



Complete Genome Sequence of *Citrobacter rodentium* Strain DBS100

Georgy Popov,^a Aline Fiebig-Comyn,^b Steven Shideler,^a Brian K. Coombes,^b  Alexei Savchenko^a

^aDepartment of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, Alberta, Canada

^bDepartment of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada

ABSTRACT *Citrobacter rodentium* strain DBS100 causes an infection of the intestines in mice. It provides an important model for human gastrointestinal pathogens, such as enteropathogenic and enterohemorrhagic *Escherichia coli*, which cause life-threatening infections. To identify the genetic determinants that are common across the enteropathogenic bacteria, we sequenced the DBS100 genome.

Citrobacter rodentium strain DBS100 is a Gram-negative murine enteropathogen (a generous gift from Brett Finlay, University of British Columbia) which is closely related to clinically important enteropathogenic and enterohemorrhagic *Escherichia coli* strains (1). The infection of mice by DBS100 is used as a model to elucidate the virulence mechanisms employed by bacterial enteropathogens to colonize the intestinal tract of their host (1). Mice orally infected with DBS100 develop colonic hyperplasia, infiltration of immune cells into the infection site, depletion of goblet cells, diarrhea, and lesions on the intestinal epithelia (1). DBS100 induces the formation of pedestal-like structures on the epithelial cell surface which support the pathogen's noninvasive attachment to the host (1). To identify genetic determinants associated with enteropathogenicity, we sequenced the DBS100 genome.

DBS100 was grown in Luria broth liquid medium at 37°C, and total genomic DNA (gDNA) was extracted using the DNeasy Ultraclean microbial kit (Qiagen, USA). The gDNA was sheared using Covaris sonication and processed with the SMRTbell template prep kit v1.0 (Pacific Biosciences, Canada) using the manufacturer's protocol, and 58,398,237 bp were sequenced with Pacific Biosciences RS II sequencing technology using one single-molecule real-time (SMRT) cell. The raw PacBio reads (genome coverage, 267×) were processed using the Hierarchical Genome Assembly Process (HGAP) workflow (2) with a cutoff of 30× to generate 6,848 corrected long subreads with minimal, average, and maximal read lengths of 500, 8,527, and 27,770 bases, respectively. The corrected long subreads were assembled into four contigs of 31,068, 42,783, 73,049, and 5,343,648 bp (genome coverage, 10.6×) using SMRT Analysis v2.3.0.140936.p5 (Pacific Biosciences). To close the circular chromosome of DBS100, we used PCR amplification of the 3,177-bp fragment between the 5' and 3' ends of the 5,343,648-bp contig using the Quick-Load Taq 2× master mix (New England BioLabs, Canada) and primers CCCTTGAACCCAGGCTACG and CGTCAACATCCGGTTATAGCG, and we sequenced this fragment using Sanger sequencing technology. Next, DBS100 gDNA was extracted using phenol chloroform treatment following purification and concentrated via ethanol precipitation (3). RNA contaminants were removed using RNase Cocktail (Invitrogen, USA) following the manufacturer's protocol. The gDNA was fragmented using Covaris sonication and processed with the NEBNext Ultra II library preparation for Illumina kit (New England Biolabs, USA) using the manufacturer's protocol, and 352.97 Mbp were sequenced with Illumina MiSeq v2 technology (2 × 150-bp paired ends). The adapter sequences at the 5' ends and 10 bases at the 3'

Citation Popov G, Fiebig-Comyn A, Shideler S, Coombes BK, Savchenko A. 2019. Complete genome sequence of *Citrobacter rodentium* strain DBS100. Microbiol Resour Announc 8:e00421-19. <https://doi.org/10.1128/MRA.00421-19>.

Editor Christina A. Cuomo, Broad Institute

Copyright © 2019 Popov 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 Alexei Savchenko, alexei.savchenko@ucalgary.ca.

Received 1 May 2019

Accepted 22 May 2019

Published 13 June 2019

ends of the raw Illumina reads were trimmed using Cutadapt v1.15 (4) to obtain 1,176,564 short reads (genome coverage, 27×) with minimal, average, and maximal lengths of 40, 127.28, and 136 bases, respectively. The trimmed short reads were analyzed for sequence quality using FastQC v0.11.5 (quality control score > 20) (5). Hybrid assembly of the Illumina-derived short reads and PacBio-derived contigs was done using Unicycler v0.4.7 (6), SPAdes v3.12.0 (7), Minimap (8), and Pilon v1.23 (9) to generate nine contigs (N_{50} , 5,232,658 bp). The following four out of the nine contigs were identical to previously reported plasmids: pCROD1 (GenBank accession number [FN543503](#)), pCROD2 ([FN543504](#)), pCROD3 ([FN543505](#)), and pCRP3 ([NC_003114](#)). The linear contigs corresponded to the large circular contig derived from the PacBio sequencing and were used to finalize the sequence of the strain DBS100 chromosome. The NCBI Prokaryotic Genome Annotation Pipeline (best-placed reference protein, v4.8) was used to annotate 5,186 genes in the 5,346,827-bp-long contig with a 54.69% GC content DBS100 chromosome. Out of the 5,186 genes, 4,788 are protein coding genes, 116 are RNA coding genes, and 282 are pseudogenes.

Data availability. The chromosomal genome sequences have been deposited in GenBank under the accession number [CP038008](#). The sequences of the PacBio-filtered reads and Illumina raw reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number [PRJNA527323](#).

ACKNOWLEDGMENT

Data were obtained using a Canadian Institutes of Health Research (CIHR) operating grant to A.S. (principal investigator) and B.K.C. (coprincipal investigator).

REFERENCES

1. Crepin VF, Collins JW, Habibzay M, Frankel G. 2016. *Citrobacter rodentium* mouse model of bacterial infection. *Nat Protoc* 11:1851–1876. <https://doi.org/10.1038/nprot.2016.100>.
2. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
3. Sambrook J, Russell DW. 2006. Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harb Protoc* 2006:pdb.prot4455. <https://doi.org/10.1101/pdb.prot4455>.
4. Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30:3506–3514. <https://doi.org/10.1093/bioinformatics/btu538>.
5. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. 2013. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front Genet* 4:288. <https://doi.org/10.3389/fgene.2013.00288>.
6. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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. <https://doi.org/10.1089/cmb.2012.0021>.
8. Li H. 2016. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics* 32:2103–2110. <https://doi.org/10.1093/bioinformatics/btw152>.
9. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.