

## Note (Genome Announcement)

# Complete genome sequence of *Betaproteobacteria* strain GR16-43 isolated from a freshwater pond in South Korea

Ahyoung Choi<sup>1\*</sup>, Kiwoon Baek<sup>2</sup>, Eu Jin Chung<sup>1</sup>, Jee-Hwan Kim<sup>3</sup>, and Gang-Guk Choi<sup>1</sup>

<sup>1</sup>Culture Techniques Research Division, <sup>2</sup>Bacterial Resources Research Division, <sup>3</sup>Bioresources Culture Collection Division, Nakdonggang National Institute of Biological Resources, Sangju 37242, Republic of Korea

## 담수에서 분리한 *Betaproteobacteria* GR16-43의 유전체 염기서열 분석

최아영<sup>1\*</sup> · 백기운<sup>2</sup> · 정유진<sup>1</sup> · 김지환<sup>3</sup> · 최강국<sup>1</sup>

<sup>1</sup>국립낙동강생물자원관 배양기술개발부, <sup>2</sup>국립낙동강생물자원관 원핵생물조사연구부, <sup>3</sup>국립낙동강생물자원관 균주보존분양부

(Received September 6, 2017; Revised September 18, 2017; Accepted September 19, 2017)

A betaproteobacterium strain GR16-43 was isolated from a surface layer of the Geomnyong Pond in Republic of Korea by a dilution-to-extinction culturing method. We report the whole genome sequence of the strain GR16-43, which contains 4,806,848 bp with a G + C content 67.12%, and to include 4,424 protein-coding genes and 47 transfer RNA genes. The genome was determined to contain the genes encoding carbon monoxide dehydrogenase, nitrate reductase, nitrite reductase, nitric oxide reductase, and the sulfur oxidation (*sox*) gene cluster, highlighting the potential importance of the bacterial group represented by the strain in the cycling of inorganic elements. These results indicate that strain GR16-43 genome showed several traits indicating adaptation of the bacteria to living in freshwater environments.

**Keywords:** Betaproteobacterium, complete genome, freshwater

Freshwater environments contain unique bacterial populations that have little overlap with those in other habitats, such as soil and marine environments. The prevalent taxa detected are *Betaproteobacteria*, *Actinobacteria*, *Alphaproteobacteria*, *Bacteroidetes*, *Cyanobacteria*, and *Verrucomicrobia* (Zwart *et al.*, 2002; Newton *et al.*, 2011). Typically, representatives of the

class *Betaproteobacteria* are the most dominant bacterial taxa detected in freshwater (Van der Gucht *et al.*, 2005). Members of *Betaproteobacteria* can utilize sulfate, sulfite, thiosulfate, sulfur, nitrate, or some other inorganic compounds. Here, we report the complete genome sequence of betaproteobacterium GR16-43 isolated from the Geomnyong Pond in Republic of Korea.

Strain GR16-43 was isolated from a surface water sample collected from the Geomnyong Pond in Republic of Korea by using a high-throughput dilution-to-extinction culturing method. A representative genomic sequence of the 16S ribosomal RNA (rRNA) gene of strain GR16-43 was compared with those of other members of the class *Betaproteobacteria* using the EzTaxon-e server (<http://eztaxon-e.ezbiocloud.net/>). The comparative sequence analysis revealed that among cultured isolates, this strain was most closely related to *Nitrosospira multififormis* ATCC 25196<sup>T</sup>, although the level of sequence similarity was low (90.5%). This strain GR16-43 is considered to represent a genus-novel taxon that is different from the genus *Nitrosospira*.

Genomic DNA was extracted from strain GR16-43 using a DNeasy blood and tissue kit (Qiagen). Genome sequencing was performed using the PacBio RS II platform (Pacific Biosciences) with a 20 kb insert SMRTbell library constructed. Sequencing analysis was performed by Chunlab Inc.. The

\*For correspondence. E-mail: [aychoi@nnibr.re.kr](mailto:aychoi@nnibr.re.kr);  
Tel.: +82-54-530-0963; Fax: +82-54-530-0969

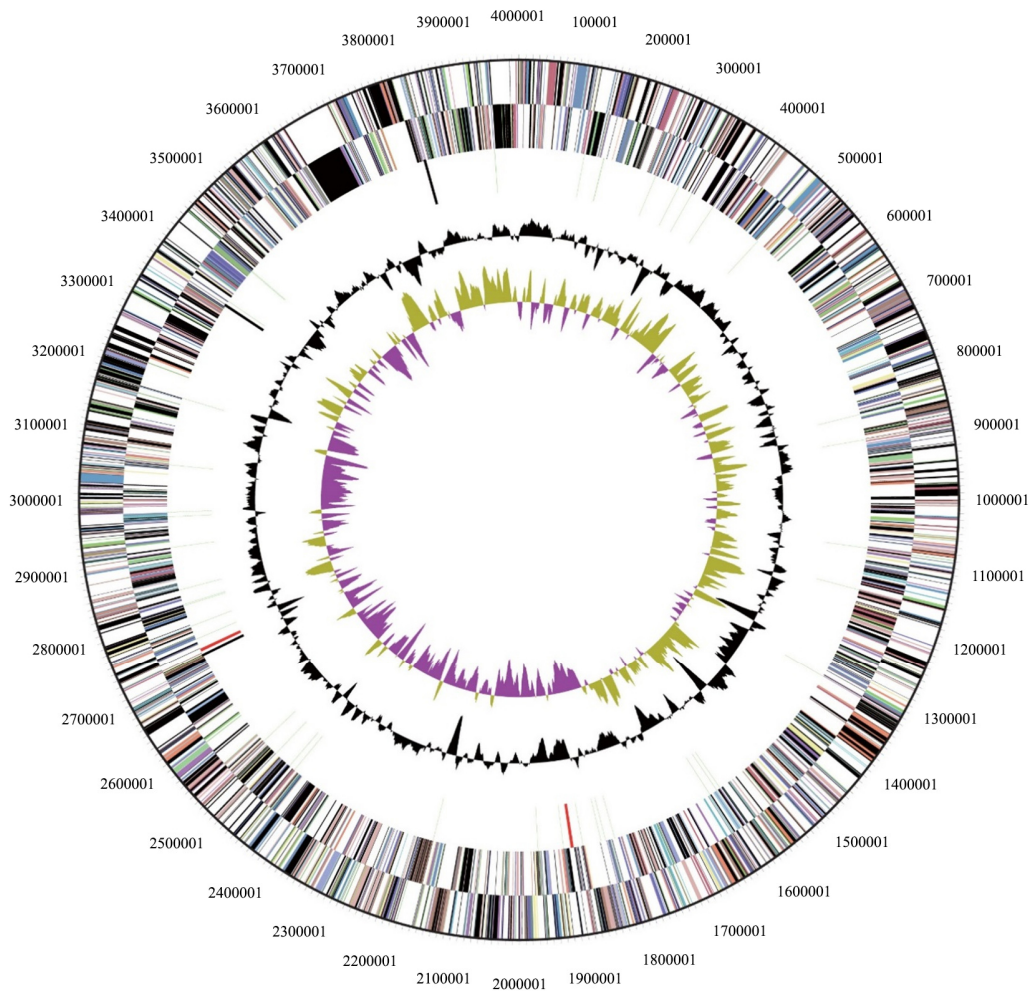
bacterial genome was assembled *de novo* into one contig, with an average genome coverage of 156.9 ×, using the PacBio SMRT Portal (2.3.0) and the hierarchical genome assembly process (Chin *et al.*, 2013). The complete genome sequence of GR16-43 is composed of a single circular chromosome. The data was then submitted to the Rapid Annotation using

Subsystem Technology (RAST) server (Aziz *et al.*, 2008) and the National Center for Biotechnology Information genome sequence database. Potential coding sequences were searched for using the Basic Local Alignment Search Tool (BLAST) against the UniProt (Wu *et al.*, 2006), Pfam (Punta *et al.*, 2012), and COG (Tatusov *et al.*, 2003) databases. rRNA and transfer RNA (tRNA) genes, as well as other miscellaneous features, were predicted using the RNAmmer 1.2 (Lagesen *et al.*, 2007) and tRNAscan-SE 1.21 (Lowe and Eddy, 1997) servers and the Rfam v12.0 database (Griffiths-Jones *et al.*, 2005). Signal peptides and transmembrane helices were predicted using the SignalP 4.1 (Petersen *et al.*, 2011) and TMHMM v2.0 (Krogh *et al.*, 2001) servers. The basic genome statistics are provided in Table 1.

The complete genome size is 4,806,848 bp with a 67.12%.

**Table 1. General features of betaproteobacterium GR16-43**

Features	Chromosome
Genome size (bp)	4,806,848
GC content (%)	67.12
rRNA genes	6
tRNA genes	47
Protein coding genes	4,424
Genes assigned to COGs	2,553



**Fig. 1. Graphical circular map of strain GR16-43.** From bottom to top: genes on forward strand (color by COG categories), genes on the reverse strand (color by COG categories), RNA genes (tRNA-green, rRNA-red, other RNAs-black), GC content, and GC skew (purple/olive).

As a result of gene prediction, this genome contains 4,424 CDSs, 47 tRNA, and 6 rRNA genes (Fig. 1). We also confirmed various genes related sulfur oxidation (*sox*) in the genome of GR16-43. The sulfur oxidation system, including sulfide oxidoreductase, sulfate reductase, and thiosulfate sulfurtransferase, is involved in the oxidation of reduced sulfur compounds. These genes are related to the *sox* cluster, *soxABDHXYZ* (GR16-43\_3249-GR16-43\_3255), which encodes the multienzyme Sox complex involving in the complete oxidation of reduced sulfur compounds to sulfate.

The genomic information reveals novel insights into sulfur oxidation in oligotrophic freshwater environments.

#### Nucleotide sequence accession number

The complete genome sequence was deposited in DDBJ/EMBL/GenBank under the accession number CP019169.

## 적 요

그람 음성이며 긴 막대 모양의 *betaproteobacteria*에 속하는 GR16-43을 한강 발원지 검릉소에서 분리하였다. GR16-43 균주에 대한 유전체분석을 실시하였으며, G+C 비율이 67.12% 인 4,806,848 bp 크기의 염기서열을 얻었다. 유전체 특징은 황산화와 관련된 다량의 유전자를 보유하고 있어 균주의 잠재적 중요성을 보여준다. 이러한 결과는 GR16-43 균주가 빈영양 담수 환경에서의 적응 연구를 위한 유전체 정보를 제공한다.

## Acknowledgements

This research was supported by a grant from the Nakdonggang National Institute of Biological Resources (NNIBR), funded by the Ministry of Environment of the Republic of Korea (The project on Development of unculturable microorganism screening method based on the new culture technology, 2017) and by the Basic Core Technology Development Program for the Oceans and the Polar Regions of the National Research Foundation (NRF) funded by the Ministry of Science, ICT and Future Planning (2016M1A5A1027453).

## References

- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., *et al.* 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75.
- Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler E.E., *et al.* 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569.
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124.
- Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580.
- Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.H., Rognes, T., and Ussery, D.W. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108.
- Lowe, T.M. and Eddy, S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964.
- Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., and Bertilsson, S. 2011. A guide to the natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. R.* **75**, 14–49.
- Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786.
- Punta, M., Coghill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., *et al.* 2012. The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–D301.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin E.V., Krylov, D.M., Mazumdr, R., Mekhedov, S.L., Nikolskaya, A.N., *et al.* 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41.
- Van der Gucht, K., Vandekerckhove, T., Vloemans, N., Cousin, S., Muylaert, K., Sabbe, K., Gillis, M., De Meester, L., and Vyverman, W. 2005. Characterization of bacterial communities in four freshwater lakes differing in nutrient load and food web structure. *FEMS Microbiol. Ecol.* **53**, 205–220.
- Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., *et al.* 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* **34**, D187–D191.
- Zwart, G., Crump, B.C., Kamst-van Agterveld, M.P., Hagen, F., and Han, S.K. 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microbiol. Ecol.* **28**, 141–155.