

# Exploring the Genome of Cheese Starter Lactic Acid Bacterium *Lactococcus lactis* subsp. *lactis* CECT 4433

Diogo Antonio Tschoeke,<sup>a</sup> Ana Paula B. Moreira,<sup>a</sup> Luciane A. Chimento Tonon,<sup>a</sup> Milene Miranda A. de Mesquita,<sup>a,b</sup> Gustavo B. Gregoracci,<sup>a,c</sup> Bruno Gomez-Gil,<sup>d</sup> Rogério Valle,<sup>e</sup> Cristiane C. Thompson,<sup>a</sup> Fabiano L. Thompson<sup>a</sup>

Instituto de Biologia, Universidade Federal do Rio de Janeiro (UFRJ), SAGE-COPPE, Rio de Janeiro, Brazil<sup>a</sup>; Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, Laboratório de Vírus Respiratórios e do Sarampo, Rio de Janeiro, Brazil<sup>b</sup>; Universidade Federal de São Paulo, Campus Baixada Santista, Departamento de Ciências do Mar, Santos, Brazil<sup>c</sup>; A.C. Unidad Mazatlán, CIAD, Mazatlán, México<sup>d</sup>; SAGE-COPPE UFRJ, Rio de Janeiro, Brazil<sup>e</sup>

**Here, we present the draft genome sequences of *Lactococcus lactis* subsp. *lactis* CECT 4433, a cheese fermentation starter strain. The genome provides further insight into the genomic plasticity, biocomplexity (including gene strain specific), and evolution of these genera.**

Received 24 September 2014 Accepted 6 October 2014 Published 13 November 2014

**Citation** Tschoeke DA, Moreira APB, Chimento Tonon LA, de Mesquita MMA, Gregoracci GB, Gomez-Gil B, Valle R, Thompson CC, Thompson FL. 2014. Exploring the genome of cheese starter lactic acid bacterium *Lactococcus lactis* subsp. *lactis* CECT 4433. *Genome Announc.* 2(6):e01142-14. doi:10.1128/genomeA.01142-14.

**Copyright** © 2014 Tschoeke et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Fabiano L. Thompson, [fabianothompson1@gmail.com](mailto:fabianothompson1@gmail.com).

The *Lactococcus lactis* subsp. *lactis* strain CECT 4433 is a member of the lactic acid bacteria (LAB), known as cheese starters. Commonly, LAB is dominant in milk and fermented dairy products, comprising a heterogeneous group of microorganisms that convert carbohydrates into lactic acid (1, 2). LAB are widely distributed in nature and occur naturally as the indigenous microflora in fermented foods, and play a major role in flavors, texture, and preservative qualities (3).

Strains of *L. lactis* have fast growth and rapid production of lactic acid in milk (1). Many industrial and artisanal starter cultures use *L. lactis* to ferment dairy products, especially hard and semi-hard cheeses, and it plays a key role in determining shelf-life, preservation, and organoleptic quality and thus influences the quality and safety of these products. The growth of *L. lactis* in milk is associated with the rapid production of lactic acid, providing flavor, assisting in curd formation, preventing the growth of pathogenic and spoilage bacteria, and creating optimal biochemical conditions for ripening (2). Due to the availability for genetic engineering and genome sequencing efforts, *L. lactis* has become a model organism (4). For example, an ethanol pathway has also been introduced into LAB hosts in order to produce ethanol (5). Here, we present the genome sequence of the *L. lactis* subsp. *lactis* strain CECT 4433.

The genome sequences of *L. lactis* subsp. *lactis* strain CECT 4433 was sequenced with the Ion PGM system, as described previously by Gorriti et al. (6). The Ion PGM 200 sequencing kit was used for all sequencing reactions. The reads obtained were assembled using *de novo* assembly in the Ion Torrent Platform based on MIRA software (7). After the MIRA assembly, we performed a second assembly based on 224 MIRA contigs, using CAP3 software (8). The genome was annotated automatically using the RAST (Rapid Annotations Using Subsystems Technology) server (9). The size of the draft genome of *L. lactis* subsp. *lactis* CECT 4433 is 2,579,153 bp, comprising 111 contigs with a G+C content of 34.92% and 28× coverage. The  $N_{50}$  contig length was

44,334 bp, the largest contig length was 145,182 bp, and the smallest was 540 bp. The numbers of CDSs was 2,643 in *L. lactis* subsp. *lactis* CECT 4433; rRNA sequences found in *L. lactis* subsp. *lactis* CECT 4433 were 9: 6 copies of 5S rRNA, 2 copies of 16S rRNA, and one 23S rRNA gene. Furthermore, 58 tRNA sequences were found in CECT 4433. Comparing the sequence comparison of CECT 4433 and *L. lactis* subsp. *lactis* IL1403<sup>T</sup>, using the SEED comparative tool, we found 394 genes only in CECT 4433. We identified specific genes related to the carbohydrate metabolism as mannose-6-phosphate isomerase, glycosyltransferase, and sucrose-6-phosphate hydrolase and genes related to amino acid metabolism as NADP-specific glutamate dehydrogenase and arginine/ornithine antiporter ArcD. However, 47.96% (188/394) of these genes were annotated as hypothetical proteins or unknown or unnamed proteins and 20.3% (80/394) were annotated as phage associated proteins (e.g., phage tail fibers, capsid, and DNA invertase).

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [JRKY00000000](https://www.ncbi.nlm.nih.gov/nuccore/JRKY00000000). The version described in this paper is version JRKY01000000.

## ACKNOWLEDGMENTS

This work was supported by CAPES, CNPq, and FAPERJ. D.A.T., A.P.B.M. and L.A.C.T. thank CNPq for Posdoc scholarships.

## REFERENCES

- 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. <http://dx.doi.org/10.1101/gr.1697R>.
- Fernández E, Alegría A, Delgado S, Martín MC, Mayo B. 2011. Comparative phenotypic and molecular genetic profiling of wild *Lactococcus lactis* subsp. *lactis* strains of the *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* genotypes, isolated from starter-free cheeses made of raw milk. *Appl. Environ. Microbiol.* 77:5324–5335. <http://dx.doi.org/10.1128/AEM.02991-10>.

3. Gao Y, Lu Y, Teng K-L, Chen M-L, Zheng H-J, Zhu Y-Q, Zhong J. 2011. Complete genome sequence of *Lactococcus lactis* subsp. *lactis* CV56, a probiotic strain isolated from the vaginas of healthy women. *J. Bacteriol.* **193**: 2886–2887. <http://dx.doi.org/10.1128/JB.00358-11>.
4. Siezen RJ, Starrenburg MJ, Boekhorst J, Renckens B, Molenaar D, van Hylckama Vlieg JE. 2008. Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl. Environ. Microbiol.* **74**:424–436. <http://dx.doi.org/10.1128/AEM.01850-07>.
5. Solem C, Dehli T, Jensen PR. 2013. Rewiring *Lactococcus lactis* for ethanol production. *Appl. Environ. Microbiol.* **79**:2512–2518. <http://dx.doi.org/10.1128/AEM.03623-12>.
6. Gorriti MF, Dias GM, Chimetto LA, Trindade-Silva AE, Silva BS, Mesquita MMA, Gregoracci GB, Farias ME, Thompson CC, Thompson FL. 2014. Genomic and phenotypic attributes of novel salinivibrios from stromatolites, sediment and water from a high altitude lake. *BMC Genomics* **15**:473. <http://dx.doi.org/10.1186/1471-2164-15-473>.
7. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, 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. <http://dx.doi.org/10.1101/gr.1917404>.
8. Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res.* **9**:868–877. <http://dx.doi.org/10.1101/gr.9.9.868>.
9. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**: D206–D214. <http://dx.doi.org/10.1093/nar/gkt1226>.