Complete Genome Sequence of *Lactobacillus acidophilus* FSI4, Isolated from Yogurt

Oleg Iartchouk, a Sergei Kozyavkin, a Valeri Karamychev, b Alexei Slesarev a,c

Novartis Institutes for BioMedical Research (NIBR), Cambridge, Massachusetts, USA a; Fidelity Systems, Inc., Gaithersburg, Maryland, USA b; Winogradsky Institute of Microbiology, RAS, Moscow, Russia c

A new *Lactobacillus acidophilus* strain, FSI4, isolated from yogurt, was isolated and sequenced in our laboratory. Our data, although supportive of previous conclusions regarding the remarkable stability of *L. acidophilus* species, indicate accumulating mutations in commercial *L. acidophilus* strains that warrant further study of the effect of damaged genes on the competitiveness of these bacteria in gut microbiota.

**Received 7 February 2015 Accepted 3 March 2015 Published 9 April 2015**

**Citation** Iartchouk O, Kozyavkin S, Karamychev V, Slesarev A. 2015. Complete genome sequence of *Lactobacillus acidophilus* FSI4, isolated from yogurt. Genome Announc 3(2): e00166-15. doi:10.1128/genomeA.00166-15.

**Copyright © 2015 Iartchouk et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.**

Address correspondence to Alexei Slesarev, alex@fidelitysystems.com.

**Lactobacillus acidophilus** is a member of the Gram-positive bacterial group capable of producing lactic acid (1–3). This capability has resulted in their use in the dairy industry, with *L. acidophilus* being the most widely commercially distributed culture (4). *L. acidophilus* is one of the first colonizers of the digestive tract of newborns, and its competitive advantage over other bacteria plays an important role in the healthy development of infants (5, 6). There are reports of multiple draft genome sequences of *L. acidophilus* strains (4, 7), but only two complete genome sequences are available (8, 9).

The complete genome sequence of *L. acidophilus* FSI4 was determined by the combined use of the Illumina and Sanger sequencing platforms. Illumina GAIIx paired-end reads (153 × coverage) were assembled with Velvet (10). Four rRNA operons and 4 types of long insertion sequence (IS) elements with pairwise similarities in the range of 96.3 to 100% were identified. Their positions in the chromosome were verified by PCR (rRNA genes) and by direct Sanger sequencing from genomic DNA (11, 12) through the IS elements, with the use of trimming technique. This technique involves the hybridization of the Sanger sequencing fragments with appropriate biotinylated complementary oligonucleotides 600 to 900 bases downstream of the primer, followed by trimming the duplex with a 4-base-cutter restriction endonuclease and running the trimmed fragments separately on a capillary sequencer.

The *L. acidophilus* FSI4 chromosome encodes 1,759 proteins, 112 RNAs, 54 riboswitches and leaders, and 98 pseudogenes. Alignment of the *L. acidophilus* genomes demonstrates that strains FSI4, NCFM, and LA-14 are extremely similar, with all genes being syntenic across all three genomes. We have found a single 9-nucleotide deletion in the NADH dehydrogenase gene in the FSI4 strain, compared to the NCFM and LA-14 strains, which results in 3 amino acid in-frame deletions in the corresponding protein. For a further comparison of the FSI4 and NCFM genomes, we first corrected sequencing errors in the NCFM genome with the Illumina data obtained by Bull et al. (4) and then analyzed the remaining differences that happened to be limited to 9 single-nucleotide substitutions. We also noticed that the majority of the identified indels and single-nucleotide polymorphisms (SNPs) (40 total) in the LA-14 genome occur in the homopolymer stretches, pointing to the possibility that they might be errors resulting from the 454 sequencing used in the LA-14 genome project (9, 13). Finally, there is a 54-bp variable region in one of the 1,522-bp IS elements of the FSI4 genome that is different in the corresponding copies of the 1,522-bp IS elements of the NCFM and LA-14 genomes.

Overall, our results, along with previous data, prove that commercial *L. acidophilus* genomes have remarkable macrostability. On the other hand, our highly accurate sequencing data reveal the presence of multiple disrupted genes in the genomes in question. We showed previously that gene disruption during the cultivation of *Bifidobacterium longum* may reduce the competitiveness of probiotics in the gut (14). However, further studies are required to clarify the effect of gene restoration on the competitive advantage of probiotics in the gut microbiota.

**Nucleotide sequence accession numbers.** The complete genome sequence of *L. acidophilus* FSI4 has been deposited in GenBank under the accession no. CP010432. The 100-bp Illumina HiSeq 2000 reads used for error correction of the *L. acidophilus* NCFM genome (accession no. CP000033) are available in the NCBI sequence read archive (SRA) under accession no. ERR386044.

**ACKNOWLEDGMENTS**

This work was supported in part by Department of Energy Small Business Innovative Research grants DE-FG02-98ER82577 and DE-FG02-00ER83009.

**REFERENCES**

1. Kandler O. 1983. Carbohydrate metabolism in lactic acid bacteria. Antonie Van Leeuwenhoek 49:209–224. http://dx.doi.org/10.1007/BF00399499.
2. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E,
Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Lee JW, Diaz-Muniz I, Dosti B, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D. 2006. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA 103:15611–15616. http://dx.doi.org/10.1073/pnas.0607117103.

3. Klaenhammer TR, Barrangou R, Buck BL, Azcarate-Peril MA, Altermann E. 2005. Genomic features of lactic acid bacteria effecting bioprocessing and health. FEMS Microbiol Rev 29:393–409. http://dx.doi.org/10.1016/j.femsre.2005.04.007.

4. Bull MJ, Jolley KA, Bray JE, Aerts M, Vandamme P, Maiden MC, Marchesi JR, Mahenthiralingam E. 2014. The domestication of the probiotic bacterium Lactobacillus acidophilus. Sci Rep 4:7202. http://dx.doi.org/10.1038/srep07202.

5. Reid G, Sanders ME, Gaskins HR, Gibson GR, Mercenier A, Rastall R, Roberfroid M, Rowland I, Cherbut C, Klaenhammer TR. 2003. New scientific paradigms for probiotics and prebiotics. J Clin Gastroenterol 37:105–118. http://dx.doi.org/10.1097/00004836-200308000-00004.

6. Konstantinov SR, Smidt H, de Vos WM, Bruijins SC, Singh SK, Valence F, Molle D, Lortal S, Altermann E, Klaenhammer TR, van Kooyk Y. 2008. S layer protein A of Lactobacillus acidophilus NCFM regulates immature dendritic cell and T cell functions. Proc Natl Acad Sci USA 105:19474–19479. http://dx.doi.org/10.1073/pnas.0810305105.

7. Falentin H, Cousin S, Clermont D, Creno S, Ma L, Chuat V, Loux Y, Rüdiger P, Bizet C, Bouchier C. 2013. Draft genome sequences of five strains of Lactobacillus acidophilus, strain CIP 76.13, isolated from humans, strains CIRM-BIA 442 and CIRM-BIA 445, isolated from dairy products, and strains DSM 20242 and DSM 9126 of unknown origin. Genome Announc 1(4):e00658-13. http://dx.doi.org/10.1128/genomeA.00658-13.

8. Altermann E, Russell WM, Azcarate-Peril MA, Barrangou R, Buck BL, McAuliffe O, Souther N, Dobson A, Duong T, Callanan M, Lick S, Hamrick A, Cano R, Klaenhammer TR. 2005. Complete genome sequence of the probiotic lactic acid bacterium Lactobacillus acidophilus NCFM. Proc Natl Acad Sci USA 102:3906–3912. http://dx.doi.org/10.1073/pnas.0409188102.

9. Stahl B, Barrangou R. 2013. Complete genome sequence of probiotic strain Lactobacillus acidophilus LA-14. Genome Announc 1(3):e00376-13. http://dx.doi.org/10.1128/genomeA.00376-13.

10. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. http://dx.doi.org/10.1101/gr.074492.107.

11. Malykh A, Polushin N, Kozyavkin S, Slesarev A. 2004. Finishing Working draft BAC projects by directed sequencing with Thermofidelase and Fimers. Methods Mol Biol 255:295–308. http://dx.doi.org/10.1385/1-59259-752-1:295.

12. Malykh A, Polushin N, Slesarev A, Kozyavkin S. 2004. Use of Fimers to eliminate polymerase chain reaction and primer-dimer artifacts and to increase yield in BAC-sequencing reactions. Methods Mol Biol 255:221–229. http://dx.doi.org/10.1385/1-59259-752-1:229.

13. Zeng F, Jiang R, Chen T. 2013. PyroHMMvar: a sensitive and accurate method to call short indels and SNPs for Ion Torrent and 454 data. Bioinformatics 29:2859–2868. http://dx.doi.org/10.1093/bioinformatics/btt512.

14. Lee JH, Karamychev VN, Kozyavkin SA, Mills D, Pavlov AR, Pavlova NV, Polouchine NN, Richardson PM, Shakhova VV, Slesarev AI, Weimer B, O’Sullivan DJ. 2008. Comparative genomic analysis of the gut bacterium Bifidobacterium longum reveals loci susceptible to deletion during pure culture growth. BMC Genomics 9:247. http://dx.doi.org/10.1186/1471-2164-9-247.