Complete genome sequence of the bacteria *Ketogulonicigenium vulgare* Y25

Xiang-hua Xiong1,2, Shuang Han1,2, Jian-hua Wang1, Ze-hui Jiang1,2, Wei Chen1,2, Ning Jia1,2, Hong-li Wei1,2, Hong Cheng1,2, Yan-xin Yang1,2, Bin Zhu1,2, Song You1,2, Jian-yong He2, Meng-xia Chen4, Chun-jie Yu1, Ying-hui Jiao4, Wei-cai Zhang1*

Laboratory of microorganism engineering, Beijing institute of biotechnology, Beijing 100071, China1; School of life science and biochemical pharmaceutical, Shenyang Pharmaceutical University, Shenyang 110016, China2; School of Preclinical Medicine, Guangxi Medical University, Nanning 530021, China3; Weisheng Pharmaceutical Co. Ltd., Shijiazhuang Pharma group, Shijiazhuang 050035, China4

*Ketogulonicigenium vulgare* is characterized by the efficient production of 2KGA from L-sorbose. *Ketogulonicigenium vulgare* Y25 is known as a 2-keto-L-gulonic acid producing strain in vitamin C industry. Here we report the finished, annotated genome sequence of *Ketogulonicigenium vulgare* Y25.

*Ketogulonicigenium vulgare* Y25 is used in production of Vitamin C, which was responsible for the conversion reaction of L-sorbose to L-Ketogulonic acid in mixed culture fermentation with *Bacillus* genus[1]. The entire genome of *Ketogulonicigenium vulgare* Y25 was sequenced to elucidate the metabolic pathway of sorbose and to obtained detailed insights into the growth potential of the organism.

The complete genome sequence of *Ketogulonicigenium vulgare* Y25 was determined by Beijing genome institute (shenjun, China) using Solexa technology. A total of 221M high-quality base pairs, giving 67.2–fold coverage of the genome, were assembled into 36 contigs using by SOAP software (http://soap.genomics.org.cn) [2]. Furthermore, the contigs was joined into 14 scaffolds using by paired-end information. Gaps between contigs were closed by custom primer walks, or by PCR amplification followed by DNA sequencing.

The genome of *Ketogulonicigenium vulgare* Y25 consists of a circular chromosome and two plasmids. The chromosome is composed of 2,776,084bp, with a G+C content of 61.72%. One plasmid contains 268,675bp, with a G+C content of 61.35% and the other contains 243,645bp, with a G+C content of 62.63%. Hence, the total size of the genome is 3,288,404bp and the average G+C content is 61.76%. There are a total of 3290 putative open reading frames (2807, 256, 227) using Glimmer, giving a coding intensity of 91.05%. 59 tRNA genes for all 20 amino acids but tyrosine and five 16S-23S-5S rRNA operons were identified.

Four genes encoding sorbose dehydrogenase were found in the chromosome. All of them were cloned and characterized. The result indicated every one could transform L-sorbose into 2-keto-gulonic acid and require pyrroloquinoline quinone as the prosthetic groups *in vitro* (unpublished). Sequence alignment analysis showed that they own high homology in nucleic acid and amin acid sequence[3-6]. It is estimated that multiple copies of sorbose dehydrogenase gene attribute to highly efficient conversion of sorbose to 2-keto-gulonic acid. A *ppqABCDE* cluster of coenzyme PQQ biosynthesis have also been isolated. It shows the same arrangement of *ppq* genes as that in other species with a small *ppqA* gene with its own promoter followed by an operon with the other four genes[7-10]. In addition, several genes coding sorbitol dehydrogenase, sorbose reductase, sorbose dehydrogenase and so on were annotated in the genome.

The Y25 genome sequence and its curated annotation are important assets to better
understand the physiology and metabolic potential of *Ketogulonicigenium vulgare* and will open up new opportunities in the functional genomics of this species.

**Nucleotide sequence accession number.** The nucleotide sequence comprising the *Ketogulonicigenium vulgare* Y25 genome was deposited in the GenBank database with the accession numbers CP002224, CP002225, CP002226.

We acknowledge the Beijing genome institute for Solexa shotgun sequencing and the analysis and annotation of the genome.

This work was funded by the National Programs for High Technology Research and Development of China (2006AA020303) and the National Key Technology R&D Program(2007BAI46B01).

Reference:


*Corresponding author. Mailing address: Dongdajie 20, Fengtai District, Beijing, 100071 China. Phone: 80-10-66948826. Fax: 80-10-63833524. E-mail: zhangweicai@hotmail.com.*

# Xiang-hua Xiong and Shuang Han contribute equally to this article.