



ELSEVIER

Contents lists available at ScienceDirect

## Data in brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Whole-genome sequence data and analysis of *Lactobacillus delbrueckii* subsp. *lactis* ACA-DC 178 isolated from Greek Kasserı cheese



Voula Alexandraki<sup>a</sup>, Maria Kazou<sup>a</sup>, Bruno Pot<sup>b</sup>,  
Effie Tsakalidou<sup>a</sup>, Konstantinos Papadimitriou<sup>a,\*</sup>

<sup>a</sup> Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Athens, Greece

<sup>b</sup> Research Group of Industrial Microbiology and Food Biotechnology (IMDO), Department of Bioengineering Sciences (DBIT), Vrije Universiteit Brussel, Brussels, Belgium

## ARTICLE INFO

## Article history:

Received 24 June 2019

Accepted 10 July 2019

Available online 16 July 2019

## Keywords:

Lactic acid bacteria

Dairy

Fermentation

Genomics

CRISPR

Genomic island

Milk

## ABSTRACT

*Lactobacillus delbrueckii* subsp. *lactis* is employed in the production of various types of cheese. Here, we report the complete genome sequence of *L. lactis* ACA-DC 178 isolated from Greek Kasserı cheese. The chromosome of ACA-DC 178 contains 2,050,316 bp with a GC content of 49.6%. A total of 2,112 genes were identified in the genome sequence including 1,752 protein-coding genes, 239 putative pseudogenes, 94 tRNA and 27 rRNA genes. According to the COG annotation, about 80% of the protein-coding genes (1,417 proteins) were assigned to at least one functional category. Approximately the 1/3 of these proteins were distributed among three categories, namely replication, recombination and repair (category L: 10.6%), translation, ribosomal structure and biogenesis (category J: 7.5%) and amino acid transport and metabolism (category E: 7.2%). Fourteen integrated GIs with a total of 159 genes were found in ACA-DC 178 genome. Several of these genes encode proteins associated with exopolysaccharide biosynthesis, amino acid transport and subunits of restriction-modification systems. One large CRISPR array of 3,197 bp containing 52 spacers, several of which are identical to phage sequences having hosts in the genus *Lactobacillus*, was also identified. The annotated genome sequence of *L. lactis* ACA-DC 178 is deposited at the European

\* Corresponding author.

E-mail address: [kpapadimitriou@aua.gr](mailto:kpapadimitriou@aua.gr) (K. Papadimitriou).

Nucleotide Archive under the accession number LS991409. Raw reads are deposited in the Sequence Read Archive (SRR8866601-3).

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

#### Specifications Table

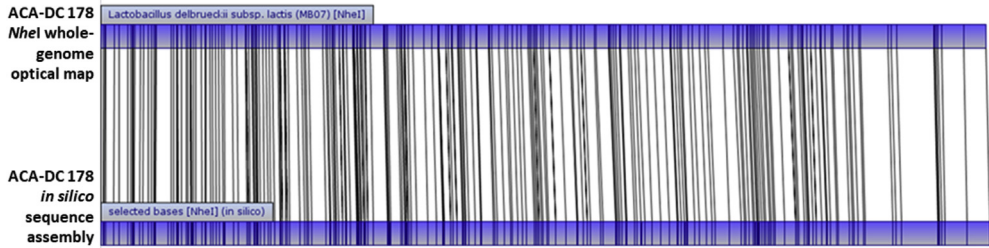
Subject	Microbiology
Specific subject area	Genome analysis
Type of data	Table, Figure
How data were acquired	Genome sequencing: Illumina HiSeq 2000 (Illumina, CA), Denovo sequence assembly: SOAPdenovo v.2.04 software, MapSolver software (OpGen Technologies, Inc., Madison, WI), Bioinformatics approaches: Rapid Annotation using Subsystem Technology (RAST) v.2.0, Prodigal, MetaGeneAnnotator, FGENESB, GenePRIMP pipeline, eggNOG-mapper v.4.5, IslandViewer 4, CRISPRFinder
Data format	Raw sequence reads and genome assembly and annotation
Parameters for data collection	Genomic DNA from pure culture
Description of data collection	Purification of genomic DNA, genome sequencing, genome assembly and annotation
Data source location	Traditional Greek Kasserli Cheese, Athens, Greece
Data accessibility	Data are deposited in the respective databases and are publicly available. The annotated whole-genome sequence of <i>L. lactis</i> ACA-DC 178 is deposited at the European Nucleotide Archive (ENA) under the accession number LS991409 ( <a href="https://www.ebi.ac.uk/ena/data/view/LS991409">https://www.ebi.ac.uk/ena/data/view/LS991409</a> ). Raw sequence reads are deposited in the Sequence Read Archive (SRA; SRR8866601-3; <a href="https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866601">https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866601</a> ; <a href="https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866602">https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866602</a> ; <a href="https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866603">https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866603</a> ).

#### Value of the Data

- *Lactobacillus delbrueckii* subsp. *lactis* is an important bacterium used in cheese production. For this reason analysis of the genome sequence of strain ACA-DC 178 will provide valuable information for its adaptation in the milk environment and its technological properties.
- Data presented here can be valuable for researchers involved in the field of genomic analysis of lactic acid bacteria and food fermentations.
- Data can be used by researchers to perform comparative and functional genomics to further shed light in the evolution, biology and technological properties of the *L. delbrueckii* species.
- Increasing the number of complete genome sequences within *L. delbrueckii* will further aid our understanding of this species.

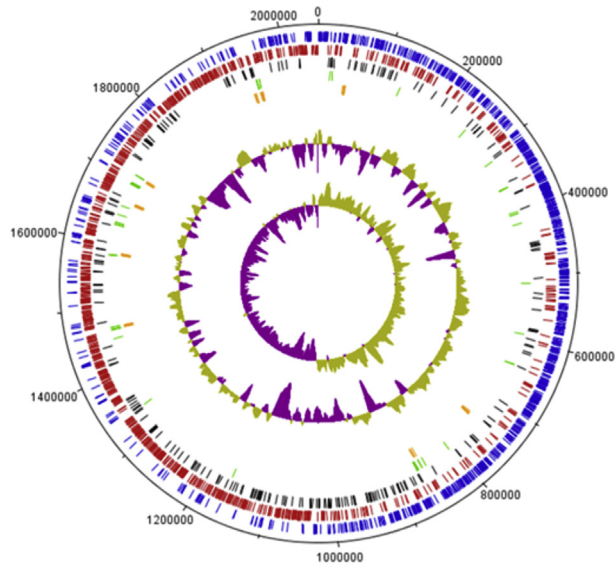
## 1. Data

In this study, we present the complete genome sequence of *L. lactis* ACA-DC 178 isolated from Greek Kasserli cheese [1,2]. The *L. delbrueckii* species consists of six subspecies, including *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus delbrueckii* subsp. *lactis*, *Lactobacillus delbrueckii* subsp. *delbrueckii*, *Lactobacillus delbrueckii* subsp. *indicus*, *Lactobacillus delbrueckii* subsp. *sunkii* and *Lactobacillus delbrueckii* subsp. *jakobsenii* [3,4]. *L. lactis* is the second subspecies used as a starter in the dairy industry along with *L. bulgaricus* within the *L. delbrueckii* species [3]. The *in silico* assembly of the ACA-DC 178 chromosome was validated against a *NheI* whole-genome optical map of the strain (Fig. 1). Our assembly presented 100% matching between the *NheI* restriction sites of the optical map and the relevant sites in our genome sequence *in silico* digested with the same enzyme. The genome was found to be 2,050,316 bp with a GC content of 49.6%. We were able to annotate a total of 2,112 genes, including



**Fig. 1.** Validation of the *L. lactis* ACA-DC 178 genome assembly. Alignment of the *in silico* genome assembly of *L. lactis* ACA-DC 178 (bottom) against a *NheI* whole-genome optical map of the strain (top).

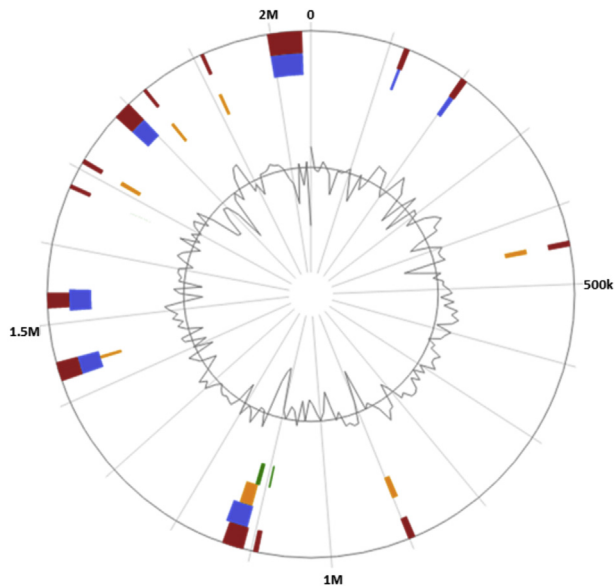
1,752 protein-coding genes, 239 putative pseudogenes, 94 tRNA and 27 rRNA genes (Fig. 2). Further analysis revealed that about 80% of the protein-coding genes (1,417 proteins) could be assigned to at least one Cluster of Orthologous Groups (COG) functional category. Most of these proteins (approximately 1/3) were distributed among three categories related to housekeeping processes, namely replication, recombination and repair (category L: 10.6%), translation, ribosomal structure and biogenesis (category J: 7.5%) and amino acid transport and metabolism (category E: 7.2%) (Table 1). Additional features of the ACA-DC 178 included 14 integrated genomic islands (GIs; Fig. 3) and a clustered regularly interspaced short palindromic repeats-CRISPR-associated (CRISPR-Cas) system (Fig. 4). The GIs carry 159 genes some of which could be assigned to functions like exopolysaccharide biosynthesis, amino acid transport and restriction-modification. The CRISPR array was relatively long, consisting of 3,197 bp and 52 spacers. Detailed analysis of the spacers identified several segments of phage sequences, which have hosts belonging to the *Lactobacillus* genus.



**Fig. 2.** Circular map of the *L. lactis* ACA-DC 178 genome. Each ring represents specific genomic features appearing from the periphery to the centre of the map: Forward CDSs (blue); Reverse CDSs (red); Pseudogenes (black); tRNA (green); rRNA (orange); %GC plot; GC skew.

**Table 1**Distribution of *L. lactis* ACA-DC 178 proteins in COG categories.

	COG	Proteins	%	Description
Information storage and processing	J	132	7.5	Translation, ribosomal structure and biogenesis
	K	100	5.7	Transcription
Cellular processes and signaling	L	186	10.6	Replication, recombination and repair
	D	19	1.1	Cell cycle control, cell division, chromosome partitioning
	M	91	5.2	Cell wall/membrane biogenesis
	N	5	0.3	Cell motility
	O	47	2.7	Posttranslational modification, protein turnover, chaperones
	T	52	3.0	Signal transduction mechanisms
	U	19	1.1	Intracellular trafficking and secretion
Metabolism	V	44	2.5	Defense mechanisms
	C	46	2.6	Energy production and conversion
	E	127	7.2	Amino acid transport and metabolism
	F	67	3.8	Nucleotide transport and metabolism
	G	102	5.8	Carbohydrate transport and metabolism
	H	30	1.7	Coenzyme transport and metabolism
	I	32	1.8	Lipid transport and metabolism
	P	73	4.2	Inorganic ion transport and metabolism
	Q	3	0.2	Secondary metabolites biosynthesis, transport and catabolism
	S	262	15.0	Function unknown
Poorly characterized	–	335	19.1	Not in COGs



**Fig. 3.** Circular map of the *L. lactis* ACA-DC 178 genome. Highlighted regions correspond to GIs. GIs are colored within the circular map according to the prediction method used: green, orange and blue were predicted by IslandPick, SIGI-HMM and IslandPath-DIMOB, respectively. The integrated GIs are presented on the periphery in red. The black line plot represents the GC content (%) of the genomic sequence.

## 2. Experimental design, materials, and methods

*L. lactis* ACA-DC 178 was grown overnight in MRS broth (Merck, Darmstadt, Germany) at 30 °C. DNA was extracted according to a previously published protocol [5]. The genome was sequenced at the Beijing Genomics Institute (BGI Co., Ltd, Hong Kong) using the Illumina HiSeq 2000 platform (Illumina,

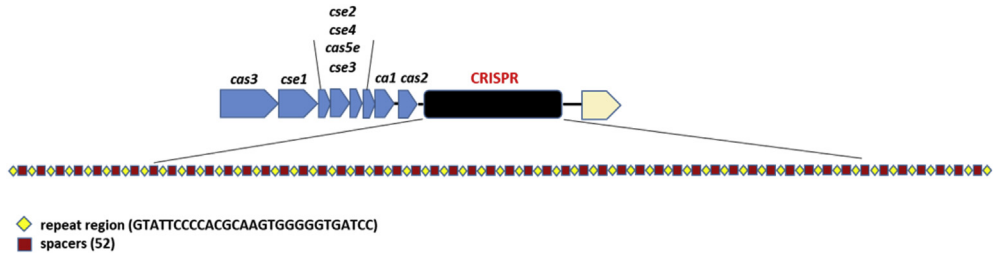


Fig. 4. Graphical presentation of the CRISPR-Cas system of *L. lactis* ACA-DC 178.

CA) employing paired-end libraries of 500 bp, 2,000 bp and 6,000 bp. The assembly of reads with SOAPdenovo v.2.04 [6] resulted in one circular chromosome that was verified against a *NheI* whole-genome optical map of the strain [7] produced at Microbion SRL (Verona, Italy). The alignment between the assembly and the optical map was performed with the MapSolver software (OpGen Technologies, Inc., Madison, WI). The ACA-DC 178 genome sequence was analyzed using Prodigal [8], MetaGeneAnnotator [9] and FGENESB [10] gene prediction programs. Genome annotation and prediction of rRNA and tRNA genes was performed with RAST v.2.0 [11] and putative pseudogenes were predicted with the GenePRIMP pipeline [12]. The results of the analysis were optimized with manual curation. COG annotations were computed using eggNOG-mapper based on eggNOG v.4.5 orthology database [13]. Further bioinformatic analysis was performed for the identification of GIs with IslandViewer 4 [14] and CRISPR with CRISPRFinder [15].

## Acknowledgments

We would like to thank Dr. Nikos Kyrpides at the Joint Genome Institute (United States Department of Energy) for the analysis of the ACA-DC 178 genome with the Gene PRIMP pipeline. The present work was co-financed by the European Social Fund and the National resources EPEAEK and YPEPTH through the Thales project.

## Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] E. Tsakalidou, R. Anastasiou, I. Vandenberghe, J. Beeumen, G. Kalantzopoulos, Cell wall bound proteinase of *Lactobacillus delbrueckii* subsp. *lactis* ACA-DC 178. Characterisation and specificity towards  $\beta$ -casein, *Appl. Environ. Microbiol.* 65 (1999) 2035–2040.
- [2] E. Tsakalidou, E. Manolopoulou, E. Kabarakis, E. Zoidou, B. Pot, K. Kersters, G. Kalantzopoulos, The combined use of whole cell protein extracts for the identification (SDS-PAGE) and enzyme activity screening of lactic acid bacteria isolated from traditional Greek dairy products, *System. Appl. Microbiol.* 17 (1994) 444–458. [https://doi.org/10.1016/S0723-2020\(11\)80062-7](https://doi.org/10.1016/S0723-2020(11)80062-7).
- [3] H. El Kafsi, J. Binesse, V. Loux, J. Buratti, S. Boudebouze, R. Dervyn, S. Kennedy, N. Galleron, B. Quinquis, J.M. Batto, B. Mouton, E. Maguin, M. van de Guchte, *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus*: a chronicle of evolution in action, *BMC Genomics* 15 (1) (2014) 407. <https://doi.org/10.1186/1471-2164-15-407>.
- [4] D.B. Adimpong, D.S. Nielsen, K.I. Sørensen, F.K. Vogensen, H. Sawadogo-Lingani, P.M. Derkx, L. Jespersen, *Lactobacillus delbrueckii* subsp. *jakobsenii* subsp. nov., isolated from dolo wort, an alcoholic fermented beverage in Burkina Faso, *Int. J. Syst. Evol. Microbiol.* 63 (2013) 3720–3726. <https://doi.org/10.1099/ijs.0.048769-0>.
- [5] D.G. Pitcher, N.A. Saunders, R.J. Owen, Rapid extraction of bacterial genomic DNA with guanidium thiocyanate, *Lett. Appl. Microbiol.* 8 (1989) 151–156. <https://doi.org/10.1111/j.1472-765X.1989.tb00262.x>.
- [6] R. Li, Y. Li, K. Kristiansen, J. Wang, SOAP: short oligonucleotide alignment program, *Bioinformatics* 24 (2008) 713–714. <https://doi.org/10.1093/bioinformatics/btn025>.
- [7] P. Latreille, S. Norton, B.S. Goldman, J. Henkhaus, N. Miller, B. Barbazuk, H.B. Bode, C. Darby, Z. Du, S. Forst, S. Gaudriault, B. Goodner, H. Goodrich-Blair, S. Slater, Optical mapping as a routine tool for bacterial genome sequence finishing, *BMC Genomics* 8 (2007) 321. <https://doi.org/10.1186/1471-2164-8-321>.

- [8] D. Hyat, G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, *BMC Bioinf.* 11 (2010) 119. <https://doi.org/10.1186/1471-2105-11-119>.
- [9] H. Noguchi, T. Taniguchi, T. Itoh, MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes, *DNA Res.* 15 (2008) 387–396. <https://doi.org/10.1093/dnares/dsn027>.
- [10] V. Solovyev, A. Salamov, Automatic annotation of microbial genomes and metagenomic sequences, in: R.W. Li (Ed.), *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, Nova Science Publishers, New York, 2011, pp. 61–78.
- [11] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: rapid annotations using subsystems technology, *BMC Genomics* 9 (2008) 75. <https://doi.org/10.1186/1471-2164-9-75>.
- [12] A. Pati, N.N. Ivanova, N. Mikhailova, G. Ovchinnikova, S.D. Hooper, A. Lykidis, N.C. Kyrpides, GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes, *Nat. Methods* 7 (2010) 455–457. <https://doi.org/10.1038/nmeth.1457>.
- [13] J. Huerta-Cepas, K. Forslund, L.P. Coelho, D. Szklarczyk, L.J. Jensen, C. von Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper, *Mol. Biol. Evol.* 34(8) (2017) 2115–2122. <https://doi.org/10.1093/molbev/msx148>.
- [14] C. Bertelli, M.R. Laird, K.P. Williams, Simon Fraser University Research Computing Group, B.Y. Lau, G. Hoad, G.L. Winsor, F.S. L. Brinkman, IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets, *Nucleic Acids Res.* 45 (2017) W30–W35. <https://doi.org/10.1093/nar/gkx343>.
- [15] I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats, *Nucleic Acids Res.* 35 (2007) W52–W57. <https://doi.org/10.1093/nar/gkm360>.