

## Genome Sequence of Strain IMCC2047, a Novel Marine Member of the *Gammaproteobacteria*<sup>∇</sup>

Ilnam Kang,<sup>1†</sup> Dongmin Kang,<sup>1†</sup> Hyun-Myung Oh,<sup>1</sup> Hana Kim,<sup>1‡</sup> Hee-Jin Kim,<sup>2</sup>  
Tae-Wook Kang,<sup>2</sup> Seon-Young Kim,<sup>2</sup> and Jang-Cheon Cho<sup>1\*</sup>

Division of Biology and Ocean Sciences, Inha University, Incheon, Republic of Korea,<sup>1</sup> and Medical Genomics Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea<sup>2</sup>

Received 4 May 2011/Accepted 10 May 2011

**Strain IMCC2047 was isolated from the Yellow Sea using dilution-to-extinction culturing. The strain was shown to occupy a distinct phylogenetic position within the *Gammaproteobacteria*. Here we present the genome sequence of strain IMCC2047, which harbors genes for various metabolic pathways, including proteorhodopsin and ribulose biphosphate carboxylase.**

Strain IMCC2047 was isolated from a surface seawater sample collected off the coastal region of the Yellow Sea, Korea. Although initially cultured by using pristine seawater-based oligotrophic liquid medium amended with minute amounts of carbon compounds (4), strain IMCC2047 could be grown as tiny colonies on a solid agar medium after subsequent culturing. Phylogenetic analysis based on 16S rRNA gene sequences showed that strain IMCC2047 formed a distinct phyletic clade and was most closely related to species of the genus *Microbulbifer* (90.1 to 92.3%) of the *Gammaproteobacteria* among the validly published species. Environmental clone sequences showing higher sequence similarity have been also reported from various marine habitats, including a mangrove ecosystem (DQ234156; 99.9%) (8), the Arctic surface sediment (EU287390; 99.0%) (7), an oxygen minimum zone (GQ350234; 98.0%) (12), and ocean crust (EU491487; 97.6%) (11). Taking into consideration the unique phylogenetic position of IMCC2047 and wide distribution of highly similar 16S rRNA gene sequences, we performed genome sequencing of strain IMCC2047.

A library for Illumina sequencing was prepared from 5 µg DNA using a Genomic DNA Sample Prep kit (Illumina Inc.) according to the manufacturer's protocol and sequenced through 76 cycles. A total of 14,205,786 single-read sequences obtained after quality check and filtering were assembled by using Velvet (13), resulting in 815 contigs (AEGLO1000001 to AEGLO1000815) with a depth of 229-fold genome coverage. The genome of IMCC2047 has 2,376,461 bp with a molar G+C ratio of 48.57%. Open reading frames (ORFs) and noncoding RNAs were predicted by using the RAST (Rapid Annotation using Subsystem Technology) server (1) and checked by BLAST analysis when needed (9). The IMCC2047 genome was predicted to contain 2,263 ORFs and 24 tRNAs.

The genome of strain IMCC2047 contained a gene encoding proteorhodopsin (PR), a light-driven proton pump widely distributed in marine bacteria (3, 5). The predicted amino acid sequence of IMCC2047 PR seems to be unique in that the glutamate residue in transmembrane helix C, which has been suggested to be crucial for proton pumping (6), was replaced by leucine. The PR sequences with similar substitutions were occasionally found in translated metagenomic sequences retrieved during the Global Ocean Sampling (GOS) expedition (10). The genome also contained genes for phosphoribulokinase and form II ribulose biphosphate carboxylase/oxygenase (2), two key enzymes of the Calvin cycle, suggesting the possibility of carbon fixation in this organism, although genes for photosynthetic reaction center or oxidation of inorganic compounds were not decisively annotated. In addition, C<sub>1</sub> compound metabolism was predicted by the presence of genes encoding enzymes such as formate dehydrogenase. Anaerobic respiration using sulfite, nitrate, and arsenate as electron acceptors was also suggested by the annotation of respective reductases. Overall, the prediction of genes associated with diverse metabolism, including phototrophy, carbon fixation, methylotrophy, and anaerobic respiration, implied that IMCC2047 could use various trophic strategies to survive in marine environments.

**Nucleotide sequence accession number.** The genome sequence of strain IMCC2047 is available in GenBank under accession number AEGLO00000000.

This study was supported by the 21C Frontier Program of Microbial Genomics and Applications and by the Korea Research Foundation Grant funded by the Korean Government (MEST, KRF-2008-521-C00263).

### REFERENCES

1. Aziz, R., et al. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**:75.
2. Badger, M. R., and E. J. Bek. 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO<sub>2</sub> acquisition by the CBB cycle. *J. Exp. Bot.* **59**:1525–1541.
3. Bějá, O., et al. 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**:1902–1906.
4. Cho, J.-C., and S. J. Giovannoni. 2004. Cultivation and growth characteristics of a diverse group of oligotrophic marine gammaproteobacteria. *Appl. Environ. Microbiol.* **70**:432–440.

\* Corresponding author. Mailing address: Division of Biology and Ocean Sciences, Inha University, Incheon, Republic of Korea. Phone: 82-32-860-7711. Fax: 82-32-232-0541. E-mail: chojc@inha.ac.kr.

† These authors contributed equally to this work.

‡ Present address: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139.

<sup>∇</sup> Published ahead of print on 20 May 2011.

5. **Fuhrman, J. A., M. S. Schwalbach, and U. Stingl.** 2008. Proteorhodopsins: an array of physiological roles? *Nat. Rev. Microbiol.* **6**:488–494.
6. **Giovannoni, S. J., et al.** 2005. Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**:82–85.
7. **Li, H., Y. Yu, W. Luo, Y. Zeng, and B. Chen.** 2009. Bacterial diversity in surface sediments from the Pacific Arctic Ocean. *Extremophiles* **13**:233–246.
8. **Liao, P.-C., B.-H. Huang, and S. Huang.** 2007. Microbial community composition of the Danshui river estuary of Northern Taiwan and the practicality of the phylogenetic method in microbial barcoding. *Microb. Ecol.* **54**:497–507.
9. **McGinnis, S., and T. L. Madden.** 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32**:W20–W25.
10. **Rusch, D. B., et al.** 2007. The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**:e77.
11. **Santelli, C. M., et al.** 2008. Abundance and diversity of microbial life in ocean crust. *Nature* **453**:653–656.
12. **Walsh, D. A., et al.** 2009. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326**:578–582.
13. **Zerbino, D. R., and E. Birney.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**:821–829.