

Draft Genome Sequence of Hydrocarbon-Degrading *Staphylococcus saprophyticus* Strain CNV2, Isolated from Crude Oil-Contaminated Soil from the Noonmati Oil Refinery, Guwahati, Assam, India

Arghya Mukherjee,^a Bobby Chettri,^b James S. Langpoklakpam,^b Arvind K. Singh,^b  Dhrubajyoti Chattopadhyay^{a*}

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

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

Here, we report the 2.6 Mb draft genome sequence of hydrocarbon-degrading *Staphylococcus saprophyticus* strain CNV2, isolated from oil-contaminated soil in Guwahati, India. CNV2 contains 2,545 coding sequences and has a G+C content of 33.2%. This is the first report of the genome sequence of an *S. saprophyticus* 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 *Staphylococcus saprophyticus* strain CNV2, isolated from crude oil-contaminated soil from the Noonmati oil refinery, Guwahati, Assam, India. *Genome Announc* 4(3):e00370-16. doi:10.1128/genomeA.00370-16.

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.

Staphylococcus saprophyticus is a Gram-negative, coagulase-negative cocci, commonly associated with urinary tract infections (1). Until now, very few studies have reported hydrocarbon-degrading attributes in an *S. saprophyticus* strain (2). However, to date no genome sequence has been available for *S. saprophyticus* inhabiting oil-contaminated environments. In the present study, we report sequencing of the entire genome of *S. saprophyticus* strain CNV2, isolated from crude oil-contaminated soil collected from the Noonmati oil refinery in Guwahati, Assam, India. Strain CNV2 was found to efficiently degrade n-hexane, n-hexadecane, diesel oil, and crude oil. Hence, the genome of this organism was sequenced to obtain better insights into the metabolic versatility and adaptability of this strain. To our knowledge, this is the first report of the genome sequence of an *S. saprophyticus* strain adapted to an oil-contaminated environment.

The genome of strain CNV2 was extracted using using an UltraClean Microbial DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) according to the manufacturer's guidelines. The purified genome was then sequenced using an Illumina HiSeq 2500 platform, generating 2,423,395 high-quality paired-end reads. Quality filtered reads were then assembled using *de novo* assemblers ABySS v3.81 (3), Edena v3.130110 (4), MaSuRCA v2.2.1 (5), SOAPdenovo2 v2.04 (6), SPAdes v3.1.1 (7), and Velvet v1.2.10 (8). Integration of the assembled contigs was carried out using CISA v1.3 (9) resulting in five contigs with a N_{50} length of 2,344,483 bp and an average length of 527,180 bp. The draft genome sequence was 2,635,899 bp in length, with a G+C content of 33.2% and 65-fold coverage. Genome annotation was carried out in the NCBI Prokaryotic Genome Annotation Pipeline which predicted the presence of 2,484 protein coding sequences (CDSs), 29 pseudogenes, 9 rRNAs, 42 tRNAs, 4 noncoding RNAs (ncRNAs), and a clustered regularly interspaced short palindromic repeat (CRISPR) array. Rapid genome annotation for CNV2 carried out

in the RAST annotation server (10) classified predicted CDSs into 398 subsystems, among which cofactors, vitamins, prosthetic groups and pigments ($s = 159$ CDSs), cell wall and capsule ($s = 108$), virulence, disease and defense ($s = 51$), RNA metabolism ($s = 125$), nucleosides and nucleotides ($s = 88$), protein metabolism ($s = 207$), DNA metabolism ($s = 77$), fatty acids, lipids, and isoprenoids ($s = 117$), stress response ($s = 75$), amino acids and derivatives ($s = 331$), and carbohydrates ($s = 290$) were the most abundant ones. The presence of an alkane-1-monooxygenase gene and large genetic investments in heavy metal and antibiotic resistance genes along with stress response genes indicate that CNV2 is highly adapted to different stress conditions, including oil contamination.

A comparison of strain CNV2 with genomes in the RAST database identified *S. saprophyticus* ATCC 15305 (score = 507) as its closest neighbor, followed by *S. equorum* strain Mu2 (score = 447) and *S. saprophyticus* KACC 16562 (score = 433). *S. epidermidis* M23864:W1 (score = 354) was identified as the eighth closest neighbor.

Nucleotide sequence accession numbers. This whole-genome shotgun sequencing project for *S. saprophyticus* strain CNV2 has been deposited in DDBJ/EMBL/GenBank under the accession no. [LUGM000000000](https://www.ncbi.nlm.nih.gov/nuccore/LUGM000000000). The version of the whole-genome sequence (WGS) described here is version LUGM01000000.

ACKNOWLEDGMENTS

The research reported in this article was supported by the NER TWINNING project (grant 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, Dhruvajyoti Chattopadhyay, and Arvind K. Singh, 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). A.M. was supported by the CSIR/UGC-NET fellowship from the UGC, Government of India.

REFERENCES

- Hovellius B, Mårdh PA. 1984. *Staphylococcus saprophyticus* as a common cause of urinary tract infections. *Rev Infect Dis* 6:328–337. <http://dx.doi.org/10.1093/clinids/6.3.328>.
- Namwong S, Tanasupawat S. 2014. Identification of *Staphylococcus* strain CH1-8 and its oil-degradation. *J Appl Pharm Sci* 4:24–29.
- 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, Formosa 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>.