

Complete Resequencing and Reannotation of the *Lactobacillus plantarum* WCFS1 Genome

Roland J. Siezen,^{a,b,c,d} Christof Francke,^{a,b,c} Bernadet Renckens,^{b,c} Jos Boekhorst,^{a,b,c,d} Michiel Wels,^{a,b,c,d} Michiel Kleerebezem,^{a,c,d} and Sacha A. F. T. van Hijum^{a,b,c,d}

TI Food and Nutrition, Wageningen, The Netherlands^a; Centre for Molecular and Biomolecular Informatics (CMBI 260), Radboud University Medical Centre, Nijmegen, The Netherlands^b; Netherlands Bioinformatics Centre (NBIC), Nijmegen, The Netherlands^c; and NIZO Food Research, Ede, The Netherlands^d

There is growing interest in the beneficial effects of *Lactobacillus plantarum* on human health. The genome of *L. plantarum* WCFS1, first sequenced in 2001, was resequenced using Solexa technology. We identified 116 nucleotide corrections and improved function prediction for nearly 1,200 proteins, with a focus on metabolic functions and cell surface-associated proteins.

Lactobacillus plantarum is a versatile facultative heterofermentative lactic acid bacterium (LAB) found in vegetables, meat, fish, and dairy products (2–4, 14, 19, 20, 32) and the gastrointestinal tract (1). *L. plantarum* WCFS1 has become one of the model strains in LAB research since the initial genome publication (25). Bioinformatics tools have been used to predict the function of its genes (7, 39, 42), reconstruct metabolic pathways (18, 43–45) and gene regulatory networks (17, 49, 51), and compare its genome with genomes of other LAB (6, 26, 52). The genomic, phenotypic, and metabolic diversity of *L. plantarum* has been previously described (31, 40, 41). *L. plantarum* has been employed as a model for LAB interactions with mammalian gut tissues in studies that provided insights into the microbial adaptation to that habitat (8–10, 27–29) and identified candidate probiotic genes (21, 22, 30, 34, 46, 47).

Resequencing performed with a Solexa GAIIX genome analyzer (BaseClear, The Netherlands) resulted in 10,783,316 reads of 50 bp (5-kb mated pairs), totaling 550 Mb (~160× coverage). RoVar software (<http://trac.nbic.nl/rovar>) was used to align Solexa reads to the *L. plantarum* WCFS1 genome sequence by the use of BLAT (24). Read alignment was allowed provided that structural variations (SVs) in the form of single nucleotide polymorphisms (SNPs) or small indels were at least 4 bp from the end of a read. SVs were allowed with a maximum of one read mismatch and a sequence depth of at least 20 reads that unanimously identified a genotype.

Improved manual annotation of encoded functions was performed using Artemis and ACT (12, 13, 36), RAST (5), ISGA (23), Pfam (15, 16), InterProScan (35), BRENDA (38), CAZy (11), TCDB (37), and ERGO (33) software and experimental evidence. *L. plantarum* supermotifs (LPSMs) (48), T-boxes (48, 50), and extracellular protein functions were as previously predicted (7, 52).

The circular chromosome (3,308,273 bp; 44.5% GC content) contains 3,042 protein-encoding genes (of which 18 are pseudogenes), 70 tRNA-encoding genes, 5 rRNA operons, 8 miscellaneous RNAs, 32 T-boxes, and 27 LPSMs. Resequencing showed 116 differences (97 single nucleotide corrections and 19 single nucleotide indels) from the published sequence (25). Thirty-eight corrections are in intergenic regions, and 78 corrections are within protein-coding sequences, leading to 55 amino acid changes and 10 corrections of the N or C terminus of encoded proteins.

Compared to the originally published *L. plantarum* WCFS1

genome (25), 27 coding sequences (CDS) or fragments have now been deleted and 34 CDS or fragments added. Annotations were improved for nearly 1,200 encoded proteins; the improvements included the addition of family information for most transcriptional regulators ($n = 190$), transporters ($n = 79$), and oxidoreductases ($n = 44$). Comparative analysis of putative secreted and cell surface-associated proteins (7, 26, 52) has led to improved annotation of 74 putative extracellular proteins (see the LAB-Secretome database at www.cmbi.ru.nl/lab_secretome/) (52). Originally, 740 CDS were annotated as corresponding to a hypothetical (membrane) protein (25), but 24 of those CDS have now been deleted and 366 have been given a general family assignment ($n = 229$) or a very specific function assignment ($n = 137$). We hope that the comprehensive curated annotation of this model LAB will be of significant use to the many *L. plantarum* researchers worldwide.

Nucleotide sequence accession number. The sequence and the annotation were deposited in EMBL/GenBank at AL935263.2 (GI:342240345), replacing versions AL935263.1 and AL935262.

ACKNOWLEDGEMENTS

We thank Paul de Vos for continued support of this project and all authors of *L. plantarum* literature who provided experimental evidence to support our reannotation.

This project was funded by the Top Institute Food and Nutrition, Wageningen, The Netherlands.

REFERENCES

- Ahrné S, et al. 1998. The normal *Lactobacillus* flora of healthy human rectal and oral mucosa. *J. Appl. Microbiol.* 85:88–94.
- Aquilanti L, et al. 2007. The microbial ecology of a typical Italian salami during its natural fermentation. *Int. J. Food Microbiol.* 120:136–145.
- Aryanta RW, Fleet GH, Buckle KA. 1991. The occurrence and growth of microorganisms during the fermentation of fish sausage. *Int. J. Food Microbiol.* 13:143–155.
- Aymerich T, Martin B, Garriga M, Hugas M. 2003. Microbial quality and direct PCR identification of lactic acid bacteria and nonpathogenic

Received 16 October 2011 Accepted 17 October 2011

Address correspondence to Roland J. Siezen, r.siezen@cmbi.ru.nl.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.06275-11

- staphylococci from artisanal low-acid sausages. *Appl. Environ. Microbiol.* 69:4583–4594.
5. Aziz RK, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
 6. Boekhorst J, et al. 2004. The complete genomes of *Lactobacillus plantarum* and *Lactobacillus johnsonii* reveal extensive differences in chromosome organization and gene content. *Microbiology* 150:3601–3611.
 7. Boekhorst J, Wels M, Kleerebezem M, Siezen RJ. 2006. The predicted secretome of *Lactobacillus plantarum* WCFS1 sheds light on interactions with its environment. *Microbiology* 152:3175–3183.
 8. Bron PA, Granelle C, Mercenier A, de Vos WM, Kleerebezem M. 2004. Identification of *Lactobacillus plantarum* genes that are induced in the gastrointestinal tract of mice. *J. Bacteriol.* 186:5721–5729.
 9. Bron PA, et al. 2004. Genetic characterization of the bile salt response in *Lactobacillus plantarum* and analysis of responsive promoters in vitro and in situ in the gastrointestinal tract. *J. Bacteriol.* 186:7829–7835.
 10. Bron PA, Meijer M, Bongers RS, de Vos WM, Kleerebezem M. 2007. Dynamics of competitive population abundance of *Lactobacillus plantarum* *ivi* gene mutants in faecal samples after passage through the gastrointestinal tract of mice. *J. Appl. Microbiol.* 103:1424–1434.
 11. Cantarel BL, et al. 2009. The Carbohydrate-Active EnZymes database (CAZY): an expert resource for glycogenomics. *Nucleic Acids Res.* 37:D233–D238.
 12. Carver T, et al. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24:2672–2676.
 13. Carver TJ, et al. 2005. ACT: the Artemis comparison tool. *Bioinformatics* 21:3422–3423.
 14. Ercolini D, Hill PJ, Dodd CE. 2003. Bacterial community structure and location in Stilton cheese. *Appl. Environ. Microbiol.* 69:3540–3548.
 15. Finn RD, et al. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.* 34:D247–D251.
 16. Finn RD, et al. 2010. The Pfam protein families database. *Nucleic Acids Res.* 38:D211–D222.
 17. Francke C, Kerkhoven R, Wels M, Siezen RJ. 2008. A generic approach to identify transcription factor-specific operator motifs; inferences for LacI-family mediated regulation in *Lactobacillus plantarum* WCFS1. *BMC Genomics* 9:145.
 18. Francke C, Siezen RJ, Teusink B. 2005. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.* 13:550–558.
 19. Gänzle MG, Vermeulen N, Vogel RF. 2007. Carbohydrate, peptide and lipid metabolism of lactic acid bacteria in sourdough. *Food Microbiol.* 24:128–138.
 20. Gardner NJ, Savard T, Obermeier P, Caldwell G, Champagne CP. 2001. Selection and characterization of mixed starter cultures for lactic acid fermentation of carrot, cabbage, beet and onion vegetable mixtures. *Int. J. Food Microbiol.* 64:261–275.
 21. Gross G, Snel J, Boekhorst J, Smits MA, Kleerebezem M. 2010. Biodiversity of mannose-specific adhesion in *Lactobacillus plantarum* revisited: strain-specific domain composition of the mannose-adhesin. *Benef. Microbes* 1:61–66.
 22. Gross G, et al. 2008. Mannose-specific interaction of *Lactobacillus plantarum* with porcine jejunal epithelium. *FEMS Immunol. Med. Microbiol.* 54:215–223.
 23. Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q. 2010. An Ergatis-based prokaryotic genome annotation web server. *Bioinformatics* 26:1122–1124.
 24. Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12:656–664.
 25. Kleerebezem M, et al. 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc. Natl. Acad. Sci. U. S. A.* 100:1990–1995.
 26. Kleerebezem M, et al. 2010. The extracellular biology of the lactobacilli. *FEMS Microbiol. Rev.* 34:199–230.
 27. Marco ML, Bongers RS, de Vos WM, Kleerebezem M. 2007. Spatial and temporal expression of *Lactobacillus plantarum* genes in the gastrointestinal tracts of mice. *Appl. Environ. Microbiol.* 73:124–132.
 28. Marco ML, et al. 2010. Convergence in probiotic *Lactobacillus* gut-adaptive responses in humans and mice. *ISME J.* 4:1481–1484.
 29. Marco ML, et al. 2009. Lifestyle of *Lactobacillus plantarum* in the mouse caecum. *Environ. Microbiol.* 11:2747–2757.
 30. Meijerink M, et al. 2010. Identification of genetic loci in *Lactobacillus plantarum* that modulate the immune response of dendritic cells using comparative genome hybridization. *PLoS One* 5:e10632.
 31. Molenaar D, et al. 2005. Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J. Bacteriol.* 187:6119–6127.
 32. Mundt JO, Hammer JL. 1968. Lactobacilli on plants. *Appl. Microbiol.* 16:1326–1330.
 33. Overbeek R, et al. 2003. The ERGO genome analysis and discovery system. *Nucleic Acids Res.* 31:164–171.
 34. Pretzer G, et al. 2005. Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. *J. Bacteriol.* 187:6128–6136.
 35. Quevillon E, et al. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33:W116–W120.
 36. Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945.
 37. Saier MH, Jr, Yen MR, Noto K, Tamang DG, Elkan C. 2009. The transporter classification database: recent advances. *Nucleic Acids Res.* 37:D274–D278.
 38. Scheer M, et al. 2011. BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* 39:D670–D676.
 39. Siezen R, et al. 2006. *Lactobacillus plantarum* gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria. *BMC Genomics* 7:126.
 40. Siezen RJ, van Hylckama Vlieg JET. 2011. Genomic diversity and versatility of *Lactobacillus plantarum*, a natural metabolic engineer. *Microb. Cell Fact.* 10:S3.
 41. Siezen RJ, et al. 2010. Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches. *Environ. Microbiol.* 12:758–773.
 42. Siezen RJ, van Enckevort FH, Kleerebezem M, Teusink B. 2004. Genome data mining of lactic acid bacteria: the impact of bioinformatics. *Curr. Opin. Biotechnol.* 15:105–115.
 43. Teusink B, et al. 2005. In silico reconstruction of the metabolic pathways of *Lactobacillus plantarum*: comparing predictions of nutrient requirements with those from growth experiments. *Appl. Environ. Microbiol.* 71:7253–7262.
 44. Teusink B, Wiersma A, Jacobs L, Notebaart RA, Smid EJ. 2009. Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation. *PLoS Comput. Biol.* 5:e1000410.
 45. Teusink B, et al. 2006. Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J. Biol. Chem.* 281:40041–40048.
 46. Troost FJ, et al. 2008. Identification of the transcriptional response of human intestinal mucosa to *Lactobacillus plantarum* WCFS1 in vivo. *BMC Genomics* 9:374.
 47. van Baarlen P, et al. 2009. Differential NF-kappaB pathways induction by *Lactobacillus plantarum* in the duodenum of healthy humans correlating with immune tolerance. *Proc. Natl. Acad. Sci. U. S. A.* 106:2371–2376.
 48. Wels M, et al. 2009. Large intergenic cruciform-like supermotifs in the *Lactobacillus plantarum* genome. *J. Bacteriol.* 191:3420–3423.
 49. Wels M, Francke C, Kerkhoven R, Kleerebezem M, Siezen RJ. 2006. Predicting cis-acting elements of *Lactobacillus plantarum* by comparative genomics with different taxonomic subgroups. *Nucleic Acids Res.* 34:1947–1958.
 50. Wels M, Groot Kormelink T, Kleerebezem M, Siezen RJ, Francke C. 2008. An in silico analysis of T-box regulated genes and T-box evolution in prokaryotes, with emphasis on prediction of substrate specificity of transporters. *BMC Genomics* 9:330.
 51. Wels M, Overmars L, Francke C, Kleerebezem M, Siezen RJ. 2011. Reconstruction of the regulatory network of *Lactobacillus plantarum* WCFS1 on basis of correlated gene expression and conserved regulatory motifs. *Microb. Biotechnol.* 4:333–344.
 52. Zhou M, Theunissen D, Wels M, Siezen RJ. 2010. LAB-Secretome: a genome-scale comparative analysis of the predicted extracellular and surface-associated proteins of lactic acid bacteria. *BMC Genomics* 11:651.