| Bacteriology | Announcement

Draft genome sequences of 14 *Lacticaseibacillus* spp. strains, representatives of a collection of 200 strains

Andrea Colautti,[1] Nicolò Rossi,[2] Carla Piazza,[2] Giuseppe Comi,[1] Lucilla Iacumin[1]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Lactobacilli have a fundamental role in the food industry as starters and probiotics, therefore, requiring special attention concerning food safety. In this work, 14 strains selected accordingly to their genetic fingerprint and physiologic characteristics are presented as representatives of a collection of 200 strains.

L actobacilli are ubiquitous lactic acid bacteria and are of particular interest, given their significant presence in fermented and non-fermented foods and in the human commensal microbiota. Thanks to their long history of human use and consumption and their use as probiotics, they have been generally recognized as safe (1). However, despite the numerous possible beneficial effects reported in the literature (2), several cases of infections caused by these bacteria have been reported over time in immunodeficient subjects, likely due to the presence of virulence genes and antibiotic-resistance genes (3).

As recently stated by EFSA (European Food Safety Authority), the whole-genome sequencing constitutes an adequate tool to taxonomically characterize and carry out a risk assessment by verifying the presence of health concern factors in microorganisms intentionally used in the food chain (4). Therefore, this work aimed to provide the genome sequence of highly diverse representative strains to have the opportunity to clarify some specific genetic traits depending on the origin of the strains and associated with the virulence factors in the strictly connected species *Lacticaseibacillus rhamnosus*, *Lacticaseibacillus paracasei, Lacticaseibacillus casei,* and *Lacticaseibacillus zeae* recently reclassified (5, 6). These strains were selected as representative from a collection of 200 strains that were characterized using RAPD (random amplification of polymorphic DNA), Rep-PCR (repetitive-element PCR), Sau-PCR, and MLST (multilocus sequence typing) based on stress-related genes. Taking into account all the techniques employed, the fingerprint analyses allowed to clusterize the strains on the bases of their genetic profile. One representative strain per each different genetic profile was selected for the whole-genome sequencing for a total of 14 strains (7–9). The culture collection was cryopreserved in MRS (De Man, Rogosa, and Sharpe) broth (Oxoid, Italy) and added with 20% glycerol until their use for the analysis. For the sequencing process, each freeze-dried strain was cultured in MRS broth at 30°C for 48 h. After centrifugation for 5 min at 5,000 × *g,* the DNA was extracted using the phenol-chloroform method (10), and genomic libraries were constructed employing the TruSeq DNA PCR-Free LT Kit (Illumina, USA) using 2.5 µg of genomic DNA, which were fragmented with a Bioruptor NGS Ultrasonicator (Diagenode, USA), followed by size evaluation using Tape Station 2200 (Agilent Technologies) (Table 1).

Library samples were loaded into a Flow Cell V3 600 cycles (Illumina, USA) according to the technical support guide. Draft genome sequencing was performed through the genomic platform consisting of a MiSeq (Illumina, UK), following the protocol of the

TABLE 1 Statistics of assembled genomes

GenBank accession no.	SRA accession no.	Organism name	Strain	Raw reads[a]	Source	Provenience[b]	Genome size[a]	Contigs[a]	N50[a]	G + C content (%)[c]	CDS[a]	Completeness (%)[d]
GCA_028878355.1	SRR17145328	L. paracasei	LMG 25883	720,624	Dairy Products	LMG	3,017,070	54	99,247	46.25	2,882	99.46
GCA_028878315.1	SRR17145327	L. paracasei	DSM 4905	755,994	Human	DSM	3,097,123	58	137,688	46.29	2,954	99.46
GCA_028878305.1	SRR17145323	L. paracasei	NRRL B-456	740,902	Unknown	ARS	3,118,403	117	102,794	46.19	2,997	99.46
GCA_028878235.1	SRR17145322	L. paracasei	M268	748,130	Dairy Products	POT	2,730,606	140	57,940	46.28	2640	99.39
GCA_028878245.1	SRR17145321	L. rhamnosus	O14	850,782	Dairy Products	POT	2,910,638	39	283,390	46.7	2,693	99.46
GCA_028878215.1	SRR17145320	L. zeae	UD2202	737,698	Dairy Products	UDI	3,038,780	42	179,246	47.97	2,778	99.46
GCA_028878255.1	SRR17145319	L. paracasei	I2	886,860	Sourdough	CAM	2,992,737	157	48,532	46.41	2,835	99.46
GCA_028878205.1	SRR17145318	L. casei	UD1001	1,742,298	Human	UDI	3,147,269	41	276,690	47.88	2,900	99.46
GCA_028878145.1	SRR17145317	L. rhamnosus	N1110	2,788,788	Human	CAM	3,068,245	84	119,405	46.57	2,848	99.46
GCA_028878115.1	SRR17145316	L. rhamnosus	N202	1,930,166	Human	CAM	2,882,421	59	123,389	46.57	2,699	99.46
GCA_028878345.1	SRR17217968	L. rhamnosus	UD193	1,639,344	Dairy Products	UDI	3,114,057	46	196,800	46.69	2,912	99.46
GCA_028878125.1	SRR17145326	L. rhamnosus	Mo2	788,482	Human	CAM	2,943,670	63	119,233	46.62	2,706	99.46
GCA_028878105.1	SRR17145325	L. paracasei	TMW 1.300	782,760	Beer	LTM	3,178,055	136	60,472	46.13	3,108	99.46
GCA_028878155.1	SRR17145324	L. paracasei	DIALYac	932,552	Dairy Products	UDI	3,037,719	95	125,133	46.24	2,950	99.46

[a]Determined using PGAP.
[b]Determined using Quast.
[c]Determined using CheckM.
[d]Provenience: LMG: BCCM/LMG, Belgian Co-ordinated Collections of Micro-organisms (BCCM), Belgium; DSM: DSM, Deutsche Sämmlung von Mikroorganismen und Zellkülturen, Braunschweig, Germany; ARS: ARS Culture (NRRL) Collection, U.S. Department of Agriculture, USA; POT: Scuola di Scienze Agrarie, Alimentari e Ambientali, Università degli Studi della Basilicata, Potenza, Italy; UDI: Dipartimento di Scienze degli Alimenti, Università degli studi di Udine, Udine, Italy; CAM: Dipartimento di Agricoltura, Ambiente e Alimenti, Università degli Studi del Molise, Campobasso, Italy; LTM: Lehrstuhl fürTechnische Mikrobiologie, Technische Universität München, Freising, Germany.

supplier (Illumina, UK). Fastq files of the 250-bp paired-end reads obtained from targeted genome sequencing of the isolated strains were used as input for genome assemblies.

The reads were analyzed and assembled with the WGA-LP pipeline (11) using the following tools included in the pipeline with default settings. The reads were qualitatively trimmed (Phred >20), and the Illumina adapters were removed using Trimmomatic v0.39 (12), validating the results with FastQC v0.11.9 (13). The presence of any contamination was verified by Kraken2 v2.0.8-b (14). The assembly was made using SPAdes v3.15.2 (15), reordering the resulting scaffolds by aligning them with reference sequences (*L. casei* 12A NZ_CP006690.1, *L. paracasei* ATCC334 NC_008526.1, and *L. rhamnosus* GG NC_013198.1) using Mauve v2.4.0 (16). The final quality of the assemblies was then evaluated by CheckM v1.1.3 (17), SamTools v1.10 (18), and Quast v5.0.2 (19). Functional annotation was carried out on the genomes using PGAP 2022-04-14.build6021 (20).

AUTHOR AFFILIATIONS

[1]Department of Agricultural, Food, Environmental and Animal Science (Di4A), University of Udine, Udine, Italy
[2]Department of Mathematics, Computer Science, and Physics, University of Udine, Udine, Italy

AUTHOR ORCIDs

Andrea Colautti http://orcid.org/0000-0003-3203-7046
Carla Piazza http://orcid.org/0000-0003-2967-8109
Lucilla Iacumin http://orcid.org/0000-0003-2967-8109

AUTHOR CONTRIBUTIONS

Andrea Colautti, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing, Data curation | Nicolò Rossi, Formal analysis, Methodology, Writing – review and editing, Conceptualization, Software | Carla Piazza, Writing – review and editing, Software, Validation | Giuseppe Comi, Writing – review and editing, Resources | Lucilla Iacumin, Writing – review and editing, Conceptualization, Funding acquisition, Project administration, Supervision

DATA AVAILABILITY

Sequences were deposited in GenBank with BioProject accession number PRJNA786620. Table 1 reports for each sample the taxonomical identification, the isolation source, GenBank accession number, sequencing and assembly statistics, and genome features of strains.

REFERENCES

1. Adams MR. 1999. Safety of industrial lactic acid bacteria. J Biotechnol 68:171–178. https://doi.org/10.1016/s0168-1656(98)00198-9
2. Colautti A, Orecchia E, Comi G, Iacumin L. 2022. Lactobacilli, a weapon to counteract pathogens through the inhibition of their virulence factors. J Bacteriol 204:e0027222. https://doi.org/10.1128/jb.00272-22
3. Colautti A, Arnoldi M, Comi G, Iacumin L. 2022. Antibiotic resistance and virulence factors in lactobacilli: something to carefully consider. Food Microbiol 103:103–934. https://doi.org/10.1016/j.fm.2021.103934
4. European Food Safety Authority (EFSA). 2021. EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. EFSA J 19:e06506. https://doi.org/10.2903/j.efsa.2021.6506
5. Zheng J, Wittouck S, Salvetti E, Franz C, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. 2020. A taxonomic note on the genus lactobacillus: description of 23 novel genera, emended description of the genus lactobacillus beijerinck 1901, and union of lactobacillaceae and leuconostocaceae. Int J Syst Evol Microbiol 70:2782–2858. https://doi.org/10.1099/ijsem.0.004107
6. Liu DD, Gu TG. 2020. Proposal to reclassify *Lactobacillus zhaodongensis*, *Lactobacillus zeae*, *Lactobacillus argentoratensis* and *Lactobacillus buchneri* subsp. *silagei* as *Lacticaseibacillus zhaodongensis* comb nov., *Lacticaseibacillus zeae* comb. nov., *Lactiplantibacillus argentoratensis* comb. nov. and *Lentilactobacillus buchneri* subsp. *silagei* comb. nov., respectively and *Apilactobacillus kosoi* as a later heterotypic synonym of *Apilactobacillus micheneri*. Int J Syst Evol Microbiol 70:6414–6417.
7. Iacumin L, Ginaldi F, Manzano M, Anastasi V, Reale A, Zotta T, Rossi F, Coppola R, Comi G. 2015. High resolution melting analysis (HRM) as a new tool for the identification of species belonging to the Lactobacillus casei group and comparison with species-specific PCRs and Multiplex PCR. Food Microbiol 46:357–367. https://doi.org/10.1016/j.fm.2014.08.007
8. Camprini L, Ginaldi F, Colautti A, Pellegrini M, Comi G, Iacumin L. 2023. Safety traits, genetic and technological characterization of L. casei/ paracasei strains. Submitted for publication (under review process)
9. Ginaldi F, Camprini L, Colautti A, Pellegrini M, Comi G, Iacumin L. 2023. Safety traits, genetic and technological characterization of L. rhamnosus strains. submitted for publication (under review process)
10. Sambrook J, Russel DW. 2001. Molecular cloning. 3rd ed. Cold Spring Harbor Laboratory Press, New York.
11. Rossi N, Colautti A, Iacumin L, Piazza C. 2021. WGA-LP: a pipeline for whole genome assembly of contaminated reads. Bioinformatics 38:846–848. https://doi.org/10.1101/2021.07.31.454518
12. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
13. Andrews S. 2010. Fastqc: a quality control tool for high throughput sequence data. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
14. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257. https://doi.org/10.1186/s13059-019-1891-0
15. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. Spades: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021
16. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704
17. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
19. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086
20. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. 2012. PGAP: pan-genomes analysis pipeline. Bioinformatics 28:416–418. https://doi.org/10.1093/bioinformatics/btr655