Metagenomes and Metagenome-Assembled Genomes from Microbiomes Metabolizing Thin Stillage from an Ethanol Biorefinery

Nathaniel W. Fortney,[a,b] Kevin S. Myers,[a,b] Abel T. Ingle,[a,b,c] Kevin A. Walters,[a,b,d] Matthew J. Scarborough,[e] Timothy J. Donohue,[a,b,d] Daniel R. Noguera[a,b,c]

aGreat Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin, USA
bWisconsin Energy Institute, University of Wisconsin-Madison, Madison, Wisconsin, USA
cDepartment of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin, USA
dDepartment of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin, USA
eDepartment of Civil and Environmental Engineering, University of Vermont, Burlington, Vermont, USA

ABSTRACT Here, we report the metagenomes from five anaerobic bioreactors, operated under different conditions, that were fed carbohydrate-rich thin stillage from a corn starch ethanol plant. The putative functions of the abundant taxa identified here will inform future studies of microbial communities involved in valorizing this and other low-value agroindustrial residues.

We are investigating how to use anaerobic microbial communities for valorizing agroindustrial residues (1–5). We reported on fermentation products when thin stillage (TS) from starch bioethanol production was fed to a set of bioreactors (4). In that study, an anaerobic bioreactor (R1$_{TS}$) was inoculated with acid-phase digester sludge from the Nine Springs Wastewater Treatment Plant (Madison, WI, USA) and provided TS as the feedstock. Four additional bioreactors (R2$_{SR-TS}$, R3$_{LowSRT}$, R4$_{T-pH}$, and R5$_{T-pH-LowSRT}$), derived from R1$_{TS}$, were operated with different temperatures, pH values, and solids retention times (SRTs), resulting in diverging microbial communities and different fermentation products (4). Genomic DNA was extracted during bioreactor operation (R1$_{TS}$, 6 samples; R2$_{SR-TS}$, 9 samples; R3$_{LowSRT}$, 6 samples; R4$_{T-pH}$, 6 samples; R5$_{T-pH-LowSRT}$, 2 samples) using a phenol-chloroform extraction method (2). DNA quantity and quality were determined using a Qubit 4 fluorometer (Thermo Fisher Scientific, USA) and NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific), respectively. DNA aliquots of 500 ng (25 samples) and 3,000 ng (4 samples) were submitted to the Joint Genome Institute (JGI) for paired-end 2 × 150-bp NovaSeq S4 (Illumina, USA) and Sequel II (Pacific Biosciences [PacBio], USA) sequencing, respectively. Illumina library preparation followed established protocols (6). PacBio sequencing library preparation included shearing of genomic DNA (g-TUBE; Covaris, LLC, USA) to 6 to 10 kb and ligation using the SMRTbell Express template preparation 2.0 kit following the manufacturer's protocol (PacBio). The resulting Illumina libraries contained between 70 million and 141 million 150-bp reads, and the PacBio libraries contained between 38 thousand and 159 thousand reads 6 to 9 kb in length. Illumina reads were filtered and error corrected using BBMap (v38.86) (mincount=2, highcountfraction=0.6) (7), assembled with metaSPAdes (v3.14.1) (8), and mapped with BBMap (v38.86) (ambiguous=random) (7) following the JGI Metagenomic Workflow (6). PacBio reads were filtered using BBtools (v38.87/38.88) (7), and CCS reads were assembled using metaFlye (v2.8.1-b1676) (3), polished with subreads using GCpp (v1.0.0-SL-release-8.0.0) (https://github.com/PacificBiosciences/gcpp), mapped using minimap2 (v2.17-r941) (4), and then binned with MetaBAT (v2:2.15) (9). The resulting metagenome-assembled genomes (MAGs) were refined by removing contigs deemed to be contaminants by ProDeGe (v2.3) (10) and a custom algorithm that compares tetranucleotide

1
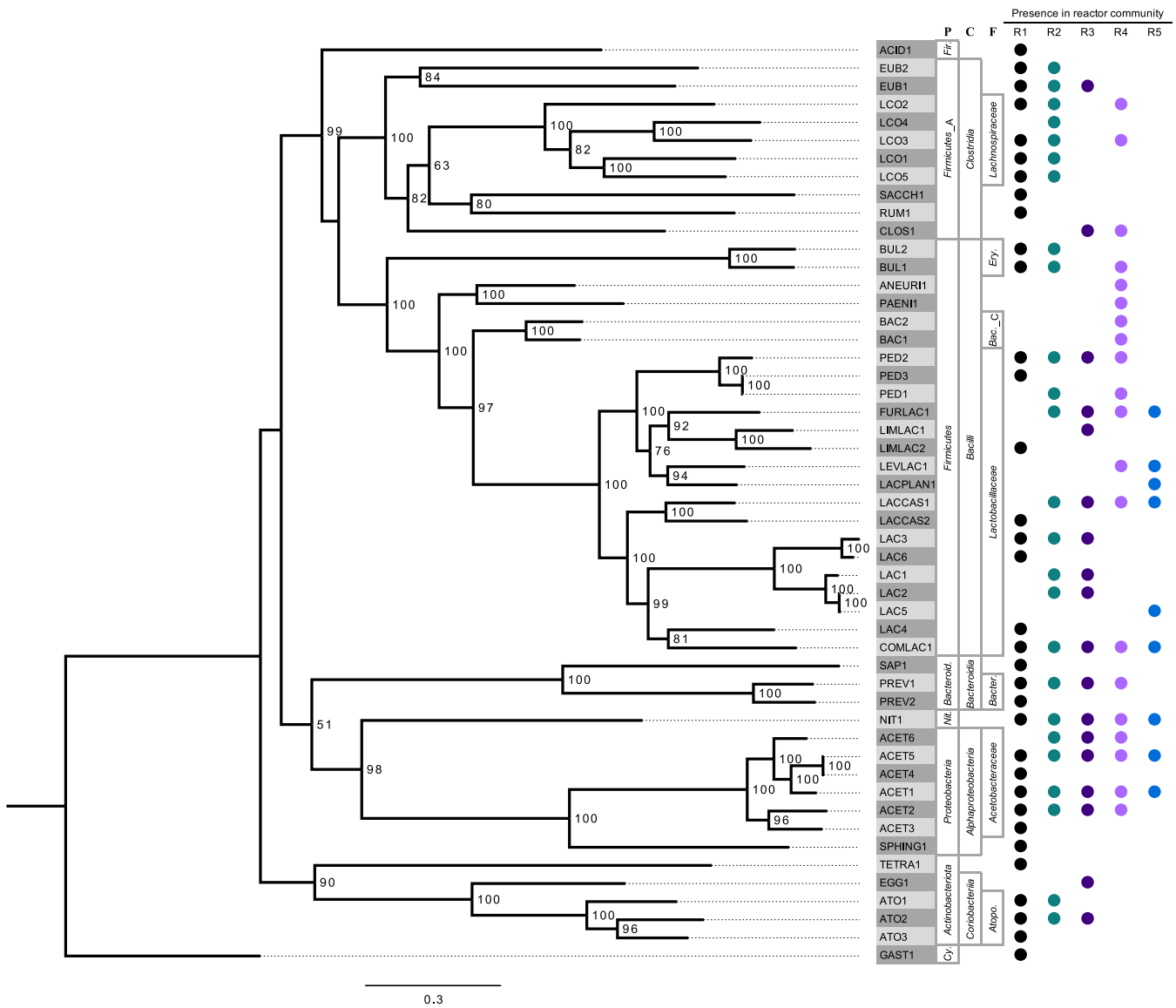
FIG 1 Phylogeny of representative MAGs and their presence in the five different bioreactors (R1 to R5), as determined by the dRep analysis (Table 1). R1, reactor 1, TS (R1$_{TS}$); R2, reactor 2, solids-removed TS (R2$_{SR-TS}$); R3, reactor 3, low SRT (R3$_{LowSRT}$); R4, reactor 4, high temperature and low pH (R4$_{T-pH}$); R5, reactor 5, low SRT, high temperature, and low pH (R5$_{T-pH-LowSRT}$). The bioreactor operating conditions are described elsewhere (4). ACET, *Acetobacter*; ACID, *Acidaminococcus*; ANEURI, *Aneurinibacillus*; ATO, *Atopobiaceae*; BAC, *Bacillus*; BUL, *Bulleidia*; CLOS, *Clostridium*; COMLAC, *Companilactobacillus*; EGG, *Eggerthellaceae*; EUB, *Eubacteriaceae*; FURLAC, *Furfurilactobacillus*; GAST, *Gastranaerophilaceae*; LAC, *Lactobacillus*; LACCAS, *Lacticaseibacillus*; LACPLAN, *Lactiplantibacillus*; LCO, *Lachnospiraceae*; LEVLAC, *Levilactobacillus*; LIMLAC, *Limosilactobacillus*; NIT, *Nitrospira*; PAENI, *Paenibacillus*; PED, *Pediococcus*; PREV, *Prevotella*; RUM, *Ruminococcus*; SACCH, *Saccharofermentans*; SAP, *Saprospiraceae*; SPHING, *Sphingobium*; TETRA, *Tetrasphaera*. Higher taxonomic levels are labeled, from left to right, phylum (P), class (C), and family (F). *Fir.*, Firmicutes_C; *Bacteroid.*, Bacteroidota; *Nit.*, Nitrospirota; *Cy.*, Cyanobacteria; *Ery.*, Erysipelotrichaceae; *Bac._C*, Bacillaceae_C; *Bacter.*, Bacteroidaceae; *Atopo.*, Atopobiaceae. The phylogenetic tree was generated in RAxML-ng with 500 bootstraps using the housekeeping gene concatenations generated by GTDB-tk, with the default selection of 120 single-copy bacterial housekeeping genes. Bootstrap values greater than 50 are shown. The scale bar indicates the number of nucleotide substitutions per sequence site.

frequency among contigs (run.GC.sh and Calculating_TF_Correlations.R [https://github.com/GLBRC/metagenome_analysis]). The MAGs obtained from individual samples were dereplicated using dRep (v3.2.2) (dereplicate command with –conW 0.5 and –N50W 5 flags for custom weighting) (11). MAG quality parameters were obtained using CheckM (v1.0.11) (12), and taxonomy was assigned using GTDB-tk (v1.5.1, database release 202) (13). MAG phylogeny was visualized using RAxML-NG (v0.9.0) (Fig. 1) (14). MAGs were annotated through the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (15).

We report a total of 266 MAGs with >75% completion, grouped in 51 clusters that represent the diversity in the individual bioreactors (Table 1). This metagenomic data

TABLE 1 Genome accession numbers and statistics

Strain name[a]	Code[b]	Reactor source[c]					Sample age (days)	SRA accession no.	No. of raw reads per sample (×1,000)	GenBank accession no.[d]	ANIm[e]	dRep[f]	GTDBtk classification
		R1	R2	R3	R4	R5							
UW_TS_ACET1_1	ACET1		X				100	SRX12729178	383	JAKVNI000000000		132.1938	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Acetobacterales; f__Acetobacteraceae; g__Acetobacter; s__Acetobacter fabarum
UW_TS_ACET1_2					X		72	SRX12687768	95,453	JALCKF000000000	0.999992	123.6701	
UW_TS_ACET1_3				X			24	SRX12665963	80,367	JALCHP000000000	0.999833	123.6696	
UW_TS_ACET1_4			X				136	SRX12664702	125,596	JALCGX000000000	0.999986	123.5262	
UW_TS_ACET1_5					X		13	SRX12687755	74,621	JALCJH000000000	0.999995	123.1801	
UW_TS_ACET1_6					X		24	SRX12686847	69,380	JALCJQ000000000	0.999991	123.1801	
UW_TS_ACET1_7				X			66	SRX12686873	104,400	JALCIB000000000	0.999984	123.1801	
UW_TS_ACET1_8				X			79	SRX12686849	79,309	JALCII000000000	0.999983	123.1801	
UW_TS_ACET1_9					X		162	SRX12687754	66,381	JALCKU000000000	0.999983	123.1801	
UW_TS_ACET1_10				X			48	SRX12670963	74,191	JALCHU000000000	0.999983	123.1801	
UW_TS_ACET1_11			X				64	SRX12658907	90,171	JALCFX000000000	0.999994	123.1801	
UW_TS_ACET1_12			X				162	SRX12686848	92,161	JALCIX000000000	0.999971	123.1801	
UW_TS_ACET1_13						X	36	SRX12729153	71,598	JAKVLI000000000	0.999989	123.1800	
UW_TS_ACET1_14						X	66	SRX12729463	81,580	JAKVLD000000000	0.999974	123.1800	
UW_TS_ACET1_15			X				202	SRX12667031	121,391	JALCHJ000000000	0.999985	123.0262	
UW_TS_ACET1_16					X		62	SRX12687723	84,579	JALCJZ000000000	0.999985	122.1801	
UW_TS_ACET1_17		X					39	SRX12657440	113,675	JALCEB000000000	0.999964	120.1678	
UW_TS_ACET1_18					X		120	SRX12687759	78,913	JALCKN000000000	0.999977	116.7101	
UW_TS_ACET1_19			X				90	SRX12660021	140,410	JALCGK000000000	0.999986	116.2101	
UW_TS_ACET1_20		X					63	SRX12729174	1,591	JAKVMC000000000	0.999976	112.8698	
UW_TS_ACET1_21			X				52	SRX12729156	634	JAKVLO000000000	0.999996	112.8689	
UW_TS_ACET1_22			X				166	SRX12729462	793	JAKVMT000000000	0.999093	107.3584	
UW_TS_ACET1_23		X					75	SRX12658615	141,269	JALCEP000000000	0.999838	100.6800	
UW_TS_ACET2_1	ACET2		X				100	SRX12729178	383	JAKVNJ000000000		132.0261	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Acetobacterales; f__Acetobacteraceae; g__Acetobacter; s__Acetobacter sp012517935
UW_TS_ACET2_2			X				166	SRX12729462	793	JAKVMU000000000	0.999967	130.6313	
UW_TS_ACET2_3		X					63	SRX12729174	1,591	JAKVMD000000000	0.999854	126.9446	
UW_TS_ACET2_4					X		62	SRX12687723	84,579	JALCJY000000000	0.999998	125.3394	
UW_TS_ACET2_5		X					75	SRX12658615	141,269	JALCEO000000000	0.999995	125.2687	
UW_TS_ACET2_6					X		24	SRX12686847	69,380	JALCJP000000000	0.999860	125.2687	
UW_TS_ACET2_7			X				90	SRX12660021	140,410	JALCGJ000000000	0.999999	125.2685	
UW_TS_ACET2_8				X			114	SRX12686875	69,693	JALCIN000000000	0.999939	125.1004	
UW_TS_ACET2_9			X				136	SRX12664702	125,596	JALCGW000000000	0.999995	124.7733	
UW_TS_ACET2_10		X					39	SRX12657440	113,675	JALCEA000000000	0.999993	124.6916	
UW_TS_ACET2_11					X		120	SRX12687759	78,913	JALCKM000000000	0.999992	124.5971	
UW_TS_ACET2_12			X				28	SRX12658904	101,250	JALCFM000000000	0.999922	124.5573	
UW_TS_ACET2_13				X			48	SRX12670963	74,191	JALCHT000000000	0.999995	114.4437	
UW_TS_ACET2_14				X			79	SRX12686849	79,309	JALCIH000000000	0.999948	109.3823	
UW_TS_ACET3_1	ACET3	X					63	SRX12729174	1,591	JAKVME000000000		131.3647	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Acetobacterales; f__Acetobacteraceae; g__Acetobacter; s__
UW_TS_ACET3_2		X					39	SRX12657440	113,675	JALCDZ000000000	0.999930	125.5032	
UW_TS_ACET3_3		X					75	SRX12658615	141,269	JALCEN000000000	0.999949	125.3846	
UW_TS_ACET4_1	ACET4	X					63	SRX12729174	1,591	JAKVMF000000000		130.1843	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Acetobacterales; f__Acetobacteraceae; g__Acetobacter; s__Acetobacter peroxydans
UW_TS_ACET5_1	ACET5		X				90	SRX12660021	140,410	JALCGI000000000		125.3562	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Acetobacterales; f__Acetobacteraceae; g__Acetobacter; s__Acetobacter peroxydans
UW_TS_ACET5_2					X		13	SRX12687755	74,621	JALCJG000000000	0.999990	125.0534	
UW_TS_ACET5_3					X		162	SRX12687754	66,381	JALCKT000000000	0.999986	123.9439	
UW_TS_ACET5_4						X	66	SRX12729463	81,580	JAKVLE000000000	0.999982	122.0555	
UW_TS_ACET5_5					X		120	SRX12687759	78,913	JALCKL000000000	0.999971	117.5738	
UW_TS_ACET5_6			X				64	SRX12658907	90,171	JALCFW000000000	0.999948	117.2484	
UW_TS_ACET5_7					X		72	SRX12687768	95,453	JALCKE000000000	1.000000	102.3029	
UW_TS_ACET5_8			X				136	SRX12664702	125,596	JALCGV000000000	0.999979	99.8305	
UW_TS_ACET6_1	ACET6				X		120	SRX12687759	78,913	JALCKK000000000		114.6918	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Acetobacterales; f__Acetobacteraceae; g__Acetobacter; s__Acetobacter indonesiensis
UW_TS_ACET6_2				X			162	SRX12686848	92,161	JALCIW000000000	0.999994	114.6918	
UW_TS_ACET6_3			X				136	SRX12664702	125,596	JALCGU000000000	0.999984	111.8759	
UW_TS_ACID1_1	ACID1	X					63	SRX12729174	15,91	JAKVMG000000000		129.8131	d__Bacteria; p__Firmicutes_C; c__Negativicutes; o__Acidaminococcales; f__Acidaminococcaceae; g__Acidaminococcus; s__Acidaminococcus provencensis
UW_TS_ANEURI1_1	ANEURI1				X		162	SRX12687754	66,381	JALCKS000000000		122.1157	d__Bacteria; p__Firmicutes; c__Bacilli; o__Aneurinibacillales; f__Aneurinibacillaceae; g__Aneurinibacillus; s__Aneurinibacillus aneurinilyticus
UW_TS_ATO1_1	ATO1		X				166	SRX12729462	793	JAKVMV000000000		131.2388	d__Bacteria; p__Actinobacteriota; c__Coriobacteriia; o__Coriobacteriales; f__Atopobiaceae; g__CADBMC01; s__CADBMC01 sp902795635
UW_TS_ATO1_2			X				202	SRX12667031	121,391	JALCHI000000000	0.999804	129.0235	
UW_TS_ATO1_3			X				136	SRX12664702	125,596	JALCGT000000000	0.999804	128.9496	

Reference genome[g]	Sequencing platform[h]	Completeness (%)	Contamination (%)	MAG size (Mbp)	No. of scaffolds	N_{50} (Mbp)	GC content (%)	No. of tRNAs	No. of 5S rRNAs	No. of 16S rRNAs	No. of 23S rRNAs
GCF_011516925.1	Sequel II	100	0.25	2.896	1	2.896	58.3	59	5	5	5
	NovaSeq S4	98.51	0.25	2.37	25	0.138	58.8	39	0	0	0
	NovaSeq S4	98.51	0.25	2.271	24	0.138	59.1	40	0	0	0
	NovaSeq S4	99	0.25	2.392	26	0.102	58.8	41	0	0	0
	NovaSeq S4	98.01	0.25	2.211	24	0.138	59.1	38	0	0	0
	NovaSeq S4	98.01	0.25	2.211	24	0.138	59.1	37	0	0	0
	NovaSeq S4	98.01	0.25	2.212	24	0.138	59.1	37	0	0	0
	NovaSeq S4	98.01	0.25	2.212	23	0.138	59.1	37	0	0	0
	NovaSeq S4	98.01	0.25	2.225	24	0.138	59.1	37	0	0	0
	NovaSeq S4	98.01	0.25	2.212	22	0.138	59.1	37	0	0	0
	NovaSeq S4	98.01	0.25	2.227	23	0.138	59.1	37	0	0	0
	NovaSeq S4	98.01	0.25	2.213	22	0.138	59.1	38	0	0	0
	NovaSeq S4	98.01	0.25	2.225	24	0.138	59.1	38	0	0	0
	NovaSeq S4	98.01	0.25	2.226	24	0.138	59.1	38	0	0	0
	NovaSeq S4	98.51	0.25	2.304	24	0.102	59	39	0	0	0
	NovaSeq S4	97.01	0.25	2.182	23	0.138	59.2	36	0	0	0
	NovaSeq S4	95.52	0.75	2.27	44	0.102	59.1	41	0	0	0
	NovaSeq S4	91.29	0.25	2.182	23	0.138	59.1	35	0	0	0
	NovaSeq S4	91.04	0.25	2.31	25	0.138	58.7	38	0	0	0
	Sequel II	82.09	0.25	2.268	2	1.51	58.2	38	2	2	2
	Sequel II	82.09	0.25	2.255	2	1.509	58.2	38	2	2	2
	Sequel II	82.59	0.25	2.086	21	0.115	58.7	38	0	1	0
	NovaSeq S4	77.79	0.25	2.174	67	0.043	58.6	35	0	0	0
GCA_012517935.1	Sequel II	100	1.24	3.367	1	3.367	54.4	60	5	5	5
	Sequel II	100	1.24	3.423	3	1.771	54.4	60	5	5	5
	Sequel II	95.02	1.24	3.213	1	3.213	54.3	51	4	4	4
	NovaSeq S4	100	0.25	3.014	30	0.123	53.9	45	0	0	0
	NovaSeq S4	100	0.25	2.931	29	0.119	53.9	45	0	0	0
	NovaSeq S4	100	0.25	3.058	30	0.119	53.9	45	0	0	0
	NovaSeq S4	100	0.25	2.917	31	0.119	53.9	46	0	0	0
	NovaSeq S4	100	0.25	3.051	37	0.11	53.9	45	0	0	0
	NovaSeq S4	100	1.24	2.906	31	0.119	53.8	45	0	0	0
	NovaSeq S4	100	1.24	2.952	35	0.115	53.8	45	0	0	0
	NovaSeq S4	100	1.24	3.078	33	0.11	53.9	46	0	0	0
	NovaSeq S4	100	1.24	3.06	37	0.108	53.8	46	0	0	0
	NovaSeq S4	89.05	0	2.664	28	0.119	54	38	0	0	0
	NovaSeq S4	84.33	0.25	2.55	35	0.108	53.8	41	0	0	0
NA	Sequel II	99.75	0	3.708	3	2.636	57.6	53	3	3	3
	NovaSeq S4	99	0	3.503	27	0.249	57.6	46	0	0	0
	NovaSeq S4	99	0	3.445	27	0.236	57.7	46	0	0	0
GCF_006539345.1	Sequel II	99.5	0.5	2.642	2	2.162	60.3	56	4	4	4
GCF_006539345.1	NovaSeq S4	100	0	2.559	29	0.117	60.7	45	0	0	0
	NovaSeq S4	100	0	2.521	34	0.102	60.9	44	0	0	0
	NovaSeq S4	99.5	0	2.52	43	0.077	60.8	45	0	0	0
	NovaSeq S4	97.71	0	2.499	46	0.074	60.9	42	1	0	1
	NovaSeq S4	94.35	0.5	2.272	72	0.039	61.1	45	0	0	0
	NovaSeq S4	94.5	0.17	2.213	83	0.037	61.2	41	0	0	0
	NovaSeq S4	76.12	0	1.815	15	0.172	60.8	31	0	0	0
	NovaSeq S4	77.59	0	2.185	89	0.028	61.1	36	1	0	1
GCF_000963945.1	NovaSeq S4	89.05	0	2.337	23	0.134	55.1	40	0	0	0
	NovaSeq S4	89.05	0	2.346	25	0.134	55.1	40	0	0	0
	NovaSeq S4	87.48	0.25	2.267	50	0.08	55.2	39	0	0	0
GCF_900291475.1	Sequel II	99.38	1.2	2.915	1	2.915	52.8	56	6	6	6
GCF_000466385.1	NovaSeq S4	99.22	1.68	5.042	94	0.088	43.6	85	0	3	6
GCA_902795635.1	Sequel II	99.85	0	1.888	1	1.888	67.4	45	2	2	2
	NovaSeq S4	100	0	1.864	4	0.635	67.5	45	0	0	0

TABLE 1 (Continued)

Strain name[a]	Code[b]	R1	R2	R3	R4	R5	Sample age (days)	SRA accession no.	No. of raw reads per sample (×1,000)	GenBank accession no.[d]	ANIm[e]	dRep[f]	GTDBtk classification
UW_TS_ATO1_4			X				64	SRX12658907	90,171	JALCFV000000000	0.999770	126.6827	
UW_TS_ATO1_5			X				52	SRX12729156	634	JAKVLP000000000	0.996217	115.9939	
UW_TS_ATO1_6		X					75	SRX12658615	141,269	JALCEM000000000	0.999687	97.0594	
UW_TS_ATO2_1	ATO2	X					63	SRX12729174	1,591	JAKVMH000000000		129.7026	d__Bacteria; p__Actinobacteriota; c__Coriobacteriia; o__Coriobacteriales; f__Atopobiaceae; g__UBA7748; s__UBA7748 sp900314535
UW_TS_ATO2_2				X			162	SRX12686848	92,161	JALCIV000000000	0.994332	117.0041	
UW_TS_ATO2_3				X			66	SRX12686873	104,400	JALCIA000000000	0.994352	116.8091	
UW_TS_ATO2_4			X				3	SRX12658692	113,360	JALCFA000000000	0.996929	114.6794	
UW_TS_ATO3_1	ATO3	X					63	SRX12729174	1,591	JAKVMI000000000		128.4379	d__Bacteria; p__Actinobacteriota; c__Coriobacteriia; o__Coriobacteriales; f__Atopobiaceae; g__Olsenella; s__Olsenella sp003862195
UW_TS_ATO3_2		X					39	SRX12657440	113,675	JALCDY000000000	0.998844	108.0398	
UW_TS_BAC1_1	BAC1				X		24	SRX12686847	69,380	JALCJO000000000		122.9222	d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales_B; f__Bacillaceae_C; g__Weizmannia; s__Weizmannia coagulans_A
UW_TS_BAC2_1	BAC2				X		62	SRX12687723	84,579	JALCJX000000000		122.1049	d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales_B; f__Bacillaceae_C; g__Heyndrickxia; s__Heyndrickxia oleronia
UW_TS_BAC2_2				X			162	SRX12687754	66,381	JALCKR000000000	0.999970	121.8388	
UW_TS_BAC2_3				X			24	SRX12686847	69,380	JALCJN000000000	0.999964	118.1412	
UW_TS_BAC2_4				X			120	SRX12687759	78,913	JALCKJ000000000	0.999897	111.7443	
UW_TS_BUL1_1	BUL1		X				100	SRX12729178	383	JAKVNK000000000		131.9919	d__Bacteria; p__Firmicutes; c__Bacilli; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Bulleidia; s__Bulleidia massiliensis_A
UW_TS_BUL1_2			X				52	SRX12729156	634	JAKVLQ000000000	0.999945	131.9795	
UW_TS_BUL1_3			X				166	SRX12729462	793	JAKVMW000000000	0.999871	131.4491	
UW_TS_BUL1_4			X				28	SRX12658904	101,250	JALCFL000000000	0.999979	112.8492	
UW_TS_BUL1_5			X				202	SRX12667031	121,391	JALCHH000000000	0.999975	112.2754	
UW_TS_BUL1_6		X					75	SRX12658615	141,269	JALCEL000000000	0.999933	107.3669	
UW_TS_BUL1_7			X				90	SRX12660021	140,410	JALCGH000000000	0.999977	107.3666	
UW_TS_BUL1_8					X		13	SRX12687755	74,621	JALCJF000000000	0.999967	107.3666	
UW_TS_BUL1_9			X				64	SRX12658907	90,171	JALCFU000000000	0.999945	107.2990	
UW_TS_BUL1_10			X				136	SRX12664702	125,596	JALCGS000000000	0.999973	107.1294	
UW_TS_BUL2_1	BUL2		X				100	SRX12729178	383	JAKVNL000000000		131.1511	d__Bacteria; p__Firmicutes; c__Bacilli; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Bulleidia; s__Bulleidia sp900319505
UW_TS_BUL2_2			X				166	SRX12729462	793	JAKVMX000000000	0.999604	130.3116	
UW_TS_BUL2_3		X					63	SRX12729174	1,591	JAKVMJ000000000	0.999992	127.6093	
UW_TS_BUL2_4		X					75	SRX12658615	141,269	JALCEK000000000	0.999989	120.6841	
UW_TS_BUL2_5			X				136	SRX12664702	125,596	JALCGR000000000	0.999890	120.6309	
UW_TS_BUL2_6			X				90	SRX12660021	140,410	JALCGG000000000	0.999982	119.8261	
UW_TS_BUL2_7			X				3	SRX12658692	113,360	JALCEZ000000000	0.999960	118.0411	
UW_TS_BUL2_8			X				202	SRX12667031	121,391	JALCHG000000000	0.999958	111.2862	
UW_TS_CLOS1_1	CLOS1				X		13	SRX12687755	74,621	JALCJE000000000		123.9054	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__Clostridium; s__Clostridium beijerinckii
UW_TS_CLOS1_2					X		62	SRX12687723	84,579	JALCJW000000000	0.999994	122.7253	
UW_TS_CLOS1_3					X		24	SRX12686847	69,380	JALCJM000000000	0.999969	120.3569	
UW_TS_CLOS1_4				X			24	SRX12665963	80,367	JALCHO000000000	0.999886	103.6952	
UW_TS_COMLAC1_1	COMLAC1	X					63	SRX12729174	1,591	JAKVMK000000000		131.0951	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Companilactobacillus;s__
UW_TS_COMLAC1_2			X				100	SRX12729178	383	JAKVNM000000000	0.999999	131.0951	
UW_TS_COMLAC1_3			X				166	SRX12729462	793	JAKVMY000000000	0.999999	131.0951	
UW_TS_COMLAC1_4			X				52	SRX12729156	634	JAKVLR000000000	0.999999	131.0951	
UW_TS_COMLAC1_5		X					75	SRX12658615	141,269	JALCEJ000000000	0.999998	127.7837	
UW_TS_COMLAC1_6			X				28	SRX12658904	101,250	JALCFK000000000	0.999998	127.7837	
UW_TS_COMLAC1_7			X				3	SRX12658692	113,360	JALCEY000000000	0.999998	127.7837	
UW_TS_COMLAC1_8			X				136	SRX12664702	125,596	JALCGQ000000000	0.999998	127.7837	
UW_TS_COMLAC1_9			X				202	SRX12667031	121,391	JALCHF000000000	0.999998	127.7837	
UW_TS_COMLAC1_10				X			66	SRX12686873	104,400	JALCHZ000000000	0.999998	127.7837	
UW_TS_COMLAC1_11				X			79	SRX12686849	79,309	JALCIG000000000	0.999998	127.7837	
UW_TS_COMLAC1_12				X			162	SRX12686848	92,161	JALCIU000000000	0.999998	127.7837	
UW_TS_COMLAC1_13					X		62	SRX12687723	84,579	JALCJV000000000	0.999998	127.7837	
UW_TS_COMLAC1_14			X				64	SRX12658907	90,171	JALCFT000000000	0.999932	126.1667	
UW_TS_COMLAC1_15					X		120	SRX12687759	78,913	JALCKI000000000	0.999999	121.2617	
UW_TS_COMLAC1_16				X			48	SRX12670963	74,191	JALCHS000000000	0.999994	121.2617	
UW_TS_COMLAC1_17					X		24	SRX12686847	69,380	JALCJL000000000	0.999992	121.2617	
UW_TS_COMLAC1_18					X		162	SRX12687754	66,381	JALCKQ000000000	0.999999	121.2617	
UW_TS_COMLAC1_19						X	66	SRX12729463	81,580	JAKVLF000000000	0.999997	121.2450	
UW_TS_COMLAC1_20			X				90	SRX12660021	140,410	JALCGF000000000	0.999999	121.2278	
UW_TS_COMLAC1_21					X		72	SRX12687768	95,453	JALCKD000000000	0.999988	121.2278	
UW_TS_COMLAC1_22						X	36	SRX12729153	71,598	JAKVLJ000000000	0.999999	121.2278	
UW_TS_EGG1_1	EGG1				X		24	SRX12665963	80,367	JALCHN000000000		120.7069	d__Bacteria; p__Actinobacteriota; c__Coriobacteriia; o__Coriobacteriales; f__Eggerthellaceae; g__RUG013; s__

Reference genome[g]	Sequencing platform[h]	Completeness (%)	Contamination (%)	MAG size (Mbp)	No. of scaffolds	N_{50} (Mbp)	GC content (%)	No. of tRNAs	No. of 5S rRNAs	No. of 16S rRNAs	No. of 23S rRNAs
	NovaSeq S4	100	0	1.85	14	0.216	67.5	45	1	1	1
	Sequel II	90.9	0	1.772	16	0.151	67.2	46	1	2	1
	NovaSeq S4	75.16	0.52	1.492	78	0.021	67.9	37	0	0	0
GCA_900314535.1	Sequel II	99.19	0	2.524	6	1.263	60	54	2	2	2
	NovaSeq S4	92.74	0.81	2.149	36	0.086	60.2	46	0	0	0
	NovaSeq S4	92.74	0.81	2.17	39	0.078	60.3	46	0	0	0
	NovaSeq S4	91.94	0	1.644	68	0.035	60.5	40	0	0	0
GCA_003862195.1	Sequel II	98.39	0	2.333	3	1.018	65.4	57	1	1	1
	NovaSeq S4	85.36	0	2.015	73	0.034	65.7	50	0	0	0
GCF_000169195.2	NovaSeq S4	98.36	0	3.03	37	0.129	47	55	7	5	5
GCF_002009555.1	NovaSeq S4	97.76	0.32	4.977	78	0.102	34.8	100	5	5	2
	NovaSeq S4	97.75	0.33	5.079	96	0.091	34.8	110	7	5	2
	NovaSeq S4	94.92	0.32	4.958	120	0.061	34.9	92	5	8	0
	NovaSeq S4	87.3	0.28	4.125	66	0.106	35.1	92	5	5	1
GCF_900343155.1	Sequel II	100	0.32	2.682	1	2.682	47.2	45	4	4	4
	Sequel II	100	0.32	2.666	1	2.666	47.2	45	4	4	4
	Sequel II	99.52	0.32	2.666	1	2.666	47.2	45	4	4	4
	NovaSeq S4	91.67	0.32	2.02	73	0.036	48	41	0	0	0
	NovaSeq S4	90.67	0.32	2.115	61	0.043	47.9	42	0	0	0
	NovaSeq S4	85.71	0.32	1.957	58	0.044	48.2	40	0	0	0
	NovaSeq S4	85.71	0.32	1.915	56	0.044	48.4	40	0	0	0
	NovaSeq S4	85.71	0.32	1.915	54	0.044	48.4	40	0	0	0
	NovaSeq S4	85.71	0.32	1.887	57	0.043	48.4	41	0	1	0
	NovaSeq S4	85.71	0.32	1.952	59	0.04	48.3	40	0	0	0
GCA_900319505.1	Sequel II	100	1.27	2.721	2	2.266	46.8	48	3	3	3
	Sequel II	100	1.27	2.715	3	1.539	46.8	48	3	3	3
	Sequel II	100	1.27	2.715	7	0.687	46.8	48	4	4	4
	NovaSeq S4	99.05	1.27	2.068	60	0.044	47.5	47	0	0	0
	NovaSeq S4	99.05	1.27	2.073	63	0.043	47.4	47	1	0	0
	NovaSeq S4	98.19	1.27	2.073	61	0.044	47.4	47	0	0	0
	NovaSeq S4	96.97	1.27	1.947	75	0.034	47.5	45	1	0	0
	NovaSeq S4	90.77	1.27	1.916	85	0.026	47.5	43	1	0	1
GCF_002006445.1	NovaSeq S4	99.19	1.61	5.479	77	0.127	29.6	84	10	2	8
	NovaSeq S4	98.39	1.61	5.468	94	0.106	29.6	76	9	2	7
	NovaSeq S4	97.58	1.61	5.333	144	0.052	29.7	83	9	2	6
	NovaSeq S4	77.42	0	2.881	28	0.179	29.8	70	6	2	5
NA	Sequel II	99.48	1.05	2.667	1	2.667	38.3	59	4	4	4
	Sequel II	99.48	1.05	2.667	1	2.667	38.3	59	4	4	4
	Sequel II	99.48	1.05	2.667	1	2.667	38.3	59	4	4	4
	Sequel II	99.48	1.05	2.667	1	2.667	38.3	59	4	4	5
	NovaSeq S4	99.48	1.05	2.581	7	0.58	38.2	25	1	0	0
	NovaSeq S4	99.48	1.05	2.582	7	0.58	38.2	25	1	0	0
	NovaSeq S4	99.48	1.05	2.542	7	0.58	38.3	24	1	0	0
	NovaSeq S4	99.48	1.05	2.568	7	0.58	38.3	25	1	0	0
	NovaSeq S4	99.48	1.05	2.581	7	0.58	38.2	25	1	0	0
	NovaSeq S4	99.48	1.05	2.582	7	0.58	38.2	25	1	0	0
	NovaSeq S4	99.48	1.05	2.589	9	0.58	38.3	26	1	0	0
	NovaSeq S4	99.48	1.05	2.582	7	0.58	38.2	25	1	0	0
	NovaSeq S4	99.48	1.05	2.582	7	0.58	38.2	25	1	0	0
	NovaSeq S4	99.48	1.05	2.521	18	0.276	38.3	24	1	0	0
	NovaSeq S4	91.62	0	1.505	4	0.844	38.1	20	1	0	0
	NovaSeq S4	91.62	0	1.531	4	0.844	38.1	21	1	0	0
	NovaSeq S4	91.62	0	1.531	4	0.844	38.1	21	1	0	0
	NovaSeq S4	91.62	0	1.505	4	0.844	38.1	20	1	0	0
	NovaSeq S4	91.62	0	1.52	5	0.838	38.1	21	1	0	0
	NovaSeq S4	91.62	0	1.518	4	0.831	38.1	21	1	0	0
	NovaSeq S4	91.62	0	1.555	5	0.831	38.1	20	1	0	0
	NovaSeq S4	91.62	0	1.518	4	0.831	38.1	21	1	0	0
NA	NovaSeq S4	99.19	1.49	1.959	56	0.045	64.8	45	2	0	0

(Continued on next page)

TABLE 1 (Continued)

Strain name[a]	Code[b]	R1	R2	R3	R4	R5	Sample age (days)	SRA accession no.	No. of raw reads per sample (×1,000)	GenBank accession no.[d]	ANIm[e]	dRep[f]	GTDBtk classification
UW_TS_EUB1_1	EUB1		X				100	SRX12729178	383	JAKVNN000000000		130.0817	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Peptostreptococcales; f__Anaerovoracaceae; g__Eubacterium_T; s__Eubacterium_T sp900766045
UW_TS_EUB1_2		X					63	SRX12729174	1,591	JAKVML000000000	0.999898	128.6116	
UW_TS_EUB1_3			X				166	SRX12729462	793	JAKVMZ000000000	0.999841	123.6557	
UW_TS_EUB1_4			X				52	SRX12729156	634	JAKVLS000000000	0.999836	122.2015	
UW_TS_EUB1_5		X					75	SRX12658615	141,269	JALCEI000000000	0.999992	122.0361	
UW_TS_EUB1_6			X				136	SRX12664702	125,596	JALCGP000000000	0.999974	120.0690	
UW_TS_EUB1_7			X				202	SRX12667031	121,391	JALCHE000000000	0.999959	119.1044	
UW_TS_EUB1_8				X			24	SRX12665963	80,367	JALCHM000000000	0.999988	117.3295	
UW_TS_EUB1_9			X				90	SRX12660021	140,410	JALCGE000000000	0.999986	116.1182	
UW_TS_EUB1_10				X			79	SRX12686849	79,309	JALCIF000000000	0.999954	110.3145	
UW_TS_EUB1_11				X			114	SRX12686875	69,693	JALCIM000000000	0.999962	107.0662	
UW_TS_EUB2_1	EUB2	X					63	SRX12729174	1,591	JAKVMM000000000		128.2474	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Eubacteriales; f__Eubacteriaceae; g__Pseudoramibacter; s__
UW_TS_EUB2_2			X				52	SRX12729156	634	JAKVLT000000000	0.999995	128.2474	
UW_TS_FURLAC1_1	FURLAC1		X				100	SRX12729178	383	JAKVNO000000000		130.8251	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Furfurilactobacillus; s__Furfurilactobacillus rossiae_A
UW_TS_FURLAC1_2			X				166	SRX12729462	793	JAKVNA000000000	0.999994	130.8251	
UW_TS_FURLAC1_3			X				52	SRX12729156	634	JAKVLU000000000	0.999999	130.7901	
UW_TS_FURLAC1_4			X				64	SRX12658907	90,171	JALCFS000000000	0.999981	124.4729	
UW_TS_FURLAC1_5				X			79	SRX12686849	79,309	JALCIE000000000	1.000000	124.4729	
UW_TS_FURLAC1_6				X			162	SRX12686848	92,161	JALCIT000000000	0.999981	124.4729	
UW_TS_FURLAC1_7					X		24	SRX12686847	69,380	JALCJK000000000	1.000000	124.4729	
UW_TS_FURLAC1_8					X		62	SRX12687723	84,579	JALCJU000000000	1.000000	124.4729	
UW_TS_FURLAC1_9					X		162	SRX12687754	66,381	JALCKP000000000	0.999999	124.4729	
UW_TS_FURLAC1_10			X				3	SRX12658692	113,360	JALCEX000000000	0.999978	120.7133	
UW_TS_FURLAC1_11						X	36	SRX12729153	71,598	JAKVLK000000000	0.999998	117.7798	
UW_TS_GAST1_1	GAST1	X					3	SRX12657373	95,824	JALCDS000000000		107.5830	d__Bacteria; p__Cyanobacteria; c__Vampirovibrionia; o__Gastranaerophilales; f__Gastranaerophilaceae; g__RUG14219; s__
UW_TS_LAC1_1	LAC1		X				100	SRX12729178	383	JAKVNP000000000		130.6879	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__Lactobacillus crispatus
UW_TS_LAC1_2			X				28	SRX12658904	101,250	JALCFJ000000000	0.999993	119.3775	
UW_TS_LAC1_3			X				3	SRX12658692	113,360	JALCEW000000000	0.999985	117.5422	
UW_TS_LAC1_4				X			66	SRX12686873	104,400	JALCHY000000000	0.999500	105.4347	
UW_TS_LAC1_5				X			162	SRX12686848	92,161	JALCIS000000000	0.999937	101.2904	
UW_TS_LAC1_6				X			114	SRX12686875	69,693	JALCIL000000000	0.999472	98.4504	
UW_TS_LAC2_1	LAC2		X				100	SRX12729178	383	JAKVNQ000000000		128.7033	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__Lactobacillus amylovorus
UW_TS_LAC2_2				X			162	SRX12686848	92,161	JALCIR000000000	0.994845	111.4521	
UW_TS_LAC3_1	LAC3		X				100	SRX12729178	383	JAKVNR000000000		122.8879	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__
UW_TS_LAC3_2		X					63	SRX12729174	1,591	JAKVMN000000000	0.999990	122.7179	
UW_TS_LAC3_3			X				3	SRX12658692	113,360	JALCEV000000000	0.999985	121.4528	
UW_TS_LAC3_4				X			24	SRX12665963	80,367	JALCHL000000000	0.996826	117.5235	
UW_TS_LAC3_5		X					75	SRX12658615	141,269	JALCEH000000000	0.999988	116.6682	
UW_TS_LAC3_6			X				90	SRX12660021	140,410	JALCGD000000000	0.999974	115.7650	
UW_TS_LAC3_7				X			48	SRX12670963	74,191	JALCHR000000000	0.995871	114.3096	
UW_TS_LAC3_8				X			162	SRX12686848	92,161	JALCIQ000000000	0.999982	112.4339	
UW_TS_LAC3_9			X				28	SRX12658904	101,250	JALCFI000000000	0.999970	112.2940	
UW_TS_LAC4_1	LAC4	X					63	SRX12729174	1,591	JAKVMO000000000		113.4839	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__RHOV01; s__RHOV01 sp003946675
UW_TS_LAC5_1	LAC5					X	66	SRX12729463	81,580	JAKVLG000000000		110.5431	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__Lactobacillus amylovorus
UW_TS_LAC6_1	LAC6	X					39	SRX12657440	113,675	JALCDX000000000		99.1641	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__
UW_TS_LACCAS1_1	LACCAS1		X				166	SRX12729462	793	JAKVNB000000000		131.8742	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lacticaseibacillus; s__Lacticaseibacillus paracasei

Reference genome[g]	Sequencing platform[h]	Completeness (%)	Contamination (%)	MAG size (Mbp)	No. of scaffolds	N_{50} (Mbp)	GC content (%)	No. of tRNAs	No. of 5S rRNAs	No. of 16S rRNAs	No. of 23S rRNAs
GCF_900766045.1	Sequel II	99.29	0.71	2.75	1	2.75	49.3	57	3	3	3
	Sequel II	97.87	0.71	2.687	1	2.687	49.3	50	3	3	3
	Sequel II	93.62	0.71	2.378	2	1.951	49.6	57	3	3	3
	Sequel II	96.45	0	2.209	13	0.318	49.9	47	2	2	2
	NovaSeq S4	98.58	0.71	2.158	39	0.094	49.9	28	1	0	0
	NovaSeq S4	97.16	0.8	2.072	50	0.062	50	22	1	0	0
	NovaSeq S4	96.45	0.71	2.116	48	0.065	49.9	36	1	0	0
	NovaSeq S4	94.33	0.95	2.07	48	0.065	49.9	24	2	0	0
	NovaSeq S4	92.91	0.71	1.952	37	0.084	50.1	28	1	0	0
	NovaSeq S4	89.01	1.06	1.899	74	0.038	50.1	22	1	0	0
	NovaSeq S4	86.52	0.71	1.873	84	0.025	50	33	0	0	0
NA	Sequel II	97.89	0.94	1.966	1	1.966	50	53	3	3	3
	Sequel II	97.89	0.94	1.966	1	1.966	50	53	3	3	3
GCF_013394045.1	Sequel II	99.46	1.36	2.553	1	2.553	43.6	60	4	4	4
	Sequel II	99.46	1.36	2.553	1	2.553	43.6	60	4	4	4
	Sequel II	99.46	1.43	2.553	1	2.553	43.6	60	4	4	4
	NovaSeq S4	98.91	1.36	2.506	23	0.176	43.5	54	2	0	0
	NovaSeq S4	98.91	1.36	2.503	24	0.176	43.5	42	1	0	0
	NovaSeq S4	98.91	1.36	2.506	24	0.176	43.5	54	2	0	0
	NovaSeq S4	98.91	1.36	2.499	24	0.176	43.6	53	2	0	0
	NovaSeq S4	98.91	1.36	2.506	24	0.176	43.5	54	2	0	0
	NovaSeq S4	98.91	1.36	2.507	25	0.176	43.5	54	2	0	0
	NovaSeq S4	95.92	1.36	2.427	29	0.124	43.6	42	2	0	0
	NovaSeq S4	91.85	1.36	1.643	13	0.209	43.4	48	2	0	0
NA	NovaSeq S4	84.62	0.85	1.714	41	0.075	32.5	34	0	1	1
GCF_002088015.1	Sequel II	98.86	0	2.31	1	2.31	36.7	66	5	5	5
	NovaSeq S4	95.09	0.65	2.131	41	0.072	36.4	52	0	0	0
	NovaSeq S4	92.49	1.41	2.114	46	0.074	36.4	49	0	0	0
	NovaSeq S4	80.51	0.76	1.961	40	0.096	36.3	46	0	0	0
	NovaSeq S4	76.26	0	1.625	25	0.101	36.8	51	0	0	0
	NovaSeq S4	76.44	2.38	1.948	98	0.025	36.2	35	0	1	0
GCF_002706375.1	Sequel II	98.12	0.06	2.106	4	1.258	37.9	64	4	4	4
	NovaSeq S4	87.87	0.65	1.639	54	0.045	37.9	34	0	0	0
NA	Sequel II	92.18	0.52	1.555	1	1.555	42.5	57	4	4	4
	Sequel II	92.03	0.52	1.555	1	1.555	42.5	57	4	4	4
	NovaSeq S4	92.18	0.52	1.402	9	0.803	42.5	24	0	0	0
	NovaSeq S4	91.41	0.54	1.299	13	0.179	42.7	26	0	0	0
	NovaSeq S4	91.6	0.52	1.387	15	0.116	42.5	20	0	0	0
	NovaSeq S4	89.08	0	1.355	10	0.218	42.4	20	0	0	0
	NovaSeq S4	89.37	2.07	1.248	17	0.134	42.7	24	0	0	0
	NovaSeq S4	85.98	0	1.281	11	0.195	42.4	18	0	0	0
	NovaSeq S4	89.32	0.65	1.215	34	0.042	42.7	21	0	0	0
GCF_003946675.1	Sequel II	89.97	3.32	2.263	26	0.124	40.7	42	2	3	3
GCF_002706375.1	NovaSeq S4	88.54	0	1.568	47	0.046	38.2	34	0	0	0
NA	NovaSeq S4	75.86	0	0.808	15	0.072	44.2	21	0	0	0

TABLE 1 (Continued)

Strain name[a]	Code[b]	R1	R2	R3	R4	R5	Sample age (days)	SRA accession no.	No. of raw reads per sample (×1,000)	GenBank accession no.[d]	ANIm[e]	dRep[f]	GTDBtk classification
UW_TS_LACCAS1_2		X					100	SRX12729178	383	JAKVNS000000000	0.999999	131.8611	
UW_TS_LACCAS1_3		X					52	SRX12729156	634	JAKVLV000000000	0.999996	131.5670	
UW_TS_LACCAS1_4		X					64	SRX12658907	90,171	JALCFR000000000	0.999993	123.6093	
UW_TS_LACCAS1_5		X					28	SRX12658904	101,250	JALCFH000000000	0.999993	123.0593	
UW_TS_LACCAS1_6		X					202	SRX12667031	121,391	JALCHD000000000	0.999992	123.0593	
UW_TS_LACCAS1_7			X				66	SRX12686873	104,400	JALCHX000000000	0.999992	122.5193	
UW_TS_LACCAS1_8			X				162	SRX12686848	92,161	JALCIP000000000	0.999997	121.4607	
UW_TS_LACCAS1_9				X			13	SRX12687755	74,621	JALCJD000000000	0.999989	120.9207	
UW_TS_LACCAS1_10			X				79	SRX12686849	79,309	JALCID000000000	0.999993	120.8893	
UW_TS_LACCAS1_11		X					3	SRX12658692	113,360	JALCEU000000000	0.999992	120.8397	
UW_TS_LACCAS1_12					X		162	SRX12687754	66,381	JALCKO000000000	0.999973	120.3119	
UW_TS_LACCAS1_13						X	36	SRX12729153	71,598	JAKVLL000000000	0.999986	120.2801	
UW_TS_LACCAS1_14					X		72	SRX12687768	95,453	JALCKC000000000	0.999981	120.1177	
UW_TS_LACCAS1_15					X		120	SRX12687759	78,913	JALCKH000000000	0.999986	119.1712	
UW_TS_LACCAS1_16			X	X			90	SRX12660021	140,410	JALCGC000000000	0.999997	118.7507	
UW_TS_LACCAS1_17				X			114	SRX12686875	69,693	JALCIK000000000	0.999985	116.7236	
UW_TS_LACCAS1_18						X	66	SRX12729463	81,580	JAKVLH000000000	0.999987	116.5279	
UW_TS_LACCAS1_19				X			24	SRX12686847	69,380	JALCJJ000000000	0.999997	107.3665	
UW_TS_LACCAS2_1	LACCAS2	X					3	SRX12657373	95,824	JALCDR000000000		109.6761	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lacticaseibacillus; s__Lacticaseibacillus songhuajiangensis
UW_TS_LACPLAN1_1	LACPLAN1					X	36	SRX12729153	71,598	JAKVLM000000000		115.2994	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactiplantibacillus; s__Lactiplantibacillus pentosus
UW_TS_LCO1_1	LCO1	X					63	SRX12729174	1,591	JAKVMP000000000		130.2972	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__UBA1066; s__UBA1066 sp900315305
UW_TS_LCO1_2			X				166	SRX12729462	793	JAKVNC000000000	0.999997	130.1962	
UW_TS_LCO1_3			X				52	SRX12729156	634	JAKVLW000000000	0.999991	127.6970	
UW_TS_LCO1_4			X				28	SRX12658904	101,250	JALCFG000000000	0.999984	120.7256	
UW_TS_LCO1_5			X				3	SRX12658692	113,360	JALCET000000000	0.999932	119.8468	
UW_TS_LCO1_6			X				202	SRX12667031	121,391	JALCHC000000000	0.999990	119.7118	
UW_TS_LCO1_7		X					75	SRX12658615	141,269	JALCEG000000000	0.999997	119.4556	
UW_TS_LCO1_8			X				64	SRX12658907	90,171	JALCFQ000000000	0.999968	116.6113	
UW_TS_LCO2_1	LCO2		X				166	SRX12729462	793	JAKVND000000000		129.5743	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__UBA4285; s__UBA4285 sp900314255
UW_TS_LCO2_2			X				3	SRX12658692	113,360	JALCES000000000	0.999911	121.9395	
UW_TS_LCO2_3			X				28	SRX12658904	101,250	JALCFF000000000	0.999998	121.9048	
UW_TS_LCO2_4				X			13	SRX12687755	74,621	JALCJC000000000	0.999988	121.3376	
UW_TS_LCO2_5			X				90	SRX12660021	140,410	JALCGB000000000	0.999999	120.9283	
UW_TS_LCO2_6			X				136	SRX12664702	125,596	JALCGO000000000	0.999999	120.2582	
UW_TS_LCO2_7			X				202	SRX12667031	121,391	JALCHB000000000	0.999998	120.2582	
UW_TS_LCO2_8			X				52	SRX12729156	634	JAKVLX000000000	0.999803	118.9415	
UW_TS_LCO2_9			X				64	SRX12658907	90,171	JALCFP000000000	0.999989	116.3856	
UW_TS_LCO2_10			X				100	SRX12729178	383	JAKVNT000000000	0.998986	112.2110	
UW_TS_LCO2_11		X					63	SRX12729174	1,591	JAKVMQ000000000	0.999869	111.2279	
UW_TS_LCO2_12		X					75	SRX12658615	141,269	JALCEF000000000	0.999293	107.6739	
UW_TS_LCO3_1	LCO3		X				100	SRX12729178	383	JAKVNU000000000		127.2784	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__CAG-791; s__CAG-791 sp900317475
UW_TS_LCO3_2			X				52	SRX12729156	634	JAKVLY000000000	0.999719	127.1687	
UW_TS_LCO3_3		X					63	SRX12729174	1,591	JAKVMR000000000	0.999973	127.1614	
UW_TS_LCO3_4			X				166	SRX12729462	793	JAKVNE000000000	0.999728	123.5975	
UW_TS_LCO3_5			X				28	SRX12658904	101,250	JALCFE000000000	0.999974	118.7097	
UW_TS_LCO3_6			X				64	SRX12658907	90,171	JALCFO000000000	0.999964	118.5090	
UW_TS_LCO3_7			X				90	SRX12660021	140,410	JALCGA000000000	0.999978	118.3242	
UW_TS_LCO3_8			X				202	SRX12667031	121,391	JALCHA000000000	0.999980	118.1897	
UW_TS_LCO3_9			X				136	SRX12664702	125,596	JALCGN000000000	0.999986	116.7790	
UW_TS_LCO3_10				X			13	SRX12687755	74,621	JALCJB000000000	0.999971	113.4499	
UW_TS_LCO4_1	LCO4		X				166	SRX12729462	793	JAKVNF000000000		125.4895	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__CAG-791; s__
UW_TS_LCO4_2			X				202	SRX12667031	121,391	JALCGZ000000000	0.999879	109.8689	
UW_TS_LCO4_3			X				136	SRX12664702	125,596	JALCGM000000000	0.999804	106.8013	
UW_TS_LCO4_4			X				90	SRX12660021	140,410	JALCFZ000000000	0.999957	96.3895	
UW_TS_LCO5_1	LCO5		X				3	SRX12658692	113,360	JALCER000000000		118.2519	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Eubacterium_H; s__Eubacterium_H sp900318755
UW_TS_LCO5_2		X					75	SRX12658615	141,269	JALCEE000000000	0.999991	101.1569	
UW_TS_LEVLAC1_1	LEVLAC1				X		13	SRX12687755	74,621	JALCJA000000000		127.3739	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Levilactobacillus; s__Levilactobacillus suantsaiihabitans

Reference genome[g]	Sequencing platform[h]	Completeness (%)	Contamination (%)	MAG size (Mbp)	No. of scaffolds	N_{50} (Mbp)	GC content (%)	No. of tRNAs	No. of 5S rRNAs	No. of 16S rRNAs	No. of 23S rRNAs
	Sequel II	99.46	0	3.008	1	3.008	46.6	60	5	5	5
	Sequel II	99.46	0	3.003	3	2.627	46.6	60	5	5	5
	NovaSeq S4	99.46	0	2.735	55	0.067	46.6	47	1	1	0
	NovaSeq S4	98.91	0	2.722	55	0.067	46.6	47	1	1	0
	NovaSeq S4	98.91	0	2.7	54	0.067	46.6	47	1	1	0
	NovaSeq S4	98.37	0	2.67	53	0.067	46.6	47	1	1	0
	NovaSeq S4	97.28	0	2.648	51	0.068	46.7	46	1	1	0
	NovaSeq S4	96.74	0	2.656	53	0.068	46.6	46	1	1	0
	NovaSeq S4	96.74	0	2.669	54	0.067	46.6	46	1	1	0
	NovaSeq S4	96.74	0	2.668	56	0.066	46.6	30	1	1	0
	NovaSeq S4	96.36	0	2.646	63	0.061	46.6	37	1	1	0
	NovaSeq S4	96.2	0	2.621	61	0.065	46.7	44	1	1	0
	NovaSeq S4	96.4	0	2.641	67	0.055	46.6	29	1	1	0
	NovaSeq S4	95.88	0.53	2.658	79	0.046	46.6	36	1	1	0
	NovaSeq S4	94.57	0	2.557	50	0.068	46.7	45	1	1	0
	NovaSeq S4	92.39	0	2.475	47	0.073	46.7	43	1	0	0
	NovaSeq S4	92.83	0	2.606	66	0.055	46.6	46	3	1	2
	NovaSeq S4	82.88	0	2.215	41	0.079	46.5	44	0	0	0
GCF_003946025.1	NovaSeq S4	87.43	1.05	1.567	39	0.057	53	30	1	0	0
GCF_003641185.1	NovaSeq S4	91.36	1.39	2.638	35	0.133	46.9	42	4	0	2
GCA_900315305.1	Sequel II	97.93	0	2.776	1	2.776	54.6	53	5	5	5
	Sequel II	97.93	0	2.65	1	2.65	54.7	53	5	5	5
	Sequel II	97.93	0	2.718	5	0.831	54.6	49	5	5	5
	NovaSeq S4	94.9	0	2.602	31	0.147	55.3	44	0	0	1
	NovaSeq S4	94.27	0	2.38	31	0.131	55.3	43	0	0	2
	NovaSeq S4	94.27	0	2.562	37	0.123	55.4	44	0	0	1
	NovaSeq S4	93.63	0	2.665	34	0.147	55.1	44	0	0	1
	NovaSeq S4	92.36	0	2.367	43	0.071	55.5	43	0	0	2
GCA_900314255.1	Sequel II	97.76	0	2.959	2	2.296	53	51	4	4	4
	NovaSeq S4	97.09	0	2.496	38	0.093	53.9	46	0	1	2
	NovaSeq S4	97.09	0	2.569	42	0.091	53.8	47	0	0	1
	NovaSeq S4	97.09	0	2.564	53	0.07	53.8	47	0	1	2
	NovaSeq S4	95.75	0	2.558	39	0.108	53.7	47	0	0	0
	NovaSeq S4	95.08	0	2.502	37	0.108	53.8	47	0	0	0
	NovaSeq S4	95.08	0	2.571	39	0.108	53.7	47	0	0	0
	Sequel II	90.38	0.67	2.817	6	0.599	53.1	51	4	4	4
	NovaSeq S4	91.95	0	2.479	45	0.077	53.8	46	0	0	1
	Sequel II	83	0	2.284	5	0.692	52.5	33	2	2	2
	Sequel II	83.67	0.34	2.263	10	0.299	53.3	37	2	3	2
	NovaSeq S4	82.33	0	2.145	27	0.117	53.4	32	0	0	0
GCA_900317475.1	Sequel II	96.15	0.69	3.332	1	3.332	51.8	47	3	3	3
	Sequel II	96.15	0.69	3.332	1	3.332	51.8	47	3	3	3
	Sequel II	96.15	0.69	3.321	1	3.321	51.8	47	3	3	3
	Sequel II	93.28	0.69	3.096	2	2.294	52	47	3	3	3
	NovaSeq S4	95.57	0.69	2.455	42	0.084	52.7	40	0	0	1
	NovaSeq S4	95.57	0.69	2.468	43	0.076	52.6	40	0	0	2
	NovaSeq S4	95.57	0.69	2.522	46	0.074	52.5	40	0	0	2
	NovaSeq S4	95	0.57	2.477	44	0.084	52.6	36	0	0	1
	NovaSeq S4	93.85	0.69	2.445	45	0.076	52.7	40	0	0	2
	NovaSeq S4	91.55	0.69	2.425	72	0.048	52.7	41	0	2	2
NA	Sequel II	96.55	1.26	4.3	4	0.931	56.6	53	5	5	5
	NovaSeq S4	88.14	0.57	2.614	113	0.025	59.8	31	0	4	0
	NovaSeq S4	84.69	0.57	2.717	119	0.023	59.6	27	0	4	0
	NovaSeq S4	75.1	0.19	2.388	135	0.019	59.7	25	0	1	1
GCA_900318755.1	NovaSeq S4	95.22	0.64	2.221	56	0.047	55.1	37	0	0	1
	NovaSeq S4	77.71	0.64	2.036	44	0.057	55.2	39	0	0	1
GCF_004745505.1	NovaSeq S4	99.38	0	2.685	9	0.462	52.5	56	3	0	2

TABLE 1 (Continued)

Strain name[a]	Code[b]	R1	R2	R3	R4	R5	Sample age (days)	SRA accession no.	No. of raw reads per sample (×1,000)	GenBank accession no.[d]	ANIm[e]	dRep[f]	GTDBtk classification
													Reactor source[c]
UW_TS_LEVLAC1_2					X		62	SRX12687723	84,579	JALCJT000000000	0.999967	113.5794	
UW_TS_LEVLAC1_3					X		72	SRX12687768	95,453	JALCKB000000000	0.999757	101.7672	
UW_TS_LEVLAC1_4						X	36	SRX12729153	71,598	JAKVLN000000000	0.999986	119.0191	
UW_TS_LIMLAC1_1	LIMLAC1		X				66	SRX12686873	104,400	JALCHW000000000		114.5011	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Limosilactobacillus; s__Limosilactobacillus mucosae
UW_TS_LIMLAC1_2			X				162	SRX12686848	92,161	JALCIO000000000	0.999945	114.3461	
UW_TS_LIMLAC2_1	LIMLAC2X						39	SRX12657440	113,675	JALCDW000000000		112.7804	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Limosilactobacillus; s__Limosilactobacillus timonensis
UW_TS_NIT1_1	NIT1	X					0	SRX12657172	120,759	JALCDO000000000		120.1879	d__Bacteria; p__Nitrospirota; c__Nitrospiria; o__Nitrospirales; f__Nitrospiraceae; g__Nitrospira_A; s__
UW_TS_PAENI1_1	PAENI1				X		120	SRX12687759	78,913	JALCKG000000000		122.2953	d__Bacteria; p__Firmicutes; c__Bacilli; o__Paenibacillales; f__Paenibacillaceae; g__Paenibacillus; s__Paenibacillus lautus_A
UW_TS_PED1_1	PED1		X				52	SRX12729156	634	JAKVLZ000000000		129.8987	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Pediococcus; s__Pediococcus acidilactici
UW_TS_PED1_2			X				28	SRX12658904	101,250	JALCFD000000000	0.999709	123.8544	
UW_TS_PED1_3					X		13	SRX12687755	74,621	JALCIZ000000000	0.999618	103.3166	
UW_TS_PED2_1	PED2		X				166	SRX12729462	793	JAKVNG000000000		129.5689	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Pediococcus; s__Pediococcus pentosaceus
UW_TS_PED2_2			X				52	SRX12729156	634	JAKVMA000000000	0.997152	129.1978	
UW_TS_PED2_3		X					63	SRX12729174	1,591	JAKVMS000000000	0.997888	127.9623	
UW_TS_PED2_4				X			48	SRX12670963	74,191	JALCHQ000000000	0.998994	119.8202	
UW_TS_PED2_5			X				28	SRX12658904	101,250	JALCFC000000000	0.997348	116.8151	
UW_TS_PED2_6		X					39	SRX12657440	113,675	JALCDV000000000	0.994612	116.4924	
UW_TS_PED2_7		X					14	SRX12657374	91,742	JALCDT000000000	0.995774	115.6940	
UW_TS_PED2_8		X					75	SRX12658615	141,269	JALCED000000000	0.995045	114.6572	
UW_TS_PED2_9				X			79	SRX12686849	79,309	JALCIC000000000	0.999421	111.8800	
UW_TS_PED2_10		X					100	SRX12729178	383	JAKVNV000000000	0.997417	110.7978	
UW_TS_PED2_11		X					90	SRX12660021	140,410	JALCFY000000000	0.998697	110.4612	
UW_TS_PED2_12					X		62	SRX12687723	84,579	JALCJS000000000	0.995410	108.9640	
UW_TS_PED2_13				X			66	SRX12686873	104,400	JALCHV000000000	0.995444	106.3706	
UW_TS_PED2_14					X		72	SRX12687768	95,453	JALCKA000000000	0.997899	99.8192	
UW_TS_PED2_15					X		24	SRX12686847	69,380	JALCJI000000000	0.995015	99.1435	
UW_TS_PED3_1	PED3	X					39	SRX12657440	113,675	JALCDU000000000		101.1763	d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Pediococcus; s__Pediococcus acidilactici
UW_TS_PREV1_1	PREV1	X					166	SRX12729462	793	JAKVNH000000000		129.5120	d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Prevotella; s__Prevotella sp002409785
UW_TS_PREV1_2			X				100	SRX12729178	383	JAKVNW000000000	0.992376	125.7923	
UW_TS_PREV1_3				X			24	SRX12665963	80,367	JALCHK000000000	0.991890	120.9675	
UW_TS_PREV1_4					X		13	SRX12687755	74,621	JALCIY000000000	0.992821	120.6188	
UW_TS_PREV1_5			X				28	SRX12658904	101,250	JALCFB000000000	0.999174	118.3461	
UW_TS_PREV1_6			X				136	SRX12664702	125,596	JALCGL000000000	0.999047	116.1356	
UW_TS_PREV1_7					X		62	SRX12687723	84,579	JALCJR000000000	0.992263	114.6652	
UW_TS_PREV1_8		X					75	SRX12658615	141,269	JALCEC000000000	0.998513	114.1342	
UW_TS_PREV1_9			X				202	SRX12667031	121,391	JALCGY000000000	0.998837	111.2630	
UW_TS_PREV1_10				X			114	SRX12686875	69,693	JALCIJ000000000	0.993799	110.3459	
UW_TS_PREV1_11			X				64	SRX12658907	90,171	JALCFN000000000	0.997702	110.1806	
UW_TS_PREV1_12			X				3	SRX12658692	113,360	JALCEQ000000000	0.999146	110.0664	
UW_TS_PREV1_13			X				52	SRX12729156	634	JAKVMB000000000	0.998067	105.5699	
UW_TS_PREV2_1	PREV2	X					3	SRX12657373	95,824	JALCDQ000000000		114.5766	d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Prevotella; s__
UW_TS_SACCH1_1	SACCH1	X					0	SRX12657172	120,759	JALCDM000000000		122.6364	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Saccharofermentanales; f__Saccharofermentanaceae; g__Saccharofermentans; s__
UW_TS_SACCH1_2		X					3	SRX12657373	95,824	JALCDP000000000	0.999783	108.5531	
UW_TS_SAP1_1	SAP1	X					0	SRX12657172	120,759	JALCDL000000000		119.2882	d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Chitinophagales; f__Saprospiraceae; g__UBA3362; s__

Reference genome[g]	Sequencing platform[h]	Completeness (%)	Contamination (%)	MAG size (Mbp)	No. of scaffolds	N_{50} (Mbp)	GC content (%)	No. of tRNAs	No. of 5S rRNAs	No. of 16S rRNAs	No. of 23S rRNAs
	NovaSeq S4	89.24	0	2.493	65	0.073	52.7	24	2	0	1
	NovaSeq S4	79.31	0	2.267	91	0.031	52.7	36	1	0	2
	NovaSeq S4	91.25	0	1.597	8	0.415	52.5	52	2	0	2
GCF_001436025.1	NovaSeq S4	91.47	0.82	1.847	60	0.04	47.1	45	3	0	2
	NovaSeq S4	88.91	0	1.684	19	0.122	47.2	40	2	1	1
GCF_900240275.1	NovaSeq S4	89.67	0.07	1.28	33	0.083	47.6	48	4	0	0
NA	NovaSeq S4	94.49	0.91	3.96	32	0.268	58.3	44	0	1	0
GCF_002407025.1	NovaSeq S4	98.63	0.8	6.678	102	0.103	51.5	44	3	5	3
GCF_000146325.1	Sequel II	98.77	0	1.931	1	1.931	42.3	56	5	5	5
	NovaSeq S4	98.77	0.1	1.77	22	0.117	42.4	52	1	0	0
	NovaSeq S4	78.6	0	1.089	17	0.101	41.8	38	1	0	0
GCF_001437285.1	Sequel II	98.77	0	1.661	1	1.661	37.4	55	5	5	5
	Sequel II	98.97	0	1.617	2	1.277	37.5	55	5	5	5
	Sequel II	98.61	0	1.649	4	0.853	37.4	56	5	5	5
	NovaSeq S4	97.12	0.71	1.436	45	0.04	37.2	49	1	0	0
	NovaSeq S4	94.42	0	1.32	44	0.035	37.3	48	1	0	0
	NovaSeq S4	93.65	0	1.306	43	0.043	37.4	48	1	0	0
	NovaSeq S4	93.24	0.62	1.351	49	0.041	37.2	49	1	0	0
	NovaSeq S4	92.31	0	1.269	49	0.034	37.3	48	1	0	0
	NovaSeq S4	89.56	0	1.276	53	0.033	37.2	40	1	0	0
	Sequel II	81.48	0	1.248	2	0.84	37.1	52	3	3	3
	NovaSeq S4	88.22	0	1.232	54	0.032	37.3	49	1	0	0
	NovaSeq S4	85.83	0	1.302	39	0.049	37.3	48	2	0	0
	NovaSeq S4	84.29	0	1.14	52	0.026	37.4	49	1	0	0
	NovaSeq S4	76.75	0	0.729	24	0.047	37.1	42	1	0	0
	NovaSeq S4	75.77	0.09	0.836	21	0.056	36.8	21	0	0	0
GCF_000146325.1	NovaSeq S4	80.26	0.62	1.482	68	0.024	42.5	23	1	1	0
GCA_002409785.1	Sequel II	97.97	1.94	3.171	1	3.171	44.3	54	4	4	4
	Sequel II	97.3	2.62	3.257	4	0.91	44.2	56	4	4	4
	NovaSeq S4	97.64	2.69	3.095	64	0.076	44.2	44	0	0	0
	NovaSeq S4	97.02	1.77	2.888	48	0.079	44.5	45	0	0	0
	NovaSeq S4	94.93	1.27	2.676	59	0.064	44.6	46	0	0	0
	NovaSeq S4	92.74	1.01	2.445	61	0.06	45	42	0	0	0
	NovaSeq S4	93.01	2.03	2.752	102	0.034	44.5	42	0	0	0
	NovaSeq S4	91.61	1.1	2.244	78	0.041	45.3	42	0	0	0
	NovaSeq S4	89.36	1.27	2.25	81	0.032	45.2	37	0	0	0
	NovaSeq S4	88.24	2.11	2.434	90	0.037	45	34	0	0	0
	NovaSeq S4	87.56	1.01	2.188	72	0.042	45.3	41	0	0	0
	NovaSeq S4	86.17	1.27	2.49	42	0.08	44.5	40	0	0	0
	Sequel II	76.69	1.32	2.744	6	0.655	44.1	43	3	3	3
NA	NovaSeq S4	89.46	0.91	2.562	33	0.205	43.6	40	1	0	0
NA	NovaSeq S4	98.23	0	1.866	30	0.076	46.7	41	2	0	0
	NovaSeq S4	86.17	0.38	1.636	70	0.027	46.8	38	2	0	1
NA	NovaSeq S4	96.78	0.14	4.309	116	0.05	33.3	34	1	2	1

TABLE 1 (Continued)

Strain name[a]	Code[b]	Reactor source[c]					Sample age (days)	SRA accession no.	No. of raw reads per sample (×1,000)	GenBank accession no.[d]	ANIm[e]	dRep[f]	GTDBtk classification
		R1	R2	R3	R4	R5							
UW_TS_SPHING1_1	SPHING1	X					0	SRX12657172	120,759	JALCDK000000000		108.9496	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae; g__Sphingobium; s__
UW_TS_RUM1_1	RUM1	X					0	SRX12657172	120,759	JALCDN000000000		110.7404	d__Bacteria; p__Firmicutes_A; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae; g__Ruminococcus; s__
UW_TS_TETRA1_1	TETRA1	X					0	SRX12657172	120,759	JALCDJ000000000		100.7487	d__Bacteria; p__Actinobacteriota; c__Actinomycetia; o__Actinomycetales; f__Dermatophilaceae; g__Tetrasphaera_A; s__Tetrasphaera_A jenkinsii

[a] Strain name assigned to each reported MAG. The abbreviation UW_TS stands for University of Wisconsin Thin Stillage bioreactor. The strain name followed by _1 denotes the representative MAG, as selected using dRep (11), for each cluster; other MAGs in the same cluster use the same strain name followed by a number in increasing order and are sorted according to the dRep scores.

[b] ACET, *Acetobacter*; ACID, *Acidaminococcus*; ANEURI, *Aneurinibacillus*; ATO, *Atopobiaceae*; BAC, *Bacillus*; BUL, *Bulleidia*; CLOS, *Clostridium*; COMLAC, *Companilactobacillus*; EGG, *Eggerthellaceae*; EUB, *Eubacteriaceae*; FURLAC, *Furfurilactobacillus*; GAST, *Gastranaerophilaceae*; LAC, *Lactobacillus*; LACCAS, *Lacticaseibacillus*; LACPLAN, *Lactiplantibacillus*; LCO, *Lachnospiraceae*; LEVLAC, *Levilactobacillus*; LIMLAC, *Limosilactobacillus*; NIT, *Nitrospira*; PAENI, *Paenibacillus*; PED, *Pediococcus*; PREV, *Prevotella*; RUM, *Ruminococcus*; SACCH, *Saccharofermentans*; SAP, *Saprospiraceae*; SPHING, *Sphingobium*; TETRA, *Tetrasphaera*.

[c] R1, reactor 1, TS (R1$_{TS}$); R2, reactor 2, solids-removed TS (R2$_{SR-TS}$); R3, reactor 3, low SRT (R3$_{LowSRT}$); R4, reactor 4, high temperature and low pH (R4$_{T-pH}$); R5, reactor 5, low SRT, high temperature, and low pH (R5$_{T-pH-LowSRT}$).

[d] NCBI GenBank accession number for each reported MAG.

[e] Average nucleotide identity (ANI) between representative MAG and other MAGs included in the same cluster by dRep. The definitions of clusters in dRep were based on having ANIs of >99%.

[f] dRep scoring calculation: $(A \times$ completeness$) - (B \times$ contamination$) + [C \times$ [contamination $\times$ (strain heterogeneity/100)]] $+ [D \times \log(N_{50})] + [E \times \log($genome size$)] + [F \times ($centrality $-$ ANI$)]$, where A to F were weighted with values of 1, 0.5, 1, 5, 0, and 1, respectively.

[g] NCBI GenBank accession number of the reference genome in GTDBtk (14) that is closest to the representative MAG; NA, not applicable, i.e., MAGs without a closely matched reference genome when using the default minimum alignment fraction of 0.65.

[h] The Sequel II platform is produced by PacBio, and the NovaSeq S4 platform is produced by Illumina.

set adds to the expanding body of knowledge about microorganisms relevant to the valorization of agroindustrial residues by fermentation (2, 16–20).

Data availability. Raw metagenomic sequence data and MAGs are available in NCBI GenBank under BioProject accession number PRJNA768492. All information on library construction and sequence can be found at https://gold.jgi.doe.gov/study?id=Gs0150020 using JGI GOLD Study identification number Gs0150020. All custom scripts are available at GitHub (https://github.com/GLBRC/metagenome_analysis).

REFERENCES

1. Scarborough MJ, Lynch G, Dickson M, McGee M, Donohue TJ, Noguera DR. 2018. Increasing the economic value of lignocellulosic stillage through medium-chain fatty acid production. Biotechnol Biofuels 11:200. https://doi.org/10.1186/s13068-018-1193-x.
2. Scarborough MJ, Lawson CE, Hamilton JJ, Donohue TJ, Noguera DR. 2018. Metatranscriptomic and thermodynamic insights into medium-chain fatty acid production using an anaerobic microbiome. mSystems 3: e00221-18. https://doi.org/10.1128/mSystems.00221-18.
3. Scarborough MJ, Myers KS, Donohue TJ, Noguera DR. 2020. Medium-chain fatty acid synthesis by "*Candidatus* Weimeria bifida" gen. nov., sp. nov., and "*Candidatus* Pseudoramibacter fermentans" sp. nov. Appl Environ Microbiol 86:e02242-19. https://doi.org/10.1128/AEM.02242-19.
4. Fortney NW, Hanson NJ, Rosa PRF, Donohue TJ, Noguera DR. 2021. Diverse profile of fermentation byproducts from thin stillage. Front Bioeng Biotechnol 9:695306. https://doi.org/10.3389/fbioe.2021.695306.
5. Ingle AT, Fortney NW, Walters KA, Donohue TJ, Noguera DR. 2021.

Reference genome[g]	Sequencing platform[h]	Completeness (%)	Contamination (%)	MAG size (Mbp)	No. of scaffolds	N_{50} (Mbp)	GC content (%)	No. of tRNAs	No. of 5S rRNAs	No. of 16S rRNAs	No. of 23S rRNAs
NA	NovaSeq S4	87.4	1.07	2.918	103	0.036	63.3	39	1	1	2
NA	NovaSeq S4	89.6	0.12	1.97	82	0.027	39.2	32	1	0	0
GCF_001046875.1	NovaSeq S4	78.29	0.45	2.848	83	0.051	68.5	41	0	0	0

Mixed acid fermentation of carbohydrate-rich dairy manure hydrolysate. Front Bioeng Biotechnol 9:724304. https://doi.org/10.3389/fbioe.2021.724304.

6. Clum A, Huntemann M, Bushnell B, Foster B, Foster B, Roux S, Hajek PP, Varghese N, Mukherjee S, Reddy TBK, Daum C, Yoshinaga Y, O'Malley R, Seshadri R, Kyrpides NC, Eloe-Fadrosh EA, Chen I-MA, Copeland A, Ivanova NN. 2021. DOE JGI Metagenome Workflow. mSystems 6:e00804-20. https://doi.org/10.1128/mSystems.00804-20.

7. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. https://sourceforge.net/projects/bbmap.

8. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.

9. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359. https://doi.org/10.7717/peerj.7359.

10. Tennessen K, Andersen E, Clingenpeel S, Rinke C, Lundberg DS, Han J, Dangl JL, Ivanova NN, Woyke T, Kyrpides N, Pati A. 2016. ProDeGe: a computational protocol for fully automated decontamination of genomes. ISME J 10:269–272. https://doi.org/10.1038/ismej.2015.100.

11. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868. https://doi.org/10.1038/ismej.2017.126.

12. 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.

13. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.

14. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

15. Tatusova T, DiCuccio M, Badretdin 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.

16. Hollister EB, Forrest AK, Wilkinson HH, Ebbole DJ, Tringe SG, Malfatti SA, Holtzapple MT, Gentry TJ. 2012. Mesophilic and thermophilic conditions select for unique but highly parallel microbial communities to perform carboxylate platform biomass conversion. PLoS One 7:e39689. https://doi.org/10.1371/journal.pone.0039689.

17. Zhu X, Feng X, Liang C, Li J, Jia J, Feng L, Tao Y, Chen Y. 2021. Microbial ecological mechanism for long-term production of high concentrations of n-caproate via lactate-driven chain elongation. Appl Environ Microbiol 87:e03075-20. https://doi.org/10.1128/AEM.03075-20.

18. Wu S-L, Sun J, Chen X, Wei W, Song L, Dai X, Ni B-J. 2020. Unveiling the mechanisms of medium-chain fatty acid production from waste activated sludge alkaline fermentation liquor through physiological, thermodynamic and metagenomic investigations. Water Res 169:115218. https://doi.org/10.1016/j.watres.2019.115218.

19. Crognale S, Braguglia CM, Gallipoli A, Gianico A, Rossetti S, Montecchio D. 2021. Direct conversion of food waste extract into caproate: metagenomics assessment of chain elongation process. Microorganisms 9:327. https://doi.org/10.3390/microorganisms9020327.

20. Han W, He P, Shao L, Lü F. 2018. Metabolic interactions of a chain elongation microbiome. Appl Environ Microbiol 84:e01614-18. https://doi.org/10.1128/AEM.01614-18.