The Journal of Microbiology, August 2009, p. 402-410
DOI 10.1007/s12275-009-0099-5

Vol. 47, No. 4

Identification and Characterization of Metagenomic Fragments from Tidal Flat Sediment

Byung Kwon Kim[1], Yoon-Dong Park[2], Hyun-Myung Oh[3], and Jongsik Chun[1]*

[1]School of Biological Sciences and Institute of Microbiology, Seoul National University, Seoul 151-742, Republic of Korea
[2]Department of Medicine, University of Illinois at Chicago, College of Medicine, Chicago, IL, USA
[3]Division of Biology and Ocean Sciences, Inha University, Incheon 402-751, Republic of Korea

(Received March 30, 2009 / Accepted April 21, 2009)

Phylogenetic surveys based on cultivation-independent methods have revealed that tidal flat sediments are environments with extensive microbial diversity. Since most of prokaryotes in nature cannot be easily cultivated under general laboratory conditions, our knowledge on prokaryotic dwellers in tidal flat sediment is mainly based on the analysis of metagenomes. Microbial community analysis based on the 16S rRNA gene and other phylogenetic markers has been widely used to provide important information on the role of microorganisms, but it is basically an indirect means, compared with direct sequencing of metagenomic DNAs. In this study, we applied a sequence-based metagenomic approach to characterize uncultivated prokaryotes from tidal flat sediment. Two large-insert genomic libraries based on fosmid were constructed from tidal flat metagenomic DNA. A survey based on end-sequencing of selected fosmid clones resulted in the identification of clones containing 274 bacterial and 16 archaeal homologs in which majority were of proteobacterial origins. Two fosmid clones containing large metagenomic DNAs were completely sequenced using the shotgun method. Both DNA inserts contained more than 20 genes encoding putative proteins which implied their ecological roles in tidal flat sediment. Phylogenetic analyses of evolutionary conserved proteins indicate that these clones are not closely related to known prokaryotes whose genome sequence is known, and genes in tidal flat may be subjected to extensive lateral gene transfer, notably between domains Bacteria and Archaea. This is the first report demonstrating that direct sequencing of metagenomic gene library is useful in underpinning the genetic makeup and functional roles of prokaryotes in tidal flat sediments.

***Keywords*:** tidal flat, metagenome, fosmid, shot-gun sequencing, phylogeny

Cultivation-independent surveys of prokaryotic 16S rRNA gene have greatly expanded our knowledge on microbial diversity on earth (Pace, 1997; Hugenholtz *et al.*, 1998). However, the information contained in this single gene is insufficient to provide the organism's physiological, biochemical, or ecological significance (Quaiser *et al.*, 2002; Treusch *et al.*, 2004). In contrast, a sequence-based survey of metagenomic libraries constructed from diverse environments has proved to be successful in addressing the functionality of prokaryotes in nature (Beja *et al.*, 2000; Rondon *et al.*, 2000; Quaiser *et al.*, 2002; Ochsenreiter *et al.*, 2003; Michael *et al.*, 2004; Treusch *et al.*, 2004). Whole genome shot-gun sequencing of metagenomic DNA has been used to dissect the microbial community structure and genetic makeup of various environments, ranging from simple mine acid drainage (Tyson *et al.*, 2004) to complex Sargasso Sea water (Venter *et al.*, 2004). Even though this approach is effective in gathering the necessary information on microbial community overall, it requires enormous cost, which is not yet available to most laboratories (Treusch *et al.*, 2004). Instead, several studies have focused on the analysis of genomic fragments of particular interest, which can be readily identified by the presence of phylogenetic markers, such as 16S rRNA gene (Quaiser *et al.*, 2002; Lopez-Garcia *et al.*, 2004; Treusch *et al.*, 2004). Although 16S rRNA genes have been regarded as the best indication of taxonomic position of DNA fragment recovered from the natural environment, the probability of obtaining such DNA molecules in practice may be very low. Ochsenreiter *et al.* (2003) estimated the presence of one archaeal 16S rRNA gene per 1.2 Giga bp DNA in soil metagenomic library. High-throughput end sequencing has been used to characterize large-insert metagenomic libraries to overcome this limitation (Breitbart *et al.*, 2002; Sebat *et al.*, 2003; Pope and Patel, 2008).

The seashore of the Yellow Sea of the Korean Peninsula consists largely of tidal flats, which have been best known as forming a dynamic area in terms of sediment erosion and deposition, and they can be distinguished from salt marsh and wetland by the degree to which these characteristics are observed (Carling, 1982). In previous studies, using both culture-dependent and independent methods, we have shown that extremely diverse bacterial and archaeal phylotypes are present in tidal flat sediment in Korea (Yi *et al.*, 2003, 2004; Kim *et al.*, 2004, 2005, 2008; Yi and Chun, 2004). Even though these studies revealed majority of prokaryotic community structure, the information is based on a single gene, i.e. the 16S rRNA gene, and the amplification of metagenomic DNA *via* PCR. In this study, we have ex-

* To whom correspondence should be addressed.
(Tel) 82-2-880-8153; (Fax) 82-2-874-8153
(E-mail) jchun@snu.ac.kr

amined the prokaryotic genetic diversity of tidal flat sediment using direct sequencing of large-insert metagenomic libraries. In addition, we present functional annotation of two large metagenomic fragments representing potentially novel phyletic lineages.

Materials and Methods

Sampling and extraction of metagenomic DNA from sediments

Approximately 20 g of soil samples were collected from the upper layer (0~2 cm) of a tidal flat sediment at Yeocha (37 36'22.3 N, 126 22'59.4 E) in Ganghwa in February 2002 and at Mohang (35 34'52.3 N, 126 30'51.3 E) in Byunsan, Republic of Korea in May 2003. Samples were stored at -80°C until DNA was extracted.

Nucleic acids (8 μg) were extracted from 20 g of soil by the direct lysis protocol for bacterial community DNA recovery (Hurst, 1997; Kim *et al.*, 2004). The ends of purified DNA were repaired to generate blunt-ended DNA according to Quaiser *et al.* (2002). To select DNA fragments of a suitable size, pulsed field gel electrophoresis was performed at 14°C, 6 V/cm for 8 h with 0.1~8.5 sec pulses using a CHEF-DRIII or CHEF-MAPPER electrophoresis systems (Bio-Rad, USA). High molecular DNA (>30 kb) was extracted from agarose gels using gelase (Epicentre).

Library construction, screening, and end sequencing

Purified metagenomic DNA (0.2 μg or 0.4 μg) was ligated into fosmid vector pEpiFOS-5 (Epicentre, USA) (Quaiser *et al.*, 2002). The ligated vectors were packaged into lambda phages and used to transfect *Escherichia coli* EPI100 (Epicentre). The resultant infected cells were spread onto Luria-Bertani Broth (LB) medium containing 12.5 μg/ml chloramphenicol. A total of 3,800 fosmid clones were transferred to 384-well microtiter plates containing 50 μl of LB medium supplemented with chloramphenicol (12.5 μg/ml) and 10% glycerol (v/v). The microtiter plates were incubated overnight at 37°C.

Fosmid DNAs containing large metagenomic fragment were isolated using the alkaline lysis miniprep method (Sambrook and Russell, 2001) or the Nucleobond BAC 100 kit (Macherey-Nagel/MN, Germany), and their ends were sequenced using T7 promoter primer; 5'-TAA TAC GAC TCA CTA TAG GG-3'. The functional and taxonomic annotation of end-sequences of fosmid clones were achieved by BLASTX search against the National Center for Biotechnology Information (NCBI) nr database.

Shot-gun sequencing of fosmid clones

To sequence fosmid clones completely, small insert DNA libraries were prepared from purified fosmid DNA using the Genome Priming System (GPS^{TM}-1, NEB), in which transposons containing the Transprimers were randomly inserted into fosmids. Subsequently, flanking regions of the inserted transposon were sequenced with the Transprimers: N; 5'-ACT TTA TTG TCA TAG TTT AGA TCT ATT TTG-3' and S; 5'-ATA ATC CTT AAA AAC TCC ATT TCC ACC CCT-3'. Assembly of the shot-gun sequencing reads was carried out using the PHRED/PHRAP/CONSED package (Ewing *et al.*, 1998; Gordon *et al.*, 2001). The gaps were closed by primer walking with sequence-derived oligonucleotides.

Bioinformatics

The prediction and functional annotation of genes in completely sequenced fosmid inserts were carried out using Artemis program (Berriman and Rutherford, 2003). The presence of tRNA genes was checked by the tRNAscan-SE server (http://lowelab.ucsc.edu/tRNAscan-SE/). Functional annotation of protein coding regions was achieved using BLASTP and PSI-BLAST programs. Transmembrane domains in identified proteins were predicted using the TMHMM server (Krogh *et al.*, 2001 http://www.cbs.dtu.dk/services/TMHMM-2.0/). Predicted proteins in fosmids were classified according to the category of the clusters of orthologous groups (COG, Tatusov *et al.*, 1997, 2000; Lopez-Garcia *et al.*, 2004; Treusch *et al.*, 2004).

Nucleotide sequences from open reading frames (ORF) identified in completely sequenced fosmid clones were aligned with homologous genes retrieved from the NCBI nr database using CLUSTAL W program (Higgins and Sharp, 1988). Manual editing was achieved using the PHYDIT program (available at http://plaza.snu.ac.kr/~jchun/phydit/), in which ambiguously aligned regions were excluded from phylogenetic analysis. A phylogenetic tree was reconstructed using the neighbor-joining method (Saitou and Nei, 1987) with the Kimura's distance model (Kimura, 1983).

Table 1. Annotation of end sequences from two metagenomic libraries

	Library	
	Byunsan	Gangwha
Clones in library	3800	2304
End-sequenced clones	314	64
Sequences with significant match to NCBI nr database[a]	255 (81.2%)	39 (60.9%)
Highest similarity to bacterial gene	236 (75.2%)	38 (59.4%)
related to *Proteobacteria*	139 (44.3%)	20 (31.3%)
related to *Bacteroidetes*	10 (3.2%)	5 (7.8%)
related to *Firmicutes*	25 (8.0%)	5 (7.8%)
Highest similarity to archaeal gene	15 (4.8%)	1 (1.6%)
Highest similarity to eukaryotic gene	4 (1.3%)	0

[a] BLASTX program with e^{-5} as cutoff was used

Nucleotide sequence accession numbers

The DNA sequences of fosmid clones YC01A01 and YC01G10 are available from GenBank database under the accession numbers DQ121375 and DQ121376, respectively.

Results and Discussion

Tidal flat sediments in the Yellow Sea of Korea are known to contain a substantial amount of humic compounds (Kim *et al.*, 2004), which is problematic in DNA isolation and subsequent cloning experiments. We employed the CsCl-EtBr density equilibrium gradient method without the polyvinyl polypyrrolidone purification step (Hurst, 1997; Kim *et al.*, 2004) and recovered *ca.* 3∼5 μg metagenomic DNA per 10 g soil sample. The mean 260/280 absorbance ratio of extracted DNA was 1.80, and DNA fragments from sediments ranged in size between 25 and 50 kb. After blunt-end cloning into the fosmid vector, 2304 *E. coli* colonies were obtained from 400 ng of soil DNA from Gangwha, and 3800 colonies from 200 ng of soil DNA from Byunsan. Restriction analysis of 26 randomly selected clones from each library indicated that the sizes of the inserted fragments ranged from 30 to 45 kb. Therefore, approximately 85 and Mb of tidal flat sediment genomic DNA were estimated to be captured in our large-insert metagenomic libraries, namely Gangwha and Byunsan, respectively.

To survey the diversity of fosmid libraries, end sequences of 378 randomly selected clones were determined and compared with sequences in the NCBI nr database. A total of 294 clones (78%) showed significant similarities to known proteins in the database (e value of $<e^{-5}$ in BLASTX search). Among them, 274 (93%) sequences showed best match to known bacterial proteins, and 16 (4.8%) to archaeal proteins (Table 1). The snapshot of microbial diversity based on the end sequences of fosmid libraries were found to be different from microbial community structure based on 16S rRNA gene-based survey (Kim *et al.*, 2004, 2005, 2008). For example, Firmicute was found to be abundant in tidal flat sediment according to end-sequence survey (Table 1), but not to previous 16S rRNA gene-based studies. The discrepancy between two surveys has also been reported in a previous study, which may result from potential biases being generated during amplification of the 16S rRNA gene or ligation of environmental DNA fragments into the fosmid or the BAC vector (Pope and Patel, 2008).

For the characterization of prokaryotes from tidal flat sediment, we attempted to determine the complete sequences of large genomic fragments. Through similarity-based annotation of end sequences, we selected two clones showing significant similarities to archaeal (hypothetical protein related to RNA polymerization of *Archaeoglobus fulgidus*) and bacterial (thioredoxin reductase of *Thermoanaerobacter tengcongensis*) proteins. Genes coding for ribosomal and transfer RNAs were not detected from both fragments.

A bacterial metagenomic fragment, designated YC01A01, was 39.4 kb long with an average G+C content of 50.5% it encoded 24 predicted ORFs (>90 amino acids Fig. 1A and Table 2). Nineteen ORFs showed significant similarities to genes of known function in the NCBI nr database, and three were homologous to hypothetical proteins in the NCBI microbial genome database. Eight of twenty-four ORFs found in YC01A01 were related to cellular processing and molecular transport. Two ORFs had no similarity to known protein sequences, of which one (ORF-18) contained putative membrane protein. ORF-14 was found to be a fused protein of dimethyladenosine transferase and Mg^{2+} transporter protein, which is the case never found in prokaryotic genomes sequenced so far, although two genes were found adjacently in the genome of *Bacteroides vulgatus*, *Bacteroides fragilis*, *Bacteroides thetaiotaomicron*, *Cytophaga hutchinsonii*, *Porphyromonas gingivalis*, *Parabacteroides distasonis*, *Gramella forsetii*, *Flavobacterium johnsoniae*, and *Flavobacterium psychrophilum* belonging to the phylum *Bacteroidetes*. Phylogenetic analysis of ORF-14 was performed using the region coding for dimethyladenosine transferase (Fig. 2). The resultant tree topologies strongly indicate that YC01A01 is derived from a bacterium representing a deep branch in the phylum *Bacteroidetes*. This conclusion is also supported by a phylo-

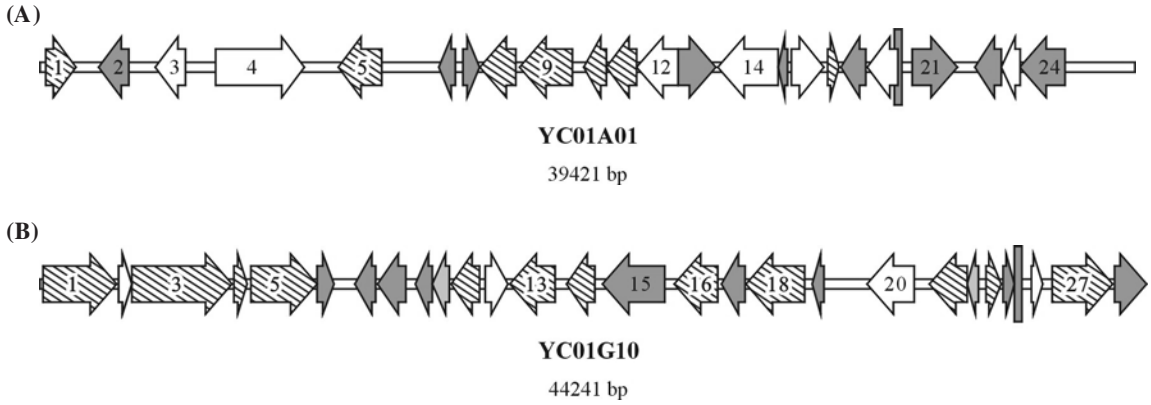

Fig. 1. Schematic representation of fosmid YC01A10 clone (A) and YC01G10 clone (B). Different shadings indicate the functional categories of the putative protein-coding genes. Diagonal strips indicate ORFs related to cellular processes, blank arrows indicate ORFs related to metabolism, and light grays indicate ORFs related to information storage. Poorly characterized proteins and hypothetical genes without homologues are shown in dark grays.

Table 2. Predicted protein-encoding genes in the metagenomic fragment YC01A01

ORF No.	Homologous to	Organism/Phylum or Domain	BLAST e-value	Position	No. of TMHs[a]	COG category	COG No.
1	Thioredoxin reductase	Thermoanaerobacter tengcongensis/ Firmicutes	1.00E-59	241..1161	1	O	COG0492
2	Hypothetical protein (predicted secreted protein)	*Shewanella oneidensi*/Gammaproteo bacteria	3.00E-90	3123..2191	0	NA[b]	
3	Trehalose-6-phosphate synthase	*Ginkgo biloba*/Eukaryote	1.00E-118	5240..4104	0	G	COG0380
4	Hydrolase, haloacid dehalogenase-like family	*Chlorobium tepidum*/Bacteriodetes	0.00E+00	6320..9565	0	RG	COG0637 COG1554
5	Signal transduction histidine kinase	*Geobacter sulfurreducens*/ Deltaproteobacteria	2.00E-43	12347..10644	7	T	COG0642
6	Hypothetical protein	*Pyrococcus abyssi*/Archaea	1.00E-111	14972..14247	0	S	COG2403
7	Similar to putative glycoprotein YajB	*Photorhabdus luminescens*/ Gammaproteobacteria	7.00E-71	15156..15773	0	S	COG3124
8	Bacterial cell division membrane protein	*Bacteroides fragilis*/Bacteriodetes	5.00E-71	17163..15802	11	D	COG0772
9	Penicillin-binding protein 2	*Geobacter sulfurreducens*/ Deltaproteobacteria	1.00E-140	19088..17226	1	M	COG0768
10	Rod shape-determining protein	*Bacteroides fragilis*/Bacteriodetes	1.00E-62	20431..19592	1	M	COG1792
11	Rod shape-determining protein MreB	*Bacillus cereus*/Firmicutes	4.00E-93	21451..20432	0	D	COG1077
12	Phosphoribosylaminoimidazole carboxy formyl formyltransferase	*Yersinia pseudotuberculosis*/ Gammaproteobacteria	0.00E+00	22992..21469	0	F	COG0138
13	ABC transporter efflux protein	*Bdellovibrio bacteriovorus*/ Deltaproteobacteria	3.00E-76	23090..24346	4	R	COG0577
14	Dimethyladenosine transferase + Mg transporter	*Pseudomonas aeruginosa*/ Gammaproteobacteria *Oceanobacillus iheyensis*/Firmicutes	1E-103/1e-74	26553..24358	5	JP	COG0030 COG2239
15	Hypothetical protein			26934..26563	0	NA	
16	Aspartate aminotransferase	*Synechococcus elongatus*/Cyanobacteria	2.00E-81	27040..28221	0	E	COG1168
17	Small heat shock protein	*Arabidopsis thaliana*/Eukaryote	2.00E-33	28338..28766	0	O	COG0071
18	Hypothetical protein			29731..28808	10	NA	
19	Glucose-6-phosphate isomerase	*Sus scrofa*/Eukaryote	1.00E-124	30802..29789	0	G	COG0166
20	Putative glucose-6-phosphate isomerase subunit	*Bacillus clausii*/Firmicutes	3.00E-30	31135..30857	0	NA	
21	Predicted aminopeptidases	*Xanthomonas axonopodis*/ Gammaproteobacteria	1.00E-101	31346..33004	0	R	COG2234
22	Uncharacterized conserved protein	*Vibrio cholerae*/Gammaproteobacteria	1.00E-118	34521..33691	0	S	COG2326
23	Uracil phosphoribosyltransferase	*Streptococcus salivarius*/Firmicutes	3.00E-84	35209..34589	0	F	COG0035
24	Putative outer membrane related protein	*Porphyromonas gingivalis*/Bacteriodetes	7.00E-55	36844..35222	0	NA	

[a] Based on TMHMM search
[b] Not assigned to COG group

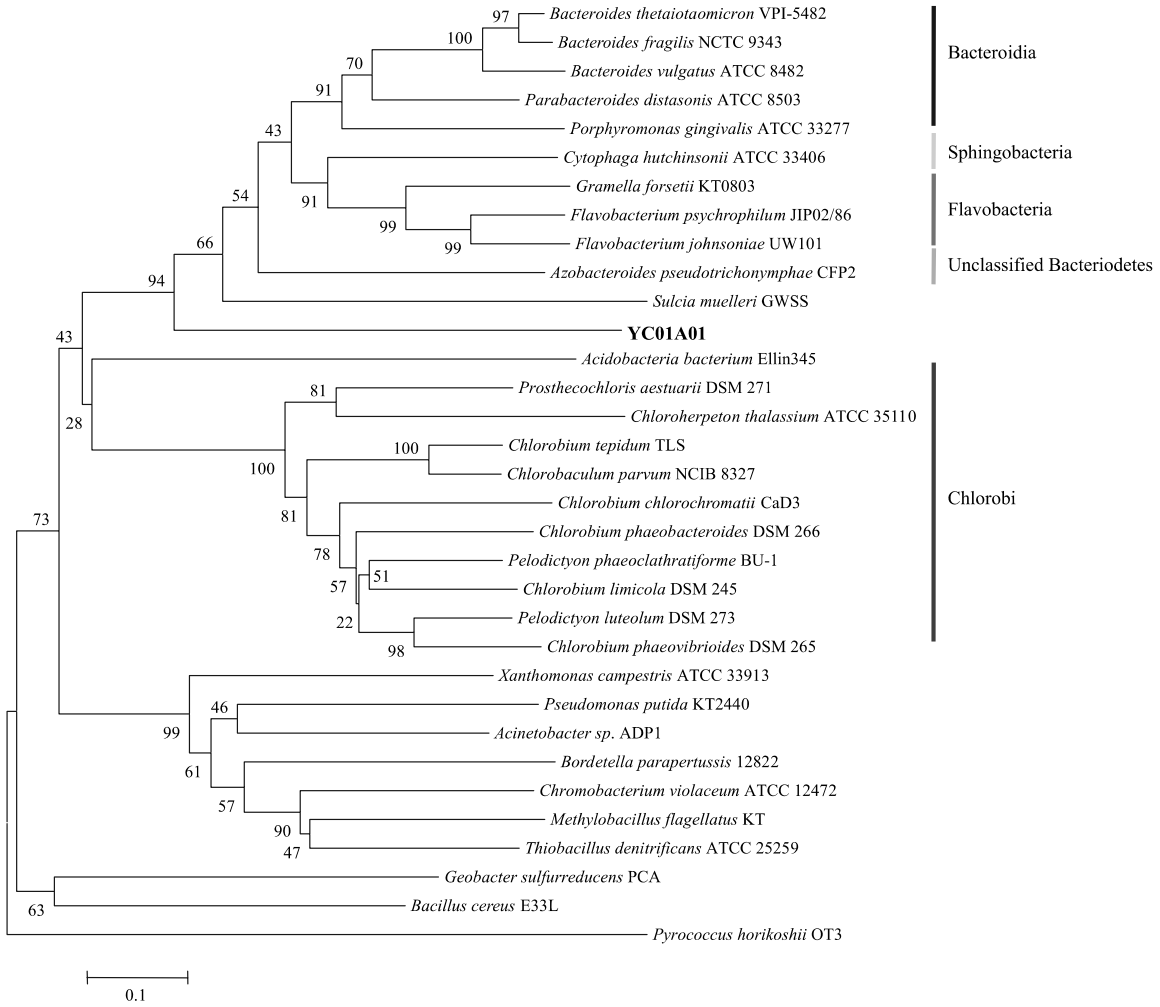

Fig. 2. Neighbor-joining tree of deduced amino acid sequence representing dimethyladenosine transferase part in ORF-14 of clone YC01A01. Numbers at the nodes indicate the levels of bootstrap support based on a neighbor-joining analysis of 1,000 resampled dataset: only values more than 50% are given. Scale bar indicates the number of substitutions per amino acid.

genetic tree based on ORF-17, coding for small heat shock protein (Fig. 3). In contrast, similarity-based functional annotation (Table 2) and phylogenetic analysis (Fig. 4) suggest that putative GTPase-coding ORF-6 has an archaeal origin, which implies the occurrence of horizontal gene transfer between domains Bacteria and Archaea. The existence of genes related to controlling cellular shape and trehalose biosynthesis suggests that YC01A01 may be originated from a rod-shaped organism that could synthesize the trehalose in a cell (Gouffi and Blanco, 2000).

The second metagenomic fragment, designated YC01G10, was 44.1 kb long with G+C content of 47.3% it encoded 28 predicted protein-coding genes (Fig. 1B and Table 3). Even though this DNA fragment was initially thought to belong to an archaeal strain, it possessed only two ORFs related to archaeal proteins, but most, i.e., 21 ORFs, to bacterial proteins, suggesting its bacterial origin rather than archaeal. Twenty ORFs showed a significant similarity to genes of known function, whereas five were homologous to uncharacterized proteins and three shared no similarity to sequences in public database. In contrast to YC01A01, most of the genes in YC01G10 were related to signal transduction system. The presence of multiple transporter systems and genes related to beta-lactamase in YC01G10 indicates that it may be from an organism resistant toampicilin. The presence of ORF-13, coding for type IV pilus biogenesis protein, suggests that it could also make pili. Both are related to defense to oxidative stress using the thioredoxin system or efflux system of ROS generator (Ma *et al.*, 1996; Paget *et al.*, 1998; Zheng *et al.*, 1998; Valko *et al.*, 2006). ORF-1 and ORF-3 were found to be paralogs as they shared 77% sequence similarity. Unlike the case for YC01A01, the lack of evolutionary conserved genes in YC01G10 hampered the identification of its taxonomic position using phylogenetic analysis of individual ORFs. Phylogenetic treeing of ORF-28, which was used for initial screening of candidate archaeal genomic fragment, indicates that ORF-28 may be recently transferred from a deep branching bacterium, but not an archaeon. Significantly higher G+C content of ORF-28 also supports this hypothesis.

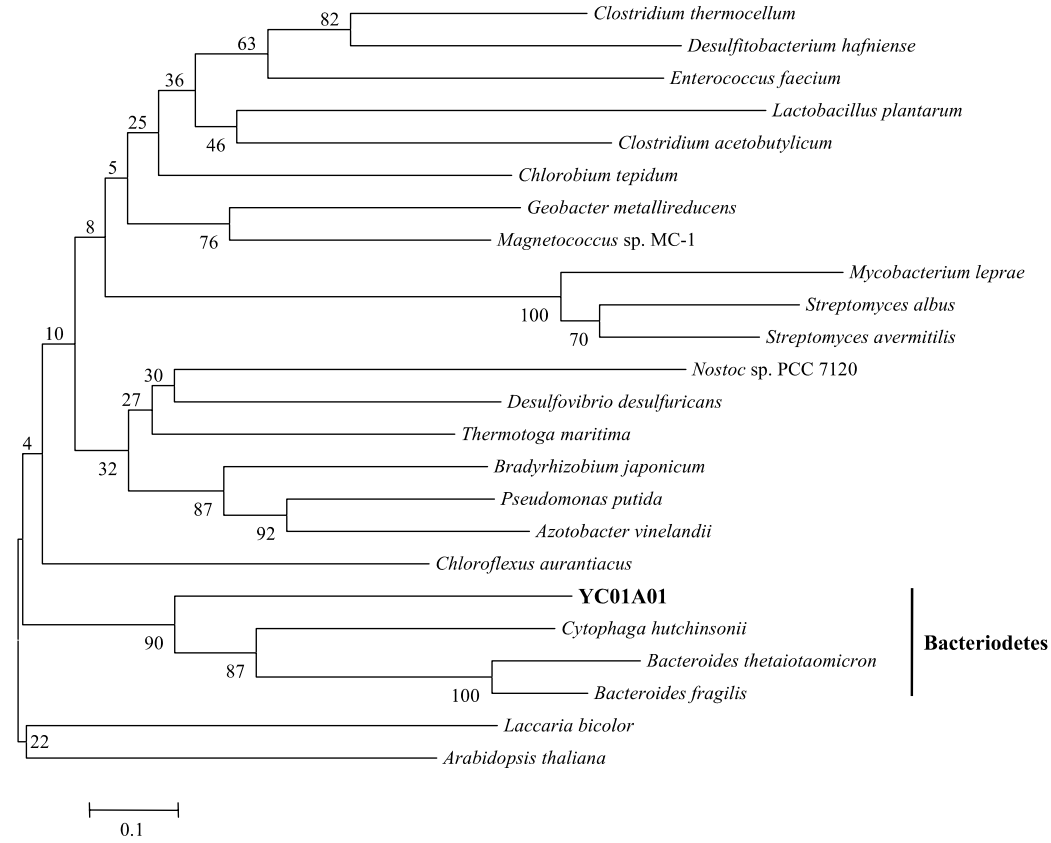

Fig. 3. Neighbor-joining tree of putative small heat shock protein (hsp20, ORF-17) implying that YC01A01 may belong to phylum *Bacteroidetes*. *Laccaria bicolor* and *Arabidopsis thaliana* were included as outgroups. Scale bar indicatesthe number of substitutions per amino acid.

It is clear from several studies that the Korean tidal flat harbors diverse organisms belonging to both domains Bacteria and Archaea (Yi *et al.*, 2003, 2004; Kim *et al.*, 2004, 2005, 2008; Yi and Chun, 2004; Lee *et al.*, 2006). However, all of previous attempts were based on PCR-mediated amplification and sequencing of phylogenetic marker genes, but not direct sequencing of metagenomic library. Even though the amount of sequenced DNA is rather limited, this study firstly surveyed microbial genetic diversity of Korean tidal flat sediments. Here, we show that our approach, i.e., end-sequencing of cloned fosmid libraries, not only is useful for underpinning microbial community structure, complementing to 16S rRNA gene-based survey, but also can provide a firm basis for understanding the genetic makeup in tidal

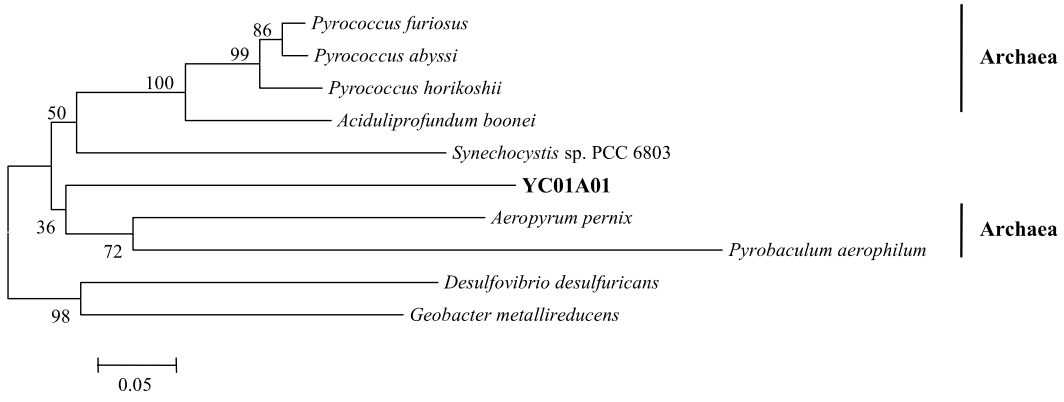

Fig. 4. Phylogenetic tree of all detected homologues to ORF 6 of YC01A01 showing that horizontal gene transfer event occurred between two prokaryotic domains. Scale bar indicates the number of substitutions.

Table 3. Predicted protein-encoding genes in the metagenomic fragment YC01G10

ORF No.	Homologous to	Organism/Phylum or Domain	BLAST e-value	Position	No. of TMHs[a]	COG category	COG No.
1	Two-component hybrid sensor and regulator	*Anabaena* sp./Cyanobacteria	0	89..2977	0	T	COG0642
2	Two-component response regulator	*Bacillus halodurans*/Firmicutes	6.00E-25	3082..3564	0	T	COG0784
3	Two-component hybrid sensor and regulator	*Anabaena* sp./Cyanobacteria	0	3610..7581	2	T	COG0642
4	Two-component response regulator	*Bacillus clausii*/Firmicutes	2.00E-24	7709..8158	0	T	COG0784
5	Hypothetical protein	*Azoarcus* sp./Betaproteobacteria	1.00E-134	8303..10927	0	T	COG2200
6	NAD(P)H dehydrogenase (quinone) activity	*Mus musculus*/Eukaryote	3.00E-46	10958..11671	0	R	COG2249
7	Putative endo alpha-1,4 polygalactosaminidase	*Streptomyces coelicolor*/Actinobacteria	2.00E-96	13364..12501	0	NA[b]	
8	Hypothetical protein	*Methanosarcina acetivorans*/Archaea	1.00E-109	14528..13386	0	S	COG2006
9	Beta-lactamase-like protein	*Mycobacterium smegmatis*/ Actinobacteria	4.00E-31	15558..14875	0	R	COG0491
10	TetR/AcrR family transcriptional regulator	*Symbiobacterium thermophilum*/ Firmicutes	6.00E-30	16178..15570	0	K	COG1309
11	Aryldialkylphosphatase related protein	*Oceanobacillus iheyensis*/Firmicutes	3.00E-57	17481..16327	0	Q	COG1228
12	Dipeptide transport system permease protein dppC	*Bacillus pseudofirmus*/Firmicutes	9.00E-73	17719..18591	6	EP	COG1173
13	Type IV pilus biogenesis protein PilB	*Geobacter sulfurreducens*/ Deltaproteobacteria	1.00E-162	20548..18689	0	N	COG2804
14	Sensor histidine kinase/response regulator	*Geobacter sulfurreducens*/ Deltaproteobacteria	2.00E-38	22105..20936	0	T	COG0642
15	Leucine-rich repeat (LRR) protein	*Dictyostelium discoideum*/Eukaryote	1.00E-137	24951..22408	0	NA	
16	Sensor histidine kinase/response regulator	*Geobacter sulfurreducens*/ Deltaproteobacteria	8.00E-79	27077..25305	3	T	COG0642
17	Hypothetical protein	*Pseudomonas aeruginosa*/ Gammaproteobacteria	2.00E-86	28145..27180	0	S	COG2984
18	TonB-dependent receptor	*Methylococcus capsulatus*/ Gammaproteobacteria	2.00E-77	30527..28179	1	P	COG1629
19	Hypothetical protein			31236..30820	0	NA	
20	Membrane protein, putative	*Treponema denticola*/Spirochaetes	2.00E-05	34944..32977	7	GEPR	COG0477
21	Beta-lactamase regulatory protein 1	*Bacillus cereus*/Firmicutes	4.00E-66	36987..35473	6	M	COG0739
22	Transcriptional regulator	*Caulobacter crescentus*/ Alphaproteobacteria	3.00E-34	37384..36989	0	K	COG3682
23	Sensory box protein/GGDEF family protein	*Pseudomonas putida*/ Gammaproteobacteria	2.00E-42	37733..38356	0	T	COG2199
24	Hypothetical protein			38434..38964	0	NA	
25	Hypothetical protein			39424..39083	0	NA	
26	Cytosine/adenosine deaminases	*Photobacterium profundum*/ Gammaproteobacteria	8.00E-47	39550..40041	0	FJ	COG0590
27	Two-component sensor histidine kinase	*Anabaena* sp./Cyanobacteria	3.00E-46	40481..42883	1	T	COG0642
28	RNA terminal phosphate cyclase operon orfB homolog, UPF0027 family	*Pyrococcus kodakaraensis*/Archaea	1.00E-169	42945..44240	0	S	COG1690

[a] Based on TMHMM search
[b] Not assigned to COG group

flat sediment. Even though it is difficult to detect lateral gene transfer events by analyzing genes in genomic fragments, the complete sequencing and functional annotation of two large metagenomic fragments suggest that such an event may take place frequently in tidal flat sediment, including transfer between domain levels. In conclusion, the physiology and genetic makeup of two prokaryotes could be partially understood through a sequence-based metagenome approach. Further sequencing efforts, such as large-scale whole shot-gun sequencing, should be applied to unravel the genetic pool of this very complex environment.

Acknowledgements

This work was supported by the National Research Laboratory Program (NRL, R0A-2005-000-10110-0) from the Ministry of Science and Technology (MOST) of the Republic of Korea.

References

Beja, O., M.T. Suzuki, E.V. Koonin, L. Aravind, A. Hadd, L.P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S.B. Jovanovich, R.A. Feldman, and E.F. DeLong. 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* 2, 516-529.

Berriman, M. and K. Rutherford. 2003. Viewing and annotating sequence data with Artemis. *Brief Bioinform.* 4, 124-132.

Breitbart, M., P. Salamon, B. Andresen, J.M. Mahaffy, A.M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. USA* 99, 14250-14255.

Carling, P.A. 1982. Temporal and spatial variation in intertidal sedimentation rates. *Sedimentol.* 29, 17-23.

Ewing, B., L. Hillier, M.C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175-185.

Gordon, D., C. Desmarais, and P. Green. 2001. Automated finishing with autofinish. *Genome Res.* 11, 614-625.

Gouffi, K. and C. Blanco. 2000. Is the accumulation of osmoprotectant the unique mechanism involved in bacterial osmoprotection? *Int. J. Food Microbiol.* 55, 171-174.

Higgins, D.G. and P.M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73, 237-244.

Hugenholtz, P., B.M. Goebel, and N.R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180, 4765-4774.

Hurst, C.J. 1997. Recovery of bacterial community DNA from soil, p. 433-434. American Society for Microbiology, Washington, D.C., USA.

Kim, B.S., B.K. Kim, J.H. Lee, M. Kim, Y.W. Lim, and J. Chun. 2008. Rapid phylogenetic dissection of prokaryotic community structure in tidal flat using pyrosequencing. *J. Microbiol.* 46, 357-363.

Kim, B.S., H.M. Oh, H. Kang, and J. Chun. 2005. Archaeal diversity in tidal flat sediment as revealed by 16S rDNA analysis. *J. Microbiol.* 43, 144-151.

Kim, B.S., H.M. Oh, H. Kang, S.S. Park, and J. Chun. 2004. Remarkable bacterial diversity in the tidal flat sediment as revealed by 16S rDNA analysis. *J. Microbiol. Biotechnol.* 14, 208-211.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution, p. 149-304. Cambridge University Press, Cambridge, USA.

Krogh, A., B. Larsson, G. von Heijne, and E.L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567-580.

Lee, M.H., C.H. Lee, T.K. Oh, J.K. Song, and J.H. Yoon. 2006. Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. *Appl. Environ. Microbiol.* 72, 7406-7409.

Lopez-Garcia, P., C. Brochier, D. Moreira, and F. Rodriguez-Valera. 2004. Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ. Microbiol.* 6, 19-34.

Ma, D., M. Alberti, C. Lynch, H. Nikaido, and J.E. Hearst. 1996. The local repressor AcrR plays a modulating role in the regulation of acrAB genes of *Escherichia coli* by global stress signals. *Mol. Microbiol.* 19, 101-112.

Michael, C.A., M.R. Gillings, A.J. Holmes, L. Hughes, N.R. Andrew, M.P. Holley, and H.W. Stokes. 2004. Mobile gene cassettes: a fundamental resource for bacterial evolution. *Am. Nat.* 164, 1-12.

Ochsenreiter, T., D. Selezi, A. Quaiser, L. Bonch-Osmolovskaya, and C. Schleper. 2003. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ. Microbiol.* 5, 787-797.

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276, 734-740.

Paget, M.S., J.G. Kang, J.H. Roe, and M.J. Buttner. 1998. sigmaR, an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2). *Embo. J.* 17, 5776-5782.

Pope, P.B. and B.K. Patel. 2008. Metagenomic analysis of a freshwater toxic cyanobacteria bloom. *FEMS Microbiol. Ecol.* 64, 9-27.

Quaiser, A., T. Ochsenreiter, H.P. Klenk, A. Kletzin, A.H. Treusch, G. Meurer, J. Eck, C.W. Sensen, and C. Schleper. 2002. First insight into the genome of an uncultivated crenarchaeote from soil. *Environ. Microbiol.* 4, 603-611.

Rondon, M.R., P.R. August, A.D. Bettermann, S.F. Brady, T.H. Grossman, M.R. Liles, K.A. Loiacono, B.A. Lynch, I.A. MacNeil, C. Minor, C.L. Tiong, M. Gilman, M.S. Osburne, J. Clardy, J. Handelsman, and R.M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66, 2541-2547.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425.

Sambrook, J. and D.W. Russell. 2001. Molecular cloning : a laboratory manual, p. 1.25-21.28. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, N.Y., USA.

Sebat, J.L., F.S. Colwell, and R.L. Crawford. 2003. Metagenomic profiling: microarray analysis of an environmental genomic library. *Appl. Environ. Microbiol.* 69, 4927-4934.

Tatusov, R.L., M.Y. Galperin, D.A. Natale, and E.V. Koonin. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33-36.

Tatusov, R.L., E.V. Koonin, and D.J. Lipman. 1997. A genomic perspective on protein families. *Science* 278, 631-637.

Treusch, A.H., A. Kletzin, G. Raddatz, T. Ochsenreiter, A. Quaiser, G. Meurer, S.C. Schuster, and C. Schleper. 2004. Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ. Microbiol.* 6, 970-980.

Tyson, G.W., J. Chapman, P. Hugenholtz, E.E. Allen, R.J. Ram, P.M. Richardson, V.V. Solovyev, E.M. Rubin, D.S. Rokhsar,

and J.F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37-43.

Valko, M., C.J. Rhodes, J. Moncol, M. Izakovic, and M. Mazur. 2006. Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem. Biol. Interact.* 160, 1-40.

Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, D. Wu, I. Paulsen, K.E. Nelson, W. Nelson, D.E. Fouts, S. Levy, A.H. Knap, M.W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.H. Rogers, and H.O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66-74.

Yi, H., K.S. Bae, and J. Chun. 2004. *Thalassomonas ganghwensis* sp. nov., isolated from tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* 54, 377-380.

Yi, H., Y.H. Chang, H.W. Oh, K.S. Bae, and J. Chun. 2003. *Zooshikella ganghwensis* gen. nov., sp. nov., isolated from tidal flat sediments. *Int. J. Syst. Evol. Microbiol.* 53, 1013-1018.

Yi, H. and J. Chun. 2004. *Hongiella mannitolivorans* gen. nov., sp. nov., *Hongiella halophila* sp. nov. and *Hongiella ornithinivorans* sp. nov., isolated from tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* 54, 157-162.

Zheng, M., F. Aslund, and G. Storz. 1998. Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science* 279, 1718-1721.