

# Draft Genome Sequence of Hydrocarbon-Degrading *Enterobacter cloacae* Strain S1:CND1, Isolated from Crude Oil-Contaminated Soil from the Noonmati Oil Refinery, Guwahati, Assam, India

Arghya Mukherjee,<sup>a</sup> Bobby Chettri,<sup>b</sup> James S. Langpoklakpam,<sup>b</sup> Arvind K. Singh,<sup>b</sup>  Dhrubajyoti Chattopadhyay<sup>a\*</sup>

Department of Biotechnology, University of Calcutta, Kolkata, West Bengal, India<sup>a</sup>; Department of Biochemistry, North-Eastern Hill University, Shillong, Meghalaya, India<sup>b</sup>

\* Present address: Dhrubajyoti Chattopadhyay, Department of Biotechnology, Amity University, Rajarhat, New Town, Kolkata, West Bengal, India.

**We report here the 4.57-Mb draft genome sequence of hydrocarbon-degrading *Enterobacter cloacae* strain S1:CND1 isolated from oil-contaminated soil in Guwahati, India. S1:CND1 contains 4,205 coding sequences and has a G+C content of 57.45%. This is the first report of the genome sequence of an *E. cloacae* adapted to an oil-contaminated environment.**

Received 22 March 2016 Accepted 23 March 2016 Published 12 May 2016

**Citation** Mukherjee A, Chettri B, Langpoklakpam JS, Singh AK, Chattopadhyay D. 2016. Draft genome sequence of hydrocarbon-degrading *Enterobacter cloacae* strain S1:CND1, isolated from crude oil-contaminated soil from the Noonmati oil refinery, Guwahati, Assam, India. *Genome Announc* 4(3):e00367-16. doi:10.1128/genomeA.00367-16.

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

Address correspondence to Dhrubajyoti Chattopadhyay, [dchattopadhyay@kol.amity.edu](mailto:dchattopadhyay@kol.amity.edu).

*Enterobacter cloacae* is a Gram-negative, rod-shaped bacterium generally associated with nosocomial infections, and is considered one of the most difficult to treat among the *Enterobacter* sp. (1). Over the years, there have been a few reports of hydrocarbon-degrading *E. cloacae* strains (2), although no genome sequence for any hydrocarbon-degrading *E. cloacae* strain is currently available. *E. cloacae* strain S1:CND1 was isolated from oil-contaminated soil collected from the Noonmati oil refinery in Guwahati, Assam, India. Strain S1:CND1 has been found to degrade alkanes, including n-hexane and n-hexadecane, and polyaromatic hydrocarbon as naphthalene, along with diesel and crude oil. Whole-genome shotgun sequencing was hence carried out to study the genetic constitution and metabolic versatility of this organism. We believe this is the first report of the draft genome sequence of a hydrocarbon-degrading *E. cloacae*.

The genomic DNA for strain S1:CND1 was extracted using an Ultra-Clean Microbial DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to the manufacturer's protocol. Isolated genomic DNA was then sequenced with an Illumina HiSeq 2500, which generated 4,580,054 paired-end reads. After quality control measures, the reads were assembled using the *de novo* assemblers ABySS v. 3.81 (3), Edena v. 3.130110 (4), MaSuRCA v. 2.2.1 (5), SOAPdenovo2 v2.04 (6), SPAdes v3.1.1 (7), and Velvet v1.2.10 (8). Assembled reads were then integrated using CISA v1.3 (9), which generated 14 contigs with a  $N_{50}$  length of 445,053 bp and an average length of 326,832.86 bp. The draft genome thus assembled was 4,575,660 bp in length with a G+C content of 57.45% and had 108-fold coverage. Genome annotation for strain S1:CND1 was carried out with the NCBI Prokaryotic Genome Annotation Pipeline, which predicted the presence of 4,205 coding sequences (CDSs), along with 16 rRNAs, 77 tRNAs, 6 noncoding RNAs (ncRNAs), and 50 pseudogenes. Rapid functional annotation for CDSs of strain S1:CND1 was carried out with the

RAST annotation server (10), which classified the CDSs into 535 subsystems. Among these, the most abundant subsystems were carbohydrates ( $s = 650$  CDSs); amino acids and derivatives ( $s = 467$ ); stress response ( $s = 157$ ); respiration ( $s = 149$ ); fatty acids, lipids, and isoprenoids ( $s = 138$ ); DNA metabolism ( $s = 117$ ); regulation and cell signaling ( $s = 150$ ); protein metabolism ( $s = 168$ ); RNA metabolism ( $s = 150$ ); membrane transport ( $s = 177$ ); virulence, disease, and defense ( $s = 100$ ); cell wall and capsule ( $s = 196$ ); and cofactors, vitamins, prosthetic groups, and pigments ( $s = 253$ ). Genome annotation revealed the presence of hydrocarbon degradation genes as alkane-1-monooxygenase, alkanesulfonate monooxygenase, naphthalene 1,2-dioxygenase, and quercetin 2,3-dioxygenase, thus underlining the extensive genetic adaptation of strain S1:CND1 to oil contamination.

A comparison of strain S1:CND1 with genomes in the RAST database identified *Escherichia coli* 88.1467 (score = 501) as its closest neighbor, followed by *E. coli* 88.0221 (score = 490) and *E. coli* 89.0511 (score = 469). *Enterobacter mori* LMG 25706 (score = 344) was identified as the 18th-closest neighbor.

**Nucleotide sequence accession numbers.** This whole-genome shotgun sequencing project for *E. cloacae* strain S1:CND1 has been deposited in DDBJ/EMBL/GenBank under the accession no. [LUGN00000000](https://www.ncbi.nlm.nih.gov/nuccore/LUGN00000000). The version of the whole-genome sequence (WGS) described here is version LUGN01000000.

## ACKNOWLEDGMENTS

The research reported in this article was supported by the NER TWINNING project (grant ID BT/306/NE/TBP/2012 dated 6/12/2012 from the Department of Biotechnology, Government of India). A.M. was supported by the CSIR/UGC-NET fellowship from the UGC, Government of India. We also hereby acknowledge the generous support provided by the Indian Oil Corporation in collection of samples.

## FUNDING INFORMATION

This work, including the efforts of Bobby Chettri, James S. Langpoklakpam, Arvind K. Singh, and Dhruvajyoti Chattopadhyay, was funded by Department of Biotechnology, Ministry of Science and Technology (DBT) (BT/306/NE/TBP/2012 dated 6/12/2012). This work, including the efforts of Arghya Mukherjee, was funded by University Grants Commission (UGC) (CSIR/UGC-NET Fellowship).

The research reported in this article was supported by the NER TWINNING project (grant ID BT/306/NE/TBP/2012, dated 6/12/2012, from the Department of Biotechnology, Government of India). Arghya Mukherjee was supported by the CSIR/UGC-NET fellowship from the UGC, Government of India.

## REFERENCES

- Davin-Regli A, Bosi C, Charrel R, Ageron E, Papazian L, Grimont PA, Cremieux A, Bollet C. 1997. A nosocomial outbreak due to *Enterobacter cloacae* strains with the *E. hormaechei* genotype in patients treated with fluoroquinolones. *J Clin Microbiol* 35:1008–1010.
- Hua X, Wu Z, Zhang H, Lu D, Wang M, Liu Y, Liu Z. 2010. Degradation of hexadecane by *Enterobacter cloacae* strain TU that secretes an exopolysaccharide as a bioemulsifier. *Chemosphere* 80:951–956. <http://dx.doi.org/10.1016/j.chemosphere.2010.05.002>.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res* 19:1117–1123. <http://dx.doi.org/10.1101/gr.089532.108>.
- Hernandez D, François P, Farinelli L, Osterås M, Schrenzel J. 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res* 18:802–809. <http://dx.doi.org/10.1101/gr.072033.107>.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677. <http://dx.doi.org/10.1093/bioinformatics/btt476>.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. <http://dx.doi.org/10.1186/2047-217X-1-18>.
- 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>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
- Lin SH, Liao YC. 2013. CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One* 8:e60843. <http://dx.doi.org/10.1371/journal.pone.0060843>.
- 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>.