Complete Genome Sequence of the Industrial Bacterium *Ketogulonicigenium vulgare* SKV

Nan Jia,[a,b] Ming-Zhu Ding,[a,b] Yu-Zhang Du,[c] Shuai Feng,[c] Feng Gao,[a,b,d] Ying-Jin Yuan[a,b]

Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, Tianjin, People's Republic of China[a]; SynBio Research Platform, Collaborative Innovation Centre of Chemical Science and Engineering (Tianjin), School of Chemical Engineering and Technology, Tianjin University, Tianjin, People's Republic of China[b]; Shandong Luwei Pharmaceutical Co., Ltd., Zibo, Shandong, People's Republic of China[c]; Department of Physics, Tianjin University, Tianjin, People's Republic of China[d]

Ketogulonicigenium vulgare has been widely used in vitamin C two-step fermentation, which converts L-sorbose to 2-keto-L-gluonic acid. Here, the complete genome of *K. vulgare* SKV, which performs better fermentation production than *K. vulgare* Hbe602, is deciphered to understand the key differences in metabolism between *K. vulgare* strains SKV and Hbe602.

Address correspondence to Ming-Zhu Ding, mzding@tju.edu.cn, or Feng Gao, fgao@tju.edu.cn.

*K*etogulonicigenium vulgare* was identified as a member of the *Proteobacteria* (1). It can convert L-sorbose to 2-keto-L-gulonic acid (2-KGA), the precursor of vitamin C (2). Mono-cultured *K. vulgare* grows poorly and the *Bacillus* spp. are usually cocultivated with it to achieve a high 2-KGA yield (3). Previously, we published the genome of *K. vulgare* Hbe602 (4). Our research showed that combinational expression of sorbose/sorbosone de-hydrogenases and cofactor pyrroloquinoline quinone in *K. vulgare* Hbe602 could enhance 2-KGA production properly (5). It was interesting to remark that industrial *K. vulgare* performed better 2-KGA production than *K. vulgare* Hbe602 when they were cocultured with *Bacillus thuringiensis* Bc601 (6).

K. vulgare strain SKV (Shandong Luwei Pharmaceutical Co., Ltd.) was cultured in 250-ml flasks at 30°C and 250 rpm for 35 h. The seed medium contains 3 g/L beef extract, 3 g/L yeast powder, 3 g/L corn steep liquor, 0.2 g/L MgSO$_4$, 1 g/L KH$_2$PO$_4$, 1 g/L urea, and 10 g/L peptone. The genomic DNA was isolated using the SDS method. The genome of *K. vulgare* SKV was sequenced by single-molecule real-time (SMRT) technology (Beijing Novogene Bioinformatics Technology Co., Ltd.). SMRT Analysis version 2.3.0 was used to filter low-quality reads, and the filtered reads were assembled to generate one contig without gaps.

The genome of *K. vulgare* SKV was annotated through the NCBI Prokaryotic Genome Annotation Pipeline (7) and using BLAST (8) against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (9) and the Clusters of Orthologous Groups (COG) of proteins database (10). The tRNAs and rRNAs were predicted by tRNAscan (11) and RNAmmer (12), respectively. The origin of replication (*oriC*) and putative DnaA boxes were identified using Ori-Finder (13). GC-Profile was used to identify the GC content variation in DNA sequences (14).

The genome of *K. vulgare* SKV consists of one circular chromosome (2,764,573 bp) and one circular plasmid (267,949 bp). The sequence difference between *K. vulgare* SKV and Hbe602 was analyzed using BLAST, and the sequence similarity between the chromosomes of *K. vulgare* SKV and Hbe602 is more than 99%.

Additionally, the plasmid in *K. vulgare* SKV is almost the same as plasmid 1 in Hbe602. The better ability of 2-KGA production in *K. vulgare* SKV may be due to the loss of plasmid 2 in Hbe602. Plasmid 2 of *K. vulgare* Hbe602 encodes 211 proteins, which are mainly related to transport systems, transcriptional regulators, and dehydrogenases. We identified that overexpression of L-sorbosone dehydrogenase (GI: 939479492) in plasmid 2 of *K. vulgare* Hbe602 produced an obvious byproduct in *K. vulgare* (15). Compared with *K. vulgare* Hbe602, the loss of dehydrogenases may lead to the higher 2-KGA production. We hope these findings can provide insight into the metabolism and gene targets for the strain improvement of *K. vulgare*.

Accession number(s). The sequence of the *K. vulgare* SKV genome has been deposited at DDBJ/EMBL/GenBank under the GenBank accession numbers CP016592 (chromosome) and CP016593 (plasmid).

ACKNOWLEDGMENT

FUNDING INFORMATION

REFERENCES

1. **Urbance JW, Bratina BJ, Stoddard SF, Schmidt TM**. 2001. Taxonomic characterization of *Ketogulonigenium vulgare* gen. nov., sp. nov. and *Ketogulonigenium robustum* sp. nov., which oxidize L-sorbose to 2-keto-L-gulonic acid. Int J Syst Evol Microbiol **51:**1059–1070. http://dx.doi.org/10.1099/00207713-51-3-1059.

2. **Yin G, He J, Ren S, Song Q, Ye Q, Lin H, Chen C, Guo X**. 1997. Production of vitamin C precursor-2-keto-L-gulonic acid from L-sorbose by a novel bacterial component system of SCB329-SCB933. J Ind Microbiol **27**:1–7.

3. **Jia N, Du J, Ding MZ, Gao F, Yuan YJ**. 2015. Genome sequence of *Bacillus endophyticus* and analysis of its companion mechanism in the *Ketogulonigenium vulgare-Bacillus* strain consortium. PLoS One **10**: e0135104. http://dx.doi.org/10.1371/journal.pone.0135104.

4. **Jia N, Ding MZ, Du J, Pan CH, Tian G, Lang JD, Fang JH, Gao F, Yuan YJ**. 2016. Insights into mutualism mechanism and versatile metabolism of *Ketogulonicigenium vulgare* Hbe602 based on comparative genomics and metabolomics studies. Sci Rep **6**:23068. http://dx.doi.org/10.1038/srep23068.

5. **Du J, Bai W, Song H, Yuan YJ**. 2013. Combinational expression of sorbose/sorbosone dehydrogenases and cofactor pyrroloquinoline quinone increases 2-keto-L-gulonic acid production in *Ketogulonigenium vulgare–Bacillus cereus* consortium. Metab Eng **19**:50–56. http://dx.doi.org/10.1016/j.ymben.2013.05.006.

6. **Jia N, Ding MZ, Gao F, Yuan YJ**. 2016. Comparative genomics analysis of the companion mechanisms of *Bacillus thuringiensis* Bc601 and *Bacillus endophyticus* Hbe603 in bacterial consortium. Sci Rep **6**:28794. http://dx.doi.org/10.1038/srep28794.

7. **Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J**. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res **44**: 6614–6624. http://dx.doi.org/10.1093/nar/gkw569.

8. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ**. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**:3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.

9. **Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M**. 2006. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res **34**: D354–D357. http://dx.doi.org/10.1093/nar/gkj102.

10. **Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA**. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4**:41. http://dx.doi.org/10.1186/1471-2105-4-41.

11. **Lowe TM, Eddy SR**. 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25**: 955–964. http://dx.doi.org/10.1093/nar/25.5.0955.

12. **Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW**. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35**:3100–3108. http://dx.doi.org/10.1093/nar/gkm160.

13. **Gao F, Zhang CT**. 2008. Ori-Finder: a web-based system for finding *oriCs* in unannotated bacterial genomes. BMC Bioinformatics **9**:79. http://dx.doi.org/10.1186/1471-2105-9-79.

14. **Gao F, Zhang CT**. 2006. GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. Nucleic Acids Res **34**:W686–W691. http://dx.doi.org/10.1093/nar/gkl040.

15. **Chen S, Jia N, Ding MZ, Yuan YJ**. 2016. Fitness analysis between the L-sorbosone dehydrogenase modules and *Ketogulonigenium vulgare* chassis. Chin J Biotechnol **32**:1–9.