UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SINGLE-CELL GENOMICS OF A CANDIDATE DIVISION TM6 – UNCULTURED BACTERIAL PHYLUM IN ZODLETONE SPRING, OKLAHOMA

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

ANNA DOLOMAN
Norman, Oklahoma
2014

SINGLE-CELL GENOMICS OF CANDIDATE DIVISION TM6 – UNCULTURED BACTERIAL PHYLUM IN ZODLETONE SPRING, OKLAHOMA

A THESIS APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

[redacted]
Dr. Lee R. Krumholz, Chair

[redacted]
Dr. Boris Wawrik

[redacted]
Dr. Mostafa S. Elshahed

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Lee Krumholz, for introducing me to the project on uncultured bacterial phyla in Zodletone Spring. I would like to express my sincere and great gratitude for his enormous patience, always wise and great advises and constant support during my research and study.

My thanks will also go to Dr. Boris Wawrik, for his great insights into the bioinformatical analysis, huge technical support and inspiration while having seemingly lost results. Thanks to Dr. Mostafa Elshahed, for in-time advices and great guidance on the roughest part of my thesis.

I owe great thanks to the Fulbright Graduate Student Scholarship and it was a great honor to be a Fulbright Scholar and come to study towards Masters' degree at the University of Oklahoma.

I would like to thank to the National Science Foundation for sponsoring this research in Zodletone Observatory. Also, many thanks to the staff of the Oscer Boomer Super Computer facility of the University of Oklahoma, who made it possible to perform memory-consuming bioinformatical calculations for my research.

My big thanks go to my awesome lab mates – Yiwei Diao, Dominique Poncelet, Luyao Wang, Chuang Li and former students – Oszan Gazioglu and Abeer Alghamdi. Their constant support and fun work atmosphere helped me a lot during my Master's degree journey.

Last, but definitely not least, my enormous thanks to my dearest parents, who were constantly supportive, even being so geographically far away from me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Previously, metagenomic studies of Zodletone Spring (south-western Oklahoma) revealed uncultured and poorly characterized bacterial phyla. Among them, a bacterial phylum "Candidate Division TM6" is present. It is widespread in a variety of natural environments and may contribute to geochemical cycling processes. We used a single-cell approach for targeting and separation of cells belonging to candidate division TM6. This approach includes design of specific primers, fluorescence *in situ* hybridization and fluorescence activated cell sorting, to obtain bacteria within this novel phylum. Extracted DNA from sorted cells was amplified using multiple displacement amplification and phylogenic analysis of sorted cells demonstrated the presence of candidate division TM6 DNA, based on 16SrRNA analysis. Amplified DNA was used for whole-genome sequencing on a MiSeq Illumina platform and generated 250 bp paired end reads. Genome *de novo* assembly was done using SPAdes 2.5.0 and Ray 2.2.0 algorithms, with subsequent phylogenetic binning of scaffolds into the TM6 assembly of 64,034 bp. Assembled genes were annotated on the RAST server and manually with the NCBI database. There were nine proteins encoded by assembled genes: ATP-dependent DNA helicase PcrA (EC 3.6.4.12), DNA polymerase III subunit gamma/tau (EC 2.7.7.7), transcriptional regulator OmpR, histidine kinase (EC 2.7.13.3), lysophospholipase (EC 3.1.1.5), cyanophycin synthase (EC 6.3.2.29), acetyltransferase 3 (EC 2.3.-.-), allophanate hydrolase (EC 3.5.1.54), RND multidrug efflux transporter and beta-lactamase (EC 3.5.2.6). All the proteins clustered distinctly from those in other different phyla. Given these results, we suggest that CD TM6 is able to use nitrogen, stored in a form of a cyanophycin polymer, by utilizing it in urea cycle.

INTRODUCTION

Microorganisms inhabit a range of environments, making Bacteria the most diverse kingdom. The diversity is currently based on 54 phyla, with 30 have cultured representatives, whereas other 24 phyla were added from year 1987 to 2003 and majority of them are uncultured and poorly characterized (Rappe and Giovannoni 2003). These uncultured phyla are referred to as – "Candidate Divisions". During the next decade, uncultured bacteria were added within already existing phyla. Currently, it is assumed that ecotrophical relationships in microbial habitats include a complex interaction between well-studied bacteria and numerous representatives of uncultured candidate divisions, which greatly outnumber the cultured bacteria (Pester, Bittner et al. 2010). In terms of the most important microorganisms related to ecosystem function, we are just looking at the top part of the iceberg.

However, we are slowly moving beneath the surface of the iceberg to gain an understanding of microbial ecotrophical relationships. One particularly efficient approach involves environmental metagenomics (Wooley, Godzik et al. 2010). Recently, a genome was assembled of candidate division TM7 from a wastewater treatment bioreactor (Albertsen, Hugenholtz et al. 2013). The genome assembly revealed the presence of heterolactic and pentose phosphate fermentation pathways, indicating that TM7 can only utilize glucose as a fermentation substrate. Such substrate preference is consistent with the presence of TM7 in other environments – human oral cavity – where TM7 was detected during the periods of high glucose concentrations (Brinig, Lepp et al. 2003). Recently, partial genomes were assembled for four candidate divisions – SR1, OD1, WWE3 and TM7 from acetate-stimulated aquifer

sediments (Kantor, Wrighton et al. 2013). Previous works on OD1 and TM7 genomes (Elshahed, Najar et al. 2005; Marcy, Ouverney et al. 2007) did not provide a complete metabolic reconstruction. All four of the above candidate divisions were defined to be oligotrophic and require the microbial community for synthesis of essential metabolites.

Candidate division TM6 is often observed in the environment and it has been reported in hypersaline biofilms on the surfaces of caves (Macalady, Jones et al. 2007), East Pacific deep-sea sediments (Li, Li et al. 2008), marine sponges (Montalvo and Hill 2011), subsurface peat layers (Serkebaeva, Kim et al. 2013), pine rhizosphere soils (Chow, Radomski et al. 2002) and murine gut microbiome (Gu, Chen et al. 2013). However, in all of the above environments, TM6 is a minor component and therefore its growth requirements have not been defined. The only study that elucidated TM6 substrate preferences used sulfate-methane transition zone sediments from Aarhus Bay, Denmark (Webster, Sass et al. 2011). Initially, TM6-related DNA was not detected after amplification, cloning and sequencing of 16S rRNA genes from sediments. However, after 6 months incubation of the sediment slurry with a mixture of sulfate, glucose and acetate, clones of TM6 were detected, indicating that growth of TM6 was limited by nutrient availability in a sulfate-methane transition zone. A genome sequencing study from a hospital sink biofilm (McLean, Lombardo et al. 2013) has suggested that TM6 lives in a symbiotic association with a free-living organism, explaining its ability to live in a variety of environments. This could also explain the difficulties in cultivation of TM6 using traditional techniques and its undefined substrate preferences. In that study, TM6 cells were sorted from a hospital sink biofilm and after genome sequencing, authors documented an assembly of 1.07 Mbp, which is consistent with this strain being

a host associated bacterium. A complete absence of pathways for synthesis of amino acids, pyrimidines and purines, were reported, but the presence of encoded cotransporting proteins sodium/proline, sodium/alanine and sodium/panthothenate suggested uptake of those compounds. However, this genome assembly is only 90% complete and therefore, some biosynthetic pathways can be missing.

Candidate division TM6 was also observed in the anoxic methane-rich Zodletone spring, located in the southwestern part of Oklahoma (Elshahed, Senko et al. 2003). Previous metagenomic studies at this spring have revealed numerous uncultured and poorly characterized bacterial phyla including SR1, OP3, WS6, TM7 and OP11 (Youssef, Couger et al. 2010). Such high bacterial diversity can be explained by the complex chemical interactions in this spring where microbial processes include sulfur disproportionation (*Desulfocapsa*), sulfate and sulfur reduction (*Desulfovibrio*), phototrophic sulfide and sulfur oxidation (*Cyanobacteria*, *Chloroflexi*) and alkane degradation (*Cytophaga*, *Planctomyces)* (Elshahed, Senko et al. 2003; Youssef, Couger et al. 2010).

This study focuses on characterization of candidate division TM6 in the Zodletone Spring sediments. We used a fluorescence *in situ* hybridization (FISH), fluorescence-activated cell sorting (FACS) and multiple displacement amplification (MDA) to target single cells of TM6 phylum. Use of fluorescence *in situ* hybridization makes it possible to visualize TM6 cell morphology. Isolated cells through fluorescence-activated cell sorting are further subjected to high-throughput sequencing to reconstruct the genome.

MATERIALS AND METHODS

Sample collection, fixation and DNA extraction

The top 10 cm of sediment were collected from the source of Zodletone Spring and 3 m downstream in early October, 2013 and kept on ice during transportation to the laboratory. Samples were fixed for FISH analysis or frozen for DNA extraction within 4 hours of sampling. Fixation involved dilution (1:1) of sediments in 8% paraformaldehyde and 2% sodium chloride in 0.1M HEPES (pH 7.0). Fixed samples were stored for 24 hours at 4°C. Then samples were differentially centrifuged in a benchtop microcentrifuge to remove non cell material (Davis, Youssef et al. 2009). Briefly, the samples were first centrifuged at 14,000xg for 10 minutes to remove fixative solution. The pellet was resuspended in 100mM sodium pyrophosphate and phosphate buffered saline (137mM sodium chloride, 2.7mM potassium chloride, 10mM disodium hydrogen phosphate and 1.8mM potassium dihydrogen phosphate) and shaken for 10 minutes on a vortexer. This was followed by a 10 minute centrifugation at 800xg to remove large sediment particles. The supernatant was then centrifuged for 20 minutes at 6,000xg to remove smaller sediment particles and a final centrifugation was done at 14,000xg for 10 minutes to collect bacterial cells. Fixed bacterial cells were resuspended in phosphate buffer saline mixture with ethanol (1:1) and stored at -20°C. DNA extraction from frozen sediments was done using Power Soil DNA Isolation Kit (Carlsbad, CA) following the manufacturers' protocol.

Cloning of TM6 16S rRNA

A TM6-specific reverse primer (487R) was designed using ARB software and was paired with the universal bacterial primer 27F (Table 1) to selectively PCR amplify

TM6 16S rRNA gene from Zodletone spring DNA. PCR mixture contained DreamTaq DNA Polymerase (Thermo Scientific, Lithuania), DreamTaq Green buffer with 20 mM MgCl2, 2mM dNTPs, forward and reverse primers (10mM). PCR conditions were as following: 94°C initial denaturation (5 min), 30 cycles of 94°C denaturation (30 sec), 60°C annealing (1 min) and 72°C extension (1 min 15 sec), followed by final extension for 15min at 72°C. The sequence of the PCR product was confirmed by Sanger sequencing. Phylogeny was determined by aligning the sequence in the SILVA 16S rRNA database (Pruesse, Peplies et al. 2012). The PCR product was then purified using Gene Jet PCR purification kit (Thermo Scientific, Lithuania) and inserted into the pCR4-TOPO TA vector and transformed into *Escherichia coli* DE3 cells (TOPO TA cloning kit , Life technologies, Grand Island, NY). Following the protocol described by Schramm and colleagues (Schramm, Fuchs et al. 2002), recombinant *E.coli* DE3 cells with TM6 16S rRNA inserts were grown in 5 ml of liquid Luria Broth with addition of kanamycin (final concentration 50μg/ml) overnight. An overnight culture was diluted (1:1000) and grown to OD_{600}=0.3-0.4. Isopropyl-beta-D-thiogalactopyranoside (1 mM) was added to the tubes for 1 hour incubation and then chloramphenicol (170μg/ml) was added to the tubes and cultures were incubated for 4 hours. All cultures were incubated at 37°C and shaken at 200 rpm. Cells were collected by centrifugation at 10,000xg for 5min and fixed with paraformaldehyde and phosphate buffer saline mixture with ethanol (1:1) as described previously and stored at -20°C.

Table 1. Probe and primers used in this study.

Primer/probe name	Primer/probe sequence (5' – 3')
487R / TM6-487	CGG TGC TTT CTC TAA TGG
27F	AGA GTT TGA TCM TGG CTC AG
1492R	TAC GGY TAC CTT GTT ACG ACT T
M13F	GTA AAA CGA CGG CCA GT
M13R	GCG GA TAA CAA TTT CAC ACA GG

Fluorescence in situ hybridization (FISH)

Zodletone spring source sediment and *E.coli* DE3 cells were additionally fixed in 75% ethanol, incubated for 5 minutes at room temperature, centrifuged at 5,000xg for 2 minutes, resuspended in 100% ethanol and incubated for another 2 minutes. Cells were centrifuged at 5,000xg for 2 minutes and then air-dried for 20 minutes. The resulting dry pellets in 0.5 ml tubes were resuspended in 10 µl of a pre-warmed hybridization buffer (0.9 M NaCl, 0.005M EDTA, 0.1M Tris, 0.01% SDS and formamide (20-70%) and pre-hybridized at 46°C in a water bath for 30 minutes. The formamide concentration in the hybridization buffer was adjusted over a range of 20% to 70%, with 10% increment to optimize hybridization conditions in the *E. coli* clones. Formamide concentration at which the brightest fluorescence peak was observed (preceded and followed by the decrease in the fluorescence intensity) was chosen as the optimum formamide concentration for FISH. The TM6-487 probe was designed using ARB (Ludwig, Strunk et al. 2004) to specifically target 16S rRNA of candidate division

TM6 and was 5' end labelled with Alexa Fluor 555 dye (Molecular Probes, Carlsbad, CA) following the manufacturers protocol (Invitrogen Protocols, 08/2011). The fluorescently labeled probe (1 µl) at 50ng/µl was added into the pre-hybridization mixture (consisted of hybridization buffer and cell material that was pre-warmed for 15 min at 46°C) and hybridized to cells in the dark at 46°C for 2 hours. After hybridization, samples were centrifuged at 5,000xg for 5 minutes and washed twice in washing buffer (20mM Tris, 5mM EDTA, 0.01% SDS and NaCl (0.9-0.008M)). Concentration of NaCl used in washing buffers depended on corresponding concentrations of formamide in hybridization buffer (SILVA Protocols, 2010). After the second wash, cell pellets were resuspended in 10 µl of 0.1M phosphate buffer (0.1M sodium phosphate dibasic and 0.1M sodium phosphate monobasic) (pH 7.2) and kept at 4°C in the dark prior prior to further analysis.

Epifluorescent microscopy

To detect the presence of bacterial cells and visualize cell morphology, the hybridized sample (1 µl) was counterstained with DAPI (4',6-diamidino-2-phenylindole), by adding 3 µl of 0.5 mg/ml DAPI to 2 ml of 0.1M phosphate buffer and 1µl of hybridized sample. The mixture was vortexed and incubated at room temperature in the dark for 30 minutes. Slides for microscopy were prepared by placing the stained cells on black 0.2µm filters using a vacuum generated by an oil rotary pump (Reiswig, Browman 1987). Both counterstained and non-stained slides were observed under 1000x magnification on an epifluorescent microscope (Olympus BX 61) using the DAPI blue filter and Alexa Fluor 555 orange filter.

Fluorescence-activated cell sorting

Cell samples after FISH were diluted in 0.5 ml of 0.1M phosphate buffer and counterstained with 0.75µl of 0.5 mg/ml DAPI dye. Counterstaining with DAPI was necessary to destinguish bacterial cells from sediment particles that were small enough to escape differential centrifugation described above. Cell sorting was performed on a Becton Dickinson Influx Cell Sorter (San Jose, CA). Fluorophores were excited with 405nm and 553nm laser and fluorescence of cells was detected at 460nm emission for DAPI stained cells and 568nm emission for Alexa Fluor 555 labeled cells. Acquisition of gates for sorting of the double-positive fluorescent cell population and process control was done using Spigot software. Sorting data was processed using Dako Cytomation Summit v4.3 software. Cells were sorted into sterile 2ml tubes that contained 0.5 ml of a sterile 0.1M phosphate buffer. After sorting, collection tubes were centrifuged at 13,000xg for 2 minutes to collect sorted cells.

Multiple displacement amplification

Tubes with sorted cells were processed under a laminar flow hood to proceed with MDA under sterile conditions. Cell lysis and whole DNA amplification by multiple displacement amplification (MDA) was done using REPLI-g Single Cell kit (QIAGEN GmbH). Briefly, the procedure includes a cell lysis at 65°C for 10 minutes, amplification at 30°C for 8 hours and heat inactivation at 65°C for 3 minutes. Reactions were held in 0.2ml tubes and concentration of product was determined using the PicoGreen assay (Invitrogen Protocols). Diluted DNA from MDA was used as a PCR template for amplification with both universal (1492R + 27F) and specific (487R + 27F) primer sets to determine the presence of 16S rRNA of candidate division TM6. Insert

positive plasmids were extracted using QIAprep Spin Miniprep Kit (Germantown, MD) and Sanger sequenced using M13F and M13R primer sequences on a vector. Phylogeny was determined by aligning the sequence in SILVA 16S rRNA database (Pruesse, Peplies et al. 2012).

Library construction and sequencing

The Illumina MiSeq sequencing platform was used to sequence libraries of DNA from MDA reactions. Libraries were constructed using the Nextera DNA sample preparation kit (San Diego, CA). Each sample was barcoded and pooled onto a single-lane generating 250 bp paired end reads.

Genome assembly, binning and annotation

Prior to assembly, reads were quality filtered (q>20) using the *sickle* algorithm (Joshi, Fass 2011) and adapter sequences were trimmed using the *cutadapt* algorithm (Martin 2011). Quality assessment was done using FastQC (Babraham Bioinformatics). Assembly was done using SPAdes 2.5.0 (Bankevich, Nurk et al. 2012) and Ray 2.2.0 (Boisvert, Laviolette et al. 2010) software. Two assemblers were used in order to be able to compare quality of assembly in general and based on assembly of 16S rRNA genes. For SPAdes assembly kmer size was set to k=127, without limit for the assembled contig length; for Ray assembly k=31, with threshold for the contig length set to 400 bp. Contigs and scaffolds statistics was calculated using QUAST (Gurevich, Saveliev et al. 2013). Assembled contigs were loaded into blastn service on the NCBI web server and blasted against bacterial 16S rRNA database. Contigs that showed similarity to bacterial taxa with as low percentage of identity as 81% were also blasted against reference TM6 candidate division sequences in the NCBI database.

CLUSTALW and TCOFFEE were used for 16S rRNA sequence alignment and construction of a phylogenetic tree. A 16S rRNA sequence of the CD TM6 isolated from the hospital sink biofilm, the only member of Candidate Division TM6 for which there is an assembled genome (McLean, Lombardo et al. 2013), was also used for construction of phylogenetic trees. Phylogenetic binning of assembled contigs was done based on differences in GC percentage and tetranucleotide frequency of TM6-related contigs and other phyla. GC percentage was calculated using custom Python script and tetranucleotide frequency was calculated using calc.kmer.pl script from Multi-Metagenome Assembly Guide (Albertsen, Hugenholtz et al. 2013). Graphs of GC frequency distribution across assembled contigs were built using the R statistical package. Filtering of candidate division TM6-related scaffolds was done using custom Python and Perl scripts, setting a threshold below or equal to the G+C content of a scaffold that contained 16S rRNA gene of CD TM6. Also, because GC content of *E.coli* genome overlaps with GC content of some of the assembled 16S rRNA of TM6 in the scaffolds, additional filtration of *E.coli* sequences in the assembly was done by manually deleting scaffolds that showed hits to the *E.coli* in the nucleotide collection database in the NCBI. Variance of tetranucleotide frequency across assembled scaffolds was calculated using formula: $Variance = (s^2/X) * 100$, where s is a standard deviation of tetranucleotide frequency for all scaffolds, X is a mean of tetranucleotide frequency for all scaffolds (Noble, Citek et al. 1998). Annotation of final scaffolds was done using unsupervised annotation on RAST server (Aziz, Bartels et al. 2008) and manual annotation using *blastp* algorithm in the NCBI database. Reference protein sequences for phylogenetic trees were retrieved from IMG database and NCBI database.

Phylogenetic trees of protein amino acid sequences were constructed in MEGA 6.06 (Tamura, Stecher et al. 2013) using maximum likelihood statistical method with 500 times bootstrapping and Jones-Taylor-Thornton model.

RESULTS

E.coli DE3 cells containing TM6 16S rRNA inserts were used to determine optimal hybridization conditions.

Microscope images of cells hybridized with a fluorescent probe at 20%-70% formamide concentration were prepared (Figure 1). The brightest fluorescence was observed at 60% formamide concentration which was used for subsequent hybridizations. Following hybridization and DAPI staining, separated cells with different morphologies were observed with DAPI (Figure 2) and a few cells showed fluorescence with the TM6-487 probe (Figure 3), which is in agreement of the low abundance of CD TM6 in Zodletone spring and source sediments (Youssef, Couger et al. 2010). Presumably, TM6 cells are very small rods.

For each of the two sampling locations we treated 2 samples of independently hybridized cells using fluorescence-activated cell sorting. Figures of sediment cell populations and specified gates for sorting based on the fluorescent characteristics are shown in Figures 4 and 5. From the Zodletone spring source, samples S1 and S2, 5 and 6 cells were sorted; and from the samples collected 3m downstream, 3m1 and 3m2, 7 and 3 cells were obtained.

MDA products from all four samples were used for separate PCR reactions (Figure 6). Bands of the appropriate length were obtained with the CD TM6 specific primer pair 487R and 27F and with the universal primers. The PCR product from the universal primers yielded around 20 clones per each sample plate and sequencing detected two different 16S rRNAs: *Syntrophus aciditrophicus* and Candidate Division TM6. The ratio of the two 16S rRNA types was 3:1 respectively.

The original MDA product was then sequenced yielding approximately 3 million reads with S1 and 6 million reads with S2; 3 million reads were obtained with 3m1 and 10 million reads with 3m2. Assembly statistics can be found in the Table 1.

Both *Syntrophus aciditrophicus* and Candidate Division TM6 16S rRNAs were identified in the assembly. Contigs with 16S rRNA of CD TM6 were extracted and a phylogenetic tree with similar sequences from the NCBI database was built (Figure 7). Sequences of CD TM6 from Zodletone spring cluster separately from all the other CD TM6 clones from different environments and represent a new class inside TM6 phylum, with regard to 85% of difference in 16S rRNA sequence (Table 3) (Luo, Rodriguez et al. 2014).

A phylogenetic tree and percentage identity matrix of CD TM6 clones from Zodletone spring (Table 3) showed 100% similarity among three of clones – S1, S2 and 3m2. Therefore, all three sets of paired-end reads were concatenated into one fastq file and used it for the new contig assembly, focusing on further scaffolding. Statistics of a new assembly with both SPAdes and Ray algorithms can be found in Table 2. Using Ray for assembly yielded higher value of N50 and assembled the largest scaffold (84313 bp), however, it lacked consistent coverage of CD TM6 16S rRNA, meaning that there were not three times higher coverage of this fragment, as would have been expected from pooling three sets of reads into one assembly. Whereas, the SPAdes assembly provided the three-fold expected increase in coverage of 16S rRNA gene of CD TM6 in scaffolds. The SPAdes assembly was chosen for further analysis.

While comparing assembled contigs with the NCBI 16S rRNA database, *Palleronia marisminoris* was also observed. On the graph of distribution of G+C

content of contigs in the assembly (Figure 8) three sets of scaffolds can be observed, with median G+C content of 39%, 52% and 65%. The mean G+C content of *Syntrophus aciditrophicus* genome (middle) is 51.46% (McInerney, Rohlin et al. 2007) and mean G+C content of *Palleronia marisminoris* genome (right) is 64.2% (Martinez-Checa, Quesada et al. 2005) that is consistent with the right peak of our assembly. The only one currently assembled genome of CD TM6 from hospital sink biofilm has a mean G+C content 36% (McLean, Lombardo et al. 2013), which is a little lower than the observed mean G+C percentage of the left peak in our assembly graph (39%).

Results describing tetranucleotide frequency distribution illustrate a correlation between G+C content and tetranucleotide frequency characteristic for a mixture of bacterial genomes (Figure 9). When the low G+C contigs were analyzed by themselves, results suggested the presence of a single bacterial genome (Figure 10). These analyses are consistent with previous studies (Noble, Citek et al. 1998).

Basic statistics of the filtered CD TM6 scaffolds after applying all the binning approaches are combined in the Table 4. From those scaffolds, all the annotated proteins with predicted functions occupied unique evolutionary positions based on phylogenetic analysis of proteins with similar functions (Figures 11 – 20). Within these nine genes two housekeeping genes were identified, an ATP-dependent DNA helicase PcrA (EC 3.6.4.12) and DNA polymerase III subunit gamma/tau (EC 2.7.7.7). Other genes are related to various cell processes including regulation of transcription (transcriptional regulator PhoB/CheY), signal transduction (histidine kinase (EC 2.7.13.3)), hydrolysis of lysophospholipids with release of fatty acids (lysophospholipase (EC 3.1.1.5)), synthesis of cyanophycin (cyanophycin synthase (EC

6.3.2.29)), acetylation (acetyltransferase 3 (EC 2.3.-.-)), degradation of urea (allophanate hydrolase (EC 3.5.1.54)), efflux of antibiotics (membrane component of RND multidrug efflux transporter) and hydrolysis of beta-lactams (beta-lactamase (EC 3.5.2.6)).

Among all these genes only one – housekeeping gene encoding DNA polymerase III subunit gamma/tau (EC 2.7.7.7) – was also found in the previously published CD TM6 assembly (McLean, Lombardo et al. 2013). A distinct bootstrap on the phylogenetic tree with protein sequences of DNA polymerase III subunit gamma/tau from representatives of various bacterial phyla (Figure 11) and protein sequences of DNA polymerases of two CD TM6 genomes – from Zodletone spring and hospital sink biofilm – supported deep-branching node. This suggests that annotated DNA polymerase III subunit gamma/tau from assembly of CD TM6 from Zodletone spring is a part of the gene set of this novel bacterium and is closely related to the known sequence of CD TM6 from the hospital sink biofilm. Since there were no other analogous genes in the genome of a CD TM6 from hospital sink biofilm that can be compared with the genes from CD TM6 from Zodletone spring, all the other phylogenetic trees illustrate distinct position of proteins from only CD TM6 from Zodletone among the various phyla (Figures 12-20).

Low bootstrap values on the nodes of phylogenetic trees can be explained by the high divergence of aligned amino acid sequences from all proteins among different phyla and short length of aligned regions (Talavera and Castresana 2007).

The longest assembled contig (6695 bp) encodes four functional proteins and one hypothetical protein. Two out of four proteins comprise a complete two-component

signal transduction system, which consists of a sensor histidine kinase (EC 2.7.13.3) (Figure 13) and response regulator component, which shows close similarity to OmpR and PhoB regulators. To further characterize the regulator, we constructed a phylogenetic tree of this protein sequence from CD TM6 assembly and sequences retrieved from NCBI protein database for both OmpR and PhoB receiver domains (Figure 16). Receiver domain from CD TM6 assembly clusters together with OmpR proteins, not PhoB. A gene encoding beta-lactamase (EC 3.5.2.6) was also observed on this contig (Figure 17). Since hydrolysis of various beta-lactams is ruled by a single enzyme, beta-lactamase, we can assume that CD TM6 from Zodletone spring exhibits resistance to this group of antibiotics. Encoded acetyltransferase 3 (EC 2.3.-.-) (Figure 18) play role in transferring acyl groups to the peptidoglycan layer resulting in the O-acetylation, which can protect cell from the lysozyme (Moynihan and Clarke 2010).

Annotated allophanate hydrolase (EC 3.5.1.54) takes part in the utilization of urea as a nitrogen source. It is interesting to note that allophanate hydrolase subunit 1 clustered on a phylogenetic tree together with allophanate hydrolase subunit 1 from members of a Firmicutes phylum (Figure 19), whereas enzyme subunit 2 clustered with protein sequences from Bacteroides phylum (Figure 20). Probably, somewhere on the evolutionary road of development of CD TM6, a crossover event with genes from both phyla took place, which resulted in such heterogeneous clustering on a phylogenetic tree.

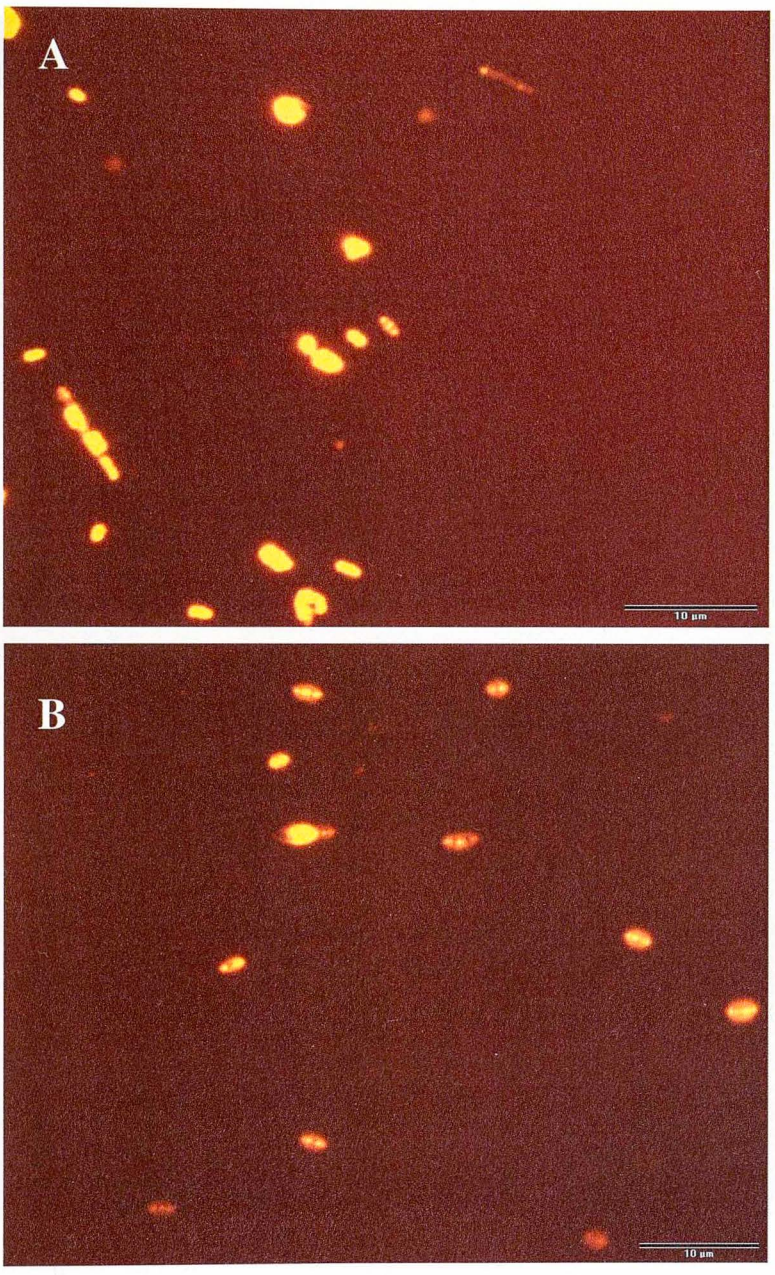

Figure 1. Fluorescence in situ hybridization of Escherichia coli DE3 cells with different formamide concentration in hybridization buffer. A) 20% formamide; B) 30% formamide; (see next page)

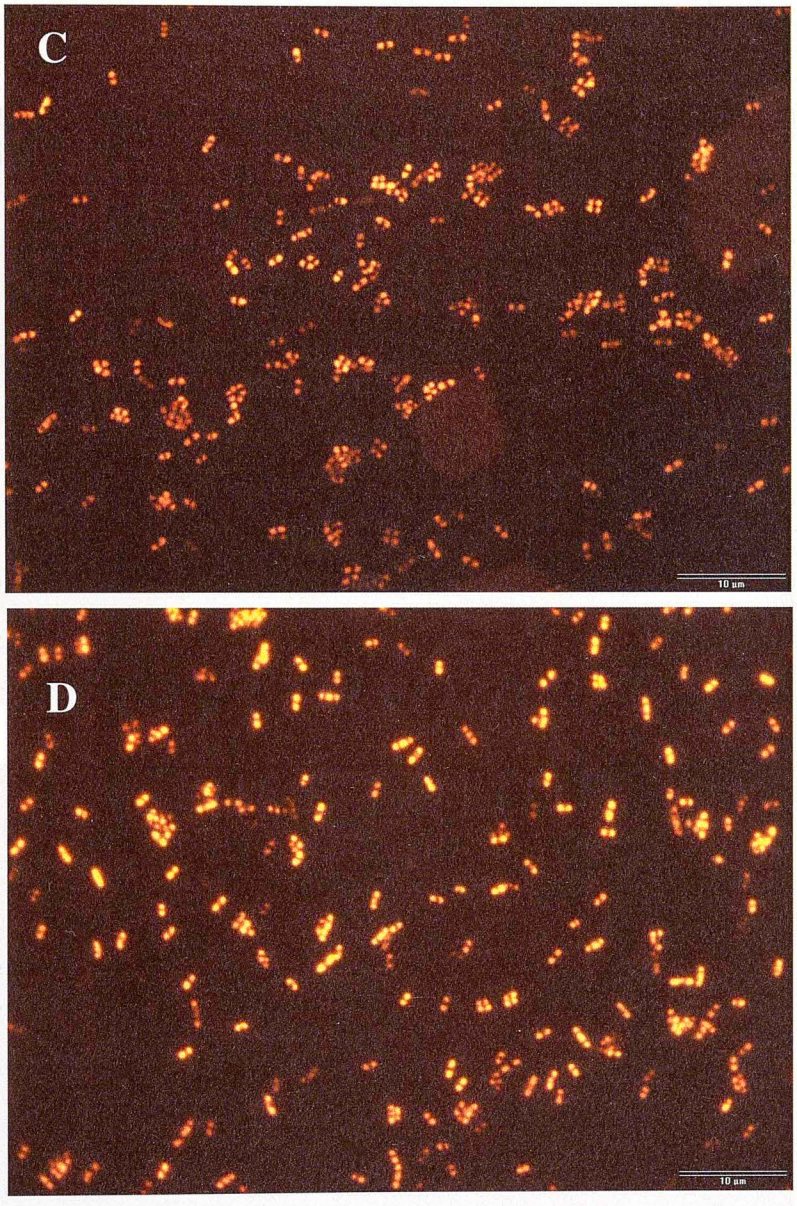

Figure 1 (cont). Fluorescence in situ hybridization of Escherichia coli DE3 cells with different formamide concentration in hybridization buffer.

C) 40% formamide; D) 50% formamide; (see next page)

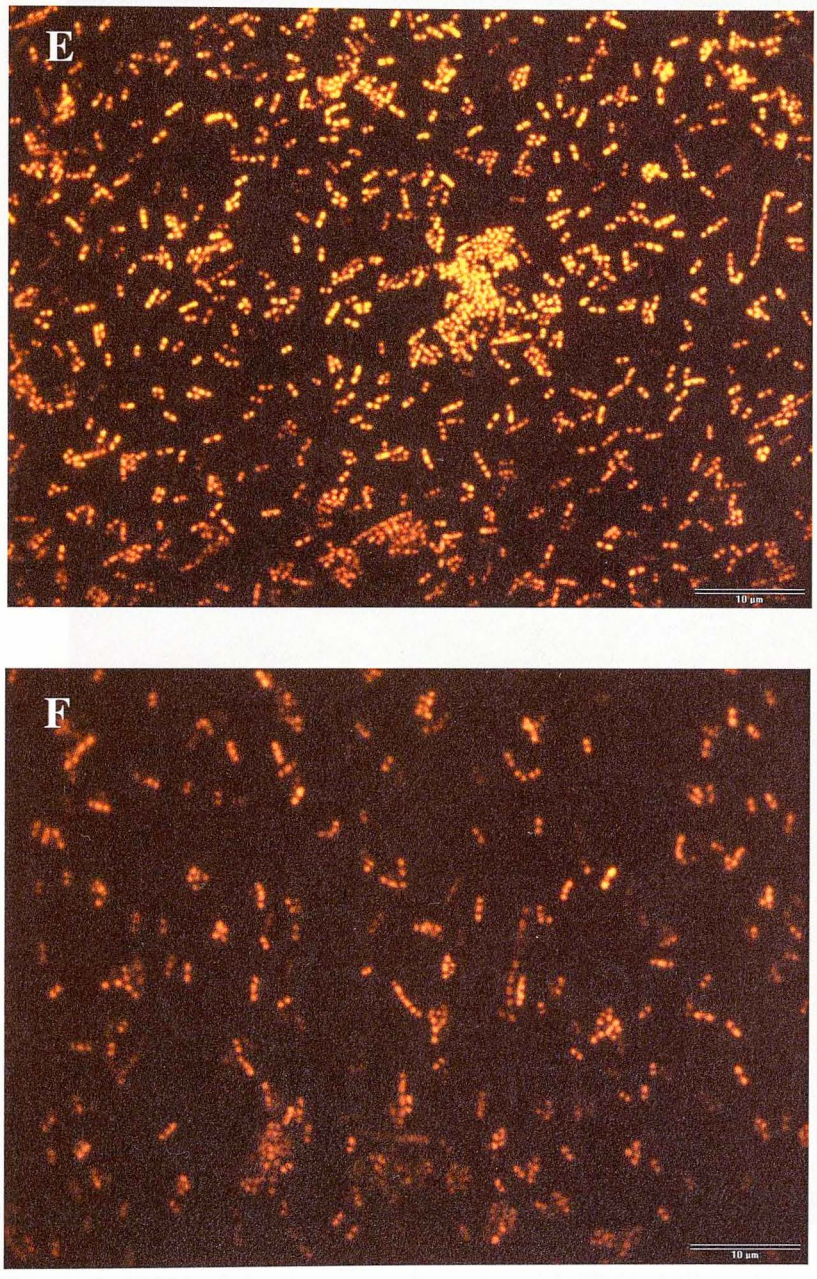

Figure 1 (cont). Fluorescence in situ hybridization of Escherichia coli DE3 cells with different formamide concentration in hybridization buffer. E) 60% formamide; F) 70% formamide.

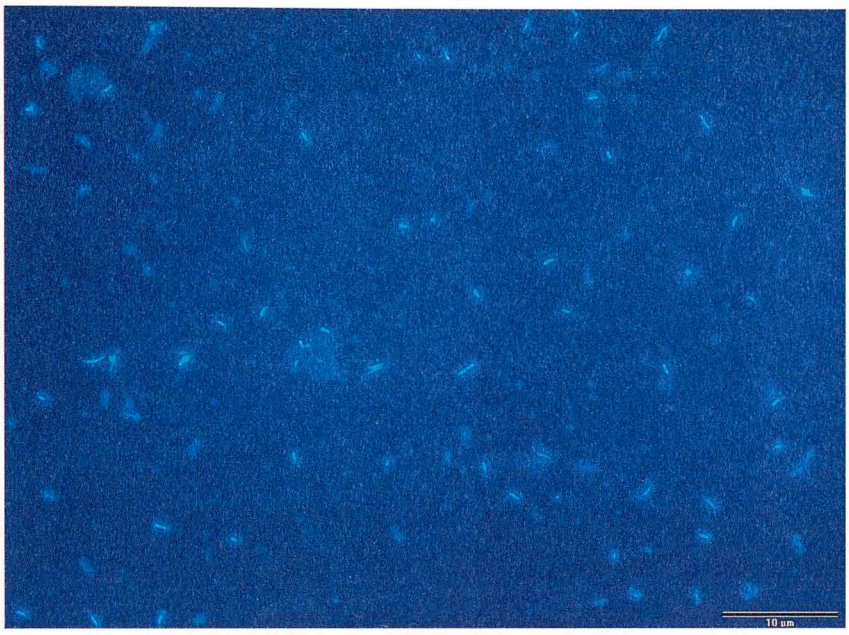

Figure 2. DAPI stained sediment sample.

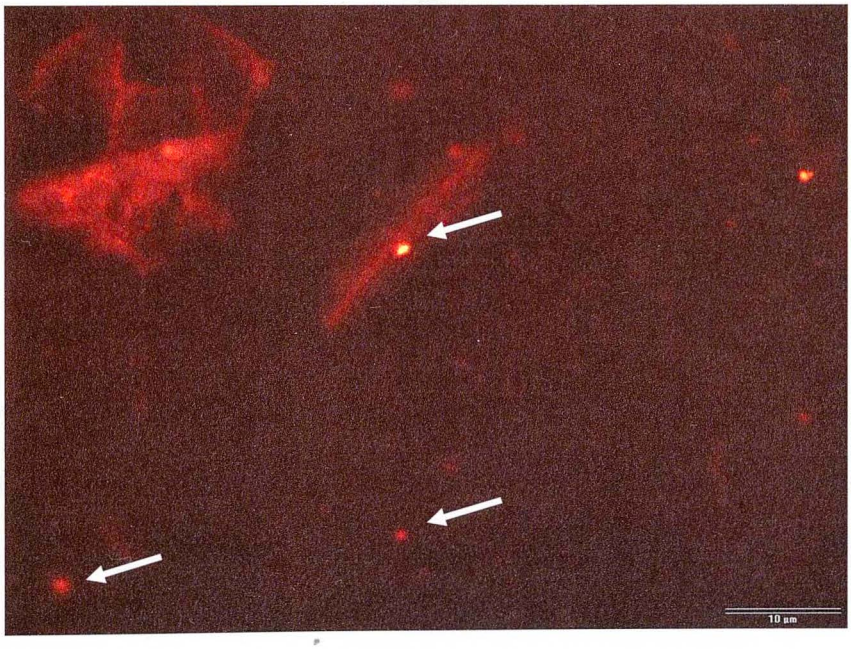

Figure 3. FISH of Zodletone sediments with TM6-487 probe. Arrows are pointing to the fluorescent CD TM6 cells.

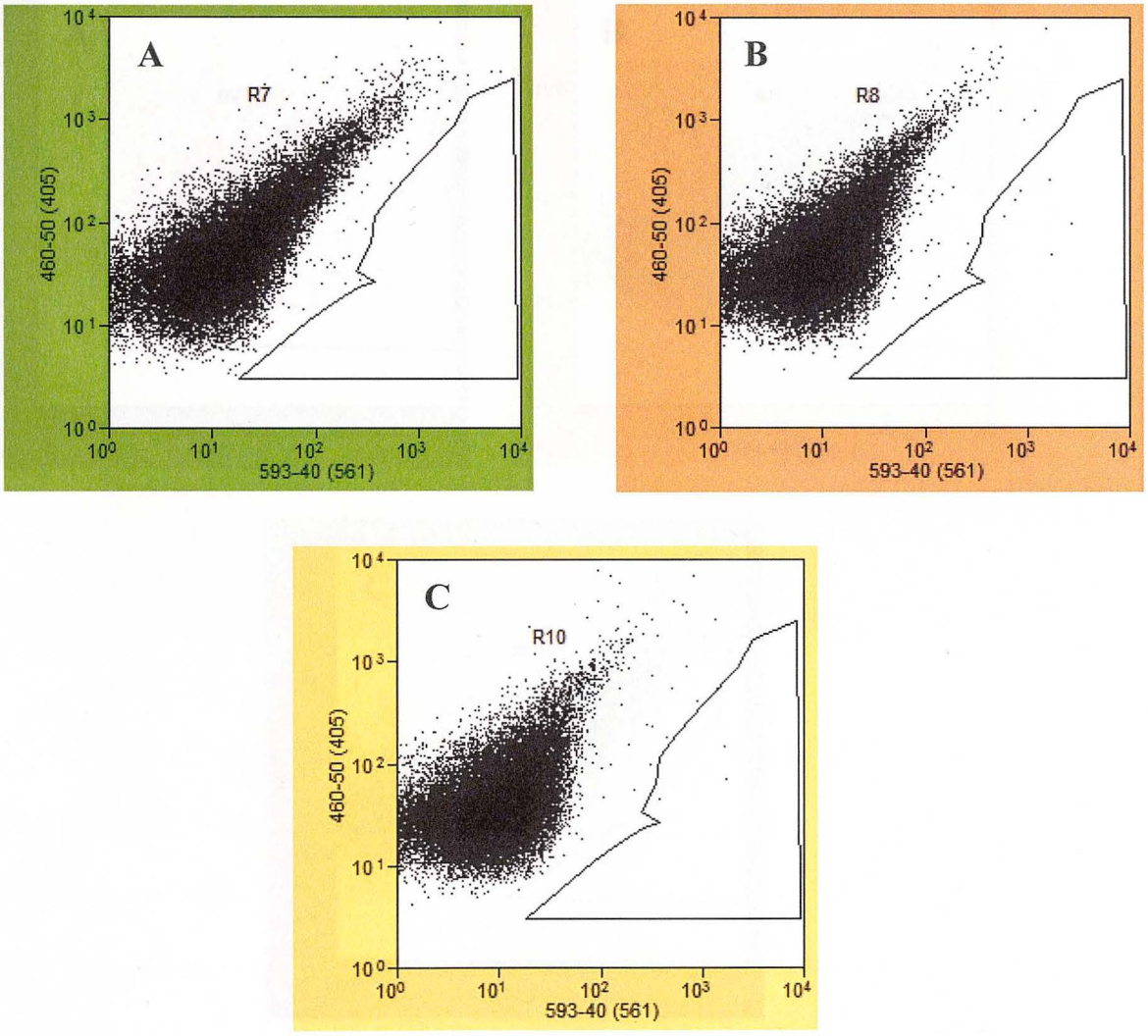

Figure 4. Fluorescence-activated cell sorting of sediment samples located 3 meters from the source of Zodletone Spring. A) Negative control, cells are stained with DAPI, no fluorescent probe added; B),C) Sediment samples with TM6487-Alexa Fluor 555 probe.

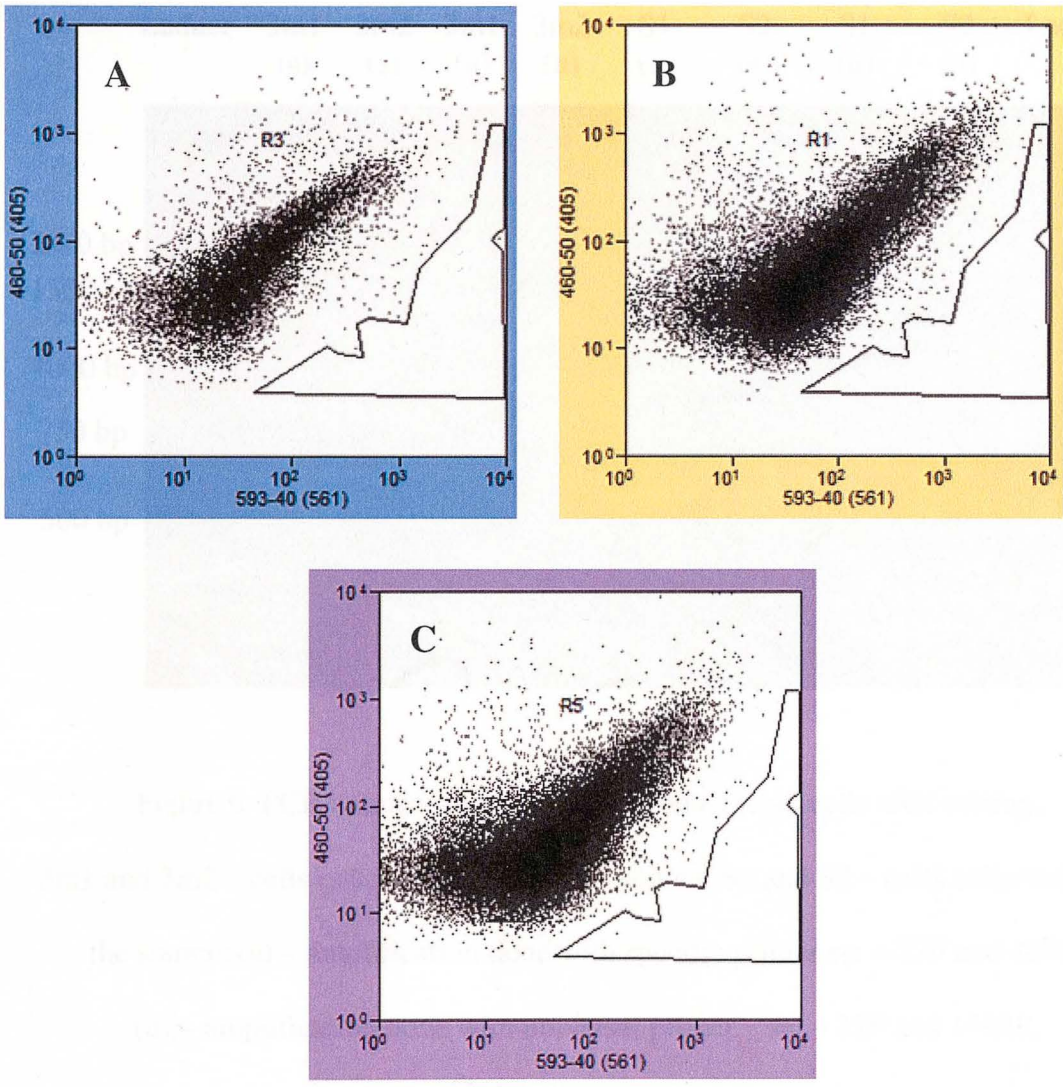

Figure 5. Fluorescence-activated cell sorting of sediment samples from the source of Zodletone Spring. A) Negative control, cells are stained with DAPI, no fluorescent probe added; B),C) Sediment samples with TM6487-Alexa Fluor 555 probe.

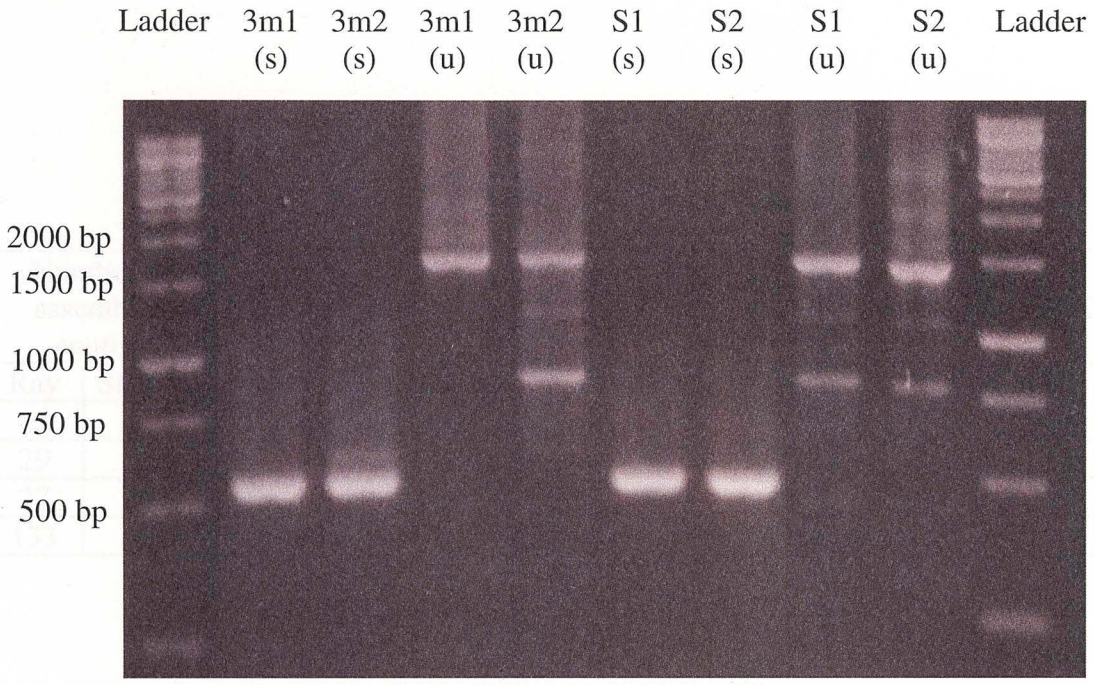

Figure 6. PCR amplification of MDA product from cells after sorting.

3m1 and 3m2 – cells collected 3 m from the source; S1 and S2 – cells collected from the source; (s) – amplification done with specific primer set – 27F and 487R; (u) – amplification done with universal primer set – 27F and 1492R.

Table 2. Assembly statistics of four sets of reads

Samples	Number of assembled contigs		Total length, bp		Largest contig, bp		N50, bp		(G+C)%, %	
	Ray	SPAdes	Ray	SPAdes	Ray	SPAdes	Ray	SPAdes	Ray	SPAdes
S1	25	157	53328	102340	6298	6081	3956	1132	46.71	48.98
S2	29	953	175737	388797	69784	14956	37873	386	50.62	53.71
3m1	27	207	43740	111615	10607	6790	2904	720	48.33	52.72
3m2	133	690	277002	398159	10283	6604	3327	735	45.48	48.09

Table 3. Assembly statistics of concatenated three sets of reads.

Number of assembled scaffolds		Total length, bp		Largest scaffold, bp		N50, bp		(G+C)%, %	
Ray	SPAdes	Ray	SPAdes	Ray	SPAdes	Ray	SPAdes	Ray	SPAdes
198	848	460754	654092	84313	21517	3875	1192	47.48	48.77

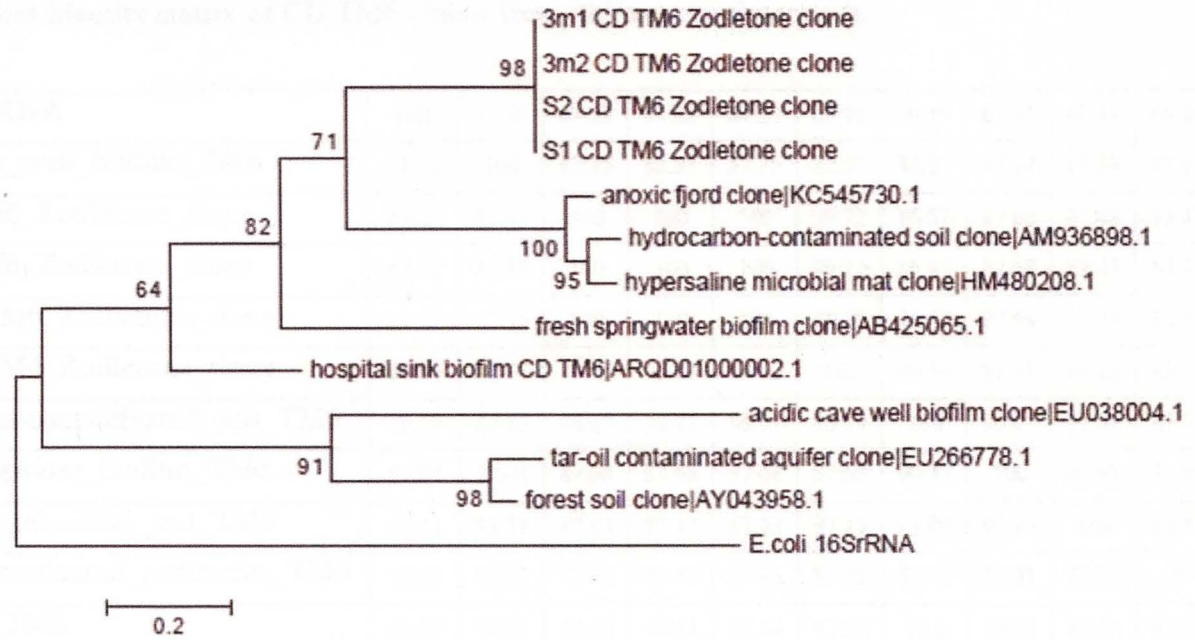

Figure 7. Phylogenetic tree of sequenced CD TM6 clones from Zodletone spring and closest 16S rRNA sequences in the NCBI 16S rRNA database.

Table 4. Percent identity matrix of CD TM6 clones from different environments.

*E.coli*_16SrRNA	100	71.51	67.22	67.22	67.22	66.94	70.19	67.04	67.41	68.91	69.19	69.27	71.43
Acidic_cave_well_biofilm_TM6	71.51	100	82.35	82.35	82.35	82.07	84.83	87.04	85.39	85.67	87.08	89.58	88.7
S1_CD_TM6_Zodletone_clone	67.22	82.35	100	100	100	99.72	86.87	87.68	87.43	83.43	87.64	86.27	85.39
S2_CD_TM6_Zodletone_clone	67.22	82.35	100	100	100	99.72	86.87	87.68	87.43	83.43	87.64	86.27	85.39
3m2_CD_TM6_Zodletone_clone	67.22	82.35	100	100	100	99.72	86.87	87.68	87.43	83.43	87.64	86.27	85.39
3m3_CD_TM6_Zodletone_clone	66.94	82.07	99.72	99.72	99.72	100	86.59	87.68	87.15	83.15	87.36	86.27	85.11
Hydrocarbon-contaminated_soil_TM6	70.19	84.83	86.87	86.87	86.87	86.59	100	90.45	91.09	87.71	88.8	87.43	91.62
Fresh_springwater_biofilm_TM6	67.04	87.04	87.68	87.68	87.68	87.68	90.45	100	92.98	87.01	89.27	89.89	91.53
Hypersaline_microbial_mat_TM6	67.41	85.39	87.43	87.43	87.43	87.15	91.09	92.98	100	85.43	87.68	89.11	90.2
Tar_oil_contaminated_sediments_TM6	68.91	85.67	83.43	83.43	83.43	83.15	87.71	87.01	85.43	100	88.52	87.92	88.83
Forest_soil_TM6	69.19	87.08	87.64	87.64	87.64	87.36	88.8	89.27	87.68	88.52	100	90.73	90.14
Anoxic_fjord_TM6	69.27	89.58	86.27	86.27	86.27	86.27	87.43	89.89	89.11	87.92	90.73	100	91.85
Hospital_sink_biofilm_TM6	71.43	88.7	85.39	85.39	85.39	85.11	91.62	91.53	90.2	88.83	90.14	90.14	100

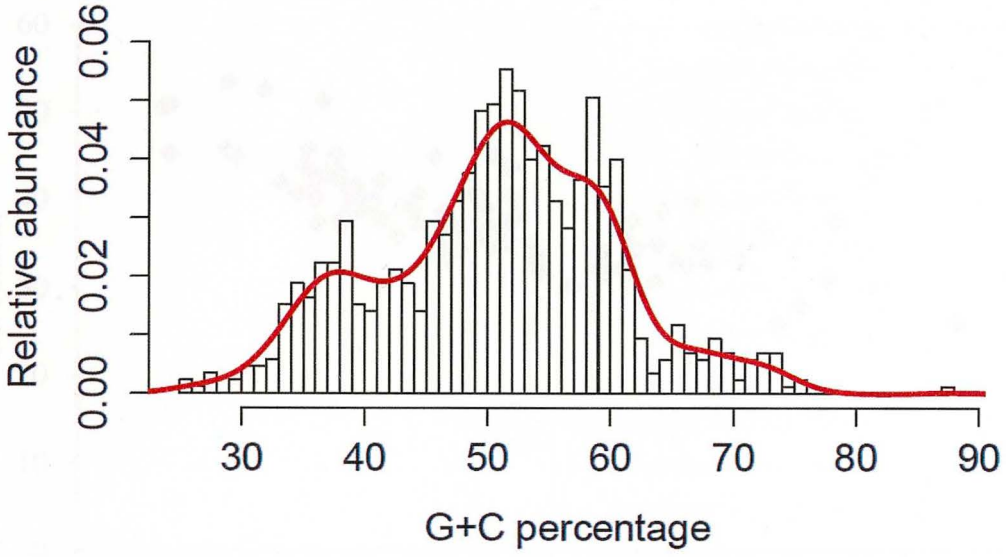

Figure 8. G+C content of assembled scaffolds in SPAdes.

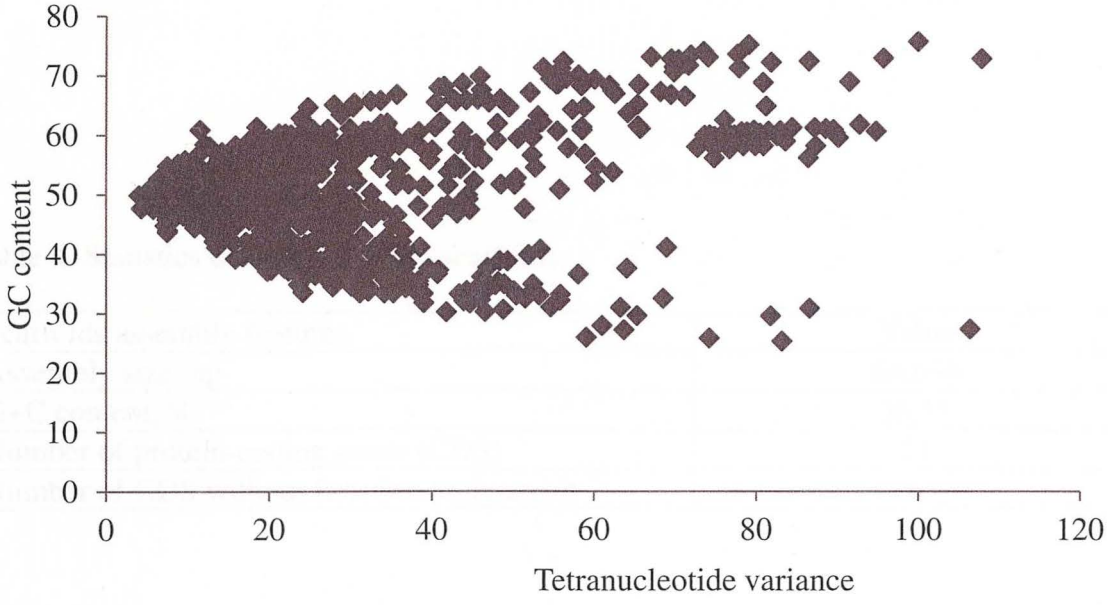

Figure 9. The GC content and tetranucleotide frequencies of unfiltered CD TM6 scaffolds (no phylogenetic binning).

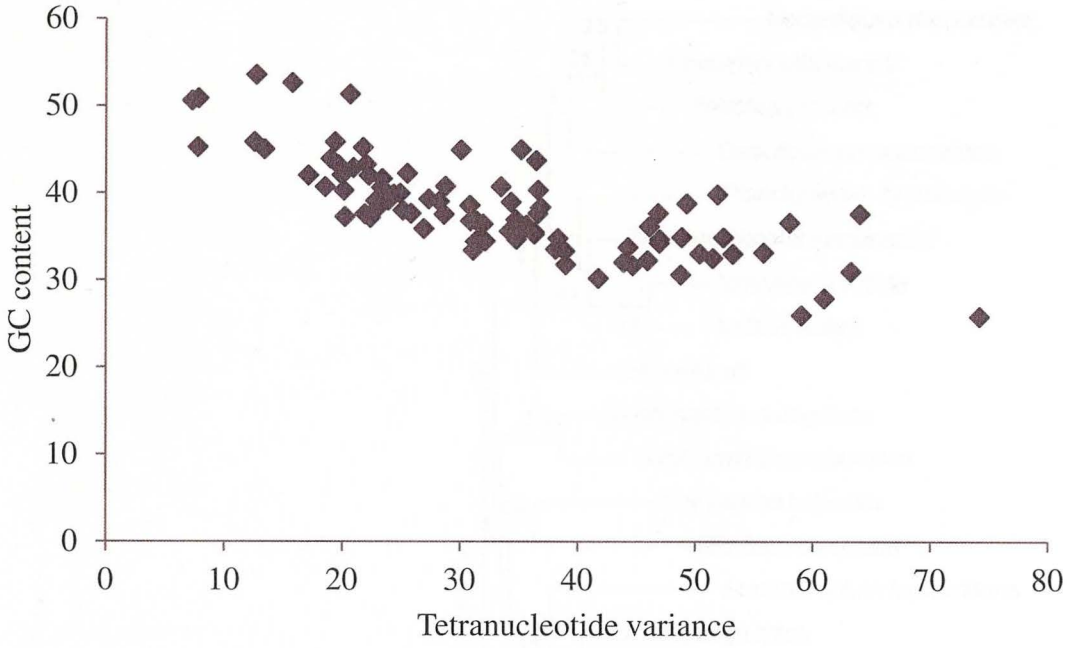

Figure 10. The GC content and tetranucleotide frequencies of CD TM6 scaffolds (after filtering and binning).

Table 4. Statistics of final CD TM6 scaffolds.

Scaffolds assembly features	Value
Assembly size, bp	64,034
G+C content, %	38.53
Number of protein-coding genes (CDS)	21
Number of CDS without function assignment	6

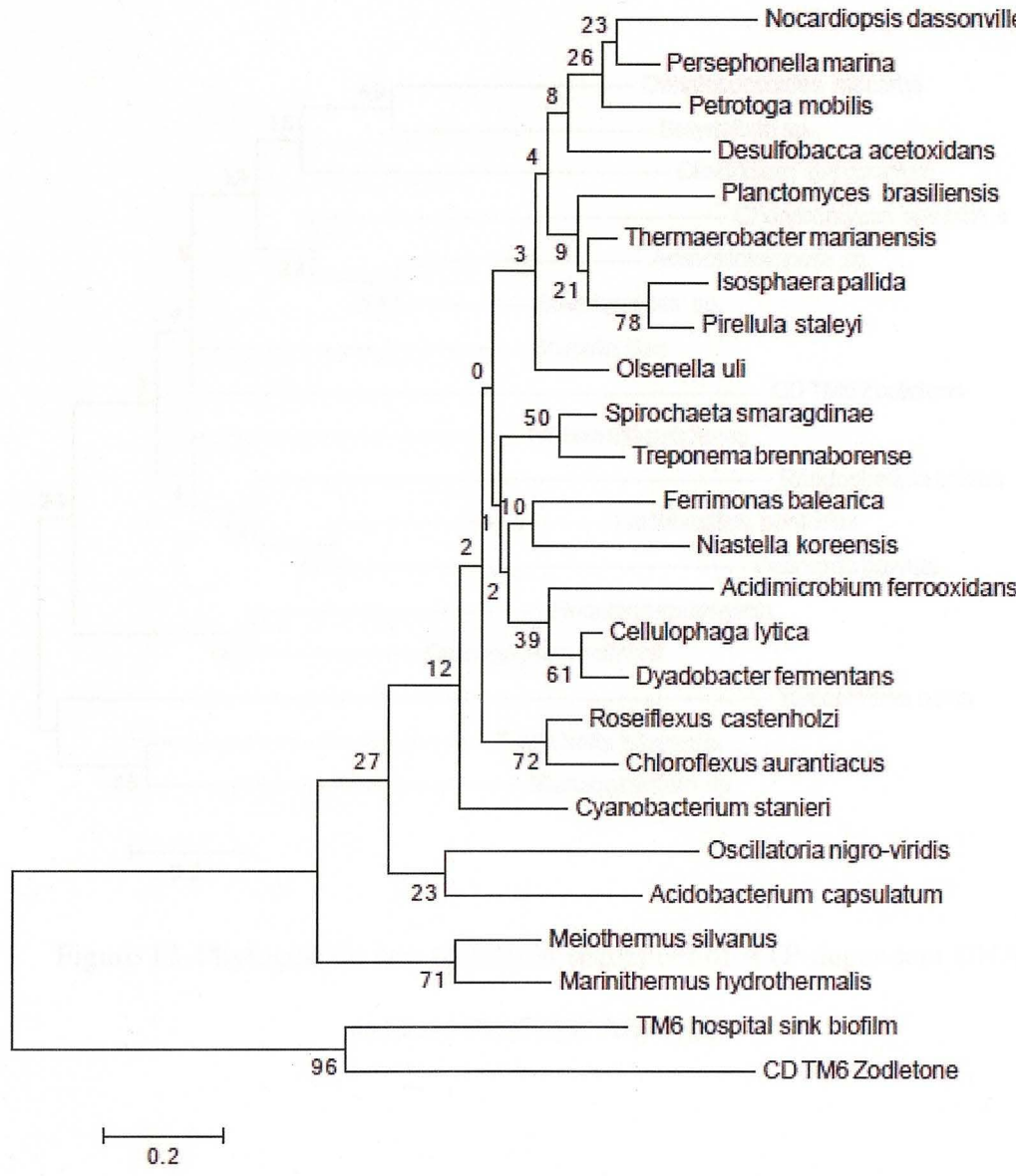

Figure 11. Phylogenetic tree of protein sequences of DNA polymerase III subunit gamma/tau (EC 2.7.7.-).

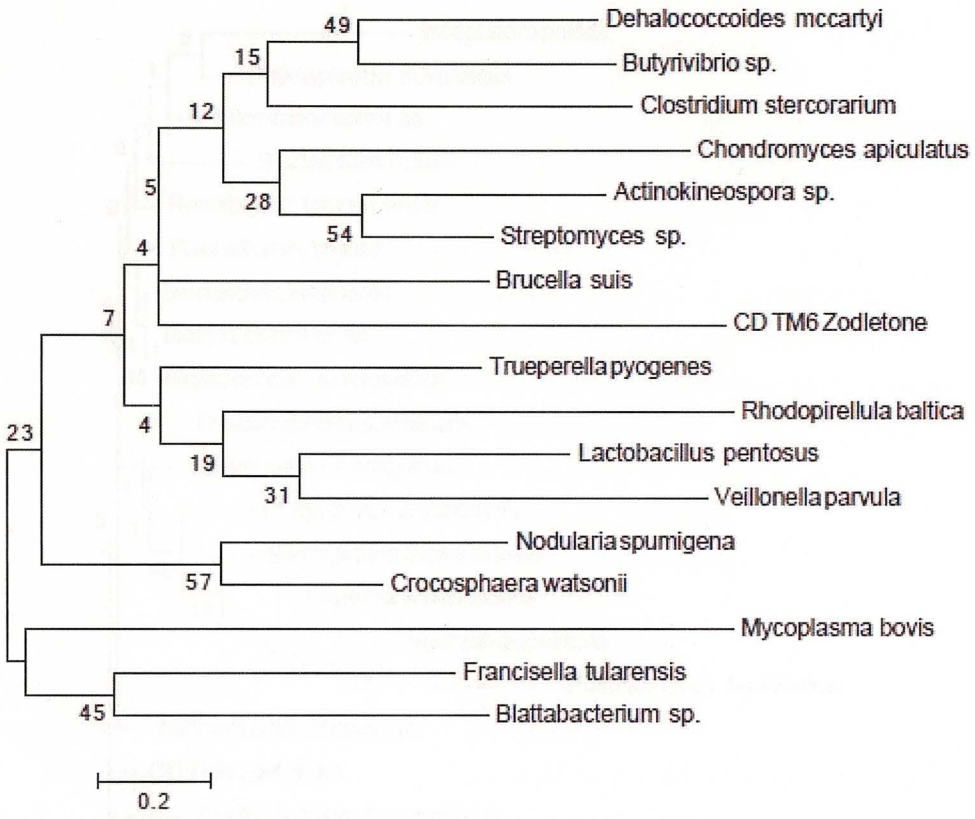

Figure 12. Phylogenetic tree of protein sequences of ATP-dependent DNA helicase PcrA (EC 3.6.4.12).

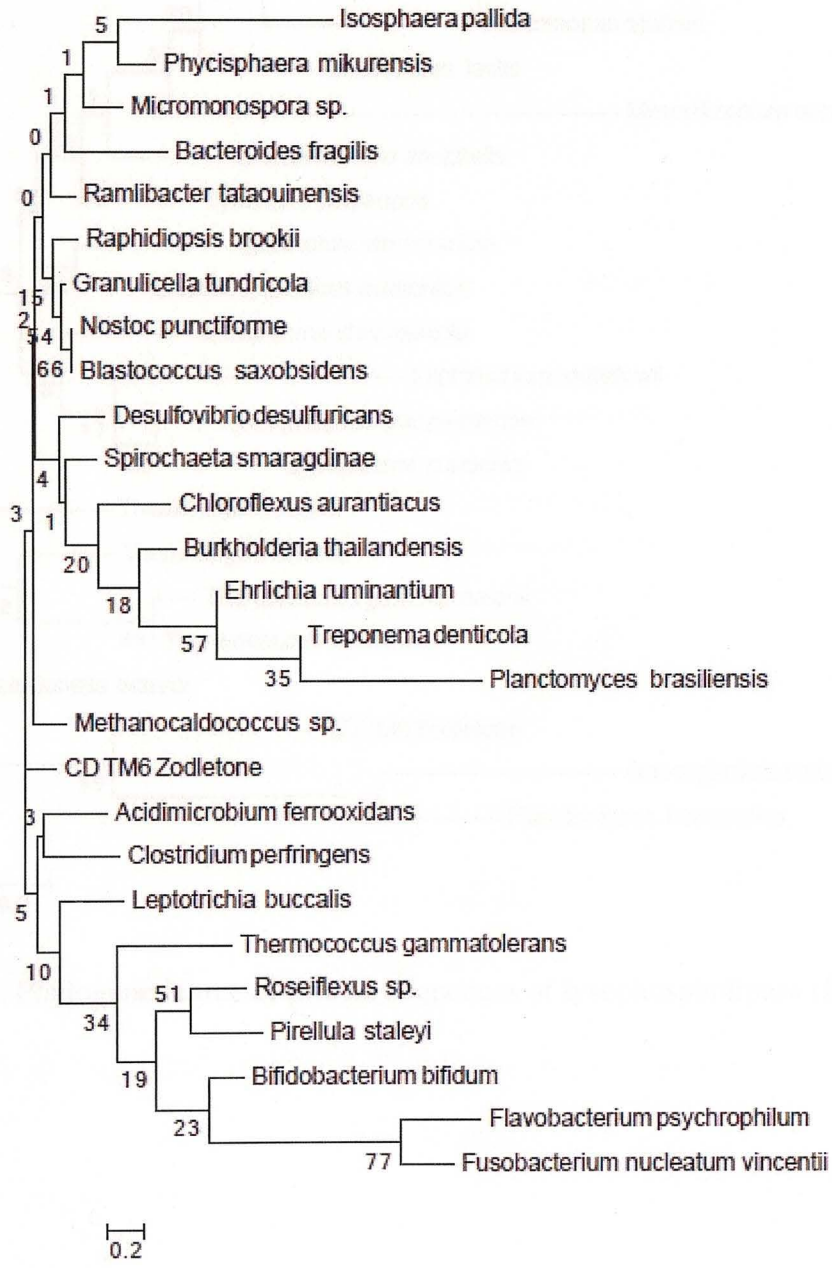

Figure 13. Phylogenetic tree of protein sequences of histidine kinase (EC 2.7.13.3).

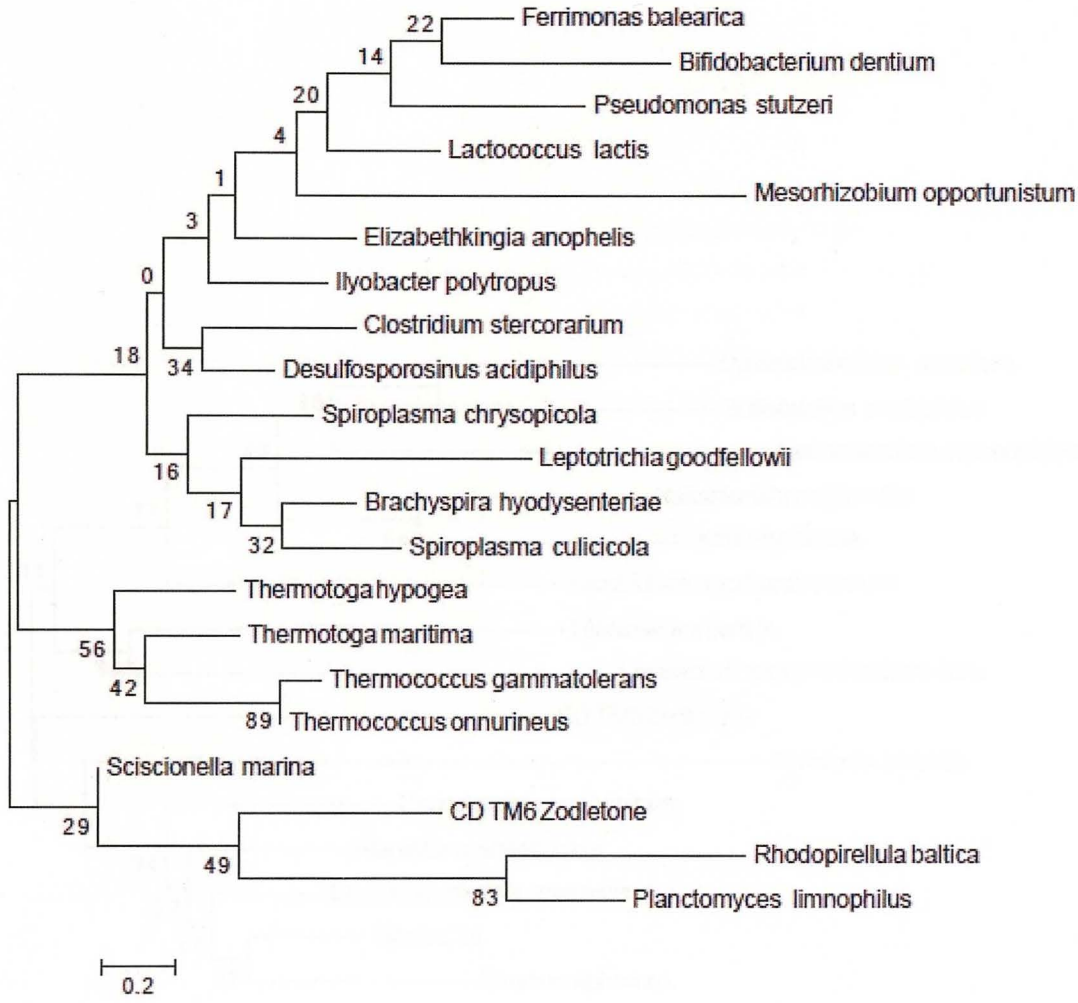

Figure 14. Phylogenetic tree of protein sequences of lysophospholipase (EC 3.1.1.5).

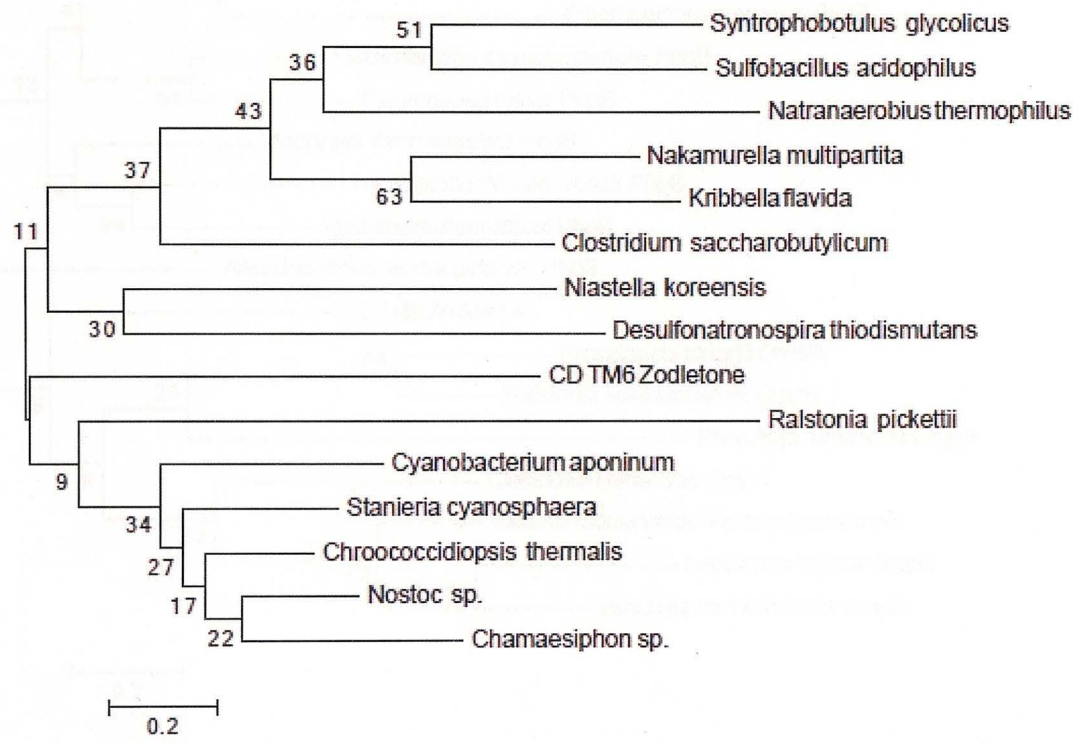

Figure 15. Phylogenetic tree of protein sequences of cyanophycin synthetase (EC 6.3.2.29).

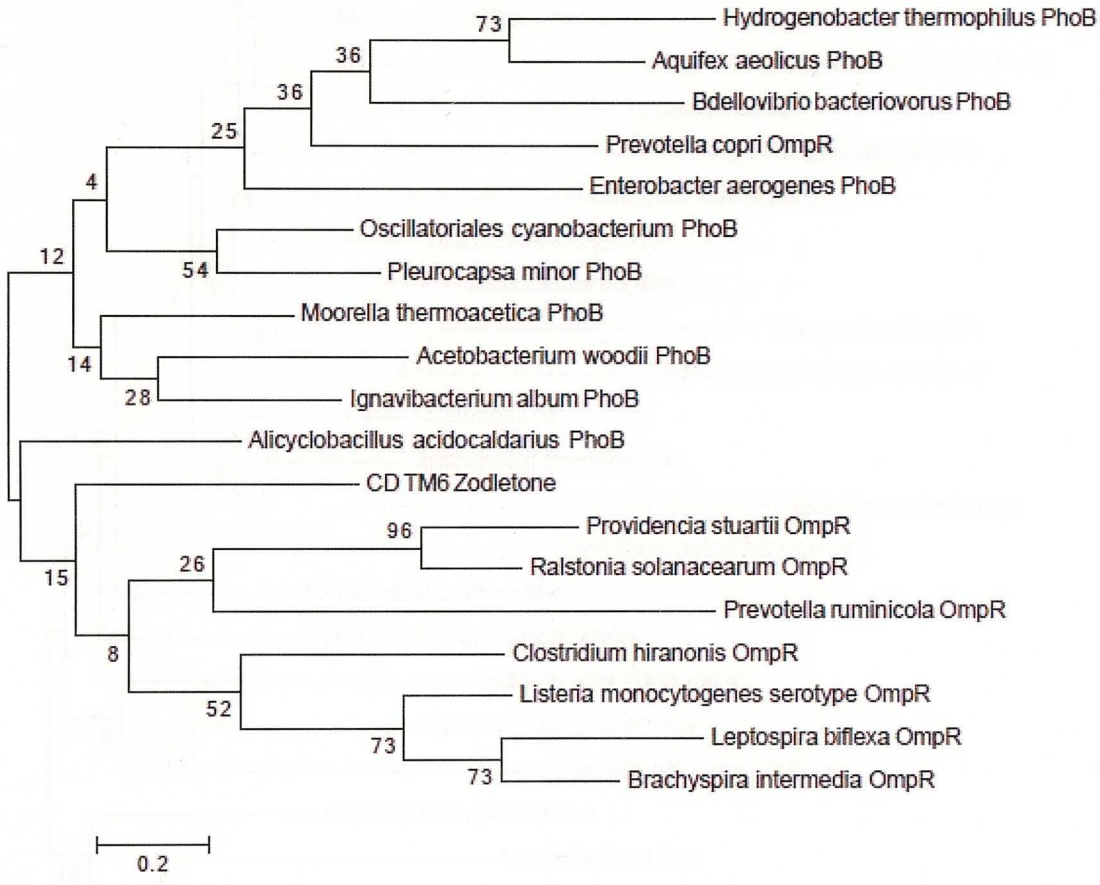

Figure 16. Phylogenetic tree of protein sequences of transcriptional regulator PhoB/CheY (EC 6.3.2.29).

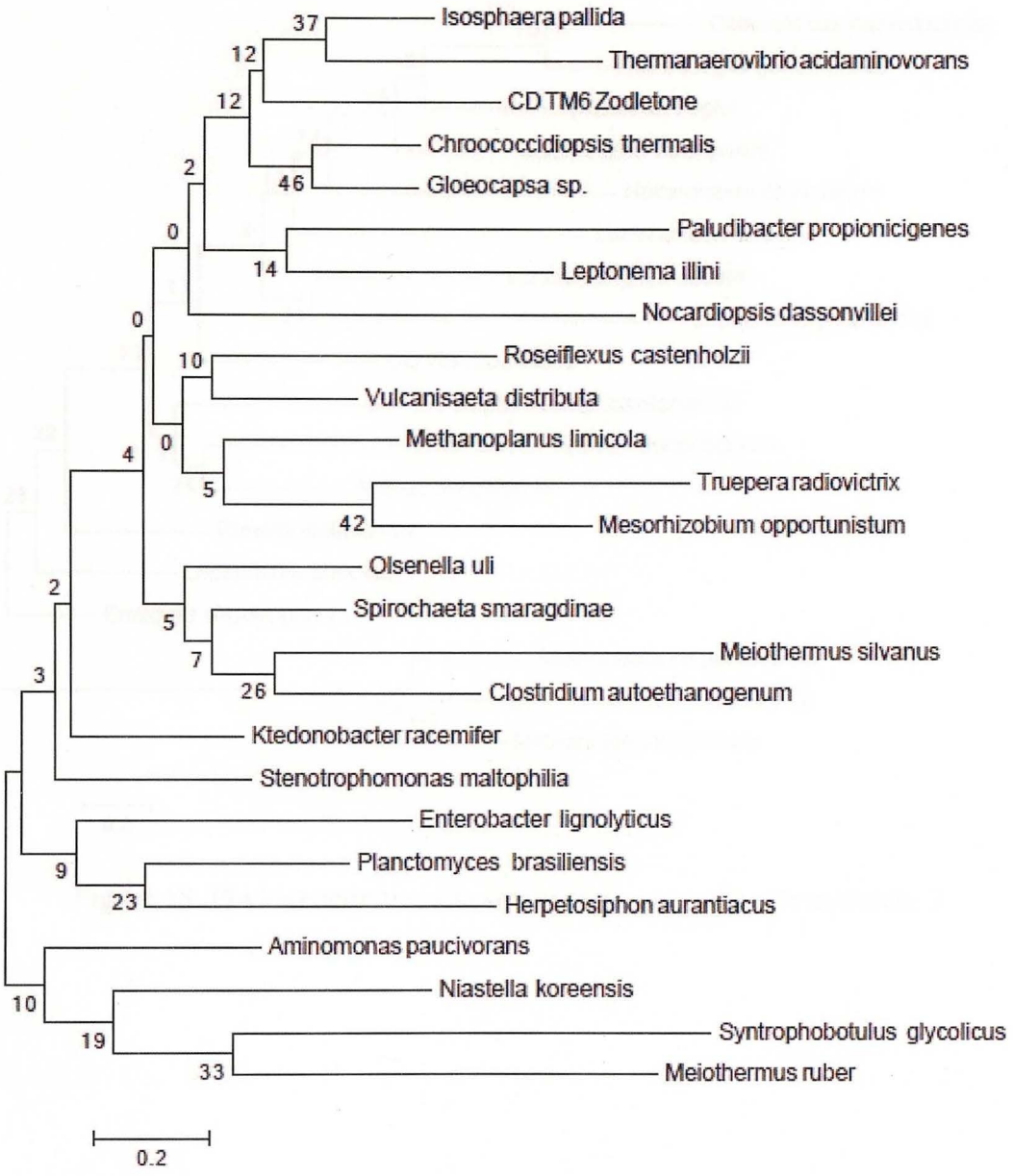

Figure 17. Phylogenetic tree of protein sequences of beta-lactamase (EC 3.5.2.6).

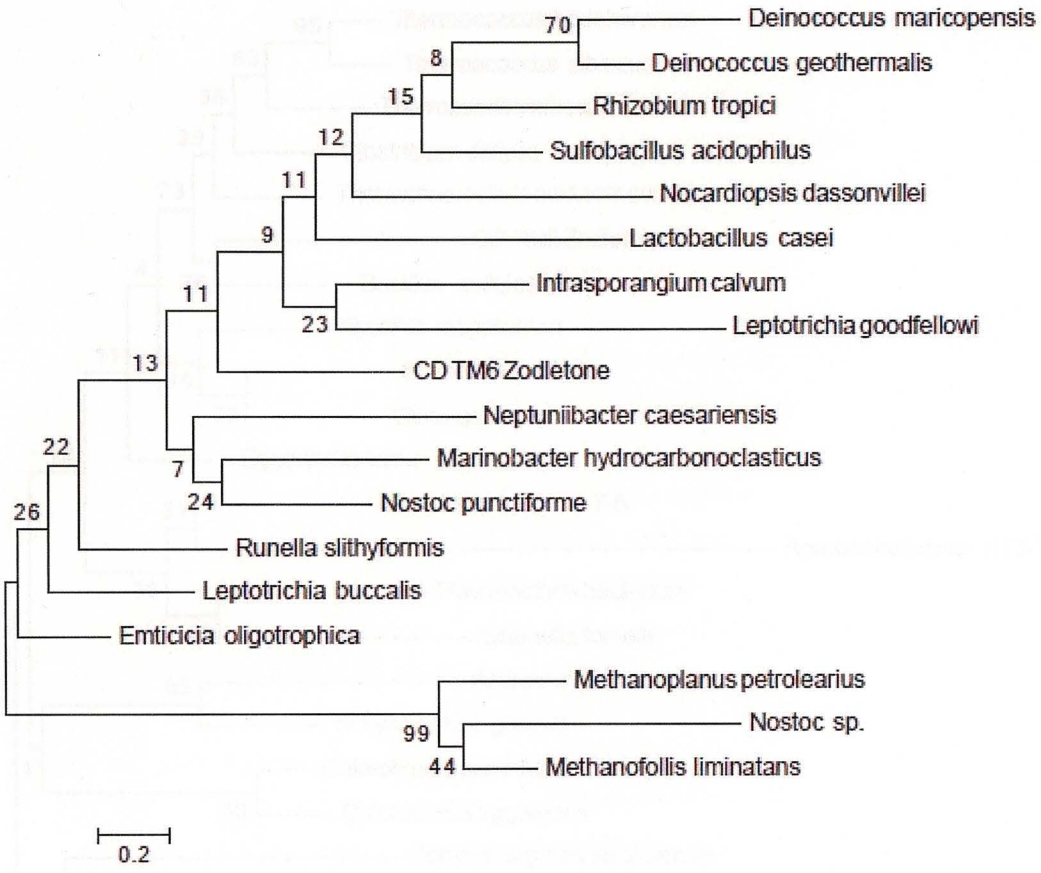

Figure 18. Phylogenetic tree of protein sequences of acyltransferase 3.

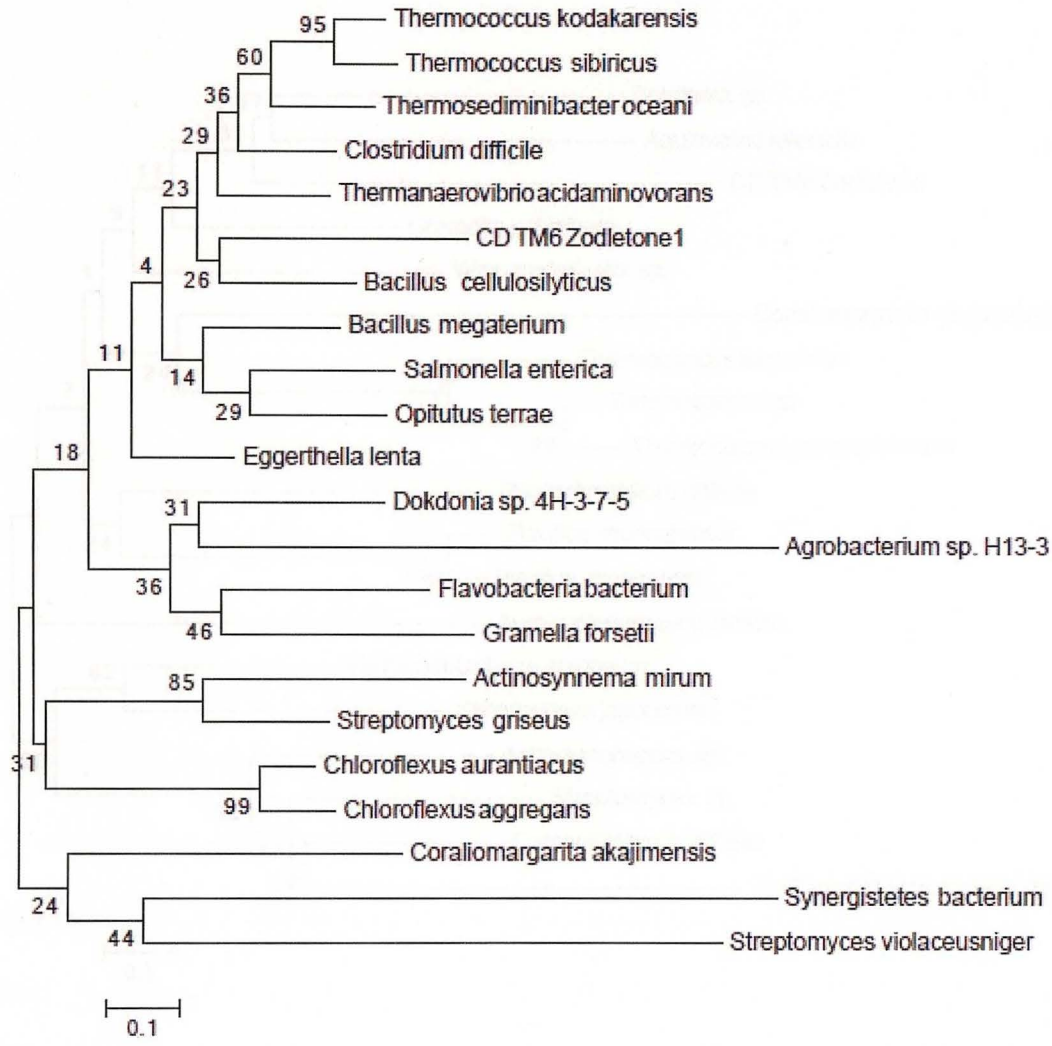

Figure 19. Phylogenetic tree of protein sequences of allophanate hydrolase subunit 1 (EC 3.5.1.54).

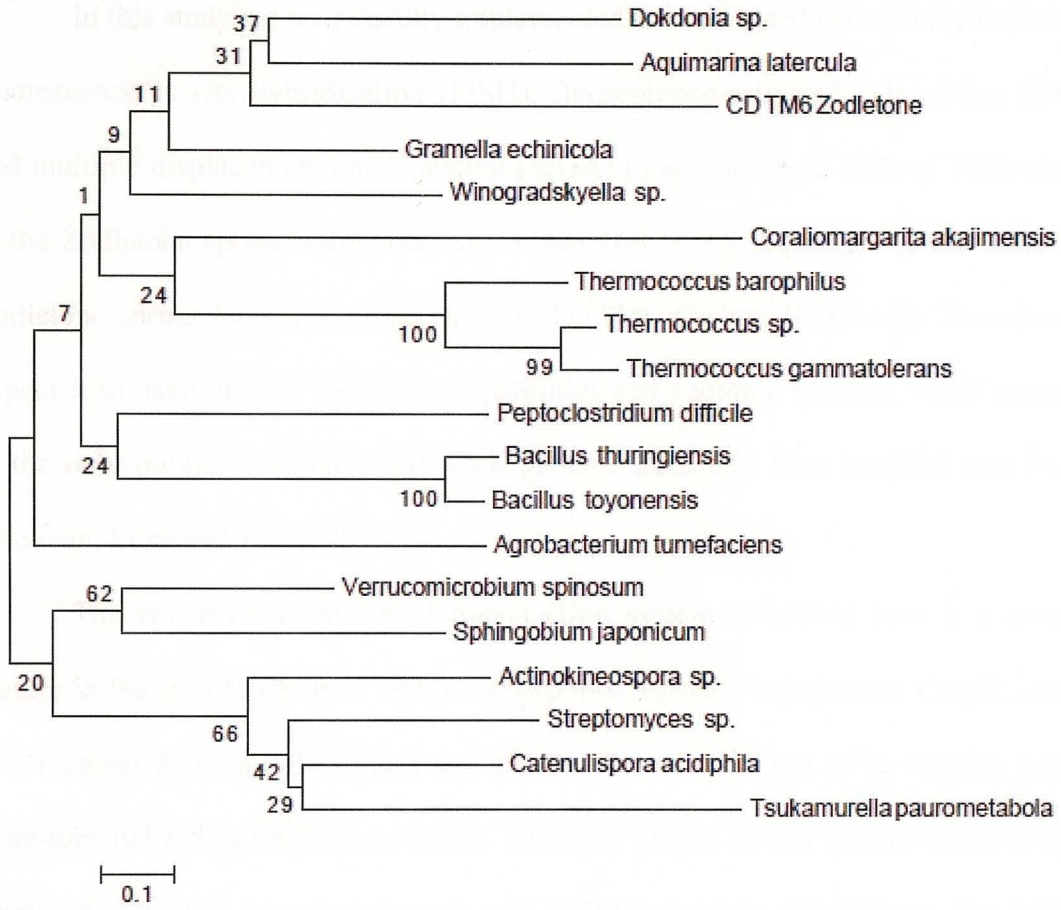

Figure 20. Phylogenetic tree of protein sequences of allophanate hydrolase subunit 2 (EC 3.5.1.54).

DISCUSSION

In this study we successfully implemented the combined use of three techniques: fluorescence *in situ* hybridization (FISH), fluorescence-activated cell sorting (FACS) and multiple displacement amplification (MDA) to access single cells of TM6 phylum in the Zodletone spring sediments. Assembled 16S rRNA sequences of CD TM6 from Zodletone spring form a separate class in the TM6 phylum (Figure 7). Therefore, we expect it to have distinct metabolic capabilities and genomic features, when compared to the only publicly available CD TM6 genome assembly from hospital sink biofilm (McLean, Lombardo et al. 2013).

The two-component signal transduction system identified here is a common feature in Bacteria (Parkinson 1993). Its response regulator component, OmpR, consists of a receiver domain (CheY-like) and effector domain that has DNA-binding winged-helix-turn-helix domain(Kenney 2002). Histidine kinase of this system consists of His Kinase A (phospho-acceptor) domain and ATP binding site (His Kinase-like ATPase) (Pirrung 1999). Together, this system alters the regulation of transcription based on the environmental influences on the sensor membrane protein.

The structure of individual domains is highly conserved for various signal transduction pathways (Laub and Goulian 2007). OmpR regulators often play a role in the transcriptional regulation of osmolarity, where environmental osmolarity defines level of transcription of porin genes (Slauch and Silhavy 1989). However, lack of *ompF* and *ompC* genes encoding porin proteins makes an assignment to osmotic function difficult. OmpR was also shown to regulate flagellar operon *flhDC*, curli fimbriae operon *csgDEFG* and the fatty acid receptor *fadL* (Feng, Oropeza et al. 2003). The

presence of a CheY-like receiver domain in OmpR regulator suggests a possible role in the chemotaxis-signal transduction and regulation of expression of chemotaxis-related proteins (CheA and CheB) (Appleby and Bourret 1998), but we have not assembled any of them in CD TM6 contigs.

There are two possible roles for the lysophospholipase (EC 3.1.1.5), depending on the location of the enzyme. If lysophospholipase is an intracellular enzyme, it can play role in the modification of the phospholipid bilayer, hydrolyzing its lysophospholipid constituent, while adjusting the membrane to respond to the stress conditions in the environment (Kingston, Subramanian et al. 2011). This lysophospholipid modification may in turn change protein-lipid interaction. An example is in proteins of mechanosensitive channels (MscL and MscS), where a change can influence the conductivity of these pore structures (Nomura, Cranfield et al. 2012). Lysophospholipases that have extracellular activity, play a major role in protection of a cell from lysophosphatidylcholine that cause lysis of bacterial cells (Weltzien 1979). For this purpose, CD TM6 would need to possess a suitable protein secretion apparatus.

Cyanophycin synthase (EC 6.3.2.29) is found in all members of Cyanobacteria phylum, where it catalyzes biosynthesis of a cyanophycin (a polymer that consists of aspartate and arginine residues) that is used as storage of nitrogen, carbon and energy (Li, Sherman et al. 2001) and in some non-cyanobacteria, like *Nitrosomonas europaea*, *Clostridium botulinum* and *Bordetella bronhiseptica*. In non-cyanobacteria, cyanophycin is accumulated in great amount during the stationary phase under phosphate-limited conditions and serves as nitrogen storage (Krehenbrink, Oppermann-Sanio et al. 2002). Therefore, it is possible that CD TM6 is using this insoluble polymer

as storage for carbon or nitrogen, or both. Annotated allophanate hydrolase (EC 3.5.1.54) converts allophanate (urea-1-carboxylate) to ammonium and carbon dioxide (Fan, Li et al. 2013). Since there is a source of arginine in cell in form of arginine-aspartate polymer cyanophycin, it is possible that allophanate hydrolase is involved in the degradation of cyanophycin. After cyanophycinases degrade cyanophycin polymer into equal amounts of aspartate and arginine, arginine can be transformed into urea by arginase (Allen, Hutchison et al. 1980). Later, urea can be utilized by activity of allophanate hydrolase.

The RND multidrug efflux transporters have an extremely broad range of functions. They are involved in heavy-metal resistance (Silver and Phung 1996; Paulsen 2003), transport of amino acids and other useful compounds (Paulsen 2003), extrusion of organic solvents (Ramos, Duque et al. 2002) (since Zodletone spring was reported to have elevated concentrations of hydrocarbons and hydrocarbon-related compounds), transport of antimicrobials (for intermicrobial competition) (Martinez, Sanchez et al. 2009), detoxification from probable endogenous toxic compounds (Neyfakh 1997), proton translocation (Lewinson and Bibi 2001), and transport of the quorum-sensing molecules (Pearson, Van Delden et al. 1999).

A question that has risen from previous study of McLean et al. (McLean, Lombardo et al. 2013) is whether CD TM6 is a free-living organism or host-associated. We observed that during epifluorescence microscopy (FISH of sediments with TM6-487 probe) fluorescent CD TM6 cells were integrated into another object (Figure 3). It is possible that the cells may be living within a host or that the cell may be incorporated into a sediment particle.

A hypothesis of protozoa-associated symbiosis, suggested by McLean and colleagues, may be possible. Some of the specific functions of annotated proteins from our assembly can be related to protozoa-CD TM6 interaction.

Presence of allophanate hydrolase enzyme in CD TM6 from Zodletone would explain how its host protozoa is dealing with waste metabolites, such as uric acid (Lawrie 1935). Usually, protozoa use contractile vacuoles to deal with excessive metabolites and regulate osmolality inside the cell; however, it doesn't mean that this function is restricted to vacuoles. If CD TM6 is indeed taking part in the utilization of uric acid from protozoa host, then CD TM6 should have additional enzymes for conversion of uric acid to allophanate. This will include urate oxidase, to convert uric acid to allantoin, and a complete pathway of conversion of allantoin to allophanate (Chisholm and Cooper 1982). Produced ammonia can be used by CD TM6 for its own needs.

The RND multidrug efflux transporter system could play a role not only in the resistance to a wide range of antimicrobials and above mentioned functions, but can also be connected with the colonization and survival of bacteria in the host organism (Piddock 2006).

Protozoa were widely found to be associated with pathogenicity and in 90% of all cases this pathogenicity is due to the symbiotic bacterial cell inside the host, which acts as an organelle (Gortz and Brigge 1998). We do not know, if CD TM6 is a pathogenic bacterium, but identified lysophospholipase may be excreted to digest membrane of the possible host, and play important role in the invasion. Also, lysophospholipase can be used to defend bacterial cell against eukaryotic detergents,

such as lysophosphatidylcholine, like it was described for members of *Legionella* species that are in natural association with a free-living amoeba (Flieger, Gong et al. 2000).

Host-associated nature of CD TM6 cells can explain its presence in environments with extreme conditions – hypersaline microbial mats (Tkavc, Gostincar et al. 2011) and hyper acidic cave biofilm (Macalady, Jones et al. 2007). On the other hand, there is still a possibility that this is a free-living bacterium, that has a well-organized stress-response machinery allowing it to live in a range of diverse environments. For example, annotated beta-lactamase and RND multidrug efflux transporter are widespread in many free-living organisms, allowing them to compete for the ecological niche in the environmental (Nikaido 1998; Fisher, Meroueh et al. 2005; Allen, Donato et al. 2010). The fact that CD TM6 hasn't been found in abundance in any of the occupied environments can be due to its oligotrophic specialist lifestyle or the unsuitable conditions for its proliferation – low concentration of sulfate (like in Aarhus bay, where addition of sulfate made CD TM6 detectable (Webster, Sass et al. 2011), low concentration of nutrients (glucose, fructose, rhamnose), or seasonal changes in the chemical composition of the studied environments (like in the soil of Canadian pine forests, where seasonal changes lead to the proliferation of previously not detected microorganisms (Axelrood, Chow et al. 2002).

Overall, it is still unclear, what is the nature of CD TM6 biochemistry and the question on its host-associated habitat remains unanswered. Use of non-cultivation approach to access members of the rare biosphere is still powerful enough, but needs certain adjustments to allow isolation of pure cells of bacterium of interest. The

contamination issue remains unresolved and only use of proper binning techniques can result in a clean assembly.

REFERENCES

Albertsen, M., P. Hugenholtz, et al. (2013). "Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes." Nat Biotechnol **31**(6): 533-538.

Allen, H. K., J. Donato, et al. (2010). "Call of the wild: antibiotic resistance genes in natural environments." Nat Rev Microbiol **8**(4): 251-259.

Allen, M. M., F. Hutchison, et al. (1980). "Cyanophycin granule polypeptide formation and degradation in the cyanobacterium Aphanocapsa 6308." J Bacteriol **141**(2): 687-693.

Appleby, J. L. and R. B. Bourret (1998). "Proposed signal transduction role for conserved CheY residue Thr87, a member of the response regulator active-site quintet." J Bacteriol **180**(14): 3563-3569.

Axelrood, P. E., M. L. Chow, et al. (2002). "Molecular characterization of bacterial diversity from British Columbia forest soils subjected to disturbance." Can J Microbiol **48**(7): 655-674.

Aziz, R. K., D. Bartels, et al. (2008). "The RAST Server: rapid annotations using subsystems technology." BMC Genomics **9**: 75.

Bankevich, A., S. Nurk, et al. (2012). "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." J Comput Biol **19**(5): 455-477.

Boisvert, S., F. Laviolette, et al. (2010). "Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies." J Comput Biol **17**(11): 1519-1533.

Brinig, M. M., P. W. Lepp, et al. (2003). "Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease." Appl Environ Microbiol **69**(3): 1687-1694.

Chisholm, G. and T. G. Cooper (1982). "Isolation and characterization of mutants that produce the allantoin-degrading enzymes constitutively in Saccharomyces cerevisiae." Mol Cell Biol **2**(9): 1088-1095.

Chow, M. L., C. C. Radomski, et al. (2002). "Molecular characterization of bacterial diversity in Lodgepole pine (Pinus contorta) rhizosphere soils from British Columbia forest soils differing in disturbance and geographic source." FEMS Microbiol Ecol **42**(3): 347-357.

Davis, J. P., N. H. Youssef, et al. (2009). "Assessment of the diversity, abundance, and ecological distribution of members of candidate division SR1 reveals a high

level of phylogenetic diversity but limited morphotypic diversity." Appl Environ Microbiol **75**(12): 4139-4148.

Elshahed, M. S., F. Z. Najar, et al. (2005). "Metagenomic analysis of the microbial community at Zodletone Spring (Oklahoma): insights into the genome of a member of the novel candidate division OD1." Appl Environ Microbiol **71**(11): 7598-7602.

Elshahed, M. S., J. M. Senko, et al. (2003). "Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring." Appl Environ Microbiol **69**(9): 5609-5621.

Fan, C., Z. Li, et al. (2013). "Structure and function of allophanate hydrolase." J Biol Chem **288**(29): 21422-21432.

Feng, X., R. Oropeza, et al. (2003). "Dual regulation by phospho-OmpR of ssrA/B gene expression in Salmonella pathogenicity island 2." Mol Microbiol **48**(4): 1131-1143.

Fisher, J. F., S. O. Meroueh, et al. (2005). "Bacterial resistance to beta-lactam antibiotics: compelling opportunism, compelling opportunity." Chem Rev **105**(2): 395-424.

Flieger, A., S. Gong, et al. (2000). "Novel phospholipase A activity secreted by Legionella species." J Bacteriol **182**(5): 1321-1327.

Gortz, H. D. and T. Brigge (1998). "Intracellular bacteria in protozoa." Naturwissenschaften **85**(8): 359-368.

Gu, S., D. Chen, et al. (2013). "Bacterial community mapping of the mouse gastrointestinal tract." PLoS One **8**(10): e74957.

Kantor, R. S., K. C. Wrighton, et al. (2013). "Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla." MBio **4**(5): e00708-00713.

Kenney, L. J. (2002). "Structure/function relationships in OmpR and other winged-helix transcription factors." Curr Opin Microbiol **5**(2): 135-141.

Kingston, A. W., C. Subramanian, et al. (2011). "A sigmaW-dependent stress response in Bacillus subtilis that reduces membrane fluidity." Mol Microbiol **81**(1): 69-79.

Krehenbrink, M., F. B. Oppermann-Sanio, et al. (2002). "Evaluation of non-cyanobacterial genome sequences for occurrence of genes encoding proteins homologous to cyanophycin synthetase and cloning of an active cyanophycin

synthetase from Acinetobacter sp. strain DSM 587." Arch Microbiol **177**(5): 371-380.

Laub, M. T. and M. Goulian (2007). "Specificity in two-component signal transduction pathways." Annu Rev Genet **41**: 121-145.

Lawrie, N. R. (1935). "Studies in the metabolism of protozoa: The nitrogenous metabolism and respiration of Bodo caudatus." Biochem J **29**(3): 588-598.

Lewinson, O. and E. Bibi (2001). "Evidence for simultaneous binding of dissimilar substrates by the Escherichia coli multidrug transporter MdfA." Biochemistry **40**(42): 12612-12618.

Li, H., D. M. Sherman, et al. (2001). "Pattern of cyanophycin accumulation in nitrogen-fixing and non-nitrogen-fixing cyanobacteria." Arch Microbiol **176**(1-2): 9-18.

Li, Y., F. Li, 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 (approximately 13 degrees N)." Extremophiles **12**(4): 573-585.

Ludwig, W., O. Strunk, et al. (2004). "ARB: a software environment for sequence data." Nucleic Acids Res **32**(4): 1363-1371.

Luo, C., R. L. Rodriguez, et al. (2014). "MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences." Nucleic Acids Res **42**(8): e73.

Macalady, J. L., D. S. Jones, et al. (2007). "Extremely acidic, pendulous cave wall biofilms from the Frasassi cave system, Italy." Environ Microbiol **9**(6): 1402-1414.

Marcy, Y., C. Ouverney, et al. (2007). "Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth." Proc Natl Acad Sci U S A **104**(29): 11889-11894.

Martinez-Checa, F., E. Quesada, et al. (2005). "Palleronia marisminoris gen. nov., sp. nov., a moderately halophilic, exopolysaccharide-producing bacterium belonging to the 'Alphaproteobacteria', isolated from a saline soil." Int J Syst Evol Microbiol **55**(Pt 6): 2525-2530.

Martinez, J. L., M. B. Sanchez, et al. (2009). "Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems." FEMS Microbiol Rev **33**(2): 430-449.

McInerney, M. J., L. Rohlin, et al. (2007). "The genome of Syntrophus aciditrophicus: life at the thermodynamic limit of microbial growth." Proc Natl Acad Sci U S A **104**(18): 7600-7605.

McLean, J. S., M. J. Lombardo, et al. (2013). "Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum." Proc Natl Acad Sci U S A **110**(26): E2390-2399.

Montalvo, N. F. and R. T. Hill (2011). "Sponge-associated bacteria are strictly maintained in two closely related but geographically distant sponge hosts." Appl Environ Microbiol **77**(20): 7207-7216.

Moynihan, P. J. and A. J. Clarke (2010). "O-acetylation of peptidoglycan in gram-negative bacteria: identification and characterization of peptidoglycan O-acetyltransferase in Neisseria gonorrhoeae." J Biol Chem **285**(17): 13264-13273.

Neyfakh, A. A. (1997). "Natural functions of bacterial multidrug transporters." Trends Microbiol **5**(8): 309-313.

Nikaido, H. (1998). "Multiple antibiotic resistance and efflux." Curr Opin Microbiol **1**(5): 516-523.

Noble, P. A., R. W. Citek, et al. (1998). "Tetranucleotide frequencies in microbial genomes." Electrophoresis **19**(4): 528-535.

Nomura, T., C. G. Cranfield, et al. (2012). "Differential effects of lipids and lyso-lipids on the mechanosensitivity of the mechanosensitive channels MscL and MscS." Proc Natl Acad Sci U S A **109**(22): 8770-8775.

Parkinson, J. S. (1993). "Signal transduction schemes of bacteria." Cell **73**(5): 857-871.

Paulsen, I. T. (2003). "Multidrug efflux pumps and resistance: regulation and evolution." Curr Opin Microbiol **6**(5): 446-451.

Pearson, J. P., C. Van Delden, et al. (1999). "Active efflux and diffusion are involved in transport of Pseudomonas aeruginosa cell-to-cell signals." J Bacteriol **181**(4): 1203-1210.

Pester, M., N. Bittner, et al. (2010). "A 'rare biosphere' microorganism contributes to sulfate reduction in a peatland." ISME J **4**(12): 1591-1602.

Piddock, L. J. (2006). "Multidrug-resistance efflux pumps - not just for resistance." Nat Rev Microbiol **4**(8): 629-636.

Pirrung, M. C. (1999). "Histidine kinases and two-component signal transduction systems." Chem Biol **6**(6): R167-175.

Pruesse, E., J. Peplies, et al. (2012). "SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes." Bioinformatics **28**(14): 1823-1829.

Pruesse, E., J. Peplies, et al. (2012). "SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes." Bioinformatics **28**(14): 1823-1829.

Ramos, J. L., E. Duque, et al. (2002). "Mechanisms of solvent tolerance in gram-negative bacteria." Annu Rev Microbiol **56**: 743-768.

Rappe, M. S. and S. J. Giovannoni (2003). "The uncultured microbial majority." Annu Rev Microbiol **57**: 369-394.

Reiswig M. H., H. I. Browman (1987). "Use of membrane filters for microscopic preparations of sponge spicules." Trans Am Microsc Soc **106**(1): 10-20.

Schramm, A., B. M. Fuchs, et al. (2002). "Fluorescence in situ hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries." Environ Microbiol **4**(11): 713-720.

Serkebaeva, Y. M., Y. Kim, et al. (2013). "Pyrosequencing-based assessment of the bacteria diversity in surface and subsurface peat layers of a northern wetland, with focus on poorly studied phyla and candidate divisions." PLoS One **8**(5): e63994.

Silver, S. and L. T. Phung (1996). "Bacterial heavy metal resistance: new surprises." Annu Rev Microbiol **50**: 753-789.

Slauch, J. M. and T. J. Silhavy (1989). "Genetic analysis of the switch that controls porin gene expression in Escherichia coli K-12." J Mol Biol **210**(2): 281-292.

Talavera, G. and J. Castresana (2007). "Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments." Syst Biol **56**(4): 564-577.

Tamura, K., G. Stecher, et al. (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." Mol Biol Evol **30**(12): 2725-2729.

Tkavc, R., C. Gostincar, et al. (2011). "Bacterial communities in the 'petola' microbial mat from the Secovlje salterns (Slovenia)." FEMS Microbiol Ecol **75**(1): 48-62.

Webster, G., H. Sass, et al. (2011). "Enrichment and cultivation of prokaryotes associated with the sulphate-methane transition zone of diffusion-controlled sediments of Aarhus Bay, Denmark, under heterotrophic conditions." FEMS Microbiol Ecol **77**(2): 248-263.

Weltzien, H. U. (1979). "Cytolytic and membrane-perturbing properties of lysophosphatidylcholine." Biochim Biophys Acta **559**(2-3): 259-287.

Wooley, J. C., A. Godzik, et al. (2010). "A primer on metagenomics." PLoS Comput Biol **6**(2): e1000667.

Youssef, N. H., M. B. Couger, et al. (2010). "Fine-scale bacterial beta diversity within a complex ecosystem (Zodletone Spring, OK, USA): the role of the rare biosphere." PLoS One **5**(8): e12414.