



# Complete Genome Sequence of *Staphylococcus edaphicus* Strain CCM 8731

Callum O. Rimmer,<sup>a</sup>  Jonathan C. Thomas<sup>a</sup>

<sup>a</sup>School of Science and Technology, Nottingham Trent University, Nottingham, United Kingdom

**ABSTRACT** Here, we present the complete genome sequence of *Staphylococcus edaphicus* strain CCM 8731, which was originally isolated from Ross Island, Antarctica. The 2,749,487-bp sequence contains 2,709 predicted genes, with a G+C content of 33.4%. The complete genome was assembled using a hybrid approach with Oxford Nanopore Technologies long-read sequencing and Illumina short-read sequencing.

**S**taphylococcal species have been isolated from a diverse array of human and environmental sources; however, few have been isolated from extreme environments (1). A complete genome for this species will accurately depict genome organization within the strain and serve as a high-quality reference for future studies (2).

*Staphylococcus edaphicus* strain CCM 8731 was isolated and cultivated by Pantůček et al. (1). We obtained *S. edaphicus* CCM 8731 as a freeze-dried sample from the Czech Collection of Microorganisms (<http://www.sci.muni.cz/ccm>). The strain was revived in Trypticase soy broth (Oxoid) and streaked onto Trypticase soy agar (TSA) for single colonies. A colony was transferred to 10 mL of brain heart infusion broth (Oxoid) and incubated overnight at 37°C. Bacterial cells were suspended in 200  $\mu$ l of buffer EB (Qiagen) and lysed by incubation at 37°C with 80  $\mu$ l of lysostaphin (240 units; Sigma). High-molecular-weight genomic DNA was extracted using Qiagen 100/G Genomic-tip columns.

Illumina library preparation was performed using the standard Illumina Nextera XT protocol, and the library was sequenced on a MiSeq system using v2 chemistry for 250-bp paired-end reads, producing approximately 110 Mbp of raw data (221,353 paired-end reads). Nanopore library preparation was performed using the ligation sequencing kit SQK-LSK109 (Oxford Nanopore Technologies) and native barcoding, with the library loaded onto a Flongle flow cell (R9.4.1). DNA was size selected using AMPure XP beads (Beckman Coulter, UK) but was not sheared. Approximately 145.5 Mbp of raw data was generated (21,236 reads, with an  $N_{50}$  value of 12.3 kbp).

Sequencing reads were assembled using default parameters for all software except where otherwise noted. Fast5 sequencing reads were basecalled using Guppy v3.6.1 (Oxford Nanopore Technologies). Adapter sequences were removed using Porechop v0.2.3 (3) with middle and end thresholds of 85% and 95%, respectively. Filtlong v0.2.1 (4) was used to filter reads based on quality and length. Overlapping reads were assembled into one contiguous sequence using Canu v1.4 (5). This assembly was polished using four iterations of Racon v1.4.2 (6) followed by Medaka v1.5 (7) and Nanopolish v0.13.2 (8). Illumina sequencing reads were trimmed using Trimmomatic v0.39 (9) to remove adapter sequences, leaving 160,478 trimmed paired-end reads. The Nanopolish-outputted genome was then error corrected with trimmed Illumina reads using Racon and Pilon v1.24 (10). The genome was manually trimmed and reorientated to *dnaA* with Circlator v1.5.5 (11) to produce a 2,749,487-bp assembly with a final combined coverage from Nanopore and Illumina data of 47.3 $\times$ .

The complete genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6 (12), which predicted 2,709 genes, 19 rRNAs, and 61 tRNAs. Assembly quality was determined by CheckM v1.1.3 (13), with a G+C content of 33.4% and completeness

**Editor** Steven R. Gill, University of Rochester School of Medicine and Dentistry

**Copyright** © 2022 Rimmer and Thomas. 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 Jonathan C. Thomas, [jonathan.thomas@ntu.ac.uk](mailto:jonathan.thomas@ntu.ac.uk).

The authors declare no conflict of interest.

**Received** 20 May 2022

**Accepted** 20 July 2022

**Published** 10 August 2022

of 99.45%. Average nucleotide identity (ANI) was calculated using fastANI v1.1 (14), showing 99.92% similarity to *S. edaphicus* CCM 8730<sup>T</sup> and 85.44% similarity to its closest relative, *Staphylococcus saprophyticus* ATCC 15305<sup>T</sup>.

**Data availability.** All data are available under BioProject accession number [PRJNA808309](https://ncbi.nlm.nih.gov/bioproject/PRJNA808309) and BioSample accession number [SAMN26082960](https://ncbi.nlm.nih.gov/biosample/SAMN26082960). The genome was deposited under GenBank accession number [CP093217](https://ncbi.nlm.nih.gov/nuccore/CP093217). Illumina raw read data and Nanopore base-called data were submitted to SRA and are available via accession numbers [SRR18070734](https://ncbi.nlm.nih.gov/sra/SRR18070734) and [SRR18070733](https://ncbi.nlm.nih.gov/sra/SRR18070733), respectively.

## REFERENCES

1. Pantůček R, Sedláček I, Indráková A, Vrbová V, Mašláňová I, Kovařovic V, Švec P, Králová S, Křištofová L, Kulikov J, Petrás P, Doškař J. 2018. *Staphylococcus edaphicus* sp. nov., isolated in Antarctica, harbors the *mecC* gene and genomic islands with a suspected role in adaptation to extreme environments. *Appl Environ Microbiol* 84:e01746-17. <https://doi.org/10.1128/AEM.01746-17>.
2. Fraser MC, Eisen JA, Nelson KE, Paulsen IT, Salzberg SL. 2002. The value of complete microbial genome sequencing (you get what you pay for). *J Bacteriol* 184:6403–6405. <https://doi.org/10.1128/JB.184.23.6403-6405.2002>.
3. Wick RR. 2017. Porechop. <https://github.com/rrwick/Porechop>.
4. Wick RR. 2017. Filtlong. <https://github.com/rrwick/Filtlong>.
5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
6. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>.
7. Oxford Nanopore Technologies. 2017. Medaka. <https://github.com/nanoporetech/medaka>.
8. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. 2017. Detecting DNA cytosine methylation using Nanopore sequencing. *Nat Methods* 14:407–410. <https://doi.org/10.1038/nmeth.4184>.
9. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
10. 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>.
11. Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>.
12. 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>.
13. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
14. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.