

Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* KF147, a Plant-Associated Lactic Acid Bacterium[∇]

Roland J. Siezen,^{1,2,3*} Jumamurat Bayjanov,³ Bernadet Renckens,^{1,3} Michiel Wels,^{1,2,3}
Sacha A. F. T. van Hijum,^{1,2,3} Douwe Molenaar,^{1,2}
and Johan E. T. van Hylckama Vlieg^{1,2,†}

NIZO Food Research, Kluyver Centre for Genomics of Industrial Fermentation, P.O. Box 20, 6710 BA Ede, Netherlands¹;
TI Food and Nutrition, P.O. Box 557, 6700 AN Wageningen, Netherlands²; and Netherlands Bioinformatics Centre,
Centre for Molecular and Biomolecular Informatics (CMBI 260), Radboud University Medical Centre,
P.O. Box 9101, 6500HB Nijmegen, Netherlands³

Received 11 March 2010/Accepted 15 March 2010

***Lactococcus lactis* is a lactic acid bacterium used in the production of many fermented dairy products. We report the complete genome sequence of *L. lactis* subsp. *lactis* KF147, a nondairy strain isolated from mung bean sprouts. The circular chromosome of 2,598,144 bp, the largest among the sequenced lactococcal strains, encodes many properties related to adaptation to the plant environment.**

Lactococcus lactis is used extensively in food fermentation for the production of cheese, sour cream, fermented milk, etc. Plant material (fermenting plant material) is a second important ecosystem occupied by *L. lactis*, where it encounters carbohydrate growth substrates other than lactose, the main carbohydrate growth substrate in milk. Our present view on genomic diversity, metabolic diversity, and adaptation of *L. lactis* is limited, as only the genomes of the dairy-derived strains *L. lactis* subsp. *lactis* IL1403 (3), *L. lactis* subsp. *cremoris* SK11 (6), and *L. lactis* subsp. *cremoris* MG1363 (10) have been reported. Therefore, we sequenced the genome of *L. lactis* subsp. *lactis* KF147, isolated from mung bean sprouts (5). Sequencing was performed (GATC Biotech AG), using a combination of traditional Sanger paired-end sequencing (13,464 reads totaling ~8 Mb; ~3-fold coverage), 454 Life Sciences GS20 pyrosequencing (278,119 reads totaling ~72 Mb; ~27-fold coverage), and Solexa sequencing (5,143,335 reads totaling ~185 Mb; ~71-fold coverage). Contigs were generated using Newbler Assembler software (454 Life Sciences) and assembled into scaffolds using paired-end reads together with mapping to the reference genome of *L. lactis* ssp. *lactis* IL1403 (GenBank accession number AE005176). Physical gaps, repeats, and assembly ambiguities were corrected through sequencing of PCR products. Automatic open reading frame (ORF) calling and annotation were first performed using Pedant-Pro (Biomax Informatics AG) and the Institute for Genome Sciences (IGS) annotation service (IGS, University of Maryland). Improved manual annotation was performed using Pfam (2, 4), InterProScan (8), and the ERGO Bioinformatics Suite (7).

The complete genome of *L. lactis* subsp. *lactis* KF147 consists of a single, circular chromosome (2,598,144 bp; 34.9% GC

content) and one plasmid, pKF147A (37,510 bp; 32.4% GC content). The chromosome contains 2,578 protein-encoding genes (of which 59 are pseudogenes), 68 tRNA-encoding genes, and 6 rRNA-encoding operons. The genome of *L. lactis* subsp. *lactis* KF147 is highly similar to that of *L. lactis* subsp. *lactis* IL1403 with respect to nucleotide sequence identity (>98%) and gene order. Our preliminary analysis of the incomplete genome (9) already showed that strain KF147 has many additional features related to growth adaptation to substrates derived from plant cell walls, including gene clusters that code for the degradation of complex plant polymers, such as xylan, arabinan, glucans, and fructans, but also for the uptake and conversion of typical plant cell wall degradation products, such as α -galactosides, β -glucosides, arabinose, xylose, galacturonate, glucuronate, and gluconate. The complete chromosome sequence of strain KF147 now shows additional features differing from strain IL1403, mainly as large inserts, of which the largest is a putative conjugative transposon of 51 kb encoding the previously described enzymes for metabolism of plant-derived α -galactosides, such as stachyose, raffinose, and melibiose (9). The latter transposon is spontaneously lost upon adaptation of *L. lactis* KF147 to growth in milk (1).

Nucleotide sequence accession number. The complete genome of *L. lactis* subsp. *lactis* KF147 has been deposited in GenBank under accession numbers CP001834 (chromosome) and CP001835 (plasmid).

We thank Bill Kelly for providing the *L. lactis* KF147 strain, Marjo Starrenburg and Iris van Swam for technical support, and the IGS annotation service (University of Maryland).

This research was partly funded by the Kluyver Centre for Genomics of Industrial Fermentation, a Centre of Excellence of the Netherlands Genomics Initiative, and partly by the Netherlands Bioinformatics Centre (NBIC).

REFERENCES

1. Bachmann, H. 2009. Regulatory and adaptive responses of *Lactococcus lactis* in situ. Ph.D. thesis. Wageningen University, Wageningen, Netherlands.
2. Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy. 2004. The Pfam protein families database. *Nucleic Acids Res.* 32:D138–D141.

* Corresponding author. Mailing address: NIZO Food Research, P.O. Box 20, 6710 BA Ede, Netherlands. Phone: 31-318659511. Fax: 31-318650400. E-mail: roland.siezen@nizo.nl.

† Present address: Danone Research, Gut and Microbiology Platform, R.D. 128-91767 Palaiseau Cedex, France.

[∇] Published ahead of print on 26 March 2010.

3. Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin. 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res.* **11**:731–753.
4. Finn, R. D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. Sonnhammer, and A. Bateman. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**:D247–D251.
5. Kelly, W. J., G. P. Davey, and L. J. Ward. 1998. Characterization of lactococci isolated from minimally processed fresh fruit and vegetables. *Int. J. Food Microbiol.* **45**:85–92.
6. Makarova, K., A. Slesarev, Y. Wolf, A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. Polouchine, V. Shakhova, I. Grigoriev, Y. Lou, D. Rohksar, S. Lucas, K. Huang, D. M. Goodstein, T. Hawkins, V. Plengvidhya, D. Welker, J. Hughes, Y. Goh, A. Benson, K. Baldwin, J. H. Lee, I. Diaz-Muniz, B. Dosti, V. Smeianov, W. Wechter, R. Barabote, G. Lorca, E. Altermann, R. Barrangou, B. Ganesan, Y. Xie, H. Rawsthorne, D. Tamir, C. Parker, F. Breidt, J. Broadbent, R. Hutkins, D. O'Sullivan, J. Steele, G. Unlu, M. Saier, T. Klaenhammer, P. Richardson, S. Kozyavkin, B. Weimer, and D. Mills. 2006. Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **103**:15611–15616.
7. Overbeek, R., N. Larsen, T. Walunas, M. D'Souza, G. Pusch, E. Selkov, Jr., K. Liolios, V. Joukov, D. Kaznadzey, I. Anderson, A. Bhattacharyya, H. Burd, W. Gardner, P. Hanke, V. Kapatral, N. Mikhailova, O. Vasieva, A. Osterman, V. Vonstein, M. Fonstein, N. Ivanova, and N. Kyrpides. 2003. The ERGO genome analysis and discovery system. *Nucleic Acids Res.* **31**:164–171.
8. Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**:W116–W120.
9. Siezen, R. J., M. J. Starrenburg, J. Boekhorst, B. Renckens, D. Molenaar, and J. E. van Hylckama Vlieg. 2008. Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl. Environ. Microbiol.* **74**:424–436.
10. Wegmann, U., M. O'Connell-Motherway, A. Zomer, G. Buist, C. Shearman, C. Canchaya, M. Ventura, A. Goesmann, M. J. Gasson, O. P. Kuipers, D. van Sinderen, and J. Kok. 2007. Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. *J. Bacteriol.* **189**:3256–3270.