



# Whole-Genome Sequencing and Annotation of Selected *Lactobacillales* Isolated from Commercial Cucumber Fermentation

Clinton A. Page,<sup>a</sup>  Ilenys M. Pérez-Díaz<sup>a</sup>

<sup>a</sup>USDA-Agricultural Research Service, Food Science and Market Quality and Handling Research Unit, Raleigh, North Carolina, USA

**ABSTRACT** We report the whole-genome sequences and annotations of 42 *Lactobacillales* isolated from commercial cucumber fermentations performed in North Carolina ( $n = 34$ ) and Minnesota ( $n = 9$ ), USA. The isolates include representatives from 12 acid-producing species.

Vegetable preservation is enabled by acidifying *Lactobacillales*. *Lactiplantibacillus plantarum*, *Lactiplantibacillus pentosus*, and *Levilactobacillus brevis* prevail in commercial cucumber fermentations, followed by *Lactococcus*, *Weissella*, *Leuconostoc*, and *Pediococcus* species (1–3). *Enterobacteriaceae* and *Enterococcaceae* may also participate in vegetable fermentations (3).

We present the genome sequences and annotations of 39 *Lactobacillales* isolated from commercial cucumber fermentations conducted in 2009 and 2010 in North Carolina (NC) and Minnesota (MN), respectively (3). These isolates were isolated from fermentations at varied time points (1, 3, 7, 14, and 30 days) and two collection tank depths (2 or 8 feet from the brine surface). Also included are *L. pentosus* strain LA0445, isolated from an anaerobic cucumber fermentation in 1983 (4), its derivative MU0445, which is deficient in malic acid decarboxylation (5), and *L. plantarum* strain T1R2b, isolated in 2020 from a low-salt cucumber fermentation with an irregular slimy brine (6). A similar slimy fermentation brine is produced by *L. plantarum* 3.2.8 (7).

All isolates were obtained from cucumber fermentation brines spiral plated on lactobacilli de Man Rogosa and Sharpe agar (MRS), supplemented with 0.0001% cycloheximide solution, and incubated at 30°C in anaerobic jars. The isolated colonies were streaked onto MRS prior to preparing frozen stocks in MRS broth supplemented with 1.5% glycerol. Pure cultures were transferred to MRS broth from frozen stocks prior to DNA extraction. The cultures were incubated at 30°C statically. DNA extraction was conducted using the Promega Wizard high-molecular-weight (HMW) extraction kit (Madison, WI). Whole-genome sequencing was performed by CosmosID (Rockville, MD). Most samples were sequenced on an Illumina NextSeq 550 platform (San Diego, CA), producing paired-end reads with a maximum length of 150 bases. Libraries for Illumina reads were prepared using the Illumina Nextera XT kit. In an earlier effort, isolates *L. plantarum* 3.2.8 and *L. pentosus* LA0445 were sequenced on a Thermo Ion S5 XL sequencer (Waltham, MA) with an average read length of 200 bp, using the Thermo Fisher Ion Xpress Plus fragment kit for the preparation of libraries. All libraries were assessed for quantity with Qubit (Thermo Fisher Scientific). The raw sequence data were trimmed for adapters and low-quality bases using BBDuk (<https://sourceforge.net/projects/bbmap>) with standard parameters (phred quality trimq = 22 and minimum length minlen = 36).

Initial assembly and annotation were performed using PATRIC (8). *De novo* assembly was performed using Unicycler version 0.4.8 (9) with a minimum contig cutoff of 300. Quality assessment of the assemblies was performed using QUAST version 5.0.2 (10),

**Citation** Page CA, Pérez-Díaz IM. 2021. Whole-genome sequencing and annotation of selected *Lactobacillales* isolated from commercial cucumber fermentation. *Microbiol Resour Announc* 10:e00625-21. <https://doi.org/10.1128/MRA.00625-21>.

**Editor** Irene L. G. Newton, Indiana University, Bloomington

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply. Address correspondence to Ilenys M. Pérez-Díaz, [ilenys.perez-diaz@ars.usda.gov](mailto:ilenys.perez-diaz@ars.usda.gov).

**Received** 24 June 2021

**Accepted** 8 October 2021

**Published** 28 October 2021

TABLE 1 Putative identification, accession numbers, and statistics for the genome sequences in this study

Organism	Genome assembly data				SRA accession no.	Assembly size (bp)	No. of contigs	Estimated coverage (x)	Total no. of reads	% GC	N <sub>50</sub> (bp)	Site of sample collection <sup>a</sup>
	GenBank accession no.	GenBank accession no.	Assembly size (bp)	No. of contigs								
<i>Lactiplantibacillus plantarum</i>												
3.2.8 (Ion Torrent sequencing)	JAGXCA0000000000	SRR14682920	3,360,349	185	525.231	8,824,794	44	67,416	NC			
3.2.8 (Illumina sequencing)	JAGYGQ0000000000	SRR14682926	3,379,107	137	183.78	4,311,400	44	90,946	NC			
7.8.4	JAGYGR0000000000	SRR14682928	3,378,788	166	75.298	1,775,922	44	114,082	NC			
T1R2b	JAGYGT0000000000	SRR14682929	3,477,739	66	207.009	5,003,588	44	370,452	NC			
<i>Lactiplantibacillus pentosus</i>												
1.2.11	JAGYGW0000000000	SRR14682931	3,634,568	116	262.253	6,694,680	46	120,249	NC			
1.2.13	JAGYX0000000000	SRR14682932	3,651,781	133	126.918	3,226,422	46	120,291	NC			
1.8.6	JAGYGY0000000000	SRR14682933	3,640,616	130	103.011	2,602,654	46	100,780	NC			
1.8.9	JAHLEP0000000000	SRR14767539	3,590,254	122	160.689	4,005,764	46	118,490	NC			
3.8.24	JAGYGZ0000000000	SRR14682934	3,662,066	136	96.685	2,478,812	46	118,379	NC			
3.2.37	JAGYHA0000000000	SRR14682939	3,655,808	111	129.527	3,282,012	46	118,381	NC			
LA0445 (Ion Torrent sequencing)	JAGXBZ0000000000	SRR14682919	3,745,923	170	325.425	6,175,578	46	90,946	NC			
<i>LA0445 (Illumina sequencing)</i>												
MU045	JAHLEN0000000000	SRR14767540	3,787,716	145	144.751	3,792,958	46	70,955	NC			
7.2.11	JAGYHB0000000000	SRR14682940	3,787,629	141	138.382	3,626,306	46	73,822	NC			
7.2.23	JAGWDS0000000000	SRR14682941	3,740,476	218	162.95	4,290,304	46	53,207	NC			
7.2.15	JAGWDT0000000000	SRR14682937	3,704,147	155	140.436	3,615,368	46	118,490	NC			
1.2.7	JAGWDU0000000000	SRR14682936	3,714,586	116	147.609	3,837,482	46	99,059	NC			
1.8.42	JAGXBD0000000000	SRR14682935	3,652,706	125	154.37	3,942,254	46	118,379	NC			
1.4.2.3	JAGXBP0000000000	SRR14682942	3,826,173	157	145.397	3,863,260	46	77,733	NC			
1.4.2.16	JAGXBP0000000000	SRR14682946	3,720,999	136	167.201	4,329,784	46	83,353	NC			
7.8.46	JAGXBC0000000000	SRR14682945	3,762,332	156	141.166	3,710,262	46	85,333	NC			
3.2.36	JAGXBS0000000000	SRR14682943	3,770,960	165	181.512	4,782,008	46	85,642	NC			
1.8.18	JAGXBT0000000000	SRR14682951	3,696,926	135	119.126	3,052,726	46	118,498	NC			
3.8.45	JAGXBU0000000000	SRR14682950	3,697,126	165	151.843	3,973,090	46	70,228	NC			
7.2.4	JAGYGS0000000000	SRR14682957	3,695,384	164	125.153	3,262,878	46	99,059	NC			
30.2.29	JAGYGU0000000000	SRR14682917	3,676,354	184	194.933	3,169,794	46	54,739	NC			
7.8.11	JAGYGV0000000000	SRR14682930	3,650,647	133	103.713	3,417,848	46	118,379	NC			
7.8.2	JAGXBV0000000000	SRR14682948	3,825,364	152	143.177	5,020,878	46	78,065	NC			
7.2.20	JAGXBW0000000000	SRR14682947	3,790,800	99	119.032	2,657,720	46	123,926	MN			
			3,795,850	140	127.716	3,830,314	46	81,129	MN			
<i>Levilactobacillus brevis</i>												
14.2.10	JAGXCC0000000000	SRR14682955	2,601,264	205	237.321	4,384,696	45	36,176	MN			
14.2.24	JAGXCD0000000000	SRR14682954	2,598,957	210	181.27	3,730,730	45	35,671	MN			
3.2.41	JAGXIY0000000000	SRR14682923	2,538,609	173	175.342	3,320,688	46	40,908	MN			
<i>Pediococcus ethanolidurans</i>												
7.8.48	JAGXJC0000000000	SRR14682925	2,119,735	183	331.606	5,082,050	37	35,566	NC			
NC579	JAGXJE0000000000	SRR14682927	2,212,777	144	303.213	4,902,456	37	39,228	NC			

(Continued on next page)

**TABLE 1** (Continued)

Genome assembly data		GenBank accession no.	SRA accession no.	Assembly size (bp)	No. of contigs	Estimated coverage (X)	Total no. of reads	% GC	N <sub>50</sub> (bp)	Site of sample collection <sup>a</sup>
Organism	GenBank accession no.									
<i>Leuconostoc citreum</i>		JAGYGO000000000	SRR14682949	1,913,902	25	375.513	5,132,798	39	398,202	MN
3.8.12										
<i>Leuconostoc fallax</i>		JAGYGP000000000	SRR14682953	1,671,397	16	262.353	5,652,968	37	1,028,303	NC
1.2.22										
<i>Leuconostoc lactis</i>		JAGXCE000000000	SRR14682952	1,712,162	25	333.01	3,977,482	43	174,610	MN
1.2.28										
<i>Leuconostoc mesenteroides</i>		JAHLEO000000000	SRR14767538	2,074,914	33	284.43	4,124,492	38	253,442	MN
1.2.47										
<i>Weissella cibaria</i>		JAGXIZ000000000	SRR14682922	2,425,944	27	240.751	4,048,644	45	225,548	NC
3.8.44										
7.8.34		JAGXJD000000000	SRR14682938	2,443,125	34	212.237	3,605,434	45	198,098	MN
<i>Weissella hellenica</i>		JAGXJA000000000	SRR14682921	1,960,091	21	453.948	6,116,734	37	322,542	NC
1.2.50										
<i>Weissella paramesenteroides</i>		JAGXJB000000000	SRR14682924	1,950,523	31	237.635	3,187,080	38	238,023	NC
3.2.24										
<i>Lactococcus lactis</i>		JAGXCF000000000	SRR14682956	2,599,249	42	304.64	5,436,048	35	322,542	NC
1.8.12										
LA0312		JAGXIX000000000	SRR14682918	2,357,047	31	323.83	5,242,810	35	300,240	NC

<sup>a</sup>Sample collection sites: NC, North Carolina; MN, Minnesota (USA).

SAMtools version 13 (11), and Pilon version 1.23 (12). The assembled genomes were annotated using RASTtk (13). The closest reference genomes were identified using Mash/MinHash with the PATRIC database (14). Upon submission to GenBank (BioProject accession number [PRJNA674638](https://doi.org/10.6026/PRJNA674638)), the assemblies were reannotated using the NCBI Prokaryotic Genome Annotation Pipeline (15). Default parameters for software were used except when noted.

A diverse range of *Lactobacillales* were detected in both culturing sources.

**Data availability.** The GenBank and Sequence Read Archive (SRA) accession numbers for each sequence are included in Table 1.

## ACKNOWLEDGMENT

Mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or North Carolina Agricultural Research Service, nor does it imply approval to the exclusion of other products that may be suitable.

## REFERENCES

1. EtcHELLS JL, Jones ID. 1946. Characteristics of lactic acid bacteria from commercial cucumber fermentations. *J Bacteriol* 52:593–599. <https://doi.org/10.1128/jb.52.5.593-599.1946>.
2. Pérez-Díaz IM. 2019. Fermented vegetables as vectors for relocation of microbial diversity from the environment to the human gut, p 91–123. *In* Azcarate-Peril M, Arnold R, Bruno-Bárceña J (ed), *How fermented foods feed a healthy gut microbiota*. Springer, Cham, Switzerland.
3. Pérez-Díaz IM, Hayes J, Medina E, Anekella K, Daughtry K, Dieck S, Levi M, Price R, Butz N, Lu Z, Azcarate-Peril MA. 2017. Reassessment of the succession of lactic acid bacteria in commercial cucumber fermentations and physiological and genomic features associated with their dominance. *Food Microbiol* 63:217–227. <https://doi.org/10.1016/j.fm.2016.11.025>.
4. Fleming HP, McFeeters RF, Daeschel MA, Humphries EG, Thompson RL. 1988. Fermentation of cucumbers in anaerobic tanks. *J Food Sci* 53:127–133. <https://doi.org/10.1111/j.1365-2621.1988.tb10192.x>.
5. Daeschel MA, McFeeters RF, Fleming HP, Klaenhammer TR, Sanozky RB. 1984. Mutation and selection of *Lactobacillus plantarum* strains that do not produce carbon dioxide from malate. *Appl Environ Microbiol* 47:419–420. <https://doi.org/10.1128/aem.47.2.419-420.1984>.
6. FideleR J. 2021. Generation of bioactive peptides and  $\gamma$ -aminobutyric acid during natural lactic acid fermentation of cucumber. Doctoral thesis. North Carolina State University, Raleigh, NC. <https://www.lib.ncsu.edu/resolver/1840.20/38713>.
7. Anekella K, Pérez Díaz IM. 2020. Characterization of robust *Lactobacillus plantarum* and *Lactobacillus pentosus* starter cultures for environmentally friendly low-salt cucumber fermentations. *J Food Sci* 85:3487–3497. <https://doi.org/10.1111/1750-3841.15416>.
8. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 48:D606–D612. <https://doi.org/10.1093/nar/gkz943>.
9. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
10. 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>.
11. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>.
12. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
13. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, III, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
14. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 17:132. <https://doi.org/10.1186/s13059-016-0997-x>.
15. Tatusova T, DiCuccio M, Badredin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.