Genome Sequence of *Klebsiella quasipneumoniae* subsp. *similipneumoniae* MB373, an Effective Bioremediator

Fozia Aslam,[a] **Azra Yasmin,**[a] **Torsten Thomas**[b]

Department of Environmental Sciences, Microbiology and Biotechnology Research Lab, Fatima Jinnah Women University, Rawalpindi, Pakistan[a]; Centre for Marine Bio-innovation & School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, New South Wales, Australia[b]

Klebsiella quasipneumoniae subsp. *similipneumoniae* MB373 was isolated from effluent of the Hattar Industrial Estate, Haripur, Pakistan. *K. quasipneumoniae* subsp. *similipneumoniae* has few cultivated/characterized members so far. Whole-genome sequencing revealed its potential for metal and toxin resistance, which further elucidated various enzymatic processes for the degradation of xenobiotics, illuminating its bioremediation applications.

Address correspondence to Azra Yasmin, azrayasmin@fjwu.edu.pk.

Abundant contaminants produced by industrial activities have resulted in the need to remediate the affected environment. For example, industrial effluents often contain various organic and inorganic pollutants, including heavy metals that have been shown to represent a risk to human health (1). Microorganisms isolated from such contaminated environments have a potential role in the remediation processes, as they often contain a diversity of genes that are involved in the degradation of xenobiotics and resistance to heavy metals or toxins (2). In this context, several bacterial genera have been reported for efflux-mediated metal tolerance/detoxification mechanisms, which have been acquired through horizontal gene transfer (3, 4).

Here, we present the genome sequence of *Klebsiella quasipneumoniae* subsp. *similipneumoniae* MB373. The strain was chosen for its high resistance against various heavy metals and antibiotics as well as for its ability to utilize and degrade organic compounds. Genomic DNA extraction and whole-genome sequencing were performed with Illumina MiSeq technology. The sequence reads obtained were of good quality, as determined by FastQC (5). Afterward, SPAdes version 3.1.0 yielded an assembly of 39 contigs (6). The N_{50} contig size was approximately 368,603 bp, showing genome coverage of about 104.0×. The assembled genome consists of 5,440,152 bp, with 57.5% G+C content.

Gene prediction and annotation were done with Rapid Annotations using Subsystems Technology (RAST) and Prokaryotic Genome Annotation Pipeline (PGAP) by NCBI (http://www.ncbi.nlm.nih.gov/genome/annotation_prok). According to PGAP, the draft genome comprised 5,343 coding sequences, with 5,136 protein-coding genes and 105 pseudogenes. There had been 102 RNA genes, out of which 13 were identified for rRNAs and 76 for tRNAs using RNAmmer (7) and tRNAscan SE (8), respectively.

RAST server results revealed 5,203 genes that were related to 583 SEED subsystems, with 58 possibly missing genes (9). Most genes were designated to the metabolism of carbohydrates (807), amino acids, and/or their derivatives (572). One hundred eighty-one genes were allocated for stress response, including universal stress protein, phage shock protein (psp) operon, and an Hfl operon. One hundred nineteen genes were assigned to the functions associated with antibiotics and toxic chemical resistances. Among metal tolerance, 22 genes were annotated for proteins related to Co-Zn-Cd, whereas nine, two, and 11 genes were identified for copper, chromium, and arsenic tolerance mechanisms, respectively.

The genome possessed 130 genes for cell regulation and signaling, which encoded BOX elements in streptococci, quorum sensing (auto inducer 2 [AI-2] transport and processing *lsrACDBFGE* operon), biofilm adhesion biosynthesis, and detoxification of tellurium compounds, etc. Furthermore, functional analysis revealed 80 genes for enzymes effectively participating in the metabolism of various aromatic compounds comprising peripheral catabolic pathways of salicylate ester, quinate, benzoate, *p*-hydroxybenzoate, chloroaromatic, as well as anaerobic decarboxylation of hydroxyaromatic compounds. Additionally, 50 genes were designated for the metabolism of central aromatic intermediates of salicylate and gentisate catabolism (six genes), catechol (eight genes), and the protocatechuate (17 genes) branch of the beta-ketoadipate pathway.

In short, the genome sequence of strain MB373 revealed various genes coding for enzymes that could be involved in the biotransformation of organic pollutants as well as resistance and detoxification of heavy metals. This makes the strain a potential candidate for bioremediation processes for environments polluted with industrial effluents.

Accession number(s). The whole-genome shotgun project of *Klebsiella quasipneumoniae* subsp. *similipneumoniae* MB373 has been deposited at DDBJ/ENA/GenBank under accession no. LYSU00000000. The version of this paper is version LYSU01000000.

REFERENCES

1. **Gikas P**. 2008. Single and combined effects of nickel (Ni(II)) and cobalt (Co(II)) ions on activated sludge and on other aerobic microorganisms: a review. J Hazard Mater **159**:187–203. http://dx.doi.org/10.1016/j.jhazmat.2008.02.048.

2. **Li AD, Li LG, Zhang T**. 2015. Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. Front Microbiol **6**:1025. http://dx.doi.org/10.3389/fmicb.2015.01025.

3. **Lima de Silva AA, de Carvalho MAR, de Souza SAL, Dias PMT, da Silva Filho RG, de Meirelles Saramago CS, de Melo Bento CA, Hofer E**. 2012. Heavy metal tolerance Cr, Ag and Hg in bacteria isolated from sewage. Braz J Microbiol **43**:1620–1631.

4. **Sundar K, Mukherjee A, Sadiq M, Chandrasekaran N**. 2011. Cr (III) bioremoval capacities of indigenous and adapted bacterial strains from Palar river basin. J Hazard Mater **187**:553–561. http://dx.doi.org/10.1016/j.jhazmat.2011.01.077.

5. **Andrews S**. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

6. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA**. 2012. SPAdes: a new genome assembly algorithm and its applications to single cell sequencing. J Comput Biol **19**:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

7. **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.

8. **Lowe TM, Eddy SR**. 1997. tRNA scan-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.

9. **Aziz RK, Bartels D, Best AA, Dejongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O**. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9**:75. http://dx.doi.org/10.1186/1471-2164-9-75.