



Discovering lactic acid bacteria by genomics

Todd Klaenhammer^{1,*}, Eric Altermann¹, Fabrizio Arigoni³, Alexander Bolotin⁴, Fred Breidt⁵, Jeffrey Broadbent⁶, Raul Cano⁷, Stephane Chaillou²⁴, Josef Deutscher⁸, Mike Gasson⁹, Maarten van de Guchte⁴, Jean Guzzo¹⁰, Axel Hartke²⁵, Trevor Hawkins¹¹, Pascal Hols¹², Robert Hutkins¹³, Michiel Kleerebezem², Jan Kok¹⁴, Oscar Kuipers¹⁴, Mark Lubbers¹⁵, Emmanuelle Maguin⁴, Larry McKay¹⁶, David Mills¹⁷, Arjen Nauta¹⁸, Ross Overbeek¹⁹, Herman Pel²⁰, David Pridmore³, Milton Saier²¹, Douwe van Sinderen²², Alexei Sorokin⁴, James Steele²³, Daniel O'Sullivan¹⁶, Willem de Vos², Bart Weimer⁶, Monique Zagorec²⁴ & Roland Siezen^{2,26,*}

¹Department of Food Science, Southeast Dairy Foods Research Center, North Carolina State University, Raleigh, NC 27695-7624, USA; ² Wageningen Center for Food Sciences, 6700AN Wageningen, The Netherlands; ³ Nestlé Research Center, Vers-chez-les-Blanc, 1000 Lausanne 26, Switzerland; ⁴Génétique Microbienne, CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France; ⁵USDA Agricultural Research Service, Department of Food Science, North Carolina State University, Raleigh, NC 27695-7624, USA; ⁶ Department of Nutrition and Food Sciences, Utah State University, Logan, UT 84322-8700, USA; ⁷ Environmental Biotechnology Institute, California Polytechnic State University, San Luis Obispo, CA 93407, USA; ⁸ Laboratoire de Génétique des Microorganismes, INRA-CNRS URA1925, F-78850 Thiverval-Grignon, France; ⁹ Institute of Food Research, Norwich, UK; ¹⁰ Laboratoire de Microbiologie, UMR INRA-Université de Bourgogne, Equipe PG2MA ENSBANA 21000 Dijon, France; ¹¹ Joint Genome Institute Production Genomic Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA; ¹² Unité de Génétique, Batiment Camoy, Place Croix du Sud, UCL, B-1348, Louvain-la Neuve, Belgium; ¹³ University of Nebraska, Department of Food Science and Technology, Lincoln, NE 68583-0919, USA; ¹⁴ Department of Molecular Genetics, University of Groningen, 9751NN Haren, The Netherlands; ¹⁵ Fonterra Research Centre [Formerly the New Zealand Dairy Research Institute], Private Bag 11 029, Palmerston North, New Zealand; ¹⁶ Department of Food Science and Nutrition, University of Minnesota, St. Paul, MN 55108, USA; ¹⁷ Department of Viticulture and Enology, University of California, Davis, CA 95616-8749, USA; ¹⁸ Corporate Research FCDF, PO Box 87, 7400AB Deventer, The Netherlands; ¹⁹ Interegated Genomics Inc, 2201W Campbell Park Dr, Chicago, IL 600612, USA; ²⁰ DSM Food Specialties, PO Box 1, 2600MA Delft, The Netherlands; ²¹ University California, San Diego, La Jolla, CA 92093-0116, USA; ²² Department of Microbiology, University College Cork, Cork, Republic of Ireland; ²³ Department of Food Science, University of Wisconsin-Madison, Madison, WI 53706-1565, USA; ²⁴ Flore Lactique et Environnement Carné, INRA-CRJ, Domaine de Vilvert, 78350 Jouy-en-Josas, France; ²⁵ Laboratoire de Microbiologie de l'Environnement, EA 956 USC INRA, IRBA Université de Caen, F-14032 Caen Cedex, France; ²⁶ Center for Molecular and Biomolecular Informatics, University of Nijmegen, PO Box 9010, 6500GL Nijmegen, The Netherlands (*Authors for correspondence; E-mail: klaenhammer@ncsu.edu, siezen@cmbi.kun.nl)

Key words: lactic acid bacteria, genomics, Gram-positive bacteria, food, health, *Lactococcus*, *Lactobacillus*, *Streptococcus*, *Pediococcus*, *Leuconostoc*, *Oenococcus*, *Propionibacterium*, *Bifidobacterium*, *Brevibacterium*

Abstract

This review summarizes a collection of lactic acid bacteria that are now undergoing genomic sequencing and analysis. Summaries are presented on twenty different species, with each overview discussing the organisms fundamental and practical significance, environmental habitat, and its role in fermentation, bioprocessing, or probiotics. For those projects where genome sequence data were available by March 2002, summaries include

a listing of key statistics and interesting genomic features. These efforts will revolutionize our molecular view of Gram-positive bacteria, as up to 15 genomes from the low GC content lactic acid bacteria are expected to be available in the public domain by the end of 2003. Our collective view of the lactic acid bacteria will be fundamentally changed as we rediscover the relationships and capabilities of these organisms through genomics.

Introduction (contributed by Todd Klaenhammer)

The lactic acid bacteria (LAB) represent a group of bacteria that are functionally related by their ability to produce lactic acid during homo- or heterofermentative metabolism. The acidification and enzymatic processes accompanying the growth of LAB impart the key flavor, texture, and preservative qualities to a variety of fermented foods. Industrial applications of the LAB rely on six key beneficial and non-pathogenic species; *Lactococcus* (milk), *Lactobacillus* (milk, meat, vegetables, cereal), *Leuconostoc* (vegetables, milk), *Pediococcus* (vegetables, meat), *Oenococcus oeni* (wine) and *Streptococcus thermophilus* (milk). Other members of the LAB, notably lactobacilli, occupy important niches in the gastrointestinal tracts of humans and animals and are considered to offer a number of probiotic benefits to general health and well being. These benefits include a positive influence on the normal microflora, competitive exclusion of pathogens, and stimulation/modulation of mucosal immunity. More recently, LAB are being used in the production of industrial chemical and biological products including biopolymers (*Leuconostoc* spp.), bulk enzymes (*Lactobacillus brevis*), ethanol, and lactic acid (*Lactobacillus casei*, *lactis*, *delbrueckii*, *brevis*) (Gold et al. 1992; Hofvendahl & Hahn-Hagerdal 2000). The LAB are also strong candidates for development as oral delivery vehicles for digestive enzymes and vaccine antigens (Wells et al. 1996; Pouwels et al. 1998; Steidler et al. 2000). Their innate acid tolerance, ability to survive gastric passage, and safety record during human consumption, are key features that can be exploited to effectively deliver biologics to targeted locations and tissues.

Realizing their practical significance in fermentation, bioprocessing, agriculture, food, and more recently, medicine, the LAB have been the subject of considerable research and commercial development over the past decade. Contributing to this explosion have been the recent efforts to determine the genome sequences of a representative collection of LAB species and strains. The first complete genome of the LAB group was published on *Lactococcus lactis* subsp. *lactis* IL1403 by Bolotin et al. (2001). The genome was

2.4 Mb in size and revealed a number of unexpected findings: biosynthetic pathways for all 20 amino acids, albeit not all are functional, a complete set of late competence genes, five complete prophages, and partial components for aerobic metabolism. Noting that some of these systems are not functional or complete, the genomic analysis of *Lactococcus* suggests an evolutionary trend toward minimization of the chromosome and elimination of unnecessary systems during adaptation to nutritionally complex environments, such as milk. At this writing, four other LAB genomes have been completed, (*L. plantarum*, *L. johnsonii*, *L. acidophilus*, *S. thermophilus*), and >20 more are in progress (Table 1) with expected completions for sequencing by the end of 2002. Additional genome sequencing is underway for microbes that are not considered as members of the LAB, but contribute important LAB-like properties either as probiotics (*Bifidobacterium longum*, *B. breve*, and *Brevibacterium linens*) or flavor adjuncts (*B. linens*, *Propionibacterium freudenreichii*). Among the total genome projects ongoing, there are several cases where genome sequences will become available for multiple strains of the same species, notably *L. lactis* (three strains), *L. casei* (two strains), *L. delbrueckii* (three strains), *S. thermophilus* (three strains), *Oenococcus oeni* (two strains) and *B. longum* (two strains).

Of the 29 genomes listed in Table 1 on LAB-type microbes, 11 genomes are being sequenced by the Department of Energy–Joint Genome Institute (JGI) in collaboration with the Lactic Acid Bacteria Genome Consortium (LABGC), composed of 10 US scientists representing seven universities in the U.S. As part of their microbial genomes program (see http://www.jgi.doe.gov/JGI_microbial/html/index.html) JGI will carry out genome sequencing of LAB species representing considerable diversity in ecological habitat (milk, meat, plants, GI tract) and roles (probiotic versus fermentation). All of the microbial genomes will undergo the same processing; first to generate a 10× coverage of each genome with a random shotgun small-insert library (3 kb), and to supplement this with ~5× coverage of a large insert (40 kb) cosmid library. These data will be incorporated into an assembly and the assembled scaffolds

Table 1. Genome sequencing projects of lactic acid bacteria

Species	Strain	genome size (Mb)	Contact person	Institution
Sequencing completed				
<i>Lactococcus lactis</i>	IL1403	2.3	A. Sorokin	INRA and Genoscope, FR
<i>Lactobacillus plantarum</i>	WCFS1	3.3	M. Kleerebezem	WCFS, NL
<i>Lactobacillus johnsonii</i>	NCC533	2.0	D.Pridmore	Nestlé, CH
<i>Lactobacillus acidophilus</i>	ATCC700396 NCFM	2.0	T. Klaenhammer, R. Cano	North Carolina State University and Cal-Poly Technical University, USA
<i>Bifidobacterium longum</i>	NCC2705	2.3	F. Arigoni	Nestlé, CH
Sequencing ongoing				
<i>Lactococcus lactis</i> ssp. <i>cremoris</i>	MG1363	2.6	O.Kuipers, M.Gasson, D. van Sinderen	University of Groningen, NL; IFR, UK; UCC, IRL
<i>Lactococcus lactis</i> ssp. <i>cremoris</i>	MG1363	2.4	A. Sorokin, A. Bolotin	INRA, FR
<i>Lactobacillus delbrueckii</i> ssp. <i>bulgaricus</i>	ATCC11842	2.3	E. Maguin, M.vd Guchte	Genoscope and INRA, FR.
<i>Lactobacillus delbrueckii</i> ssp. <i>bulgaricus</i>	DN-100107	2.1	T. Smokvina	Danone Vitapole, FR
<i>Lactobacillus sakei</i>	23K	1.9	M. Zagorec, S. Chaillou	INRA, FR.
<i>Lactobacillus casei</i>	BL23	2.6	J. Deutscher, A. Hartke	INRA/CNRS, Caen University, FR
<i>Lactobacillus helveticus</i>	CNRZ32	2.4	J. Steele	University of Wisconsin, USA
<i>Lactobacillus rhamnosus</i>	HN001	2.4	M. Lubbers	Fonterra Research Center (formerly NZDRI) and ViaLactia BioSciences, NZ
<i>Streptococcus thermophilus</i> .	LMG18311	1.9	P. Hols	UCL, Belgium
<i>Streptococcus thermophilus</i>	CNRZ1066	1.8	A.Bolotin	INRA, FR, Integrated Genomics Inc, USA
<i>Oenococcus oeni</i>	IOEB84.13	1.8	J. Guzzo	Universities Dijon and Bordeaux-INRA- GENOME Express, FR
<i>Bifidobacterium breve</i> .	NCIMB8807	2.4	D. van Sinderen	University College Cork (UCC), IRL
<i>Propionibacterium</i> <i>freudenreichii</i>	ATCC6207	2.6	H. Pel, J. Sikkema	DSM Food Specialties, Friesland Coberco Dairy Foods, NL
Joint Genome Institute/Lactic Acid Bacteria Genome Consortium (JGI/LABGC)				
<i>Lactobacillus gasseri</i>	ATCC33323	1.8	T. Klaenhammer	North Carolina State University, USA
<i>Lactobacillus casei</i>	ATCC334	>2.2	J. Broadbent	Utah State University, USA
<i>Lactobacillus delbrueckii</i> ssp. <i>bulgaricus</i>	ATCCBAA-365	2.3	J. Steele	University of Wisconsin, Madison, USA
<i>Lactobacillus brevis</i>	ATCC367	2.0	M. Saier	University of California, San Diego, USA
<i>Lactococcus lactis</i> ssp. <i>cremoris</i>	SK11	2.3	B. Weimer, L. McKay	Utah State University and University of Minnesota, USA
<i>Leuconostoc mesenteroides</i>	ATCC 8293		F. Breidt, H.P. Fleming	USDA, North Carolina State University, USA
<i>Oenococcus oeni</i>	PSU1 ATCC BAA-331	1.8	D. Mills	University of California, Davis, USA
<i>Pediococcus pentosaceus</i>	ATCC25745	2.0	J. Steele, J. Broadbent	University of Wisconsin, Madison, Utah State University, USA
<i>Streptococcus thermophilus</i>	ATCC BAA-491	1.8	R. Hutkins	University of Nebraska, USA
<i>Brevibacterium linens</i>	BL2/ ATCC 9174	3.0	B. Weimer	Utah State University, USA
<i>Bifidobacterium longum</i>	DJ010A	2.1	D. O'Sullivan	University of Minnesota, USA

ordered by PCR primer walking to fill gaps. The scaffolds will undergo an automated annotation which will be performed by Oak Ridge National Laboratories.

It is noteworthy that as these sequences are generated, they will be placed in the public domain on the JGI website for common use. The timely and public availability of genome information for various LAB species will catapult our collective efforts to carry on with comparative and functional genomic analyses of the LAB group.

The explosion of available genome sequences for LAB will accelerate their exploitation in both traditional and non-traditional arenas. While phylogenetically closely related by their small genomes (~2–4 Mb), the LAB occupy a diverse set of ecological niches suggesting that considerable genetic adaptation has occurred during their evolution. Comparison of the genome sequences of multiple LAB species and strains is expected to provide a critical view of microbial adaptation and genetic events leading to their adaptation to specialized environments. Comparative genomics among the microbes sequenced thus far has already illustrated that essential housekeeping gene functions are widely conserved among microbes and horizontal gene transfer commonly occurs. An expected outcome of comparative genomics of LAB will be the definition of conserved and unique genetic functions in LAB that enable core functions, e.g., production of lactic acid, proteolytic and peptidase activities, survival at low pH, stress tolerance, production of antimicrobials, transport systems, cell signaling, and attachment/retention in dynamically mobile environments.

It is anticipated that IS-elements, bacteriophages, and mobile genetic elements provide the major routes through which horizontal gene transfer occurs, and the roadmap to the most interesting and practically significant genetic regions that underscore the unique and beneficial properties of the LAB. It is well documented that the LAB undergo conjugation, exist in phage-contaminated environments where gene transfer may occur by transduction, and harbor sets or remnants of competence genes for transformation (Bolotin et al. 2001). Understanding gene transfer, particularly in environments where LAB coexist, or compete, will provide one important view of their evolution, adaptation, and potential for unique applications. Conjugation has played a key role in the evolution and adaptation of *L. lactis* to a milk environment noting that many attributes for growth in milk, including lactose and casein utilization and bacteriophage res-

istance, are encoded by conjugative plasmid DNA in this species (Broadbent 2001).

With the availability of genomic maps of more than 25 LAB, comparative genomic analysis will identify critical similarities and differences and is expected to identify 'islands of adaptability', defined as key genetic regions that may be instrumental in the evolution of the various species to their specialized habitats, and their functions within those environments. Comparisons of food-grade LAB with other related Gram-positive pathogens (*Enterococcus faecalis*, *Streptococcus agalactiae*, *Streptococcus equi*, *Streptococcus pneumoniae*, *Streptococcus mutans*, *Streptococcus pyogenes*, *Listeria monocytogenes*) has already revealed many common features and will most certainly define the essential genetic differences between pathogens, non-pathogens, and commensals. Over the next year, our scientific community will be fundamentally empowered by the availability of numerous related, yet distinct genomes, to make these comparative analyses and define the similarities and differences that characterize the genomes of LAB.

This chapter presents a collection of summaries on the LAB organisms that are now in genomic sequencing or analysis. Each summary describes the organism and its roles in the environment or bioprocessing, the status of the sequencing effort as of March 2002 and, for some, selected interesting features that have been uncovered (Tables 2 and 3). Our view of the LAB will be fundamentally changed as we discover the relationships and capabilities of these organisms through genomics.

SUMMARIES OF SEQUENCING PROJECTS

Lactococcus lactis subsp. *cremoris* SK11 (contributed by Larry L. McKay and Bart Weimer)

Lactococci are mesophilic LAB that were first isolated from green plants. However, today they are used extensively in food fermentations, which represent about 20% of the total economic value of fermented foods produced throughout the world. This group of bacteria, previously designated the lactic streptococci (*Streptococcus lactis* subsp. *lactis* or *S. lactis* subsp. *cremoris*) was placed in this new taxon in 1987 by Schleifer. Lactococci gained notable interest because many of their functions important for successful fermentations are linked to plasmid DNA

Table 2. Genome statistics (March 2002)

Species	Strain	Genome		ORFs		Prophages		IS elements		RNA		Plasmids
		Size (Mb)	Contigs %GC	Intact	Remnant	Total intact	Different types	rRNA	tRNA	operons	loci number	
<i>L. lactis</i>	IL1403	2.366	1 35.4	2310	5 1	43	6 (IS981, IS982, IS983, IS904, IS905, IS1077)	6	62	62	None	
<i>L. lactis</i>	MG1363	~2.5	458 37.1 (>1 kb)	~2500	0 6	62	8 (IS981, IS982, IS905, IS904, IS946, IS1077G, ISS1, IS1675)	6	?	?	None	
<i>L. plantarum</i>	WCFS1	3.308	1 44.5	~3050	2 2	12	2 (IS1165, IS1182)	5	5	62	1.9 (39.5%) 2.4 (34.3%) 36.1 (40.8%)	
<i>L. johnsonii</i>	NCC533	2.022	1 34.6	1857	3 0	15	6 (ISL3, IS30)	6	4	87 ?	None	
<i>L. acidophilus</i>	ATCC700396	1.99	1 34.7	1979	0 3	6	1 (IS3)	6	6	?	None	
<i>L. gasseri</i>	ATCC33323	1.84	67 35.1	1803	1 1	3	3 (ISL3, IS1397, IS1167)	>2	20	20	None	
<i>L. helveticus</i>	CNRZ32	~2.4	~1000 37.1	> 2,300	? ≥1	?	> 5 (IS1201, ISL3, ISLh1, IS3, IS30)	≥3	?	?	None	
<i>L. delbrueckii</i> ssp. <i>bulgaricus</i>	ATCC11842	1.8	3 50	?	? ?	Several		?	6 or 7	?	None	
<i>L. sakei</i>	23K	1.772 (96%)	<150 41.4	>1792	? ?	12	3 (IS1520-types, new IS30-like, new IS150-like)	7	6	>58	None	
<i>L. casei</i>	BL23	~2.6	563 46	?	? Several	18	5 (IS1, IS3, IS30)	≥5	?	?	None	
<i>O. oeni</i>	IOEB84.13	1.754	63 37.9	1784	4	3		2	2	43	3.9	
<i>P. freudenreichii</i>	ATCC6207	2.64	12 67.4	2552	? Many	?	?	2 (?)	2	44	None	
<i>B. breve</i>	NCIMB8807	~2.43	140 ~53	?	0 ~35	?	(IS2606, IS1383, IS1088, IS10, IS21)	~4	?	?	None	
<i>B. longum</i>	ATCC9174	2.256	1 60.1	~1730	Some?	16	5 (IS3, IS21, IS30, IS256, IS200/IS605)	4	3	57	3.6 (64.8%)	

