RESEARCH ARTICLE

Draft genome sequence of *Thauera* sp. DTG from a denitrifying quinoline degrading microbial consortium

Guojun Wu, Menghui Zhang, Chaochun Wei, Yun Wang, Xiangyu Yao, Liping Zhao and Xiaojun Zhang*

State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: The draft genome sequence of *Thauera* sp. DTG was reconstructed from a metagenome of a denitrifying quinoline degrading microbial consortium. The organism is most closely related to *Thauera aminoaromatic* S2 and *Thauera* sp. MZ1T and is a facultative anaerobe. It is predicted to take the central role for quinoline denitrifying degradation.

Keywords: genome sequence, denitrifying, quinoline

*Correspondence to: Xiaojun Zhang, State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University, Shanghai, 200240, China ; Email: xjzhang68@sjtu.edu.cn

Received: November 25, 2015; **Accepted:** March 8, 2016; **Published Online:** April 8, 2016

Citation: Wu G, Zhang M, Wei C, *et al.* 2016, Draft genome sequence of *Thauera* sp. DTG from a denitrifying quinoline degrading microbial consortium. *Applied Environmental and Biotechnology*, vol.1(1): 38–43. http://dx.doi.org/10.26789/AEB.2016.01.012.

1. Introduction

Quinoline is a heterocyclic aromatic organic compound with the chemical formula C_9H_7N. It is often reported as an environmental contaminant associated with facilities processing oil shale or coal, and has also been found at legacy wood treatment sites or in the factory of manufacturing other products, where it is used mainly as an intermediate. Owing to the solubility in water, it has significant potential for spread in the environment, which may aggravate water contamination.

An increased incidence of liver vascular tumors has been observed in rats and mice orally exposed to quinolone[1]. EPA has provisionally classified quinoline as a Group C, possible human carcinogen (http://www2.epa.gov/iris).

We monitored the performance of a coking wastewater treatment plant for years. The most of COD was removed from the treatment system, which used A2/O craft. But quinoline was reluctant to be degraded. Therefore, we setup a lab scale bioreactor to accumulate a high efficient quinoline degrading microbial community[2]. The bioreactor demonstrates a high efficiency for removing the quinoline from the synthetic wastewater under denitrifying condition at temperature ranging from 15°C to 25°C. The synthetic wastewater used quinoline as the sole carbon source and supplemented with mineral salt and nitrate.

The consortia in the quinoline degrading bioreactor was composed of diverse bacteria including these from family of Rhodocyclaceae, Desulfobacteraceae, Nocardiaceae, Syntrophaceae, Ignavibacteriaceae, Spirochaetaceae, Comamonadaceae, Flavobacteriaceae, Desulfomicrobiaceae, etc. Interestingly, the abundance of bacteria from the genus of *Thauera* (member of Rhodocyclaceae) could be as high as 45%. The high abundance of *Thauera* genus in the bioreactor implied its important role in the quinoline degradation process, as indicated in the previous work[2,3]. We constructed a phylogenetic tree using the 16S rRNA gene sequences retrieved from metagenome or direct sequen-

cing of 16S rRNA amplicon (Figure 1). The bacteria of this genus were diverse in the bioreactor, especially in the early stage of the construction of the bioreactor. Several attempts of isolation on the most abundant *Thauera* bacteria had failed. The only three *Thauera* isolates, strains Q4, Q20-C, and 3-35, nevertheless showed low abundance in the bioreactor community. Those three strains of isolates didn't show any quinoline degrading effect in the medium containing quinoline as sole carbon source[3]. However, our previous results showed that the quinoline degradation efficiency is significantly correlated with the abundance of *Thauera* bacteria in the consortium[4]. We speculated that the most predominant *Thauera* bacteria should play an important role in the quinoline degradation within the denitrifying bioreactor.

Figure 1 showed that the diversity of *Thauera* genus in the bioreactor was significantly lower after 5 years running. Only one OTU, which is homologue of the most abundant *Thauera* OTU detected in the early stage, was present in the bioreactor nevertheless in high abundance after undergoing long time running. To further understand the community, especially the role of the *Thauera* bacteria, the metagenome of the bioreactor was sequenced using pyrosequencing technology. We assembled a *Thauera* genome by using the metagenome data. The function predication of this genome was performed to understand the role of *Thauera* bacteria in the bioreactor.

The total sequencing reads of metagenome is 1,224,289, which account for 714,921,886 bp. After the quality control, 73.86% reads (904223 reads, 524,158,639 bp) left for subsequent analysis.

The contig set S2, which was derived from the reference-based assembly using *Thauera aminoaromatic* S2 as the reference, contained 2104 contigs and was ~2.89 Mbp in total. The high quality reads, which could not be mapped on the reference genome, were de novo assembled into contig set NS2. The NS2 set contained 19,198 contigs and was 49.87 Mbp in total.

We visualized the tetranucleotide frequency patterns of all contigs with more than 2 Kbp in both S2 and NS2 sets (Figure 2A and 2B) and used this map in combination with the information of contig coverage (Figure 2C, supplementary table 1) to do contig binning and subsequential filtering, thus reconstructing the draft genome of *Thauera* sp. DTG (see 'Methods' section for detail). The draft genome for DTG included 2,407 contigs (> 500 bp), The sum of assembled reads of these contigs was about 26,196,753 bp in total. The *N*50 of contig lengths is 2,456 bp. The contigs were reordered using *Thauera aminoaromatic* S2 as reference. The draft genome of *Thauera* sp. DTG contained 4,346,234 bp in total. Human Microbiome Project (HMP) constructed the core gene groups, which comprise single copy genes conversed among all sequenced genomes in Bacteria. Thus evaluating whether genes from the core groups are present in the draft genomes can give a good indication of the completeness of the draft genomes. Among all 66 bacterial core groups, 64 existed in the assembled draft genome, which indicated a good completeness of *Thauera* sp. DTG. The annotation was performed using Prokka, which predicted a total of 4751 protein-coding genes and 4 ribosomal RNAs.

The two most similar recorded genomes are *Thauera aminoaromatic* S2 and *Thauera* sp. MZ1T. The identities of assembled *Thauera* sp. DTG with these two genomes are 95.95% and 95.25% respectively. *Thauera aminoaromatic* S2 was also active bacteria from an aromatic wastewater treatment community[5]. *Thauera* sp. MZ1T was reported in 2004. It is a zoogloeal cluster-forming organism and was isolated from an industrial wastewater treatment system[6]. Its analogues were reported for aromatic compounds degradation.

The genome of *Thauera* sp. DTG contains relative higher number of genes related with environmental information processing. For instance, it contains complete PhoR-PhoB (phosphate starvation response) two-component regulatory system, EnvZ-OmpR (osmotic stress response) two-component regulatory system, CusS-CusR (copper tolerance) two-component regulatory system, NarX-NarL (nitrate respiration) two-component regulatory system, RegB-RegA (redox response) two-component regulatory system. The genes for aromatic compounds degradation also prevails in its genome as listed in Table 1. Genes for Benzoyl-CoA denitrifying degradation, including benzoyl-CoA reductase subunit C (K04112) and 6-oxocyclohex-1-ene-carbonyl-CoA hydrolase (K07539) for benzonate anaerobic degradation imply its anaerobic degrading capacity[7,8]. All denitrification genes were found in the DTG genome, including narG, narI, narJ, napB, nirS, norB, norC and nosZ gene. *Thauera* bacteria act as a denitrifier had been directly detected by metagenomic sequencing in the wastewater treatment environment[9]. Therefore, the predominant one of *Thauera* bacteria in the reactor has the potential for quinoline or other aromatic compounds

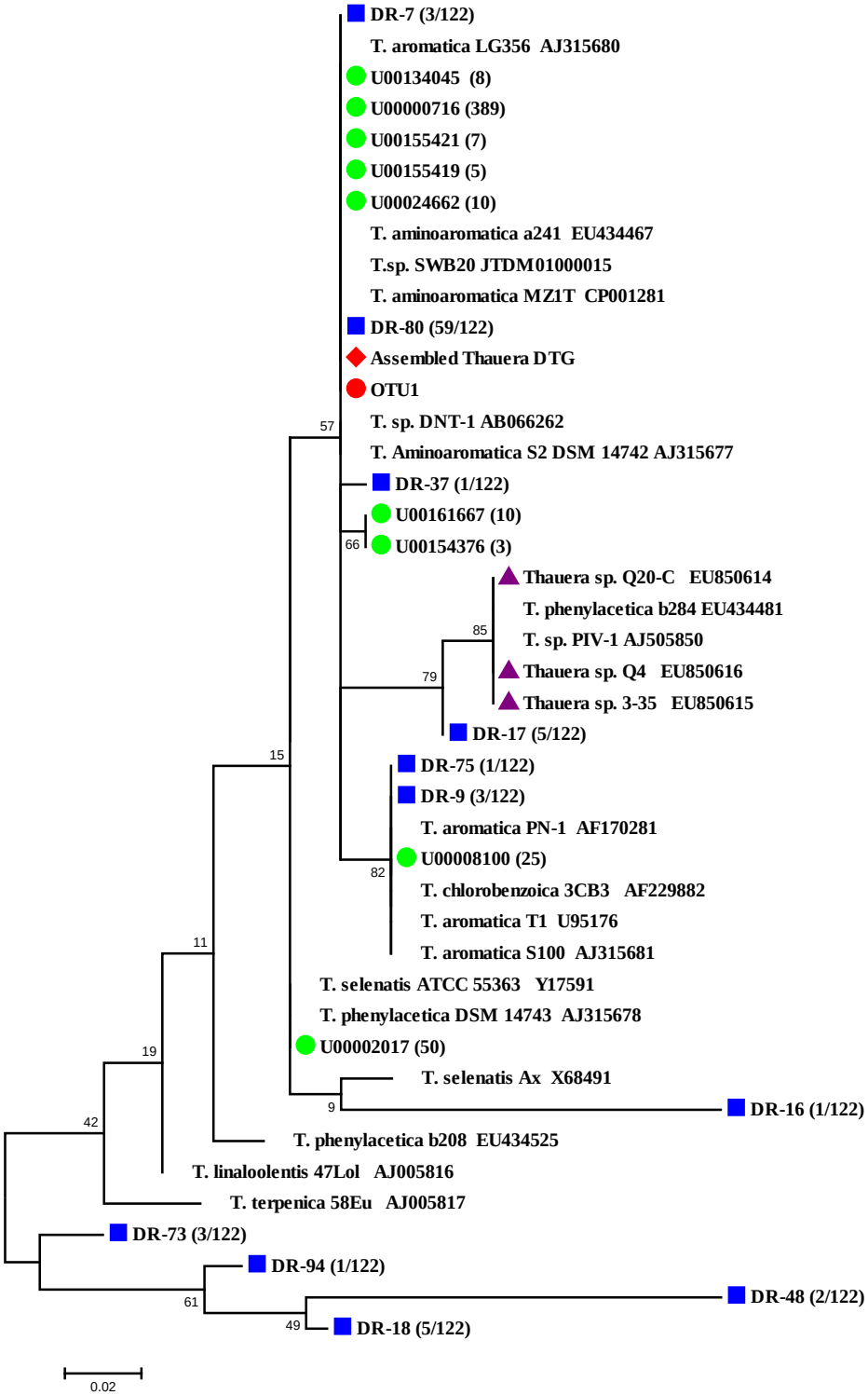

Figure 1. Neighbor-joining phylogenetic tree indicating major species in the *Thauera* dominated consortium and their closest GenBank matches. 16S rRNA genes from *Thauera* DTG draft genome showed 100% identity with *Thauera aminoaromatica* DSM14742 and sequence of *Thauera* clone DR80 (highest abundance in very early stage bioreactor sample), U00000716 (highest abundance in 3 years bioreactor sample), and also OTU1 (highest abundance in 5 years bioreactor sample). Blue solid square indicates sequences from clone library constructed using early stage bioreactor sample[4]; green diamond indicates 16S rRNA high throughput sequencing reads of sample from bioreactor running for 3 years; pink solid circle indicates 16S rRNA high throughput sequencing reads of sample from bioreactor running for 5 years; red diamond indicates the 16S rRNA gene sequences of assembled *Thauera* sp. DTG genome; purple triangle indicates 16S rRNA gene sequences of three isolates.

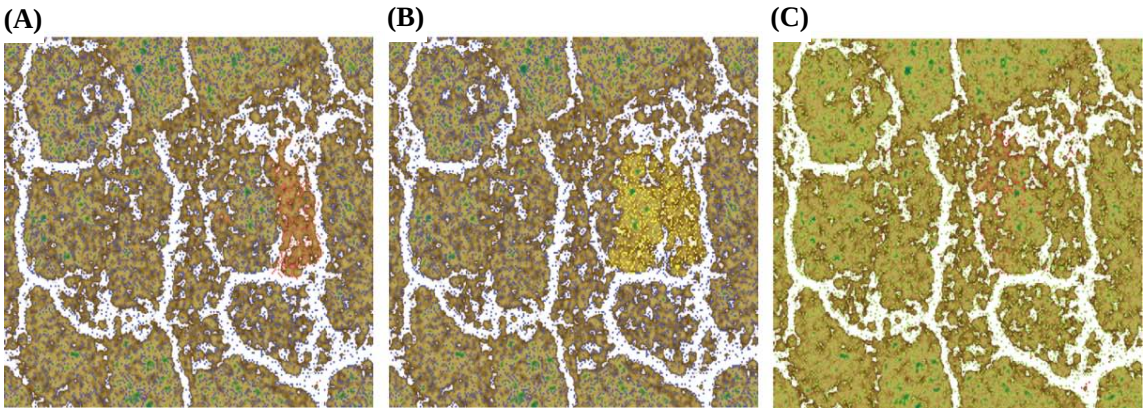

Figure 2. ESOM map of tetranucleotide frequency of contigs. The Gradient of the map is "earthcolor". Lighter (tend to white) indicate more difference in tetranucleotide patterns. The dots represent 2 Kbp contig regions and dot colors stand for different meanings: **(A)** purple for contigs in NS2 set and red for contigs in S2 set; **(B)** yellow for the contig bin containing S2 set and contigs in NS2 set with similar tetranucleotide frequency to S2; **(C)** green for contigs with coverage within the coverage range (2.59±7.76, median±2 s.d.) of S2 set and red for contigs exceeding that range.

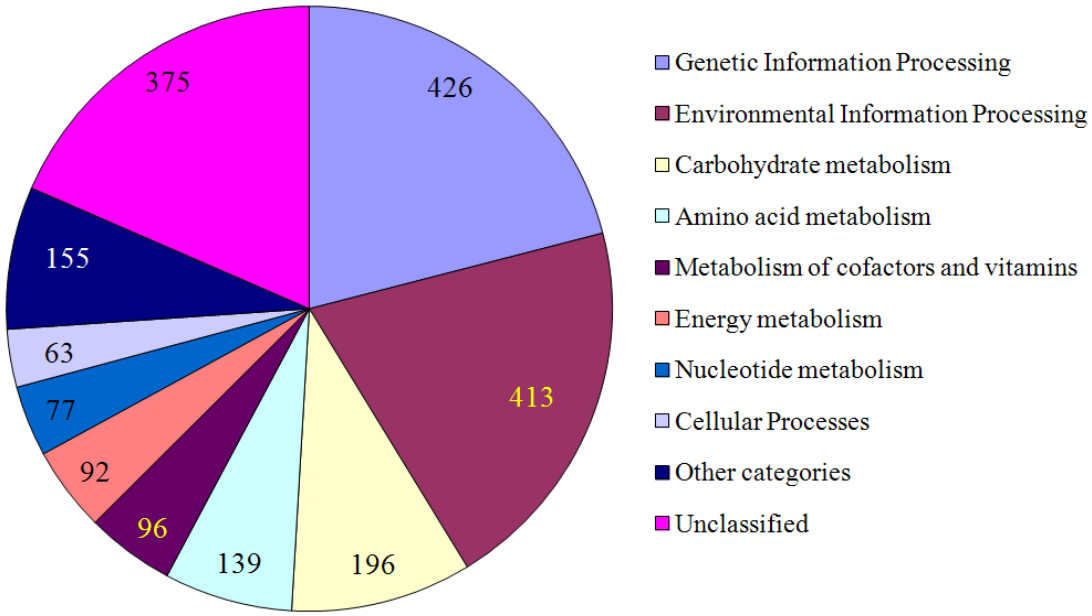

Figure 3. Annotated pathway of assembled *Thauera* sp. DTG genome based on KEGG database (42.7% of 4751 predicted genes were annotated)

denitrifying degradation. However, high proportion of genes haven't be annotated, the function and important of this bacteria in the bioreactor still need more investigation.

2. Methods

2.1 DNA Extraction and Sequencing

Biofilm sample was collected from the surface of supporting materials in the lab scale quinoline degrading bioreactor. The sample was then centrifuged at 10000 ×g for 10 minutes. The pellet was re-suspended in the buffer Z (10mM Tris-HCl, 150mM NaCl, pH 8.0). The 1.5 mL eppendoff tube with the sample was then performed repeated frozen-thaw treatment. In detail, −80°C for 10 minutes, then transfer to 60°C water bath for 2 minutes. After that, keep tube on ice for 1 minute. Sample was mixed on vortex two times in highest speed for 5 minutes. Between each vertex, the mixture was kept on ice for 5 minutes. Afterwards, the standard SDS lysis and Phenol/Chloroform extraction was used before ethanol precipitation. The DNA solution was treated with RNase to remove RNA.

The purified DNA sample was used for sequencing

Table 1. The KO of aromatic compounds degradation found in the assembled genome

KO ID	KO Function	Number of KO in DTG genome
K00141	xylC; benzaldehyde dehydrogenase (NAD)	1
K01055	pcaD; 3-oxoadipate enol-lactonase	1
K01617	dmpH, xylI, nahK; 2-oxo-3-hexenedioate decarboxylase	1
K04112	bcrC, badD; benzoyl-CoA reductase subunit C	1
K05549	benA-xylX; benzoate/toluate 1,2-dioxygenase alpha subunit	1
K05550	benB-xylY; benzoate/toluate 1,2-dioxygenase beta subunit	1
K05783	benD-xylL; dihydroxycyclohexadiene carboxylate dehydrogenase	1
K07539	oah; 6-oxocyclohex-1-ene-carbonyl-CoA hydrolase	1
K10217	dmpC, xylG; aminomuconate-semialdehyde /2-hydroxymuconate-6-semialdehyde dehydrogenase	1
K13953	adhP; alcohol dehydrogenase, propanol-preferring	1
K16246	dmpP, poxF; phenol hydroxylase P5 protein	1
K16249	dmpK, poxA; phenol hydroxylase P0 protein	1
K18364	bphH, xylJ, tesE; 2-oxopent-4-enoate/cis-2-oxohex-4-enoate hydratase	1
K18366	bphJ, xylQ, nahO, tesF; acetaldehyde/propanal dehydrogenase	2

library construction according to the guild for metagenomics from Roche brochure. Pyrosequencing was finished in Bohao Ltd. Company in Shanghai.

2.2 Data Quality Control

454 Replicate Filter[10] was used to de-duplicate the raw reads. Prinseq[11] was employed (a) to trim the reads from the 3′ end until reaching the first nucleotide with a quality threshold of 20; and (b) to remove read pairs if either read was shorter than 200 bp or contained 'N' bases.

2.3 The Assembly of *Thauera* sp. DTG Genome

Firstly, the high quality reads were assembled into contigs (contig set S2) by gsMapper V2.9 (Roche Inc.) using *Thauera aminoaromatic S2* (IMG: 19005) as the reference genome with parameters set at 50bp for overlap length, 95% for nucleotide identity. The contigs shorter than 500bp were discarded. The reads, which were not aligned to *Thauera aminoaromatic S2* at the previous stage, were de novo assembled into contigs (contig set NS2) by gsAssemby V2.9 (Roche Inc.) with the same parameters as applied in gsMapper. The contigs shorter than 500bp were discarded.

The tetranucleotide signatures of all contigs > 2 kb in the both two contigs sets were calculated following the ESOM analysis pipeline[12]. The ESOM topographic map was generated by ESOM-1.1 software[13] to cluster and visualize tetranucleotide frequency patterns with parameters set as K-batch for training algorithm, 200 for rows in map, 240 for columns in map, 40 for start value for radius and RobustZT for data normalization. The contigs in set NS2, which had similar tetranucleotide signatures to S2, were added into set S2. For a more accurate clustering, coverage information of all contigs was calculated (from the outputs of gsMapper and gsAssembly). The contigs, which were added into S2 from NS2 at previous stage, were discarded if their coverage exceed the coverage range (median ± 2 s.d.) of S2 set. The final S2 set was considered as the draft genome of *Thauera* sp. DTG.

Mauve aligner[14] was used to reorder the contigs of the *Thauera* sp. DTG draft genome. The annotation was performed using Prokka[15] and BlastKOALA (http://www.kegg.jp/blastkoala/). The "Bacterial Core Gene Evaluation" protocol of HMP (http://hmpdacc.org/resources/tools_protocols.php) was adapted to test the completeness of the draft genome of *Thauera* sp. DTG.

2.4 Nucleotide Sequence Accession Numbers

This whole-genome shotgun project including the metagenomic data and the draft genome of *Thauera* sp. DTG has been deposited at EBI under PRJEB11730. The 16S rRNA gene sequences obtained from the sequencing of bioreactor samples and used for constructing the *Thauera* phylogenetic tree has been deposited at European Nucleotide Archive under the accession numbers LN907802-LN907822.

Conflict of Interest and Funding

No conflict of interest was reported by all authors. This study was financially supported by the National

Natural Science Foundation of China (NSFC 21177086), Shanghai International Collaboration Program (1223-0706800).

References

1. Asakura S, Sawada S, Sugihara T, *et al.* 1997, Quinoline-induced chromosome aberrations and sister chromatid exchanges in rat liver. *Environmental and Molecular Mutagenesis*, vol.30(4): 459–467. http://dx.doi.org/10.1002/(SICI)1098-2280(1997)30:4<459::AID-EM11>3.0.CO;2-C.
2. Zhang X, Yue S, Zhong H, *et al.* 2011, A diverse bacterial community in an anoxic quinoline-degrading bioreactor determined by using pyrosequencing and clone library analysis. *Applied Microbiology and Biotechnology*, vol.91(2): 425–434. http://dx.doi.org/10.1007/s00253-011-3296-1.
3. Mao Y, Zhang X, Xia X, *et al.* 2010, Versatile aromatic compound-degrading capacity and microdiversity of *Thauera* strains isolated from a coking wastewater treatment bioreactor. *Journal of Industrial Microbiology & Biotechnology*, vol.37(9): 927–934. http://dx.doi.org/10.1007/s10295-010-0740-7.
4. Liu B, Zhang F, Feng X, *et al.* 2006, *Thauera* and *Azoarcus* as functionally important genera in a denitrifying quinoline-removal bioreactor as revealed by microbial community structure comparison. *FEMS Microbiology Ecology*, vol.55(2): 274–286. http://dx.doi.org/10.1111/j.1574-6941.2005.00033.x.
5. Mechichi T, Stackebrandt E, Gad'on N, *et al.* 2002, Phylogenetic and metabolic diversity of bacteria degrading aromatic compounds under denitrifying conditions, and description of *Thauera phenylacetica* sp. nov., *Thauera aminoaromatica* sp. nov., and *Azoarcus buckelii* sp. nov. *Archives of Microbiology*, vol.178(1): 26–35. http://dx.doi.org/10.1007/s00203-002-0422-6.
6. Allen M S, Welch K T, Prebyl B S, *et al.* 2004, Analysis and glycosyl composition of the exopolysaccharide isolated from the floc-forming wastewater bacterium *Thauera* sp. MZ1T. *Environmental Microbiology*, vol.6(8): 780–790. http://dx.doi.org/10.1111/j.1462-2920.2004.00615.x.
7. Boll M and Fuchs G, 1995, Benzoyl-coenzyme A reductase (dearomatizing), a key enzyme of anaerobic aromatic metabolism. ATP dependence of the reaction, purification and some properties of the enzyme from *Thauera aromatica* strain K172. *European Journal of Biochemistry*, vol.234: 921–933. http://dx.doi.org/10.1111/j.1432-1033.1995.921_a.x.
8. Laempe D, Jahn M and Fuchs G, 1999, 6-Hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenase and 6-oxocyclohex-1-ene-1-carbonyl-CoA hydrolase, enzymes of the benzoyl-CoA pathway of anaerobic aromatic metabolism in the denitrifying bacterium *Thauera aromatica*. *European Journal of Biochemistry*, vol. 263(2): 420–429. http://dx.doi.org/10.1046/j.1432-1327.1999.00504.x.
9. Mao Y, Xia Y, Wang Z, *et al.* 2014, Reconstructing a Thauera genome from a hydrogenotrophic-denitrifying consortium using metagenomic sequence data. *Applied Microbiology and Biotechnology*, vol.98(15):6885–6895. http://dx.doi.org/10.1007/s00253-014-5756-x.
10. Gomez-Alvarez V, Teal T K and Schmidt T M, 2009, Systematic artifacts in metagenomes from complex microbial communities. *The ISME Journal*, vol.3: 1314–1317. http://dx.doi.org/10.1038/ismej.2009.72.
11. Schmieder R and Edwards R, 2011, Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, vol.27(6): 863–864. http://dx.doi.org/10.1093/bioinformatics/btr026.
12. Dick G J, Andersson A F, Baker B J, *et al.* 2009, Community-wide analysis of microbial genome sequence signatures. *Genome Biology*, vol.10: R85. http://dx.doi.org/10.1186/gb-2009-10-8-r85.
13. Ultsch A and Mörchen F, 2005, *ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM.* Technical Report No. 46, Department of Mathematics and Computer Science, University of Marburg, Germany.
14. Rissman A I, Mau B, Biehl B S, *et al.* 2009, Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics*, vol.25(16): 2071–2073. http://dx.doi.org/10.1093/bioinformatics/btp356.
15. Seemann T, 2014, Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, vol.30(14): 2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.