AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of the Type Strain *Citrobacter rodentium* DSM 16636

Anika Wahl,[a] **Klaus Neuhaus**[a]

[a]Core Facility Microbiome, ZIEL-Institute for Food & Health, Technical University Munich, Freising, Germany

ABSTRACT The type strain *Citrobacter rodentium* DSM 16636 was characterized in 1995. This species is widely used in rodents to study the virulence of locus-of-enterocyte-effacement-type pathogens, such as enterohemorrhagic *Escherichia coli*. The type strain had not been sequenced yet. Here, we report the closed genome (5.3 Gbp) and its plasmid (39.3 kbp).

Citrobacter rodentium was first described in 1995 by Schauer et al. (1) as mesophilic, Gram-negative, rod-shaped enterobacterium, originally isolated from mice. Strain CDC1843-73, first classified as an atypical *Citrobacter freundii* strain, was later defined as the type strain, DSM 16636 (1).

C. rodentium is a natural enteric mouse pathogen and it causes transmissible murine colonic hyperplasia. Due to its similarity in patho-mechanisms to enterohemorrhagic and enteropathogenic *Escherichia coli*, as it also has a locus of enterocyte effacement (LEE), this species is widely used as a surrogate for human gastrointestinal diseases in mouse models. Other core virulence factors are shared as well (2–4). For in-depth analysis, complete genomes are important. However, while two other *C. rodentium* genomes are published (i.e., ICC 168 and DBS 100 [2, 5]), the type-strain genome was only available in fragments.

The type strain is deposited in several strain collections (e.g., DSM 16636, ATCC 51116, NBRC 105723, CIP 104675, CCUG 30795, JCM 14073, CCM 7398), including in the Weihenstephan Strain Collection as WS 4383. Here, we analyzed the strain from the Weihenstephan Strain Collection, which had been previously obtained from DSMZ (Braunschweig, Germany). *C. rodentium* was grown aerobically in liquid lysogeny broth at 37°C. Total genomic DNA was extracted using phenol/chloroform extraction with CTAB (cetrimonium bromide) (6). Coextracted RNA was digested with RNase A (20 mg/mL; Thermo Fisher Scientific, USA) according to the manufacturer's protocol.

The extracted DNA (without previous shearing) was sequenced by SNPsaurus (Oregon, USA). Sequencing was performed using Pacific Biosciences (PacBio) RS II sequencing technology using one single-molecule real-time (SMRT) cell. SMRTbell libraries were made from genomic DNA (gDNA) using the Express Template prep kit 2.0 from PacBio according to the manufacturer's protocol v2.0. Samples were pooled into a single multiplexed library and size-selected using Sage Sciences' BluePippin system according to the manufacturer's recommendations using the 0.75% DF marker S1 high-pass 6-kb to 10-kb v3 run protocol and S1 marker. A size selection cutoff of 8,000 bp (start value) was used. The size-selected SMRTbell library was annealed and bound according to the SMRT Link setup and sequenced on a Sequel II instrument.

Next, 45,641 raw PacBio reads were converted to fasta format with SAMtools v1.9 (7). Genome assembly was performed with Flye v2.8.3-b1695 with the following parameters: –plasmids –iterations 2 –asm-coverage 120 (8, 9); the results were assembled into a genome with a total length of 5,381,137 bp with a mean coverage of 84×, consisting of two contigs. One contig consists of the genome (N_{50}, 5,341,875 bp; coverage,

Editor Irene L. G. Newton, Indiana University, Bloomington

Address correspondence to Klaus Neuhaus, neuhaus@tum.de.

The authors declare no conflict of interest.

Received 28 December 2021
Accepted 24 March 2022
Published 5 April 2022

82×; GC content, 54.63%), and the other contig is a plasmid named pCRTS (for plasmid of *C. rodentium*'s type strain) with 39,262 bp and a coverage of 384×. Thus, we expect about 4 to 5 plasmids per genome under the growth conditions used here. Genome annotation was performed by NCBI with PGAP v5.3 (10, 11). The prediction resulted in 4,992 coding sequences, including 22 rRNAs, 2 repeat regions, 88 tRNAs, 1 transfer-messenger RNA (tmRNA), and 4,326 unique protein-coding genes.

Data availability. The raw sequence reads have been deposited at the NCBI Sequence Read Archive under the SRA accession number SRR18162050 (BioProject number PRJNA759082, BioSample number SAMN21155799). The GenBank accession numbers are CP082833 (genome) and CP082834 (pCRTS). For all sequences, the first versions are described in this paper.

REFERENCES

1. Schauer DB, Zabel BA, Pedraza IF, O'Hara CM, Steigerwalt AG, Brenner DJ. 1995. Genetic and biochemical characterization of *Citrobacter rodentium* sp. nov. J Clin Microbiol 33:2064–2068. https://doi.org/10.1128/jcm.33.8.2064-2068.1995.
2. Petty NK, Bulgin R, Crepin VF, Cerdeño-Tárraga AM, Schroeder GN, Quail MA, Lennard N, Corton C, Barron A, Clark L, Toribio AL, Parkhill J, Dougan G, Frankel G, Thomson NR. 2010. The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. J Bacteriol 192:525–538. https://doi.org/10.1128/JB.01144-09.
3. 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.
4. Collins JW, Keeney KM, Crepin VF, Rathinam VAK, Fitzgerald KA, Finlay BB, Frankel G. 2014. *Citrobacter rodentium*: infection, inflammation and the microbiota. Nat Rev Microbiol 12:612–623. https://doi.org/10.1038/nrmicro3315.
5. Popov G, Fiebig-Comyn A, Shideler S, Coombes BK, Savchenko A. 2019. Complete genome sequence of *Citrobacter rodentium* strain DBS100. Microbiol Resour Announc 8:1–2. https://doi.org/10.1128/MRA.00421-19.
6. 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.
7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
8. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17:1103–1110. https://doi.org/10.1038/s41592-020-00971-x.
9. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
10. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
11. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.