



HAL
open science

Complete genome sequence of *Sporosarcina pasteurii* type strain DSM33

Matthew Tuttle, Blake Stamps, John Chodkowski, Marco Mechan-Llontop,
Ashley Shade, Chia-Suei Hung, Maneesh Gupta, Michael Carter

► **To cite this version:**

Matthew Tuttle, Blake Stamps, John Chodkowski, Marco Mechan-Llontop, Ashley Shade, et al..
Complete genome sequence of *Sporosarcina pasteurii* type strain DSM33. *Microbiology Resource
Announcements*, 2024, 40, pp.2379 - 2396. 10.1007/s10706-021-02041-1 . hal-04795333

HAL Id: hal-04795333

<https://hal.science/hal-04795333v1>

Submitted on 21 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 **Complete genome sequence of *Sporosarcina pasteurii* Type Strain DSM33**

2
3 **Authors**

4
5 Matthew J. Tuttle^{a,b}, Blake W. Stamps^a, John L. Chodkowski^c, Marco E. Mechan-Llontop^c, Ashley Shade^d,
6 Chia-Suei Hung^a, Maneesh K. Gupta^a, Michael S. Carter^{a#}

7
8 **Affiliations**

9
10 a. Materials and Manufacturing Directorate, Air Force Research Lab, Wright-Patterson Air Force Base,
11 Dayton, Ohio, USA

12 b. Biological and Nanoscale Technologies Division, UES Inc., a BlueHalo Company, Dayton, Ohio, USA

13 c. Department of Microbiology, Immunology and Genetics, Michigan State University, East Lansing,
14 Michigan, USA

15 d. Universite Claude Bernard Lyon 1; CNRS, UMR 5557 Ecologie Microbienne, 16 rue Dubois 69622
16 Villeurbanne, FRANCE

17
18 #Address correspondence to Michael S. Carter, michael.carter.90@us.af.mil.

19
20 **Running title**

21
22 Complete genome of *Sporosarcina pasteurii* DSM33

23

24 **Abstract**

25

26 Here, we present the complete genome sequence of *Sporosarcina pasteurii* Type Strain DSM33, a model
27 organism for microbially-induced calcium carbonate precipitation. This genome consists of a single 3.3
28 Mb chromosome and is an improvement upon draft *S. pasteurii* genome sequences currently available
29 in public databases.

30

31 **Announcement**

32

33 *Sporosarcina pasteurii* is a Gram positive endospore-forming alkaliphilic bacterium first isolated in 1889
34 (1, 2). Most strains produce high concentrations of urease (3). *S. pasteurii* DSM33 (also known as ATCC
35 11859) and other *S. pasteurii* strains have emerged as models for the production of biocement via
36 microbially-induced calcium carbonate precipitation, given their high levels of urease activity (4–6). *S.*
37 *pasteurii* DSM33 is the most commonly used strain, so here we sequenced and assembled its complete
38 genome (Figure 1).

39

40 *S. pasteurii* DSM33 was obtained from the German Collection of Microorganisms and Cell Cultures
41 (Leibniz Institute DSMZ). DSM33 was grown in brain-heart infusion broth (MilliporeSigma) amended
42 with 2% (w/v) urea (MilliporeSigma) for 18 h at 30°C with 250 rpm shaking prior to DNA extraction. For
43 Nanopore sequencing (Oxford Nanopore Technologies), DNA was isolated via phenol-chloroform
44 extraction (7). Nanopore sequencing libraries were prepared using a Ligation sequencing kit (SQK-
45 LSK109; Oxford Nanopore Technologies). Libraries were sequenced using a FLO-MIN111 flow cell on a
46 MinION sequencer. Base calling was performed using Guppy (high-accuracy model; v5.0.16; Oxford
47 Nanopore Technologies), reads were quality controlled using MinIONQC (v1.4.1; $Q \geq 10$) (8), and
48 adapters trimmed using Porechop (v0.2.4) (9). A total of 76442 raw reads were obtained with an N50 of
49 22709 bp. For Illumina sequencing, DNA was extracted from an independent sample of DSM33 using a
50 DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer's protocol for Gram positive bacteria.
51 Illumina libraries were prepared using a SeqWell purePlex™ DNA Library Prep Kit, per the manufacturer's
52 directions. After preparation, the library pool quality was assessed using a QuBit BR assay (ThermoFisher
53 Scientific) for concentration and insert size using an Agilent TapeStation D5000 kit. The sample was then
54 denatured, diluted to a load concentration of 10 pM, and 250 bp paired-end reads were sequenced
55 using an Illumina MiSeq 500 cycle V2 kit, yielding 6.27 million raw reads. Illumina reads were trimmed
56 for quality, adapters, and length (minimum length 150 bp, minimum Q score 30) using FastP. DNA
57 assembly of Nanopore reads was first performed using the Tricycler pipeline (v0.5.3) (10) using default
58 parameters with Flye (v2.9.1-b1780) (11), Raven (v1.8.1) (12), and miniasm (v0.3-r179) (13) as the
59 assemblers. Post consensus, in Tricycler, the assembly was polished with the Illumina reads using
60 Polypolish (v0.5.0) and POLCA (within mascara v4.1.0). Annotations were generated using the NCBI
61 Prokaryotic Genome Annotation Pipeline (PGAP; v6.7) (14).

62

63 The final genome assembly comprised a complete genome sequence consisting of a single circular
64 chromosome 3.3 Mb in size with a G+C content of 39.2%. Average nanopore read coverage was 109-
65 fold. Tricycler assembly (10) circularized the genome which was rotated by the assembler to place *dnaA*
66 at the start of the contig. Genome annotation resulted in 3115 protein-coding sequences (CDSs), 8 rRNA
67 operons, 69 tRNAs, and 9 ncRNAs.

68

69 **Data availability**

70

71 The genome sequence of *S. pasteurii* DSM33 is available in GenBank under accession number CP160452.
72 Raw sequence reads are available under the Sequence Read Archive (SRA) accession numbers
73 SRR28831121 (Illumina) and SRR28967830 (Oxford Nanopore Technologies).
74

75 **Acknowledgments**

76
77 This work was supported by the Air Force Research Laboratory Seedlings for Disruptive Capabilities
78 Program and has been cleared for public release under reference number AFRL-2024-2064.
79

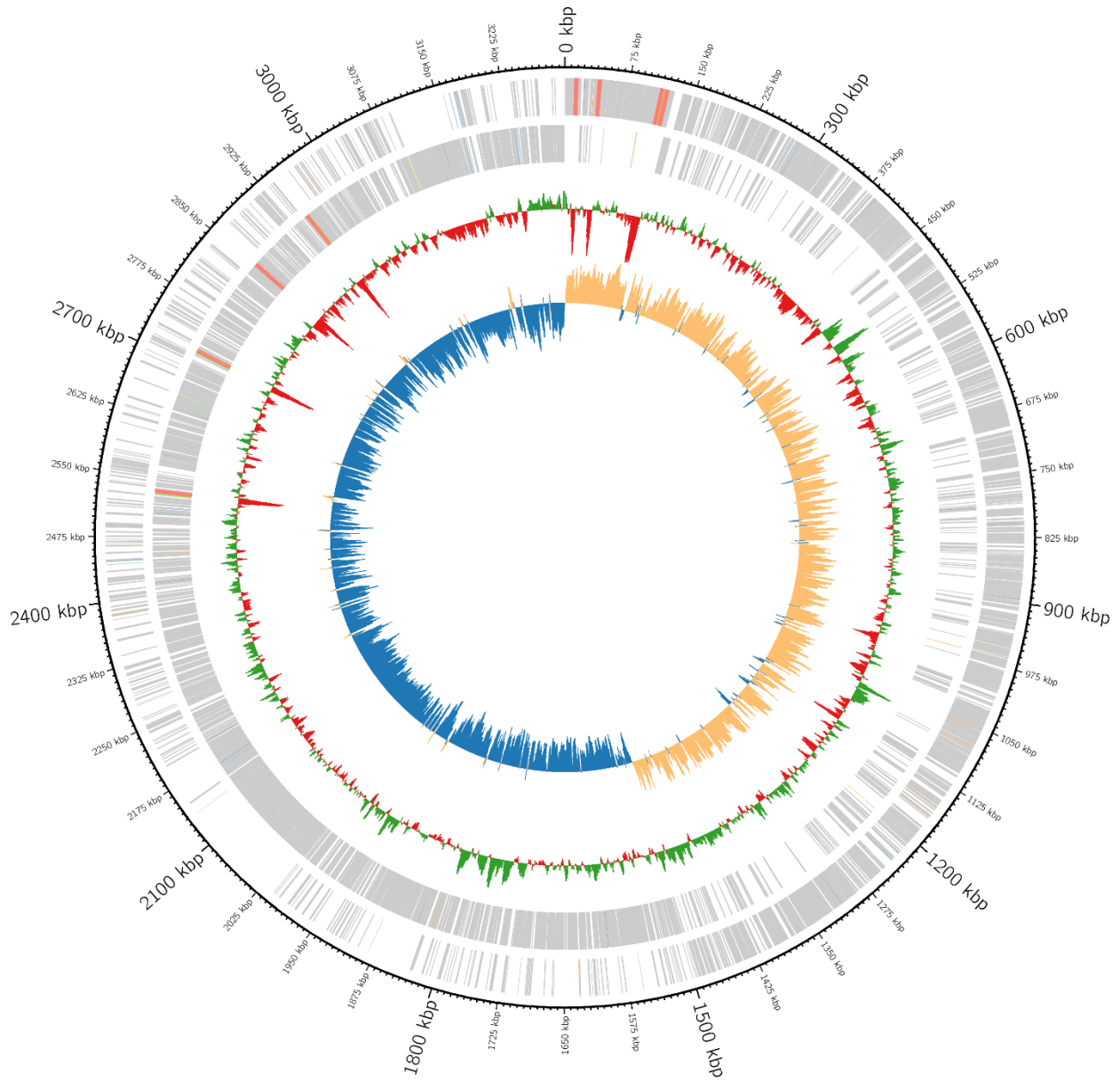
80 **References**

- 81
- 82 1. Kluuyver and van Niel 1936, 401AL emend. Yoon, Lee, Weiss, Kho, Kang and Park 2001b, 1085.
83 2015. *Sporosarcina*, p. 1–7. *In* Bergey’s Manual of Systematics of Archaea and Bacteria.
 - 84 2. Miquel P. 1889. ÉTUDE SUR LA FERMENTATION AMMONIACALE ET SUR LES FERMENTS DE L’URÉE.
85 *Ann Microgr* 2:13–33.
 - 86 3. Mobley HLT, Hausinger RP. 1989. Microbial Ureases: Significance, Regulation, and Molecular
87 Characterization 53:85–108.
 - 88 4. Chuo SC, Mohamed SF, Setapar SHM, Ahmad A, Jawaid M, Wani WA, Yaqoob AA, Ibrahim MNM.
89 2020. Insights into the current trends in the utilization of bacteria for microbially induced calcium
90 carbonate precipitation. *Materials* 13:1–28.
 - 91 5. Pacheco VL, Bragagnolo L, Reginatto C, Thomé A. 2022. Microbially Induced Calcite Precipitation
92 (MICP): Review from an Engineering Perspective. *Geotech Geol Eng*
93 <https://doi.org/10.1007/s10706-021-02041-1>.
 - 94 6. Carter MS, Tuttle MJ, Mancini JA, Martineau R, Hung C-S, Gupta MK. 2023. Microbially Induced
95 Calcium Carbonate Precipitation by *Sporosarcina pasteurii*: a Case Study in Optimizing Biological
96 CaCO₃ Precipitation. *Appl Environ Microbiol* 0:e01794-22.

- 97 7. Carter MS, Tuttle MJ, Gupta MK. 2024. Phenol-chloroform DNA extraction from *Sporosarcina*
98 *pasteurii*. protocols.io <https://doi.org/dx.doi.org/10.17504/protocols.io.kqdg324eev25/v1>.
- 99 8. Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2019. MinIONQC: fast and simple
100 quality control for MinION sequencing data. *Bioinformatics* 35:523–525.
- 101 9. Wick, Ryan R. 2018. Porechop: adapter trimmer for Oxford Nanopore reads (0.2.4).
- 102 10. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Tricycler:
103 consensus long-read assemblies for bacterial genomes. *Genome Biol* 22:266.
- 104 11. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat
105 graphs. 5. *Nat Biotechnol* 37:540–546.
- 106 12. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. 5. *Nat Comput*
107 *Sci* 1:332–336.
- 108 13. Wick RR, Holt KE. 2019. Benchmarking of long-read assemblers for prokaryote whole genome
109 sequencing. *F1000Research* 8:2138.
- 110 14. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt
111 KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res*
112 44:6614–6624.
- 113 15. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. 2021. Bakta: rapid and
114 standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb*
115 *Genomics* 7:000685.

- 116 16. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009.
117 Circos: An information aesthetic for comparative genomics. *Genome Res* 19:1639–1645.

118



119
 120 **Figure 1. Circular genome map of the complete *S. pasteurii* DSM33 genome.** Moving from the outer
 121 circle inward, the five circular tracks represent: (i) sequence location in kbp, (ii) forward strand genome
 122 features, (iii) reverse strand genome features, (iv) GC content relative to average, and (v) GC skew.
 123 Forward and reverse strand features include CDSs (gray), tRNA/tRNAs (green), rRNAs (red), ncRNAs
 124 (orange), and ncRNA regions (blue). This plot was generated by Bakta (15) using Circos (16).