

**UCC Library and UCC researchers have made this item openly available.
Please [let us know](#) how this has helped you. Thanks!**

| | |
|------------------------------------|--|
| Title | Complete genome sequence of Lactococcus lactis subsp. Cremoris 3107, host for the model Lactococcal P335 Bacteriophage TP901-1 |
| Author(s) | Erazo Garzon, Andrea; Mahony, Jennifer; Bottacini, Francesca; Kelleher, Philip; van Sinderen, Douwe |
| Publication date | 2019-01-17 |
| Original citation | Garzon, A. E., Mahony, J., Bottacini, F., Kelleher, P. and van Sinderen, D. (2019) 'Complete genome sequence of Lactococcus lactis subsp. cremoris 3107, host for the model lactococcal P335 bacteriophage TP901-1', Microbiology Resource Announcements, 8(3), e01635-18 (3pp). doi: 10.1128/MRA.01635-18 |
| Type of publication | Article (peer-reviewed) |
| Link to publisher's version | https://mra.asm.org/content/ga/8/3/e01635-18.full.pdf http://dx.doi.org/10.1128/MRA.01635-18 Access to the full text of the published version may require a subscription. |
| Rights | © 2019, Erazo Garzon 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 |
| Item downloaded from | http://hdl.handle.net/10468/7536 |

Downloaded on 2023-02-05T08:30:45Z



Complete Genome Sequence of *Lactococcus lactis* subsp. *cremoris* 3107, Host for the Model Lactococcal P335 Bacteriophage TP901-1

Andrea Erazo Garzon,^a Jennifer Mahony,^{a,b} Francesca Bottacini,^{a,b} Philip Kelleher,^{a,b} Douwe van Sinderen^{a,b}

^aSchool of Microbiology, University College Cork, Cork, Ireland

^bAPC Microbiome Ireland, University College Cork, Cork, Ireland

ABSTRACT The complete genome sequence of *Lactococcus lactis* subsp. *cremoris* 3107, a dairy starter strain and a host for the model lactococcal P335 bacteriophage TP901-1, is reported here. The circular chromosome of *L. lactis* subsp. *cremoris* 3107 is among the smallest genomes of currently sequenced lactococcal strains. *L. lactis* subsp. *cremoris* 3107 harbors a complement of six plasmids, which appears to be a reflection of its adaptation to the nutrient-rich dairy environment.

The dairy fermentation industry relies on starter or adjunct cultures, predominantly composed of lactic acid bacteria (LAB), to produce high-quality end products (1). Like all bacteria, LAB strains are susceptible to infection by bacteriophages (phages), which are ubiquitous in the dairy environment and largely insensitive to pasteurization treatments (2). A major economic concern for the dairy fermentation industry is phage attack of starter and adjunct strains during the fermentation process, which may negatively impact the final product quality and production regimes; this may thus lead to significant economic losses (3). Among the LAB, strains of the industrially significant species *Lactococcus lactis* are particularly susceptible to infection by phages (4, 5). *Lactococcus lactis* subsp. *cremoris* 3107 is the host of the model lysogenic lactococcal P335 phage TP901-1 (6). Because this *L. lactis* strain is host to an important model phage, its genome was sequenced to better characterize genes involved in phage-host interactions. The genome consists of a single 2.4-Mb chromosome (36% GC content) and six plasmids ranging in size from 2 to 60 kb, and analysis of the genes carried on both the chromosome and plasmids suggests that the plasmids are required for metabolism in the nutrient-rich dairy environment.

For PacBio sequencing, *L. lactis* subsp. *cremoris* 3107 cell pellets (from an overnight culture; see below) containing 10⁹ CFU were provided to GATC Biotech Ltd. (Germany) to perform chromosomal DNA extraction, library construction, and single-molecule real-time (SMRT) sequencing on Pacific Biosciences RS (run 1) and RS II (run 2) sequencing platforms. The library for PacBio sequencing was prepared using the SMRTbell template prep kit with 8- to 12-kb inserts, according to the manufacturer's instructions (Pacific Biosciences, Menlo Park, CA, USA). *De novo* genome assembly of the SMRT sequencing data was performed using the RS_HGAP_Assembly.2 protocol (default parameters) implemented in the Pacific Biosciences SMRT Analysis portal (version 2.3.1). Quality filtering was performed automatically during assembly using the SMRT Portal P-filter module. Two SMRT cells were used for PacBio sequencing to achieve an initial assembly of the derived 52,984 filtered reads into 23 contigs, with an *N*₅₀ contig length of 337,497 bp and an average reference coverage of 60.37×.

For Illumina-based sequencing, 5 μg chromosomal DNA from *L. lactis* subsp. *cremoris* 3107 was extracted using phenol-chloroform-based extractions, as previously described (7), following overnight growth of the strain at 30°C in M17 broth (Oxoid, UK)

Citation Erazo Garzon A, Mahony J, Bottacini F, Kelleher P, van Sinderen D. 2019. Complete genome sequence of *Lactococcus lactis* subsp. *cremoris* 3107, host for the model lactococcal P335 bacteriophage TP901-1. *Microbiol Resour Announc* 8:e01635-18. <https://doi.org/10.1128/MRA.01635-18>.

Editor Julia A. Maresca, University of Delaware

Copyright © 2019 Erazo Garzon 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 Douwe van Sinderen, d.vansinderen@ucc.ie.

Received 30 November 2018

Accepted 12 December 2018

Published 17 January 2019

TABLE 1 Genome features of *Lactococcus lactis* subsp. *cremoris* 3107

| <i>L. lactis</i> subsp. <i>cremoris</i> 3107 chromosome or plasmid | Length (bp) | GC content (%) | Genome coverage (×) | |
|---|-------------|----------------|---------------------|-----------|
| | | | PacBio | Illumina |
| Chromosome | 2,378,982 | 35.82 | 65.01 | 846.88 |
| p3107A | 50,160 | 35.64 | 228.24 | 1,633.72 |
| p3107B | 60,216 | 33.38 | 230.17 | 1,550.52 |
| p3107C | 26,709 | 37.63 | 455.12 | 2,467.09 |
| p3107D | 2,232 | 33.56 | 615.77 | 11,381.27 |
| p3107E | 18,170 | 34.81 | 491.65 | 4,443.82 |
| p3107F | 4,199 | 31.6 | 332.51 | 7,180.93 |

supplemented with 0.5% (vol/vol) glucose. Genomic libraries were constructed using the TruSeq DNA PCR-free LT kit (Illumina) and 2.5 µg of genomic DNA, which was fragmented with a Bioruptor next-generation sequencing (NGS) ultrasonicator (Diagenode, USA), followed by size evaluation using the TapeStation 2200 system (Agilent Technologies). Library samples were loaded into a flow cell version 3 (600 cycles; Illumina), and draft genome sequencing was performed on a MiSeq genomic platform (Illumina, UK) at GenProbio srl (Parma, Italy). Fastq files of the paired-end reads obtained from the genome sequencing were used as input for genome assemblies through the MEGAnnotator pipeline (8). The MIRA program (version 4.0.2) was used for *de novo* assembly of the genome sequence (9). The Illumina data were mapped on corresponding PacBio scaffolds to provide confidence in the generated sequence quality and to resolve base conflicts using Bowtie2 version 2.2.7, achieving a mapping coverage of 847× for the chromosome and an average coverage of 4,776× for the plasmids. Remaining low-quality regions or sequencing conflicts were resolved by primer walking and Sanger sequencing of PCR products (Eurofins MWG Operon, Germany).

Putative protein-encoding genes were identified using the prediction software Prodigal version 2.0 (10). Protein-encoding genes were automatically annotated using BLASTP version 2.2.26 (E value cutoff, 0.0001) sequence alignments against the non-redundant (nr) protein database curated by the NCBI (<ftp://ftp.ncbi.nih.gov/blast/db/>). Following automatic annotation, the obtained open reading frames (ORFs) were manually inspected and refined using the genome browser and annotation tool Artemis version 16 (11). Finally, ORF annotations were refined further where necessary using alternative functional searches using the PFAM database (12) and the Clusters of Orthologous Groups (COG) database (13). Predicted tRNA and rRNA genes were identified using tRNA-scan-SE version 1.4 (<http://lowelab.ucsc.edu/tRNAscan-SE/>) and RNAmmer version 1.2 (<http://www.cbs.dtu.dk/services/RNAmmer/>), respectively. Using Artemis version 16, the predicted RNA-specifying loci were manually added to the genome.

The complete genome content of *L. lactis* subsp. *cremoris* 3107 is represented by a single circular chromosome plus six plasmids (Table 1). The *L. lactis* subsp. *cremoris* 3107 genome is predicted to contain 2,380 protein-encoding genes, of which 101 are pseudogenes. The genome of *L. lactis* subsp. *cremoris* 3107 contains 164 transposase-encoding genes, including 20 copies of IS712H and 31 copies of IS982B. The high number of transposons and pseudogenes within this relatively small lactococcal chromosome suggests that the *L. lactis* subsp. *cremoris* 3107 genome has undergone significant genome decay while adapting to its environment. The *L. lactis* subsp. *cremoris* 3107 plasmids encode various traits for adaptation to the nutrient-rich dairy environment, such as lactose metabolism, making this strain suitable as a starter or adjunct culture.

Data availability. The complete chromosome and plasmids of *L. lactis* subsp. *cremoris* 3107 were deposited in GenBank under accession no. [CP031538](#) (chromosome), [CP031539](#) (p3107A), [CP031540](#) (p3107B), [CP031541](#) (p3107C), [CP031542](#) (p3107D), [CP031543](#) (p3107E), and [CP031544](#) (p3107F). The SMRT and Illumina raw reads were deposited in SRA under BioProject no. [PRJNA438435](#).

ACKNOWLEDGMENTS

J. Mahony is the recipient of a Starting Investigator Research Grant (SIRG) funded by Science Foundation Ireland (SFI) (reference no. 15/SIRG/3430), and D. van Sinderen is the recipient of an SFI Investigator award (reference no. 13/IA/1953).

REFERENCES

1. Teuber M. 1993. Lactic acid bacteria, p 325–366. *In* Rehm H-J, Reed G (ed), *Biotechnology set*, 2nd ed. VCH, Weinheim, Germany.
2. Deveau H, Labrie SJ, Chopin M-C, Moineau S. 2006. Biodiversity and classification of lactococcal phages. *Appl Environ Microbiol* 72:4338–4346. <https://doi.org/10.1128/AEM.02517-05>.
3. Coffey A, Ross RP. 2002. Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application, p 303–321. *In* Konings W, Kuipers OP, Huis in 't Veld JHJ (ed), *Lactic acid bacteria: genetics, metabolism and applications*. Springer, Dordrecht, Netherlands.
4. Ainsworth S, Zomer A, de Jager V, Bottacini F, van Hijum SA, Mahony J, van Sinderen D. 2013. Complete genome of *Lactococcus lactis* subsp. *cremoris* UC509.9, host for a model lactococcal P335 bacteriophage. *Genome Announc* 1:e00119-12.
5. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, Weissenbach J, Ehrlich SD, Sorokin A. 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* 11:731–753. <https://doi.org/10.1101/gr.GR-1697R>.
6. Braun V, Jr, Hertwig S, Neve H, Geis A, Teuber M. 1989. Taxonomic differentiation of bacteriophages of *Lactococcus lactis* by electron microscopy, DNA-DNA hybridization, and protein profiles. *Microbiology* 135:2551–2560. <https://doi.org/10.1099/00221287-135-9-2551>.
7. Sambrook J, Russell DW. 2006. Purification of nucleic acids by extraction with phenol: chloroform. *CSH Protoc* 2006:pdb.prot4455.
8. Lugli GA, Milani C, Mancabelli L, van Sinderen D, Ventura M. 2016. MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. *FEMS Microbiol Lett* 363:fnw049. <https://doi.org/10.1093/femsle/fnw049>.
9. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 14:1147–1159. <https://doi.org/10.1101/gr.1917404>.
10. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
11. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945. <https://doi.org/10.1093/bioinformatics/16.10.944>.
12. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A. 2010. The Pfam protein families database. *Nucleic Acids Res* 38:D211–D222. <https://doi.org/10.1093/nar/gkp985>.
13. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <https://doi.org/10.1186/1471-2105-4-41>.