Genome Sequence of *Lactobacillus ingluviei*, a Bacterium Associated with Weight Gain in Animals

Vicky Merhej, Fabrice Armougom, Catherine Robert, and Didier Raoult
Aix Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, INSERM 1095, Marseille, France

We report the draft genome sequence of *Lactobacillus ingluviei* strain Autruche 4 (CSURP209) isolated from an ostrich. *L. ingluviei* is associated with weight gain in mice. This genome sequence may help us understand the obesity-induced mechanisms of intestinal bacteria.

*Lactobacillus ingluviei* is a lactic acid bacterium found in the gastrointestinal tracts of pigeons (3) and ostriches (5) whose weight-increasing properties have recently been described in chicks and ducks (2) and in mice (1). The genome of this bacterium in comparison with other sequenced *Lactobacillus* species may help us understand the implication of *Lactobacillus* as a growth promoter and in the development of obesity (7).

The sequence data were obtained using a combination of two sequence technologies: the 454 GS FLX Titanium pyrosequencing system (Roche Diagnostic, Branchford, CT) (6) with a shotgun library and a 3-kb paired-end library and SOLiD version 4 with a 150-bp paired-end library (Life Technologies). The 454 system paired-end and shotgun sequencing generated 293,489 reads and 211,630 reads, respectively (totaling 505,119 reads and 412,459,355 bp). The reads from paired-end and SOLiD sequencing were contaminated with *Staphylococcus haemolyticus*. The removal of *Staphylococcus* sequences was performed by mapping the shotgun reads against the paired-end assembly sequences using Newbler version 2.0.0.1 (Roche) and CLC Genomics Workbench v4.7.2 (CLC bio, Aarhus, Denmark). Eight scaffolds containing 40 contigs were assigned to *Lactobacillus*, corresponding to 444,493 reads from the 454 system (158,889,712 bp) and 478,391 reads from SOLiD (20,065,630 bp), which represents 71-fold genome coverage. The functional annotation of predicted genes was achieved using Prodigal to predict open reading frames (ORFs) and BLASTp (4).

The draft genome includes 1,979,644 bases (G+C content of 50.90%). It comprises 1,927 predicted genes, including 68 predicted RNA genes (4 rRNAs and 64 tRNAs).

Among the *Lactobacillus* species with publically available genome sequences, *Lactobacillus fermentum* IFO 3956 (accession no. NC_010610) is the closest to *L. ingluviei* based on their similarities in genomic sequences. The comparison of *L. fermentum* and *L. ingluviei* revealed a large shuffling of gene order but highly similar gene contents, with the exception of 112 protein-encoding genes that are present in *L. ingluviei* but not in *L. fermentum*. Among them, there are 11 genes that code for cell wall, membrane, and envelope biogenesis, 21 genes that code for amino acid transport and metabolism, and 13 genes that code for poorly characterized proteins.

Comparative genome analysis with other published *Lactobacillus* sequences revealed some unique predicted proteins: an exopoly saccharide biosynthesis protein, a multiple antibiotic transporter, isopropylmalate-homocitrate-citramalate synthases, a lipoate synthase, and hypothetical proteins. The genome sequence of *L. ingluviei* provides a good resource for studying the physiological role of *Lactobacillus* as a growth promoter.

**Nucleotide sequence accession number.** This Whole Genome Shotgun project has been deposited in GenBank/EMBL under accession no. CAKF00000000.

**ACKNOWLEDGMENTS**

This work was funded by the “IHU Méditerranée Infection.” We thank Fatima Drissi for help with genome assembly.

**REFERENCES**


Received 4 July 2012 Accepted 6 August 2012