



# Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* SD96

 Robin Dorau,<sup>a</sup>  Jun Chen,<sup>a</sup> Peter Ruhdal Jensen,<sup>a</sup> Christian Solem<sup>a</sup>

<sup>a</sup>National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark

**ABSTRACT** The genome of *Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* SD96, a strain used for cheese production, is presented. SD96 is refractory to phage attack, which is a desired property for starter bacteria. Its 10 plasmids provide industrially important traits, such as lactose and citrate metabolism, proteolytic activity, and phage resistance.

*Lactococcus lactis*, considered to be the model organism of lactic acid bacteria, has a long history of use in the dairy industry. Many mesophilic starter cultures used for cheese production contain several lactic acid bacteria, which are usually *Lactococcus* and *Leuconostoc* species (1). *Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* is present in many mesophilic starter cultures and is responsible for a buttery aroma and gas formation from citrate. It has been shown that large genetic variety exists among lactococci, even between members of the same subspecies (2, 3).

Here, we present the whole genome of *L. lactis* subsp. *lactis* bv. *diacetylactis* SD96, a bacteriophage-insensitive mutant of strain SD16, which was originally isolated from an Italian cheese by starter culture provider Sacco S.r.l. To date, none of the phages contained in the comprehensive phage collection at Sacco S.r.l. has been found to be able to infect SD96, and dairies using SD96 have not reported any phage attack for this strain. Because of the contributions of plasmid-encoded genes to industrially relevant phenotypes, we were particularly interested in the plasmid content of this strain.

DNA for sequencing was isolated from an early-stationary-phase culture in M17 broth supplemented with 0.5% lactose, which was inoculated from a single colony. Total DNA was isolated using a genomic DNA purification kit (product number K0512) from Thermo Scientific. Sequencing was carried out on both PacBio Sequel and Illumina HiSeq 4000 platforms, using two distinct DNA preparations. Illumina sequencing using a PE150 PCR-free library provided 8.78 million raw paired-end 150-bp reads (SRA accession number [SRX6686433](https://www.ncbi.nlm.nih.gov/sra/SRX6686433)). The sequences were processed and assembled using various software programs with default settings unless otherwise stated. Raw Illumina reads were processed using Geneious Prime (version 2019.1.3) (4). Briefly, adapter sequences and low-quality sequences were trimmed from both ends; finally, all reads above 50 bp were kept. This process resulted in 7.25 million sequences, with an average length of 138 bp. From the PacBio Sequel project using a 20-kb library, we received 331,855 raw reads, with an average length of 7,575 bp and a maximal read length of 44,814 bp (SRA accession number [SRX6711783](https://www.ncbi.nlm.nih.gov/sra/SRX6711783)). For *de novo* assembly, the PacBio reads alone were processed and assembled using the Canu assembler (version 1.8) (5), which provided 13 linear scaffolds. The Circlator tool (version 1.5.5) (6) provided a closed chromosome and four plasmids. Another six plasmids were closed by identifying repeating regions on the scaffolds using the Repeat Finder plugin (version 1.0) in Geneious, followed by manual evaluation and trimming. Furthermore, the presence of replication-associated genes was confirmed for all 10 plasmids by sequence comparison with previously characterized *L. lactis* plasmids. Two scaffolds (4,944 and 5,967 bp)

**Citation** Dorau R, Chen J, Jensen PR, Solem C. 2020. Complete genome sequence of *Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* SD96. *Microbiol Resour Announc* 9:e01140-19. <https://doi.org/10.1128/MRA.01140-19>.

**Editor** David A. Baltrus, University of Arizona

**Copyright** © 2020 Dorau 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 Christian Solem, [chso@food.dtu.dk](mailto:chso@food.dtu.dk).

**Received** 7 October 2019

**Accepted** 18 December 2019

**Published** 16 January 2020

could not be closed or associated with the genome and did not contain replication-associated genes. Finally, the circularized sequences were polished using Pilon (version 1.23), for which the processed Illumina paired-end reads had been mapped to the circular sequences using Bowtie2 (version 2.3.0) in Geneious. The genome and plasmids were annotated using the RAST server (version 2.0) (7).

The 2.42-Mbp chromosome of SD96 has a GC content of 35.3%. The 10 plasmids have an average GC content of 34.4% (minimum, 30.3%; maximum, 36.5%) and sizes of 64.9, 49.6, 37.4, 25.4, 11.7, 8.3, 7.9, 7.3, 4.6, and 2.7 kbp (GenBank accession numbers CP043518 to CP043528). Plasmids associated with protein degradation, lactose and citrate metabolism, and phage defense, as well as two cryptic plasmids with unknown purpose, were identified.

**Data availability.** The genome sequences of *L. lactis* subsp. *lactis* biovar diacetylactis SD96, including its 10 plasmids, are available in GenBank under accession numbers CP043523 (chromosome), CP043525 (pSD96\_01), CP043526 (pSD96\_02), CP043527 (pSD96\_03), CP043528 (pSD96\_04), CP043524 (pSD96\_05), CP043518 (pSD96\_06), CP043519 (pSD96\_07), CP043520 (pSD96\_08), CP043521 (pSD96\_09), and CP043522 (pSD96\_10). Raw data have been deposited in the Sequence Read Archive (SRA) under accession numbers SRX6686433 (Illumina data) and SRX6711783 (PacBio Sequel data).

## ACKNOWLEDGMENT

We thank Sacco S.r.l. for kindly providing us with *L. lactis* subsp. *lactis* bv. diacetylactis SD96.

## REFERENCES

1. Leroy F, De Vuyst L. 2004. Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends Food Sci Technol* 15:67–78. <https://doi.org/10.1016/j.tifs.2003.09.004>.
2. Erkus O, de Jager VC, Spus M, van Alen-Boerrigter IJ, van Rijswijck IM, Hazelwood L, Janssen PW, van Hijum SA, Kleerebezem M, Smid EJ. 2013. Multifactorial diversity sustains microbial community stability. *ISME J* 7:2126–2136. <https://doi.org/10.1038/ismej.2013.108>.
3. Frantzen CA, Kleppen HP, Holo H. 2018. *Lactococcus lactis* diversity in undefined mixed dairy starter cultures as revealed by comparative genome analyses and targeted amplicon sequencing of *epsD*. *Appl Environ Microbiol* 84:e02199-17. <https://doi.org/10.1128/AEM.02199-17>.
4. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
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. 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>.
7. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma 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. <https://doi.org/10.1186/1471-2164-9-75>.