

Draft Genome Sequence of the *Lactobacillus mucosae* Strain Marseille

Fatima Drissi, Vicky Merhej, Caroline Blanc-Tailleux, Didier Raoult

Aix Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm 1095, Marseille, France

***Lactobacillus mucosae* strain Marseille, isolated from stool samples of a child suffering from a malnutrition disorder called Kwashiorkor, produces bacteriocin and seems to have specific carbohydrate and lipid metabolisms different from those of other *Lactobacillus* organisms. The draft genome sequence of this strain is presented here.**

Received 19 June 2015 Accepted 24 June 2015 Published 30 July 2015

Citation Drissi F, Merhej V, Blanc-Tailleux C, Raoult D. 2015. Draft genome sequence of the *Lactobacillus mucosae* strain Marseille. *Genome Announc* 3(4):e00841-15. doi: 10.1128/genomeA.00841-15.

Copyright © 2015 Drissi 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 Didier Raoult, didier.raoult@gmail.com.

Lactobacillus mucosae is a heterofermentative lactic acid bacterium found in the gastrointestinal tract (1). It has been shown to have the ability to adhere to gastrointestinal mucus and to produce antimicrobial agents (2, 3), enabling the bacterium to colonize the gut efficiently and to inhibit the growth of pathogenic bacteria.

The genome of *L. mucosae* Marseille was sequenced using the MiSeq technology (Illumina Inc., San Diego, CA, USA) with the mate pair strategy. Illumina reads were assembled with SPAdes software (4, 5). Contigs obtained were combined by use of SSPACE (6) and Opera software v1.2 (7) helped by GapFiller v1.10 (8) to reduce the set. Noncoding genes and miscellaneous features were predicted using RNAmmer (9), ARAGORN (10), Rfam (11), Pfam (12), and Infernal (13). Coding DNA sequences (CDSs) were predicted using Prodigal (14), and functional annotation was achieved using BLAST+ (15) and HMMER3 (16) against the UniProtKB database (17). The BUR database and the KEGG automatic annotation server (KAAS) were used to annotate genes encoding bacteriocins (18) and to determine the metabolic profile of the genome (19), respectively.

The draft genome of *L. mucosae* Marseille consists of 12 contigs of sizes ranging between 2,027 and 539,623 bp. The genome consists of a single chromosome of 2,369,669 bp with 46.18% G+C content. Of the 2,282 predicted genes, 2,199 were protein-coding genes, 14 were rRNAs genes, and 69 were tRNA genes. A total of 1,535 genes (69.80%) were assigned as putative function, 62 genes (2.82%) were identified as ORFans (open reading frames [ORFs] with no detectable homology to other ORFs in the database), and the 481 remaining genes were annotated as hypothetical proteins (21.87%). A phylogenetic tree produced from the 16S rRNA genes revealed that strain Marseille is most closely related to *Lactobacillus mucosae* LM1 (NCBI reference sequence accession no. NZ_CP011013.1). A comparison between these two strains shows an identity of 97%, but a higher number of genes involved in carbohydrate transport and metabolism in strain Marseille than in LM1 (156 [8%] versus 131 [7%]).

The genome of strain Marseille includes genes encoding 12 different bacteriocins, ranging in size from 38 to 67 amino acids. Moreover, the *L. mucosae* Marseille genome has a specific mucus-binding protein (*mub*) gene, with significant homology (100%

coverage and 100% similarity) to the *L. mucosae* LM1 *mub* gene. This gene is a common characteristic in *L. mucosae* species and has antimicrobial effects through cell surface protection (20). Interestingly, the Marseille genome encodes two enzymes that are not usually present in *Lactobacillus* genomes, an endoglucanase that participates in the degradation of cellulose to glucose and a maltooligosyl trehalose synthase that catalyzes the conversion of maltooligosaccharide into the nonreducing saccharide. Its genome also contains more genes involved in lipid transport and metabolism than other *Lactobacillus* genomes and has several genes encoding phospholipid phosphatase. Strain Marseille has a particular carbohydrate and lipid metabolism that may be attributable to the fact that it lives in the gastrointestinal tract of a malnourished host (M. Million, M. Tidjani Alou, S. Khelaifia, D. Bachar, J. C. Lagier, N. Dione, S. Brah, P. Hugon, V. Lombard, F. Armougom, J. Fromonot, C. Robert, C. Michelle, A. Diallo, A. Fabre, R. Guieu, C. Sokhna, B. Henrissat, P. Parola, and D. Raoult, submitted for publication).

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited in the ENA under the accession no. [CVQW000000000](https://ena.ebi.ac.uk/ena/record/EV000000000/).

ACKNOWLEDGMENTS

This work was funded by the “IHU Méditerranée Infection.” We thank Xegen Company (<http://www.xegen.fr>) for automating the genomic annotation process. Vicky Merhej was supported by a Chairs of Excellence program from the CNRS (Centre National de Recherche Scientifique).

REFERENCES

- London LE, Price NP, Ryan P, Wang L, Auty MA, Fitzgerald GF, Stanton C, Ross RP. 2014. Characterization of a bovine isolate *Lactobacillus mucosae* DPC 6426 which produces an exopolysaccharide composed predominantly of mannose residues. *J Appl Microbiol* 117:509–517. <http://dx.doi.org/10.1111/jam.12542>.
- Fakhry S, Manzo N, D’Apuzzo E, Pietrini L, Sorrentini I, Ricca E, De Felice M, Baccigalupi L. 2009. Characterization of intestinal bacteria tightly bound to the human ileal epithelium. *Res Microbiol* 160:817–823. <http://dx.doi.org/10.1016/j.resmic.2009.09.009>.
- Bleckwedel J, Terán LC, Bonacina J, Saavedra L, Mozzi F, Raya RR. 2014. Draft genome sequence of the mannitol-producing strain *Lactobacillus mucosae* CRL573. *Genome Announc* 2(6):e01292-14. <http://dx.doi.org/10.1128/genomeA.01292-14>.

4. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 20: 714–737. <http://dx.doi.org/10.1089/cmb.2013.0084>.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
6. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27:578–579. <http://dx.doi.org/10.1093/bioinformatics/btq683>.
7. Gao S, Sung WK, Nagarajan N. 2011. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-end sequences. *J Comput Biol* 18:1681–1691. <http://dx.doi.org/10.1089/cmb.2011.0170>.
8. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with Gap-Filler. *Genome Biol* 13:R56. <http://dx.doi.org/10.1186/gb-2012-13-6-r56>.
9. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
10. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <http://dx.doi.org/10.1093/nar/gkh152>.
11. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res* 31:439–441. <http://dx.doi.org/10.1093/nar/gkg006>.
12. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301. <http://dx.doi.org/10.1093/nar/gkr1065>.
13. Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25:1335–1337. <http://dx.doi.org/10.1093/bioinformatics/btp157>.
14. 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. <http://dx.doi.org/10.1186/1471-2105-11-119>.
15. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <http://dx.doi.org/10.1186/1471-2105-10-421>.
16. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <http://dx.doi.org/10.1371/journal.pcbi.1002195>.
17. The UniProt Consortium. 2011. Ongoing and future developments at the universal protein resource. *Nucleic Acids Res* 39:D214–D219. <http://dx.doi.org/10.1093/nar/gkq1020>.
18. Drissi F, Buffet S, Raoult D, Merhej V. 2015. Common occurrence of antibacterial agents in human intestinal microbiota. *Front Microbiol* 6:441. <http://dx.doi.org/10.3389/fmicb.2015.00441>.
19. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185. <http://dx.doi.org/10.1093/nar/gkm321>.
20. Roos S, Karner F, Axelsson L, Jonsson H. 2000. *Lactobacillus mucosae* sp. nov., a new species with *in vitro* mucus-binding activity isolated from pig intestine. *Int J Syst Evol Microbiol* 50:251–258. <http://dx.doi.org/10.1099/00207713-50-1-251>.