

GEM: An Integrated Information Bank for Microbial Genome Research

KIM Jihyun F., HUR Cheol-Goo, JEONG Haeyoung, YOON Sung Ho, KANG Ho-Young, CHO Hwan-Gue¹, PARK Seung-Hwan and OH Tae Kwang²

Division of Genomics & Proteomics, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea

¹Graphics Application Lab., Dep. of Computer Science, Pusan National University, Korea

²Division of 21C Frontier Microbial Genomics and Applications Center, KRIBB, Daejeon, Korea

Genome Encyclopedia of Microbes (GEM <http://www.gem.re.kr>) is a web portal with an integrated database system established to expedite the microbial genome research by providing well-processed microbial genome information and analysis tools. It is composed of three different directories. The genome projects section contains sequence databases of microbial genomes being sequenced in our team and analyzed through a pipelined genome annotation system. Currently, *Paenibacillus polymyxa* E681, *Hahella chejuensis* KCTC 2396, and *Leuconostoc citreum* KM20 are in service. The genome resources section has MatGene Blast, Cozyme DB, and Target DB. MatGene Blast is designed to search through an in-house database of completely sequenced microbial genomes with a query protein sequence. Cozyme DB, which is an integrated version of COG, KEGG, and ENZYME databases, offers metabolic pathway information as well as functional data for a given query protein sequence. Target DB contains groups of industrially or pharmaceutically important genetic entities, and shows the characteristics of target groups such as motif, phylogeny, chromosomal location, etc. They will include pathogenicity islands, esterase/lipase genes and genes involved in biosynthesis of polyketides or non-ribosomal peptides, and type III protein secretion. We are also developing tools useful for analysis of microbial genomes. Ophiron is a program for operon prediction, and a program presenting a circular view of the genome map will allow various needs of users. Finally, GEM will function as the repository for the outputs of microbial functional genomic research projects in the 21C Frontier Microbial Genomics and Applications Center Program.

Functional Genomics of a Symbiotic Thermophile, *Symbiobacterium toebii* Isolated from Korean Compost

KIM Kwang Kim¹, BAE Jin-Woo², SONG Jae Jun², HONG Seung-Pyo¹, LEE Seung-Goo², YOON Sun-Woo², POO Haryoung², KITAKAWA Masanari³, and SUNG Moon-Hee^{1,4}

¹BioLeaders Corporation, Daejeon, Korea

²KRIBB, Daejeon, Korea

³Dragon Genomics Co. Ltd., Mie, Japan

⁴Dept. of Bio & Nanochemistry, Kookmin Univ., Seoul, Korea.

Symbiobacterium toebii is a symbiotic thermophile which grows only with *Geobacillus toebii*. For its axenic culture, the cell free extract of *Geobacillus toebii* have to be supplied into the medium. Its genome consists of a circular chromosome of 3,474,049 base pairs containing 3,942 predicted protein coding sequences (CDs) and there was not an extra-chromosomal element (ECE). An average G+C content of the genomic DNA was 68.8 mol%. Of these protein-coding sequences, 1,546 (40%) protein coding sequences have functional assignments and 2,396 (60%) have no match to any protein database. Of these protein-coding genes, a lot of gene families were repeatedly expanded by duplication. The genome contains at least 26 transposases and insertion sequence (IS) elements, indicating that transposase has played an important evolutionary role by horizontal gene transfer. The genome of *S. toebii* has the complete respiratory apparatus using nitric oxide (NO) as an electron acceptor and eight signal peptidase genes, as well as several genes for the secretory machinery. It also contains at least 70 putative ATP-binding transport proteins and more than 83 putative transcriptional regulators. Comparison with other sequenced microorganisms revealed that lots of gene families in *S. toebii* were closely related with those of *B. subtilis* and *B. halodurans*.