 | COG0637 | [R] |
| rRNA | 8619–8504 | | 5S rRNA gene | *Desulfurivibrio alkaliphilus* | *Deltaproteobacteria* | *Proteobacteria* | 89/103(86%) | – | 2E-25 | – | – |
| rRNA | 11736–8718 | | 23S rRNA gene | *Syntrophus aciditrophicus* | *Deltaproteobacteria* | *Proteobacteria* | 2351/3049(77%) | – | 0 | – | – |
| tRNA | 11955–11883 | | tRNA-Ala | *Candidatus* Kuenenia stuttgartiensis | *Planctomycetacia* | *Planctomycetes* | 64/73(88%) | – | 5E-16 | – | – |
| rRNA | 13749–12226 | | 16S rRNA gene | *Ignavibacterium album* | *Ignavibacteria* | *Chlorobi* | 1355/1464(93%) | – | 0 | – | – |
| 7 | 15468–14029 | 479 | polysialic acid transport protein | *Candidatus* Cloacamonas acidaminovorans | – | candidate division WWE1 | 122/393(31%) | 186/393(47%) | 5E-33 | COG1596 | [M] |
| 8 | 16512–15499 | 337 | dTDP-glucose 4,6-dehydratase | *Chloroherpeton thalassium* | *Chlorobia* | *Chlorobi* | 219/335(65%) | 261/335(78%) | 2E-164 | COG1088 | [M] |
| 9 | 17240–16515 | 241 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 10 | 17364–18251 | 295 | dTDP-4-dehydrorhamnose reductase | *Methanobacterium* sp. | *Methanobacteria* | *Euryarchaeota* | 106/296(36%) | 167/296(56%) | 2E-45 | COG1091 | [M] |
| 11 | 18309–19160 | 283 | apurinic endonuclease Apn1 | *Sphaerobacter thermophilus* | *Thermomicrobia* | *Chloroflexi* | 145/279(52%) | 194/279(70%) | 4E-104 | COG0648 | [L] |
| 12 | 20087–19173 | 304 | hypothetical protein | *Parabacteroides johnsonii* | *Bacteroidia* | *Bacteroidetes* | 107/236(45%) | 160/236(68%) | 4E-63 | – | – |
| 13 | 21712–20273 | 479 | replicative DNA helicase | *Thermodesulfatator indicus* | *Thermodesulfobacteria* | *Thermodesulfobacteria* | 232/436(53%) | 317/436(73%) | 2E-156 | COG0305 | [L] |
| 14 | 22610–21873 | 245 | phage SPO1 DNA polymerase-like protein | *Deferribacter desulfuricans* | *Deferribacteres* | *Deferribacteres* | 101/164(62%) | 125/164(76%) | 6E-70 | COG1573 | [L] |
| 15 | 23887–22685 | 400 | phosphopantothenoylcysteine decarboxylase/phosphopantothenate/cysteine ligase | *Runella slithyformis* | *Sphingobacteria* | *Bacteroidetes* | 212/399(53%) | 277/399(69%) | 2E-132 | COG0452 | [H] |
| 16 | 24466–24176 | 96 | hypothetical protein | *Leeuwenhoekiella blandensis* | *Flavobacteria* | *Bacteroidetes* | 35/86(41%) | 49/86(57%) | 4E-7 | – | – |
| 17 | 25045–24470 | 191 | guanylate kinase | *Thermotoga maritima* | *Thermotogae* | *Thermotogae* | 87/173(50%) | 123/173(71%) | 7E-58 | COG0194 | [F] |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| 18 | 25909–25049 | 286 | hypothetical protein | *Chlorobium phaeobacteroides* | *Chlorobia* | *Chlorobi* | 120/291(41%) | 173/291(59%) | 6E-59 | COG1561 | [S] |
| 19 | 28229–26178 | 683 | outer membrane protein assembly complex, YaeT protein | *Polymorphum gilvum* | *Alphaproteobacteria* | *Proteobacteria* | 156/658(24%) | 271/658(41%) | 7E-22 | COG4775 | [M] |
| 20 | 28923–28324 | 199 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 21 | 29608–28913 | 231 | ErfK/YbiS/YcfS/YnhG family protein | *Methylomonas methanica* | *Gammaproteobacteria* | *Proteobacteria* | 64/186(34%) | 97/186(52%) | 2E-19 | COG3034 | [S] |
| 22 | 29715–30899 | 394 | CapA family protein | *Pedobacter* sp. | *Sphingobacteria* | *Bacteroidetes* | 143/323(44%) | 191/323(59%) | 2E-77 | COG2843 | [M] |
| 23 | 31491–30937 | 184 | ribosome recycling factor | *Burkholderia cenocepacia* | *Betaproteobacteria* | *Proteobacteria* | 97/174(56%) | 133/174(76%) | 8E-64 | COG0233 | [J] |
| 24 | 32249–31509 | 246 | uridylate kinase | *Prosthecochloris aestuarii* | *Chlorobia* | *Chlorobi* | 156/234(67%) | 201/234(86%) | 1E-115 | COG0528 | [F] |
| 25 | 33244–32372 | 290 | elongation factor Ts | *Sphingobacterium spiritivorum* | *Sphingobacteria* | *Bacteroidetes* | 133/284(47%) | 188/284(66%) | 3E-79 | COG0264 | [J] |
| 26 | 34269–33259 | 336 | 30S ribosomal protein S2 | *Chloroherpeton thalassium* | *Chlorobia* | *Chlorobi* | 149/237(63%) | 191/237(81%) | 2E-107 | COG0052 | [J] |
| 27 | 34525–34379 | 48 | 30S ribosomal protein S9 | *Thermodesulfovibrio yellowstonii* | *Nitrospira* | *Nitrospira* or *Nitrospirae* | 40/48(83%) | 44/48(92%) | 7E-21 | COG0103 | [J] |
| W5-102b, 38 856 bp, 59.3% G+C content | | | | | | | | | | | |
| 1 | 3–863 | 286 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| rRNA | 1089–973 | | 5S rRNA gene | *Pelobacter carbinolicus* | *Deltaproteobacteria* | *Proteobacteria* | 98/107(84%) | – | 5E-26 | – | – |
| rRNA | 4177–1205 | | 23S rRNA gene | *Conexibacter woesei* | *Actinobacteria* | *Actinobacteria* | 2322/2940(79%) | – | 0 | – | – |
| rRNA | 6001–4472 | | 16S rRNA gene | *Rubrobacteridae bacterium* | *Actinobacteria* | *Actinobacteria* | 1291/1500(86%) | – | 0 | – | – |
| 2 | 6877–6443 | 144 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 3 | 7192–7956 | 254 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 4 | 9291–8056 | 411 | tyrosyl-tRNA synthetase | *Conexibacter woesei* | *Actinobacteria* | *Actinobacteria* | 214/393(54%) | 276/393(70%) | 1E-134 | COG0162 | [J] |
| 5 | 10619–9357 | 420 | hypothetical protein | *Selenomonas noxia* | *Negativicutes* | *Firmicutes* | 39/79(49%) | 52/79(66%) | 6E-09 | COG3872 | [R] |
| 6 | 11493–10810 | 227 | thymidine kinase | *Anoxybacillus flavithermus* | *Bacilli* | *Firmicutes* | 119/188(63%) | 150/188(80%) | 8E-84 | COG1435 | [F] |
| 7 | 12452–11589 | 287 | 50S ribosomal protein L31 | *Actinosynnema mirum* | *Actinobacteria* | *Actinobacteria* | 48/64(75%) | 54/64(84%) | 1E-26 | COG0254 | [J] |
| 8 | 12515–13306 | 263 | cytochrome c biogenesis protein transmembrane region | *Nocardiopsis dassonvillei* | *Actinobacteria* | *Actinobacteria* | 85/233(36%) | 118/233(51%) | 3E-24 | COG0785 | [O] |
| 9 | 13315–14034 | 239 | alpha/beta superfamily-like protein | *Streptosporangium roseum* | *Actinobacteria* | *Actinobacteria* | 37/96(39%) | 52/96(54%) | 1E-08 | COG0596 | [R] |
| 10 | 14849–14148 | 233 | MIP family channel protein | *Meiothermus silvanus* | *Deinococci* | *Deinococcus-Thermus* | 103/207(50%) | 133/207(64%) | 9E-53 | COG0580 | [G] |
| 11 | 16301–15036 | 421 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 12 | 16622–17665 | 347 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 13 | 18566–17883 | 227 | hypothetical protein | *Haladaptatus paucihalophilus* | *Halobacteria* | *Euryarchaeota* | 47/125(38%) | 65/125(52%) | 6E-11 | – | – |
| 14 | 19922–18843 | 359 | iron-sulfur cluster binding protein | *Rhodothermus marinus* | *Sphingobacteria* | *Bacteroidetes* | 113/309(37%) | 156/309(50%) | 6E-52 | COG1600 | [C] |
| 15 | 20986–19973 | 337 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 16 | 22685–21117 | 522 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 17 | 24373–22793 | 526 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| 18 | 24622–25650 | 342 | alcohol dehydrogenase | *Anaeromyxobacter dehalogenans* | *Deltaproteobacteria* | *Proteobacteria* | 156/341(46%) | 222/341(65%) | 2E-102 | COG0604 | [CR] |
| 19 | 25870–26580 | 236 | hypothetical protein | No significant similarities found | – | – | – | – | – | – | – |
| 20 | 27724–26645 | 359 | hypothetical protein | *Halalkalicoccus jeotgali* | *Halobacteria* | *Euryarchaeota* | 141/341(41%) | 201/341(59%) | 3E-79 | COG4948 | [MR] |
| 21 | 28129–27803 | 108 | hypothetical protein | *Conexibacter woesei* | *Actinobacteria* | *Actinobacteria* | 53/101(52%) | 71/101(70%) | 2E-28 | COG2151 | [R] |
| 22 | 28269–29063 | 264 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 23 | 29209–30000 | 263 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 24 | 31329–30010 | 439 | transcription termination factor Rho | *Rubrobacter xylanophilus* | *Actinobacteria* | *Actinobacteria* | 242/379(64%) | 299/379(79%) | 1E-168 | COG1158 | [K] |
| 25 | 31896–32417 | 173 | hypoxanthine phosphoribosyltransferase | *Thermomicrobium roseum* | *Thermomicrobia* | *Chloroflexi* | 57/92(62%) | 73/92(79%) | 5E-32 | | |
| 26 | 32422–33006 | 194 | hypoxanthine phosphoribosyltransferase | *Bacteroides capillosus* | *Bacteroidia* | *Bacteroidetes* | 79/160(49%) | 118/160(74%) | 5E-53 | COG0634 | [F] |
| 27 | 33093–33356 | 87 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 28 | 33455–33703 | 82 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 29 | 33961–33725 | 78 | peptidase M23 | *Conexibacter woesei* | *Actinobacteria* | *Actinobacteria* | 45/77(58%) | 54/77(70%) | 1E-23 | COG0739 | [M] |
| 30 | 34167–36929 | 920 | DNA polymerase III, alpha subunit | *Sphaerobacter thermophilus* | *Thermomicrobia* | *Chloroflexi* | 503/886(57%) | 632/886(71%) | 0 | COG0587 | [L] |
| 31 | 37053–38729 | 558 | gamma-glutamyltransferase | *Shewanella sediminis* | *Gammaproteobacteria* | *Proteobacteria* | 262/544(48%) | 354/544(65%) | 1E-161 | COG0405 | [E] |
| W5-51b, 31 991 bp, 60.2% G+C content | | | | | | | | | | | |
| 1 | 1316–3 | 437 | ATP-dependent Clp protease ATP-binding subunit ClpC | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 368/448(82%) | 409/448(91%) | 0 | COG0542 | [O] |
| 2 | 2431–1358 | 357 | ATP: guanido phosphotransferase | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 215/357(60%) | 269/357(75%) | 4E-148 | COG3869 | [E] |
| 3 | 2922–2431 | 163 | hypothetical protein | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 82/169(49%) | 117/169(69%) | 1E-42 | COG3880 | [S] |
| 4 | 3623–2922 | 233 | ABC-type antimicrobial peptide transport system, ATPase component | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 128/221(58%) | 161/221(73%) | 1E-81 | COG1136 | [V] |
| 5 | 4801–3620 | 393 | ABC-type transport system, involved in lipoprotein release, permease component | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 213/394(54%) | 277/394(70%) | 7E-133 | COG4591 | [M] |
| 6 | 6360–4867 | 497 | lysyl-tRNA synthetase (class II) | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 268/496(54%) | 341/496(69%) | 4E-172 | COG1190 | [J] |
| 7 | 7349–6357 | 330 | protein chain release factor B | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 192/324(59%) | 253/324(78%) | 3E-137 | COG1186 | [J] |
| 8 | 7755–7564 | 63 | hypothetical protein | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 34/63(54%) | 48/63(76%) | 8E-19 | – | – |
| 9 | 9156–8170 | 328 | response regulator receiver modulated diguanylate cyclase | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 137/309(44%) | 191/309(62%) | 5E-85 | COG2199 | [T] |
| 10 | 10304–9312 | 330 | hypothetical protein | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 101/303(33%) | 159/303(52%) | 1E-36 | – | – |
| 11 | 11315–11752 | 145 | hypothetical protein | *Geobacter* sp. | *Deltaproteobacteria* | *Proteobacteria* | 54/123(44%) | 76/123(62%) | 3E-27 | COG2165 | [NU] |
| 12 | 11765–13942 | 725 | metal dependent phosphohydrolase | *Candidatus* Koribacter versatilis | *Acidobacteria* | *Acidobacteria* | 105/195(54%) | 132/195(68%) | 2E-60 | COG2206 | [T] |
| 13 | 14158–14691 | 177 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| rRNA | 14982–14867 | | 5S rRNA gene | *Desulfurivibrio alkaliphilus* | *Deltaproteobacteria* | *Proteobacteria* | 99/112(88%) | – | 2E-31 | – | – |
| rRNA | 18026–15048 | | 23S rRNA gene | *Pelobacter carbinolicus* | *Deltaproteobacteria* | *Proteobacteria* | 2255/3016(75%) | – | 0 | – | – |
| rRNA | 19786–18243 | | 16S rRNA gene | *Gemmatimonas aurantiaca* | *Gemmatimonadetes* | *Gemmatimonadetes* | 1312/1553(84%) | – | 0 | – | – |
| 14 | 20376–22232 | 618 | 2-oxoglutarate ferredoxin oxidoreductase alpha subunit | *Plesiocystis pacifica* | *Deltaproteobacteria* | *Proteobacteria* | 378/609(62%) | 474/609(78%) | 0 | COG0674 | [C] |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| 15 | 22237–23262 | 341 | 2-oxoglutarate ferredoxin oxidoreductase beta subunit | *Plesiocystis pacifica* | *Deltaproteobacteria* | *Proteobacteria* | 201/334(60%) | 253/334(76%) | 1E-149 | COG1013 | [C] |
| 16 | 23416–23544 | 42 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 17 | 25699–23660 | 679 | TonB-dependent receptor | *Rhodospirillum centenum* | *Alphaproteobacteria* | *Proteobacteria* | 333/683(49%) | 455/683(67%) | 0 | COG1629 | [P] |
| 18 | 27309–26206 | 367 | hypothetical protein | *Vibrio cholerae* | *Gammaproteobacteria* | *Proteobacteria* | 161/354(45%) | 229/354(65%) | 3E-106 | COG0500 | [QR] |
| 19 | 27308–27889 | 193 | peptidase M15A | *Nitrosococcus halophilus* | *Gammaproteobacteria* | *Proteobacteria* | 86/183(47%) | 113/183(62%) | 5E-43 | – | – |
| tRNA | 28221–28133 | | tRNA-Ser | *Thauera* sp. | *Betaproteobacteria* | *Proteobacteria* | 81/89(91%) | – | 4E-25 | – | – |
| 20 | 28405–28671 | 88 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 21 | 28733–31126 | 797 | mechanosensitive ion channel | *Bacillus* sp. | *Bacilli* | *Firmicutes* | 109/274(40%) | 167/274(61%) | 9E-58 | COG3264 | [M] |
| 22 | 31787–31158 | 209 | methyltransferase type 11 | *Methylacidiphilum infernorum* | *Verrucomicrobiae* | *Verrucomicrobia* | 80/197(41%) | 126/197(64%) | 1E-49 | COG0500 | [QR] |
| 23 | 31990–31775 | 71 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| W4-21b, 38 544 bp, 49.0% G+C content | | | | | | | | | | | |
| 1 | 281–3 | 92 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 2 | 1120–278 | 280 | 3-demethylubiquinone-9 3-methyltransferase | *Oscillochloris trichoides* | *Chloroflexi* | *Chloroflexi* | 98/234(42%) | 131/234(56%) | 2E-38 | COG0500 | [QR] |
| 3 | 1710–1117 | 197 | hypothetical protein | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 75/173(43%) | 102/173(59%) | 4E-30 | – | – |
| 4 | 1815–2837 | 340 | major facilitator transporter | *Desulfatibacillum alkenivorans* | *Deltaproteobacteria* | *Proteobacteria* | 108/318(34%) | 178/318(56%) | 1E-50 | COG0477 | [GEPR] |
| 5 | 4548–2854 | 564 | hypothetical protein | *Trichodesmium erythraeum* | – | *Cyanobacteria* | 85/284(30%) | 132/284(46%) | 4E-12 | COG1413 | [C] |
| 6 | 5552–4560 | 330 | thiamine biosynthesis protein | *Geobacter metallireducens* | *Deltaproteobacteria* | *Proteobacteria* | 148/304(49%) | 201/304(66%) | 1E-76 | COG0482 | [J] |
| 7 | 7229–6279 | 316 | 2-dehydropantoate 2-reductase | *Desulfotomaculum acetoxidans* | *Clostridia* | *Firmicutes* | 128/315(41%) | 190/315(60%) | 3E-66 | COG1893 | [H] |
| 8 | 8203–7226 | 325 | alcohol dehydrogenase zinc-binding domain protein | *Candidatus Poribacteria* | – | *Poribacteria* | 170/325(52%) | 234/325(72%) | 1E-98 | COG0604 | [CR] |
| 9 | 8699–8265 | 144 | D-tyrosyl-tRNA(Tyr) deacylase | *Thermodesulfovibrio yellowstonii* | *Nitrospira* | *Nitrospira* | 81/144(56%) | 104/144(72%) | 2E-38 | COG1490 | [J] |
| 10 | 8890–8696 | 64 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 11 | 9078–9557 | 159 | hypothetical protein | *Thermovibrio ammonificans* | *Aquificae* | *Aquificae* | 56/155(36%) | 78/155(50%) | 3E-08 | – | – |
| 12 | 10707–9547 | 386 | hypothetical protein | *Rubrobacter xylanophilus* | *Actinobacteria* | *Actinobacteria* | 156/375(42%) | 215/375(57%) | 2E-74 | COG1600 | [C] |
| 13 | 11803–10736 | 355 | RNA methyltransferase, TrmA family | *Coraliomargarita akajimensis* | *Opitutae* | *Verrucomicrobia* | 115/361(32%) | 172/361(48%) | 5E-37 | COG2265 | [J] |
| 14 | 12485–11931 | 184 | peptide deformylase | *Sorangium cellulosum* | *Deltaproteobacteria* | *Proteobacteria* | 78/183(43%) | 109/183(60%) | 7E-33 | COG0242 | [J] |
| 15 | 13162–12533 | 209 | DNA repair protein RadC | *Vibrio vulnificus* | *Gammaproteobacteria* | *Proteobacteria* | 105/209(50%) | 130/209(62%) | 2E-47 | COG2003 | [L] |
| 16 | 15248–13644 | 534 | single-stranded-DNA-specific exonuclease RecJ | *Thermoanaerobacter wiegelii* | *Clostridia* | *Firmicutes* | 179/451(40%) | 292/451(65%) | 1E-103 | COG0608 | [L] |
| 17 | 17546–15342 | 734 | protein-export membrane protein SecD | *Desulfurivibrio alkaliphilus* | *Deltaproteobacteria* | *Proteobacteria* | 307/709(43%) | 439/709(62%) | 5E-152 | COG0342 | [U] |
| 18 | 17814–17587 | 75 | preprotein translocase subunit YajC | *Desulfovibrio vulgaris* | *Deltaproteobacteria* | *Proteobacteria* | 39/70(56%) | 49/70(70%) | 3E-12 | COG1862 | [U] |
| 19 | 19083–17950 | 377 | queuine tRNA-ribosyltransferase | *Peptostreptococcus stomatis* | *Clostridia* | *Firmicutes* | 199/353(56%) | 253/353(72%) | 2E-117 | COG0343 | [J] |
| 20 | 20195–19137 | 352 | S-adenosylmethionine:tRNA ribosyltransferase-isomerase | *Tepidanaerobacter* sp. | *Clostridia* | *Firmicutes* | 183/353(52%) | 248/353(70%) | 5E-97 | COG0809 | [J] |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| 21 | 20791–20234 | 185 | putative Rossmann fold nucleotide-binding protein | *Thermus thermophilus* | *Deinococci* | *Deinococcus-Thermus* | 54/150(36%) | 76/150(51%) | 1E-15 | COG1611 | [R] |
| tRNA | 21052–20979 | | tRNA-Arg | *Rhodothermus marinus* | *Sphingobacteria* | *Bacteroidetes* | 63/72(88%) | – | 2E-14 | – | – |
| 22 | 22186–21128 | 352 | SpoIID/LytB domain-containing protein | *Candidatus* Poribacteria sp. | – | *Poribacteria* | 105/269(39%) | 159/269(59%) | 2E-46 | COG2385 | [D] |
| 23 | 23149–22775 | 124 | response regulator receiver protein | *Desulfobulbus propionicus* | *Deltaproteobacteria* | *Proteobacteria* | 45/121(37%) | 69/121(57%) | 2E-10 | COG2204 | [T] |
| 24 | 24154–23270 | 294 | cobalt-zinc-cadmium resistance protein CzcD | *Rubrobacter xylanophilus* | *Actinobacteria* | *Actinobacteria* | 138/288(48%) | 180/288(63%) | 9E-63 | COG1230 | [P] |
| 25 | 24583–24245 | 112 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 26 | 25179–24865 | 104 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 27 | 26282–25245 | 345 | hypothetical protein | *Planctomyces maris* | *Planctomycetacia* | *Planctomycetes* | 160/307(52%) | 204/307(66%) | 8E-84 | COG1611 | [R] |
| 28 | 26682–26284 | 132 | amino acid-binding ACT domain protein | *Chthoniobacter flavus* | *Spartobacteria* | *Verrucomicrobia* | 57/128(45%) | 85/128(66%) | 3E-25 | COG4747 | [R] |
| rRNA | 27052–26935 | | 5S rRNA gene | *Candidatus* Methylomirabilis oxyfera | – | candidate division NC10 | 97/117(83%) | – | 7E-24 | – | – |
| rRNA | 30165–27134 | | 23S rRNA gene | *Desulfococcus oleovorans* | *Deltaproteobacteria* | *Proteobacteria* | 2315/3058(76%) | – | 0 | – | – |
| tRNA | 30546–30475 | | tRNA-Ala | *Persephonella marina* | *Aquificae* | *Aquificae* | 68/73(93%) | – | 2E-20 | – | – |
| tRNA | 30723–30650 | | tRNA-Ile | *Coraliomargarita akajimensis* | *Opitutae* | *Verrucomicrobia* | 65/74(88%) | – | 1E-16 | – | – |
| rRNA | 32424–30865 | | 16S rRNA gene | *Thioalkalivibrio thiocyanoxidans* | *Gammaproteobacteria* | *Proteobacteria* | 1063/1375(77%) | – | 0 | – | – |
| 29 | 34044–32716 | 442 | DNA polymerase | *Syntrophus aciditrophicus* | *Deltaproteobacteria* | *Proteobacteria* | 216/441(49%) | 289/441(66%) | 2E-115 | COG0389 | [L] |
| 30 | 34624–34031 | 197 | LexA repressor | *Syntrophus aciditrophicus* | *Deltaproteobacteria* | *Proteobacteria* | 95/196(48%) | 136/196(69%) | 2E-46 | COG1974 | [KT] |
| 31 | 36302–35109 | 397 | radical SAM domain protein | *Thermodesulfovibrio yellowstonii* | *Nitrospira* | *Nitrospira* | 94/339(28%) | 154/339(45%) | 1E-31 | COG0535 | [R] |
| 32 | 38077–36437 | 546 | radical SAM domain protein | *Beggiatoa* sp. | *Gammaproteobacteria* | *Proteobacteria* | 125/320(39%) | 185/320(58%) | 2E-61 | COG1032 | [C] |
| 33 | 38263–38102 | 53 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| W5-15b, 36 707 bp, 47.4% G+C content | | | | | | | | | | | |
| 1 | 3–713 | 236 | adenylate/guanylate cyclase | *Rhizobium etli* | *Alphaproteobacteria* | *Proteobacteria* | 111/233(48%) | 146/233(63%) | 1E-57 | COG2114 | [T] |
| 2 | 1968–799 | 389 | UBA/THIF-type NAD/FAD-binding protein | *Terriglobus saanensis* | *Acidobacteria* | *Acidobacteria* | 245/395(62%) | 296/395(75%) | 2E-164 | COG0476 | [H] |
| 3 | 2350–2147 | 67 | cold shock protein E, CspE | *Buchnera aphidicola* | *Gammaproteobacteria* | *Proteobacteria* | 39/65(60%) | 51/65(78%) | 3E-23 | COG1278 | [K] |
| 4 | 2495–3652 | 385 | class V aminotransferase | *Syntrophobacter fumaroxidans* | *Deltaproteobacteria* | *Proteobacteria* | 150/369(41%) | 227/369(62%) | 2E-92 | COG0520 | [E] |
| 5 | 3653–4195 | 180 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 6 | 4220–4417 | 65 | phage shock protein C, PspC | *Methanosarcina mazei* | *Methanomicrobia* | *Euryarchaeota* | 37/61(61%) | 46/61(75%) | 3E-15 | COG1983 | [KT] |
| 7 | 4434–5279 | 281 | hypothetical protein | *Magnetococcus* sp. | – | *Proteobacteria* | 147/280(53%) | 200/280(71%) | 5E-105 | COG3298 | [L] |
| 8 | 5287–6222 | 311 | rhodanese domain-containing protein | *Pseudomonas aeruginosa* | *Gammaproteobacteria* | *Proteobacteria* | 173/309(56%) | 230/309(74%) | 1E-128 | COG1054 | [R] |
| 9 | 6292–6429 | 45 | PhoH family protein | *Desulfobacterium autotrophicum* | *Deltaproteobacteria* | *Proteobacteria* | 30/37(81%) | 36/37(97%) | 9E-13 | COG1875 | [T] |
| 10 | 6471–6866 | 131 | hypothetical protein | *Bacillus* sp. | *Bacilli* | *Firmicutes* | 71/130(55%) | 84/130(65%) | 1E-43 | COG3011 | [S] |
| 11 | 6913–8514 | 533 | class V aminotransferase | *Desulfitobacterium hafniense* | *Clostridia* | *Firmicutes* | 235/525(45%) | 335/525(64%) | 4E-154 | COG0075 | [E] |
| 12 | 10172–8580 | 530 | alpha-isopropylmalate/homocitrate synthase family transferase | *Geobacter sulfurreducens* | *Deltaproteobacteria* | *Proteobacteria* | 313/518(60%) | 412/518(80%) | 0 | COG0119 | [E] |

to be continued

Continued from Table A1

| ORF | Position | Size/aa | Predicted function | Closest relatives | | | | | | COG No. | COG category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Organism | Class | Phylum/Domain | Identity | Positives | E value | | |
| 13 | 11277–10162 | 371 | aspartate kinase | *Geobacter* sp. | *Deltaproteobacteria* | *Proteobacteria* | 226/368(61%) | 290/368(79%) | 2E-151 | COG0527 | [E] |
| 14 | 11849–11355 | 164 | ATP-binding protein | *Butyrivibrio proteoclasticus* | *Clostridia* | *Firmicutes* | 60/107(56%) | 79/107(74%) | 5E-36 | COG0802 | [R] |
| 15 | 13417–11864 | 517 | carbohydrate kinase | *Thermincola* sp. | *Clostridia* | *Firmicutes* | 229/507(45%) | 332/507(65%) | 3E-144 | COG0063 | [G] |
| 16 | 13918–14886 | 322 | radical SAM domain-containing protein | *Syntrophobacter fumaroxidans* | *Deltaproteobacteria* | *Proteobacteria* | 116/291(40%) | 186/291(64%) | 1E-81 | COG0535 | [R] |
| 17 | 15221–15057 | 54 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 18 | 16799–15357 | 480 | cysteinyl-tRNA synthetase | *Geobacter bemidjiensis* | *Deltaproteobacteria* | *Proteobacteria* | 260/480(54%) | 338/480(70%) | 0 | COG0215 | [J] |
| 19 | 17719–16928 | 263 | short-chain dehydrogenase/reductase SDR | *Sphaerobacter thermophilus* | *Thermomicrobia* | *Chloroflexi* | 149/256(58%) | 188/256(73%) | 3E-102 | COG1028 | [IQR] |
| rRNA | 18209–19730 | | 16S rRNA gene | *Parachlamydia acanthamoebae* | *Chlamydiae* | *Chlamydiae* | 1003/1325(76%) | – | 0 | – | – |
| tRNA | 19819–19892 | | tRNA-Ile | *Sphingomonas suberifaciens* | *Alphaproteobacteria* | *Proteobacteria* | 70/74(95%) | – | 1E-23 | – | – |
| tRNA | 19933–20005 | | tRNA-Ala | *Candidatus* Kinetoplastibacterium desouzaii | *Betaproteobacteria* | *Proteobacteria* | 71/73(97%) | – | 8E-26 | – | – |
| rRNA | 20121–23081 | | 23S rRNA gene | *Pelobacter carbinolicus* | *Deltaproteobacteria* | *Proteobacteria* | 2369/2968(80%) | – | 0 | – | – |
| 20 | 23808–24557 | 249 | hypothetical protein | *Hippea maritima* | *Deltaproteobacteria* | *Proteobacteria* | 171/231(74%) | 210/231(91%) | 8E-144 | COG0655 | [R] |
| 21 | 24563–24820 | 85 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 22 | 24965–25609 | 214 | glutathione S-transferase | *Zea mays* | – | Eukarya | 68/168(40%) | 96/168(57%) | 3E-32 | COG0625 | [O] |
| 23 | 25633–26100 | 155 | hypothetical protein | *Collimonas fungivorans* | *Betaproteobacteria* | *Proteobacteria* | 71/134(53%) | 86/134(64%) | 2E-40 | COG4312 | [S] |
| 24 | 26213–26407 | 64 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 25 | 26542–28659 | 705 | RNA polymerase sigma factor RpoD | *Rhodospirillum rubrum* | *Alphaproteobacteria* | *Proteobacteria* | 490/716(68%) | 558/716(78%) | 0 | COG0568 | [K] |
| tRNA | 28780–28852 | | tRNA-Met | *Azospirillum* sp. | *Alphaproteobacteria* | *Proteobacteria* | 66/73(90%) | – | 9E-19 | – | – |
| 26 | 30595–29450 | 381 | RecB family nuclease | *Nitrosomonas* sp. | *Betaproteobacteria* | *Proteobacteria* | 100/369(27%) | 150/369(41%) | 1E-16 | – | – |
| 27 | 30804–30592 | 70 | hypothetical protein | *Legionella pneumophila* | *Gammaproteobacteria* | *Proteobacteria* | 47/63(75%) | 53/63(84%) | 1E-25 | – | – |
| 28 | 31520–31236 | 94 | DNA-binding protein HU | *Parvularcula bermudensis* | *Alphaproteobacteria* | *Proteobacteria* | 55/94(59%) | 71/94(76%) | 7E-34 | COG0776 | [L] |
| 29 | 34537–31808 | 909 | lytic murein transglycosylase family protein | *Rhodospirillum centenum* | *Alphaproteobacteria* | *Proteobacteria* | 274/623(44%) | 364/623(58%) | 4E-145 | COG0741 | [M] |
| 30 | 34536–34862 | 108 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 31 | 34922–35098 | 58 | hypothetical protein | no significant similarities found | – | – | – | – | – | – | – |
| 32 | 35200–36456 | 418 | poly-gamma-glutamate synthesis protein | *Nitrosospira multiformis* | *Betaproteobacteria* | *Proteobacteria* | 255/399(64%) | 301/399(75%) | 0 | COG2843 | [M] |
| 33 | 36476–36706 | 76 | nucleoside-triphosphatase | *Desulfonatronospira thiodismutans* | *Deltaproteobacteria* | *Proteobacteria* | 33/62(55%) | 43/62(73%) | 3E-14 | COG1618 | [F] |