



## Sequencing the genome of *Lactobacillus casei* BL23: current status and preliminary results

Manuel Zúñiga<sup>1</sup>, Vicente Monedero<sup>1</sup>, Alain Mazé<sup>2,3</sup>, Grégory Boël<sup>2,3</sup>, Axel Hartke<sup>3</sup>, Alexa Bourand<sup>2</sup>, Josef Deutscher<sup>2</sup>

<sup>1</sup>Departamento de Biotecnología, Instituto de Agroquímica y Tecnología de Alimentos, Consejo Superior de Investigaciones Científicas, Paterna (Valencia), España

<sup>2</sup>Microbiologie et Génétique Moléculaire, CNRS/INRA, Grignon, France

<sup>3</sup>Laboratoire de Microbiologie de l'Environnement, Université de Caen, Caen, France

*Lactobacillus casei* is a species belonging to the group of lactic acid bacteria of considerable interest in the food industry for milk fermentation and cheese production. Furthermore, *L. casei* has attracted intense interest as a probiotic. *Lactobacillus casei* ATCC 334 was sequenced by the Lactic Acid Bacteria Genome Sequencing Consortium. However, a number of studies suggested that there were significant differences between this strain and *Lactobacillus casei* BL23. In particular, estimations of their genome sizes by real-time quantitative PCR indicated a genome size of 3.1 Mb ( $\pm$  200 kb) for BL23 and 2.8 Mb ( $\pm$  200 kb) for ATCC 334 (actual size 2.93 Mb). This prompted us to start the sequencing of this strain. To this end, a shotgun cloning and sequencing strategy was followed. After assembly of contigs, gaps are being filled by PCR. So far, 2.97 Mb in 96 contigs have been sequenced. The preliminary analysis of the sequence shows that genes present in both organisms are usually more than 90% identical. However, the genome of BL23 contains numerous insertions which probably account for most physiological differences between the two strains. We have focused our studies first on genes involved in sugar metabolism, because many genes present only in BL23 encode proteins for carbohydrate uptake systems, including the PEP:carbohydrate phosphotransferase system, ABC transporters and ion symport systems. We have characterized operons for lactose, myo-inositol, maltose, sorbitol and sorbose utilization present in BL23, which are either incomplete, plasmid-encoded or absent in ATCC 334. In addition, other uncharacterized putative sugar utilization gene clusters not present in ATCC 334 have been identified in BL23. Phage-related gene clusters are present in both organisms, but are quite differently organized and located at different positions in the genome. Interestingly, only strain BL23 contains a CRISPR region and CRISPR-associated genes, which mediate phage resistance. BL23 also contains an about 21 kb region for exopolysaccharide synthesis, which is absent from strain ATCC 334. Among others, this region contains the gene for a tyrosine kinase, enzymes known to regulate exopolysaccharide synthesis and assembly, and a duplication of the *rmlACBD* genes whose products are involved in the synthesis of exopolysaccharide precursors. Various types of transposases are frequently found inserted in both organisms, but again are mostly found at different positions. In contrast, both strains contain five rRNA regions, which are located at identical positions in the genome and nearly identical.