

Article

Genome Characterization of *Lactiplantibacillus plantarum* Strain UTNGt2 Originated from *Theobroma grandiflorum* (White Cacao) of Ecuadorian Amazon: Antimicrobial Peptides from Safety to Potential Applications

Gabriela N. Tenea * and Clara Ortega

Table S1. Assembly summary results.

Sample	Contigs	Total contig bases	N50	Max length	Min length	Mean length
UTNGt2	31	3,264,448	186,152	307,684	1,169	105,304

Contigs: The number of contigs assembled; Total bases of contigs: The total length of contigs; N50: 50% of all contig bases come from contigs longer than this value; Max length: The length of the longest contig; Min length: The length of the shortest contig; Mean length: The average length of contigs assembled.

Table S2. Mapping overall results.

Library name	Total reads	Mapped reads	Coverage (%)	Depth	Ins.size (Std.)
UTNGt2	5,258,542	5,258,086 (99.99%)	100.00	160.15	454.46 (93.64)

Legend: Library name: Sample's library name; Total reads: Total number of reads; Mapped reads: Total number of mapped reads; Coverage (%): The percentage of mapped sites ($\geq 1x$); Depth: Average mapping depth; Ins.size (Std.): The length between adapters and standard deviation of predicted length.

Table S3. BUSCO analysis results. Used Lineage: bacteria_odb9 (number of species: 3663, number of BUSCOs: 148).

Status	# of BUSCOs	Percentage (%)
Complete BUSCOs (C)		
Complete and single-copy BUSCOs (S)	142	95.95
Complete and duplicated BUSCOs (D)	2	1.35
Fragmented BUSCOs (F)	1	0.68
Missing BUSCOs (M)	3	2.03
Total BUSCO groups searched	148	100

Status: A quantitative assessment list of the completeness in terms of expected gene content. The following two conditions were used to create a status: a. Expected range of scores; b. Expected range of length alignments. If both conditions are met, it is classified as Complete (These complete busco matches are either single-copy or duplicated). If length alignments is not met, it is classified

as Fragmented. If both conditions are not met, it is classified as Missing. # of BUSCOs: Identified count in sample; Percentage: Identified percentage in sample.

Table 4. List of reference ID genomes, species, and lineage used for phylogenetic analysis retrieved from the gyType database.

Reference ID	query1/ Gt2	Species	Lineage
GCA_0000.0114/98.97/99.143745.1	01/99.07/99.07	<i>Lactobacillus plantarum</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Lactobacillus s__Lactobacillus plantarum
GCA_0030.0342/95.08/95.641165.1	68/95.48/95.49	<i>Lactobacillus plantarum</i> subsp. <i>argenteratensis</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Lactobacillus s__Lactobacillus plantarum
GCA_0030.0883/85.70/87.641145.1	70/86.10/86.28	<i>Lactobacillus paraplantarum</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Lactobacillus s__Lactobacillus paraplantarum
GCA_0030.1519/79.24/82.641185.1	55/79.69/79.98	<i>Lactobacillus pentosus</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Lactobacillus s__Lactobacillus pentosus
GCA_0050.1823/74.22/80.405405.1	23/74.95/75.10	<i>Lactobacillus plajomi</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Lactobacillus s__Lactobacillus plajomi
GCA_0010.1868/76.72/80.039045.1	49/77.19/77.41	<i>Lactobacillus herbarum</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Lactobacillus s__Lactobacillus herbarum
GCA_004.02301/67.00/#/354495.1	67.80/68.71	<i>Pediococcus pentosaceus</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Pediococcus s__Pediococcus pentosaceus
GCA_000.02438/65.33/#/769695.1	65.84/68.11	<i>Oenococcus alcoholitolerans</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Oenococcus s__Oenococcus alcoholitolerans
GCA_000.02438/64.27/#/092505.1	64.81/66.30	<i>Leuconostoc kimchii</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Leuconostoc s__Leuconostoc kimchii
GCA_000.02438/64.40/#/014445.1	65.04/66.41	<i>Leuconostoc mesenteroides</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Leuconostoc s__Leuconostoc mesenteroides
GCA_003.02438/64.65/#/363255.1	65.65/66.69	<i>Weissella thailandensis</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Weissella s__Weissella thailandensis
GCA_012.02438/64.66/#/396435.1	65.78/67.26	<i>Weissella hellenica</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Weissella s__Weissella hellenica
GCA_900.02438/66.87/#/111205.1	67.73/68.55	<i>Pediococcus ethanolidurans</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Pediococcus s__Pediococcus ethanolidurans
GCA_000.02438/64.73/#/160595.1	65.56/67.56	<i>Leuconostoc mesenteroides</i> subsp. <i>cremoris</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Leuconostoc s__Leuconostoc mesenteroides
GCA_001.02438/65.04/#/932615.1	65.97/67.41	<i>Weissella jogaejeotgali</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Weissella s__Weissella jogaejeotgali
GCA_005.02630/64.94/#/405345.1	65.65/66.40	<i>Enterococcus viikkiensis</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Enterococcaceae g__Enterococcus s__Enterococcus viikkiensis
GCA_008.02630/60.97/#/086625.1	61.64/61.91	<i>Acetobacterium bakii</i>	k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales f__Eubacteriaceae g__Acetobacterium s__Acetobacterium bakii
GCM100.02630/66.95/#/08912	67.91/68.76	<i>Pediococcus parvulus</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Pediococcus s__Pediococcus parvulus
GCA_000.02630/60.42/#/577815.1	61.42/62.58	<i>Clostridium saudiense</i>	k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales f__Clostridiaceae g__Clostridium s__Clostridium saudiense
GCA_000.02630/65.30/#/260375.1	65.40/67.13	<i>Leuconostoc carnosum</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Leuconostoc s__Leuconostoc carnosum
GCA_002.02630/67.26/#/982135.1	68.08/69.18	<i>Pediococcus inopinatus</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Pediococcus s__Pediococcus inopinatus
GCA_001.02630/66.72/#/437255.1	67.53/68.57	<i>Pediococcus damnosus</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Pediococcus s__Pediococcus damnosus
GCA_000.02960/59.41/#/373965.1	60.76/60.23	<i>Uliginosibacterium gangwonense</i>	k__Bacteria p__Proteobacteria c__Betaproteobacteria o__Rhodocyclales f__Zoogloeaceae g__Uliginosibacterium s__Uliginosibacterium gangwonense

GCA_000 0.2960/60.02/#/ 012545.1 60.57/64.33	<i>Methanosphaera stadtmanae</i>	k__Archaea p__Euryarchaeota c__Methanobacteria o__Methanobacteriales f__Methanobacteriaceae g__Methanosphaera s__Methanosphaera stadtmanae
GCA_000 0.2960/60.48/#/ 320405.1 61.63/62.32	<i>Clostridium celatum</i>	k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales f__Clostridiaceae g__Clostridium s__Clostridium celatum
GCA_000 0.2960/61.15/#/ 258405.1 61.77/63.54	<i>Ignavibacterium album</i>	k__Bacteria p__Ignavibacteriae c__Ignavibacteria o__Ignavibacteriales f__Ignavibacteriaceae g__Ignavibacterium s__Ignavibacterium album
GCA_000 0.2960/61.20/#/ 015285.1 61.75/61.94	<i>Psychromonas ingrahamii</i>	k__Bacteria p__Proteobacteria c__Gammaproteobacteria o__Alteromonadales f__Psychromonadaceae g__Psychromonas s__Psychromonas ingrahamii
GCA_000 0.2960/64.21/#/ 166735.2 65.08/66.82	<i>Leuconostoc inhae</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Leuconostocaceae g__Leuconostoc s__Leuconostoc inhae
GCA_000 0.2960/61.74/#/ 367405.1 62.87/64.78	<i>Rickettsia prowazekii</i>	k__Bacteria p__Proteobacteria c__Alphaproteobacteria o__Rickettsiales f__Rickettsiaceae g__Rickettsia s__Rickettsia prowazekii
GCA_000 0.2960/61.00/#/ 297255.1 61.63/63.63	<i>Nitrolancea hollandica</i>	k__Bacteria p__Chloroflexi c__Thermomicrobia o__Sphaerobacterales f__Sphaerobacteraceae g__Nitrolancea s__Nitrolancea hollandica
GCA_000 0.2960/61.23/#/ 378365.1 62.32/63.75	<i>Megamonas rupellensis</i>	k__Bacteria p__Firmicutes c__Negativicutes o__Selenomonadales f__Selenomonadaceae g__Megamonas s__Megamonas rupellensis
GCA_000 0.2960/60.89/#/ 173355.1 61.61/62.61	<i>Anaerococcus hydrogenalis</i>	k__Bacteria p__Firmicutes c__Tissierellia o__Tissierellales f__Peptoniphilaceae g__Anaerococcus s__Anaerococcus hydrogenalis
GCA_000 0.2960/61.18/#/ 378225.1 62.19/64.02	<i>Methylophilus methylotrophus</i>	k__Bacteria p__Proteobacteria c__Betaproteobacteria o__Nitrosomonadales f__Methylophilaceae g__Methylophilus s__Methylophilus methylotrophus
GCA_000 0.2960/61.16/#/ 153185.1 61.76/63.31	<i>Reinekea blandensis</i>	k__Bacteria p__Proteobacteria c__Gammaproteobacteria o__Oceanospirillales f__Saccharospirillaceae g__Reinekea s__Reinekea blandensis
GCA_000 0.2960/62.74/#/ 008045.1 63.34/63.83	<i>Rickettsia typhi</i>	k__Bacteria p__Proteobacteria c__Alphaproteobacteria o__Rickettsiales f__Rickettsiaceae g__Rickettsia s__Rickettsia typhi
GCA_000 0.2960/62.04/#/ 257395.1 63.18/64.64	<i>Diplorickettsia massiliensis</i>	k__Bacteria p__Proteobacteria c__Gammaproteobacteria o__Legionellales f__Coxiellaceae g__Diplorickettsia s__Diplorickettsia massiliensis
GCA_000 0.2960/63.94/#/ 376985.1 64.47/65.11	<i>Streptococcus henryi</i>	k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Streptococcaceae g__Streptococcus s__Streptococcus henryi
GCA_000 0.2960/61.40/#/ 300235.2 62.05/63.59	<i>Thalassospira xiamenensis</i>	k__Bacteria p__Proteobacteria c__Alphaproteobacteria o__Rhodospirillales f__Rhodospirillaceae g__Thalassospira s__Thalassospira xiamenensis
GCA_0000.2960/61.44/#/# 299895.1 /63.61	<i>Idiomarina xiamenensis</i>	k__Bacteria p__Proteobacteria c__Gammaproteobacteria o__Alteromonadales f__Idiomarinaceae g__Idiomarina s__Idiomarina xiamenensis
GCA_000 0.2960/60.49/#/ 190595.1 61.66/62.73	<i>Cellulophaga lytica</i>	k__Bacteria p__Bacteroidetes c__Flavobacteriia o__Flavobacteriales f__Flavobacteriaceae g__Cellulophaga s__Cellulophaga lytica

Table S5. EggNOG category distribution result.

Eggnog	Description	Count	Ratio (%)
J	Translation, ribosomal structure and biogenesis	150	5.0573
A	RNA processing and modification	0	0
K	Transcription	251	8.4626
L	Replication, recombination and repair	139	4.6864
B	Chromatin structure and dynamics	0	0
D	Cell cycle control, cell division, chromosome partitioning	25	0.8429
Y	Nuclear structure	0	0
V	Defense mechanisms	70	2.3601
T	Signal transduction mechanisms	73	2.4612
M	Cell wall/membrane/envelope biogenesis	166	5.5968
N	Cell motility	3	0.1011
Z	Cytoskeleton	0	0
W	Extracellular structures	0	0
U	Intracellular trafficking, secretion, and vesicular transport	23	0.7755
O	Posttranslational modification, protein turnover, chaperones	63	2.1241
C	Energy production and conversion	110	3.7087

G	Carbohydrate transport and metabolism	279	9.4066
E	Amino acid transport and metabolism	201	6.7768
F	Nucleotide transport and metabolism	86	2.8995
H	Coenzyme transport and metabolism	72	2.4275
I	Lipid transport and metabolism	56	1.8881
P	Inorganic ion transport and metabolism	137	4.619
Q	Secondary metabolites biosynthesis, transport and catabolism	17	0.5732
R	General function prediction only	300	10.1146
S	Function unknown	745	25.118
Total	-	2966	100

count: number of genes; ratio (%): % of genes

Table S6. Antibiotic susceptibility of the UTNGt2 strain.

Antimicrobial E-test strip	MIC (mg/L)	Susceptibility*	EFSA cut-off values (mg/L) <i>Lactobacillus obligate</i> Heterofermentative [EFSA, 2012]
Amoxicillin	0.5	S	4
Amoxicillin: clavulanic acid	0.016	S	4
Ampicillin	0.5	S	2
Cefotaxime	0.002	S	16
Erythromycin	2	S	1
Gentamycin	16	S	16
Penicillin	0.25	S	0.25
Tetracycline	32	S	32**

*The microbiological breakpoints reported by the FEEDAP were used to categorize lactobacilli as susceptible or resistant. The strains showing a MIC higher than the EFSA breakpoint were considered resistant. Susceptible (S): a bacterial strain is defined as susceptible when it is inhibited at a concentration of a specific antimicrobial equal to or lower than the established cut-off value ($S \leq x$ mg / L). Resistant (R): a bacterial strain is defined as resistant when it is not inhibited at a concentration of a specific antimicrobial above the established cut-off value ($R > x$ mg / L). ** MIC reference for *L. plantarum*.

Table S7. Summary of Prophage regions identified on the UTNGt2 genome using the PHASTER webserver.

Contig	Region	Region length	Completeness	Score	# Total proteins	Region position	Most common Phage	GC%
2	1	3.3 kb	incomplete	20	6	214956-218324	PHAGE_Xantho_Carpasina_NC_047962(1)	49.04
3	1	6.2 kb	incomplete	10	6	156134-162384	PHAGE_Bacill_G_NC_023719(3)	47.85
5	1	9.4 kb	incomplete	10	11	121969-131437	PHAGE_Vibrio_YC_NC_048709(1)	44.83
6	3	39.9 kb	intact	150	50	16543-56481	PHAGE_Lactob_Sha1_NC_019489(27)	40.79

Legend: region: the number assigned to the region; region length: the length of the sequence in that region; completeness: a prediction of whether the region contains a intact or incomplete prophage based on the criteria: intact (score > 90%); questionable (score 70-90); incomplete (score < 70%); score: the score of the region based on the mentioned criteria; # total proteins: the number of ORFs present in the region; Region position: the start and end positions of the region on the bacterial chromosome; Most common phage: the phages (s) with the highest number of proteins most similar to those in the region; GC%: the percentage of GC nucleotides of the region.

Table S8. Description of the Cluster gene regions of UTNGt2 detected showing similarity on the MIBig database.

Reference ID	Similarity score	Type	Compound(s)	Organism
<u>BGC000061</u> 7.1	0.22	RiPP	coagulin	<i>Bacillus coagulans</i>
<u>BGC000140</u> 7.1	0.16	RiPP	bicereucin	<i>Bacillus cereus</i> SJ1
<u>BGC000193</u> 0.1	0.16	RiPP	pallidocin	<i>Aeribacillus pallidus</i>
<u>BGC000186</u> 2.1	0.16	RiPP	geocillicin	<i>Aeribacillus pallidus</i>
<u>BGC000055</u> 8.1	0.16	RiPP	sublancin 168	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168
<u>BGC000186</u> 3.1	0.15	RiPP, Terpene	bacillicin CER074	<i>Bacillus mycoides</i>
<u>BGC000186</u> 1.1	0.15	RiPP	bacillicin BAG2O	<i>Bacillus cereus</i> BAG2O-1
<u>BGC000048</u> 4.1	0.15	RiPP	glycocin F	<i>Lactobacillus plantarum</i>
<u>BGC000050</u> 0.1	0.14	RiPP	carnolysin A1, carnolysin A2	<i>Carnobacterium maltaromaticum</i>
<u>BGC000050</u> 4.1	0.14	RiPP	cytolysin ClyLI, cytolysin ClyLs	Plasmid pAD1
<u>BGC000064</u> 7.1	0.45	Terpene	carotenoid	<i>Rhodobacter sphaeroides</i>
<u>BGC000064</u> 8.1	0.23	Terpene	carotenoid	<i>Myxococcus xanthus</i>
<u>BGC000065</u> 6.1	0.23	Terpene	zeaxanthin	<i>Xanthobacter autotrophicus</i> Py2
<u>BGC000063</u> 7.1	0.23	Terpene	carotenoid	<i>Corynebacterium glutamicum</i>
<u>BGC000064</u> 5.1	0.21	Terpene	carotenoid	<i>Halobacillus halophilus</i> DSM 2266
<u>BGC000122</u> 7.1	0.21	Terpene	isorenieratene	<i>Streptomyces collinus</i> Tu 365
<u>BGC000063</u> 3.1	0.21	Terpene	carotenoid	<i>Streptomyces avermitilis</i>
<u>BGC000064</u> 0.1	0.2	Terpene	carotenoid	<i>Enterobacteriaceae</i> bacterium DC404
<u>BGC000063</u> 6.1	0.2	Terpene	carotenoid	<i>Brevibacterium linens</i>

<u>BGC000063</u> <u>0.1</u>	0.2	Terpene	(2R,3S,3'S)-2-hydroxyastaxanthin	<i>Paracoccus haeundaensis</i>
<u>BGC000028</u> <u>6.1</u>	0.19	Polyketide	viguiepinol	<i>Streptomyces sp.</i> KO-3988
<u>BGC000055</u> <u>4.1</u>	0.16	RiPP	SRO15-3108	<i>Streptomyces filamentosus</i> NRRL 15998
<u>BGC000155</u> <u>1.1</u>	0.16	RiPP	citrulassin E	<i>Streptomyces glaucescens</i>
<u>BGC000038</u> <u>6.1</u>	0.16	NRP	malleobactin A, malleobactin B, malleobactin C, malleobactin D	<i>Burkholderia thailandensis</i> E264
<u>BGC000055</u> <u>1.1</u>	0.16	RiPP	SapB	<i>Streptomyces coelicolor</i> A3(2)
<u>BGC000020</u> <u>5.1</u>	0.16	Polyketide	bryostatin	<i>Candidatus Endobugula sertula</i>
<u>BGC000048</u> <u>9.1</u>	0.16	RiPP	enterocin AS-48	<i>Enterococcus faecalis</i>
<u>BGC000050</u> <u>4.1</u>	0.16	RiPP	cytolysin ClyLI, cytolysin ClyLs	Plasmid pAD1
<u>BGC000050</u> <u>0.1</u>	0.16	RiPP	carnolysin A1, carnolysin A2	<i>Carnobacterium maltaromaticum</i>
<u>BGC000049</u> <u>0.1</u>	0.16	RiPP	garvicin ML	<i>Lactococcus garvieae</i> DCC43

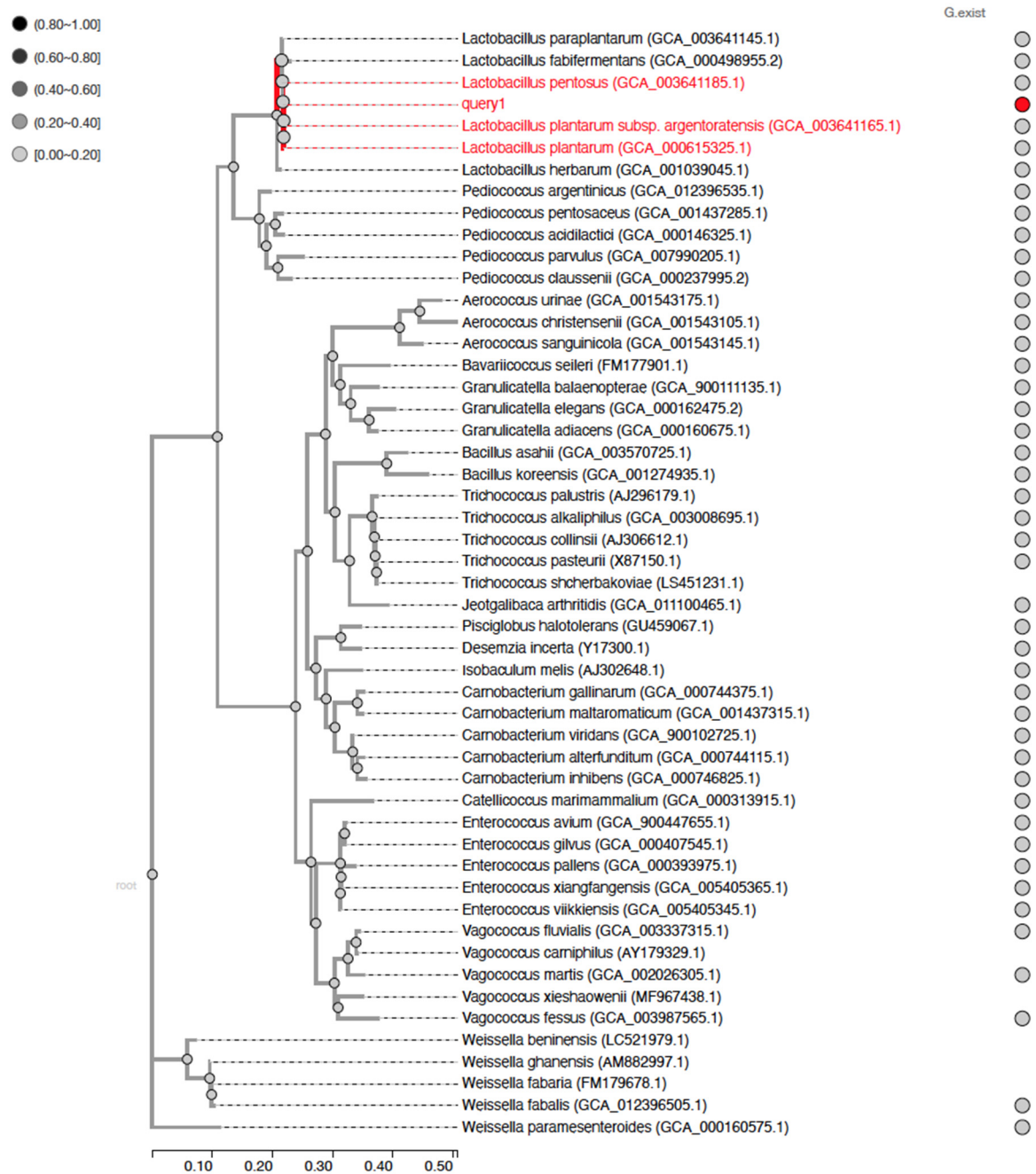


Figure S1. Phylogenetic tree based on the multiple alignment result of the target and selected 16S gene sequences to the gcType 16S rDNA gene database. Query 1: UTNGt2.

