Complete Genome Sequence of the Sourdough Isolate *Lactobacillus zymae* ACA-DC 3411

Maria Kazou,a Voula Alexandraki,a Bruno Pot,b Effie Tsakalidou,a Konstantinos Papadimitrioua

Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Athens, Greecea; Research Group of Industrial Microbiology and Food Biotechnology (IMDO), Vrije Universiteit Brussel, Brussels, Belgiumb

**ABSTRACT** *Lactobacillus zymae* is a Gram-positive lactic acid bacterium belonging to the *Lactobacillus brevis* clade. Here, we report the first complete genome sequence of *L. zymae* ACA-DC 3411, which was isolated from traditional Greek wheat sourdough. Whole-genome analysis may reveal adaptive traits of strain ACA-DC 3411 in the sourdough ecosystem.

*Lactobacillus zymae* is a heterofermentative lactic acid bacterium (LAB) species found in fermented foods (1–4), which was transferred from the *Lactobacillus buchneri* clade to the *Lactobacillus brevis* clade, according to a recent 16S rRNA phylogenetic analysis of lactobacilli (5). *L. zymae* ACA-DC 3411 was isolated from traditional Greek wheat sourdough manufactured without baker’s yeast (3, 4). Sourdough has a complex microflora consisting of LAB and yeast species, with lactobacilli being among the most significant group of microorganisms in sourdough fermentation. LAB are mainly involved in dough acidification, whereas yeasts and heterofermentative LAB species participate in the leavening process (6). Analysis of the ACA-DC 3411 genome could prove useful to understand its adaptation in the sourdough environment.

Whole-genome sequencing was performed using the Illumina HiSeq 2000 platform and three paired-end libraries with insert sizes of 500 bp, 2,000 bp, and 6,000 bp at the Beijing Genomics Institute (BGI Co., Ltd., Hong Kong). After filtering, the reads were assembled with the SOAPdenovo version 2.04 software, and the resulting contigs were placed into superscaffolds (7, 8). The assembly was validated using the whole-genome optical map of the strain (9). The map was generated at Microbion SRL (Verona, Italy), and the alignment between the assembly and the optical map was created with the Argus optical mapping system (OpGen Technologies, Inc., Madison, WI). Prediction of protein-coding genes was carried out using Prodigal (10), MetaGeneAnnotator (11) FGENESB (12), and RAST version 2.0, with RAST also being used for the genome annotation and prediction of rRNA and tRNA genes (13). Furthermore, genes were evaluated with the GenePRIMP pipeline for annotation anomalies, including putative pseudogenes (14). Functional annotation of the genome was performed with the WebMGA server (15), the IslandViewer 4 Web-based resource (16), the Phobius Web server (17), and the Pfam database (18) for COG annotation, genomic islands, genes with signal peptides and transmembrane helices, and genes with Pfam domains, respectively.

The genome sequence of ACA-DC 3411 consisted of 2,734,129 bp, with a G+C content of 52.9%. A total of 2,584 genes were identified in the genome, including 2,424 protein-coding genes, 91 potential pseudogenes, 15 rRNA genes, and 54 tRNA genes. According to the COG results, 1,930 protein-coding genes (approximately 80%) were assigned to a putative functional category, with the most abundant being related to
replication, recombination, and repair (14%). Moreover, 19 integrated genomic islands were predicted in the ACA-DC 3411 genome, containing a total of 265 genes potentially acquired through horizontal gene transfer. Fifty-six of these genes code for hypothetical proteins, and the rest are of variable function. Additionally, the analysis revealed that the genome contains also 285 protein-coding genes with signal peptides, 545 with transmembrane helices, and 2,012 with Pfam domains. Further analysis of the ACA-DC 3411 genome may reveal the technological potential of the strain for sourdough fermentation.

Accession number(s). The genome sequence of *L. zymae* ACA-DC 3411 is deposited at the European Nucleotide Archive under the accession number LT854705.

ACKNOWLEDGMENTS

We thank Nikos Kyrpides at the Joint Genome Institute (U.S. Department of Energy) for analysis of the ACA-DC 3411 genome with the GenePRIMP server.

The present work was cofinanced by the European Social Fund and the National resources EPEAEK and YEPETH through the Thales project.

REFERENCES

1. Park JY, Jeong SJ, Kim JH. 2014. Characterization of a glutamate decarboxylase (GAD) gene from *Lactobacillus zymae*. Biotechnol Lett 36:1791–1799. https://doi.org/10.1007/s10529-014-1539-9.

2. Cheng L, Luo J, Li P, Yu H, Huang J, Luo L. 2014. Microbial diversity and flavor formation in onion fermentation. Food Funct 5:2338–2347. https://doi.org/10.1039/c4fo00196f.

3. De Vuyst L, Schrijvers V, Paramithiotis S, Hoste B, Vancanneyt M, Swings J, Kalantzopoulos G, Tsakalidou E, Messens W. 2002. The biodiversity of lactic acid bacteria in Greek traditional wheat sourdoughs is reflected in both composition and metabolite formation. Appl Environ Microbiol 68:6059–6069. https://doi.org/10.1128/AEM.68.12.6059-6069.2002.

4. Vancanneyt M, Neyes P, De Wachter M, Engelbeen K, Snauwaert C, Cleenwerck I, Van der Meulen R, Hoste B, Tsakalidou E, De Vuyst L, Swings J. 2005. *Lactobacillus acidifarinae* sp. nov. and *Lactobacillus zymae* sp. nov., from wheat sourdoughs. Int J Syst Evol Microbiol 55:615–620. https://doi.org/10.1099/ijs.0.63274-0.

5. Salvetti E, Torriani S, Felix GE. 2012. The genus *Lactobacillus*: a taxonomic update. Probiotics Antimicrob Proteins 4:217–226. https://doi.org/10.1007/s12602-012-9117-8.

6. Corsetti A, Settanni L. 2007. Lactobacilli in sourdough fermentation. Food Res Int 40:539–558. https://doi.org/10.1016/j.foodres.2006.11.001.

7. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18. https://doi.org/10.1186/2047-217X-1-18.

8. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:2013–2014. https://doi.org/10.1093/bioinformatics/btp336.

9. Latreille P, Norton S, Goldman BS, Henkhaus J, Miller N, Barbazuk B, Bode HB, Darby C, Du Z, Forst S, Gaudriault S, Goodner B, Goodrich-Blair H, Slater S. 2007. Optical mapping as a routine tool for bacterial genome sequence finishing. BMC Genomics 8:321. https://doi.org/10.1186/1471-2164-8-321.

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. Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res 15:387–396. https://doi.org/10.1093/dnares/dsn027.

12. Solovyev V, Salamov A. 2011. Automatic annotation of microbial genomes and metagenomic sequences, p 61–78. In Li RW (ed.), Metagenomics and its applications in agriculture, biomedicine and environmental studies. Nova Science Publishers, New York, NY.

13. Aziz RK, Bartels D, Best AA, DeLongh M, Disz T, Edwards RA, Formsmoa 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.

14. Pati A, Ivanova NN, Mikhailova N, Ochinnkova G, Hooper SD, Lykidis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods 7:455–457. https://doi.org/10.1038/nmeth.1457.

15. Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable Web server for fast metagenomic sequence analysis. BMC Genomics 12:444. https://doi.org/10.1186/1471-2164-12-444.

16. Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N, McKellar AG, Langille MG, Brinkman FS. 2015. Island-Viewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Res 43:W104–W108. https://doi.org/10.1093/nar/gkv401.

17. Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius Web server. Nucleic Acids Res 35:W429–W432. https://doi.org/10.1093/nar/gkm256.

18. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Pponder SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344.