

# GENOME SEQUENCE OF *LACTOBACILLUS GASSERI* 2016: COMPARATIVE GENOME ANALYSIS, UNIQUE GENES, RESPONSIBLE FOR THE PROBIOTIC PROPERTIES.

Karlyshev A.<sup>1</sup>, Melnikov V.<sup>2</sup>, Khlebnikov V.<sup>3</sup>, Kosarev I.<sup>3\*</sup>, Sakulin V.<sup>3</sup>, Kulikova N.<sup>3</sup>, Vasilenko R.<sup>3</sup>, Uversky V.<sup>3</sup>, Lyubetsky V.<sup>4</sup>, Netrusov A.<sup>2</sup>, Sukhikh G.<sup>5</sup>, Abramov V.<sup>3</sup>

<sup>1</sup>*School of Life Sciences, Faculty of Science, Engineering and Computing; Kingston University, Kingston upon Thames, Penrhyn Road, KT1 2 EE, United Kingdom*

<sup>2</sup>*Moscow State University, Moscow 119992, Russia*

<sup>3</sup>*Institute of Immunological Engineering, Lyubuchany, Moscow Region, 142380, Russia*

<sup>4</sup>*Institute for information transmission problems of the Russian academy of science (Kharkevich institute) 19, B. Karetnyi, Moscow, Russian Federation, 127994*

<sup>5</sup>*Research Center of Obstetrics Gynecology and Neonatology, Ministry of Health and Social Development of the Russian Federation, 4, Oparin street, Moscow, 117997, Russia*

\*e-mail: kosarev-52@mail.ru

Probiotic bacteria including *Lactobacillus* spp were found to be very promising for treatment of urogenital infections. We present a draft genome sequence of *L.gasseri* 2016m obtained using IonTorrent PGM. Sequencing on 314 chip produced 371,671 reads with the total size of 42.2 million bases. Assembly of reads using IonTorrent assembler plugin produced 440 contigs (0.5 – 24.2 kb; total size: 1,609,472 bases, GC content: 35.2%). At the time of preparation of materials for this publication, the only complete genome sequence of *Lactobacillus gasseri* available was that for strain ATCC 33323 (accession number, NC\_008530, genome size 1,894,360 bp, 35.3% GC) reported in 2006. Mapping of reads onto the genome of the reference strain using CLC Genomics Workbench program allowed generation of ten consensus sequences with a size range from 1,302 to 795,849 bases. The unmapped reads were assembled onto 19 contigs between 1,048 and 11,426 bases in size. The combined size of mapped and unmapped contigs was 1,879,515 bases, corresponding to 99.2 % of the reference genome. The contigs were verified by mapping the reads to contigs. The total number of reads mapped was 35,186,093 (94.67%) corresponding to 21.25 fold genome coverage. Mapping of reads onto the reference genome revealed a large number of common genes, including a gene encoding helveticin (100% identity). Typical missing genes corresponded to integrases and recombinases related to various prophages as well as transposons. The assembled unmapped reads were further analysed in order to identify the genes that are present in strain 2016, but absent in the reference strain. These contigs were analysed using a batch Blastx utility of CLC Genomics Workbench and a non-redundant amino acid sequence database. Products of some genes revealed high level of similarity to cell surface proteins found in other *Lactobacillus* spp, including a putative cell wall surface anchor family protein (*L. rhamnosus* LRHMDP2, 67% identity in 543 amino acids), cell surface protein (*L. antri* DSM 16041, 99% identity in 477 amino acids) and mannose-specific adhesin, LPXTG-motif cell wall anchor (*L. plantarum* WCFS1, 68% identity in 484 amino acids).

1. R. A. MacPhee, et al. (2010). Probiotic strategies for the treatment and prevention of bacterial vaginosis. *Expert Opin. Pharmacother.* **11**:2985-2995. doi: 10.1517/14656566.2010.512004.
2. R. Barrons, and D. Tassone. (2008). Use of *Lactobacillus* probiotics for bacterial genitourinary infections in women: A review. *Clin. Ther.* **30**:453-468. doi: 10.1016/j.clinthera.2008.03.013.
3. K. Makarova, et al. (2006). Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **103**:15611-15616. doi: 10.1073/pnas.0607117103.