Due to its extensive industrial applications in various fields, such as foods, cosmetics, pharmaceuticals, transport fuels, and agrochemical industries (1–3), 2,3-butanediol (2,3-BD) is a promising biobased bulk chemical. Many microorganisms, such as Klebsiella pneumoniae, Klebsiella oxytoca, Enterobacter cloacae, and Serratia marcescens (1, 2, 4–6), are known to be able to produce 2,3-BD. Among all these strains, K. pneumoniae was the first to be identified (1, 7) and is one of the best organisms that has shown the potential for industrial 2,3-BD production because of its more complete fermentation, broad substrate spectrum (hexoses, pentose, certain disaccharides, and uronic acid derived from the hydrolysates of hemicellulosic and cellulosic materials), cultural adaptability, and high efficiency (1, 7, 8). As a commercial producer of 2,3-BD (9, 10), given the limited genetic information on the metabolic mechanism underlying the formation of 2,3-BD, research about this strain is mainly focused on fermentative conditions and fermentation substrates rather than metabolic engineering. We therefore sequenced and analyzed the whole-genome of strain K. pneumoniae CICC10011 to provide the genetic basis for the production of 2,3-BD at a high titer.

Here, we present the draft genome sequence of strain K. pneumoniae CICC10011, obtained using the Illumina HiSeq 2500 system at the Chinese National Human Genome Center, Shanghai, China. The reads were trimmed and de novo assembled with Velvet version 1.2.03 (11). Open reading frames (ORFs) were identified using the program Glimmer 3.02 (http://ccb.jhu.edu/software/glimmer/index.shtml). These ORFs were further annotated with comparison to the NCBI nr, KEGG, and Clusters of Orthologous Groups (COG) databases. The draft genome sequence of K. pneumoniae CICC10011 was annotated with the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP). rRNAs were predicted by RNamer (12), and tRNAs predicted by tRNAscan (13).

The draft genome sequence of K. pneumoniae CICC10011 comprises 4,883,939 bp, which was assembled into 58 contigs. The N50 quality measurement of the contigs was 162,338 bp, with a G+C content of 58.4%, and the largest contig assembled was 856,167 bp. The genome sequence was annotated using the PGAAP. The draft genome sequence of CICC10011 contains 5,603 genes, including 3 rRNA genes (5S rRNA, 16S rRNA, and 23S rRNA), and 39 tRNA genes. The annotation results showed that 4,234 proteins have clear biological functions; of these, 3,173 proteins have KEGG orthologs, and 4,433 proteins have COG classifications.

The genome sequence of this promising strain provides significant opportunities to further investigate the genetic and regulatory mechanisms underlying the formation of 2,3-BD, to explain the genetic reasons for its high productivity and biomass, to analyze the byproducts that may hinder the accumulation of 2,3-BD, and to thoroughly understand the genetic, biological, and physiological characteristics of widely used K. pneumoniae strains.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. LBCM00000000. The version described in this paper is the first version, LBCM01000000.

ACKNOWLEDGMENTS

This work was financially supported by the National Science Foundation for Distinguished Young Scholars of China (no. 21225626), the National Natural Science Foundation of China (no. 21376002 and 21476111), the National High Technology Research and Development Program of China (no. 2011AA02A207), and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

REFERENCES


