Lactobacillus crispatus can persist in the vertebrate gastrointestinal tract and is among the prevalent species of the Lactobacillus-dominated human vaginal microbiota (2, 9, 13, 14). It belongs to the so-called acidophilus group (3), which has attracted interest because some of its species are important factors in the production of fermented foods (12) and some can, at least transiently, colonize the human host (2, 9, 13, 14). Moreover, some specific strains, mainly L. acidophilus NCFM and L. johnsonii NCC 533, have received prominence as intestinal-health-promoting microbes (4). Although the genomes of seven members of the acidophilus complex have been sequenced to date (12), the genome sequences of L. crispatus and other predominant lactobacillar species in the urogenital flora have mostly remained obscure. Vaginal lactobacilli can have an important role in controlling the health of the host (2, 14). They can, for example, positively influence and stabilize the host’s vaginal microbiota via the production of compounds that are acidic or exert a direct inhibiting action toward pathogenic bacteria (2, 14). In addition to the antimicrobial compounds, the competitive exclusion of pathogens is another mechanism by which the host’s microbiota can be balanced (2). L. crispatus ST1 was originally isolated from the crop of a chicken, and PCR profiling of L. crispatus isolates has verified it to be an abundant colonizer of the chicken crop (6, 8). It also displays a strong protein-dependent adhesion to the epithelial cells of the human vagina and has been shown to inhibit the adhesion of avian pathogenic Escherichia coli (6, 7).

The genome was sequenced (18× coverage) using a 454 pyrosequencer with GS FLX chemistry (Roche). The contig order was confirmed and gaps were filled by sequencing PCR fragments from the genomic DNA template using ABI 3730 and Big Dye chemistry (Applied Biosystems). Genomic data were processed using the Staden Package (11) and gsAssembler (Roche). Coding sequences (CDSs) were predicted using Glimmer3 (5) followed by manual curation of the start sites. The remaining intergenic regions were reanalyzed for missed CDSs by using BlastX (1). Annotation transfer was performed based on a BlastP search, followed by Blannotator analysis using default settings (http://ekhidna.biocenter.helsinki.fi/poxo/blannotator) and manual verification. Orthologous groups between the different lactobacillar proteomes were identified using OrthoMCL (10).

The genome of L. crispatus ST1 consists of a single circular chromosome 2.04 Mbp in size, with an overall G+C content of 37%, without any plasmids. There are 64 tRNA genes, 4 rRNA operons, and 2 CRISPR loci. Out of the 2,024 predicted CDSs, a putative function was assigned to 77%, whereas 10% of the CDSs were annotated as conserved and 13% as novel. Based on the orthologous grouping, 302 (15%) of the CDSs encoded by ST1 have no detectable homologs in any of the Lactobacillus proteomes published to date. The genome sequence of L. crispatus ST1 was deposited in EMBL under accession number FN692037.

This research was supported by the Academy of Finland (general research projects 114498, 109849, 116507, and 123900 and ERA-NET PathoGenoMics project 118982) and the European Union Network of Excellence EuroPathoGenomics.

REFERENCES