

Complete Genome Sequence of *Lysinibacillus sphaericus* B1-CDA, a Bacterium That Accumulates Arsenic

Aminur Rahman,^{a,b} Noor Nahar,^a Jana Jass,^b Björn Olsson,^a Abul Mandal^a

Systems Biology Research Center, School of Bioscience, University of Skövde, Skövde, Sweden^a; The Life Science Center, School of Science and Technology, Örebro University, Örebro, Sweden^b

Here, we report the genomic sequence and genetic composition of an arsenic-resistant bacterium, *Lysinibacillus sphaericus* B1-CDA. Assembly of the sequencing reads revealed that the genome size is ~4.5 Mb, encompassing ~80% of the chromosomal DNA.

Received 3 August 2015 Accepted 28 November 2015 Published 21 January 2016

Citation Rahman A, Nahar N, Jass J, Olsson B, Mandal A. 2016. Complete genome sequence of *Lysinibacillus sphaericus* B1-CDA, a bacterium that accumulates arsenic. *Genome Announc* 4(1):e00999-15. doi:10.1128/genomeA.00999-15.

Copyright © 2016 Rahman et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Abul Mandal, abul.mandal@his.se.

The arsenic-resistant strain B1-CDA was isolated from arsenic-contaminated land in Bangladesh (1). Sequencing of the genomic DNA of B1-CDA was performed by an Illumina HiSeq 2500 PE100 sequencer with a single sequencing index. The genome assembly started with Illumina 100-bp paired-end reads of genomic DNA with an insert length of 300 bp. The read quality was checked using FastQC (2). The raw reads were quality trimmed and corrected using Quake (3). Properly paired reads ≥ 30 bp in length were selected from the pool of corrected reads, and the remaining singleton reads were considered single-end reads. Both types of reads were then used in k -mer-based *de novo* assembly by employing SOAPdenovo (4). The set of scaffolds with the largest N_{50} was identified by evaluating k -mers ranging from 29 to 99. The optimal scaffold sequences were further subjected to gap closing by utilizing the corrected paired-end reads. The resulting scaffolds of length ≥ 300 bp were chosen as the final assembly (5).

A total of 11,105,899 pairs of reads were generated by Illumina deep sequencing. Analysis of the raw reads with FastQC showed that the average per base Phred score was ≥ 32 for all positions, and the mean per sequence Phred score was 38. The overall G+C content was 38%. After quality trimming, error correction, and removal of the TruSeq adapter sequence, 10,940,654 read pairs (98.5%) and 145,888 single-end sequences remained for further analysis. The set of scaffold sequences with maximal N_{50} (507,225 bp) was produced at a k -mer of 91. The corresponding scaffold sequences were subjected to gap closure using the corrected paired-end reads, and the resulting scaffolds (≥ 300 bp) were defined as the final assembly. The final assembly was 4,509,276 bp, and it consisted of 31 scaffolds ranging from 314 bp to 1,145,744 bp.

The assembled genome sequence was annotated with RAST (6). The RAST analysis pipeline uses tRNAscan-SE to predict tRNA genes (7) and the Glimmer algorithm to predict protein-coding genes (8). Predictions of tRNA-, rRNA-, and protein-coding genes were performed based on 77 RAST-predicted tRNA genes. RAST resulted in 11 rRNA genes, including seven 5S, one

16S, and three 23S genes. A total of 4,513 protein-coding genes were predicted using the Glimmer algorithm, of which 2,671 protein-coding genes were annotated by RAST's automated homology analysis and assigned to functional categories. GeneMark (9) and FgenesB (10) algorithms were also applied, yielding 4,562 and 4,323 genes, respectively. The functional annotation by RAST and Blast2GO (11) indicated that B1-CDA contains many genes, which are responsive to metal ions, like arsenic, cobalt, copper, iron, nickel, potassium, manganese, and zinc. All protein-coding sequences resulting from GeneMark were used by Blast2GO for functional annotation. Based on the phylogenetic trees inferred by using the neighbor-joining method (12) presented in the MEGA6 software (13), B1-CDA resembles *Lysinibacillus sphaericus* G10, R-27024, and CICR-X12.

In summary, strain B1-CDA demonstrates the presence of several metal-responsive genes that might be utilized in bioremediation of toxic metals in polluted environments.

Nucleotide sequence accession numbers. The genome sequence of B1-CDA strain has been deposited in GenBank under the accession number [LJYY000000000](https://www.ncbi.nlm.nih.gov/nuclink/LJYY000000000). The version described in this paper is the first version, LJYY000000000.1.

ACKNOWLEDGMENTS

This research has been funded mainly by the Swedish International Development Cooperation Agency (SIDA) (grant no. AKT-2010-018) and partly by the Nilsson-Ehle Foundation (The Royal Physiographic Society in Lund) in Sweden.

FUNDING INFORMATION

This research has been funded mainly by the Swedish International Development Cooperation Agency (SIDA; grant number AKT-2010-018) and partly by the Nilsson-Ehle Foundation (The Royal Physiographic Society in Lund) in Sweden.

REFERENCES

- Rahman A, Nahar N, Nawani NN, Jass J, Desale P, Kapadnis BP, Hossain K, Saha AK, Ghosh S, Olsson B, Mandal A. 2014. Isolation and

- characterization of a *Lysinibacillus* strain B1-CDA showing potential for bioremediation of arsenics from contaminated water. *J Environ Sci Health A Tox Hazard Subst Environ Eng* 49:1349–1360. <http://dx.doi.org/10.1080/10934529.2014.928247>.
2. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
 3. Kelley DR, Schatz MC, Salzberg SL. 2010. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol* 11:R116. <http://dx.doi.org/10.1186/gb-2010-11-11-r116>.
 4. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20:265–272. <http://dx.doi.org/10.1101/gr.097261.109>.
 5. Rahman A, Nahar N, Nawani NN, Jass J, Ghosh S, Olsson B, Mandal A. 2015. Comparative genome analysis of *Lysinibacillus* B1-CDA, a bacterium that accumulates arsenics. *Genomics*. 106:384–392. <http://dx.doi.org/10.1016/j.ygeno.2015.09.006>.
 6. 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>.
 7. 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>.
 8. Salzberg SL, Delcher AL, Kasif S, White O. 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 26: 544–548. <http://dx.doi.org/10.1093/nar/26.2.544>.
 9. Borodovsky M, McIninch J. 1993. GenMark: parallel gene recognition for both DNA strands. *Comput Chem* 17:123–133. [http://dx.doi.org/10.1016/0097-8485\(93\)85004-V](http://dx.doi.org/10.1016/0097-8485(93)85004-V).
 10. Salamov AA, Solovyev VV. 2000. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res* 10:516–522. <http://dx.doi.org/10.1101/gr.10.4.516>.
 11. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435. <http://dx.doi.org/10.1093/nar/gkn176>.
 12. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
 13. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725–2729. <http://dx.doi.org/10.1093/molbev/mst197>.