

Supplementary Material

Whole Genome Sequencing and Metagenomics Reveals Diversity and Prevalence of Soil *Listeria* spp. in the Nantahala National Forest

Jia Wang^a, Claire N. Schamp^a, Lauren K. Hudson^a, Harleen K. Chaggar^a, Daniel W. Bryan^a, Katie N. Garman^b, Mark Radosevich^c, Thomas G. Denes^{a,*}

^aDepartment of Food Science, The University of Tennessee, Knoxville, Tennessee 37996, USA

^bTennessee Department of Health, Nashville, Tennessee 37243, USA

^cDepartment of Biosystems Engineering and Soil Science, The University of Tennessee, Knoxville, Tennessee 37996, USA

Jia Wang and Claire N. Schamp contributed equally to this work.

*Corresponding Author

Email address: tgdenes@utk.edu

15
16

Table S1. Summary of features of genomes of *Listeria* spp. isolated from soil samples at different altitudes in the Nantahala National Forest*

Isolate or representative strain	Species	Sampling altitude (ft)	No. Contigs	Total length (bp)	GC content (%)	Completeness (%)	Contamination (%)	Other identical isolates	
UTK S2-0001	<i>Listeria cossartiae</i> subsp. <i>cayugensis</i>	1500	10	2859587	38.62	99.17	0.55	UTK S2-0002 UTK S2-0003	UTK S2-0018
UTK S2-0007 ^a	<i>Listeria swaminathanii</i>	1500	11	2820687	38.59	99.17	0	UTK S2-0008 UTK S2-0009 UTK S2-0026 UTK S2-0027 UTK S2-0028	UTK S2-0029 UTK S2-0043 UTK S2-0063 UTK S2-0064
UTK S2-0071 ^a	<i>Listeria swaminathanii</i>	4500	11	2820713	38.59	99.17	0	UTK S2-0017	
UTK S2-0014	<i>Listeria marthii</i>	2500	9	2776380	38.62	99.45	0.83	—	
UTK S2-0015	<i>Listeria booriae</i>	3500	55	3969179	41.57	96.69	1.1	—	
UTK S2-0019	<i>Listeria cossartiae</i> subsp. <i>cayugensis</i>	1500	10	2859380	38.62	99.17	0.55	—	
UTK S2-0020	<i>Listeria cossartiae</i> subsp. <i>cayugensis</i>	1500	10	2859285	38.62	99.17	0.55	—	
UTK S2-0031	<i>Listeria cossartiae</i> subsp. <i>cayugensis</i>	2500	9	2732998	38.72	99.17	0.55	—	
UTK S2-0038	<i>Listeria cossartiae</i> subsp. <i>cayugensis</i>	1500	8	2732291	38.71	99.17	0.55	—	
UTK S2-0058	<i>Listeria monocytogenes</i>	1500	14	2843579	37.91	99.45	0	—	
UTK S2-0059	<i>Listeria monocytogenes</i>	1500	14	2843947	37.91	99.45	0	UTK S2-0021 UTK S2-0056	UTK S2-0057
UTK S2-0061	<i>Listeria cossartiae</i> subsp. <i>cayugensis</i>	1500	10	2728973	38.75	99.45	0.55	UTK S2-0004 UTK S2-0006 UTK S2-0022	UTK S2-0023 UTK S2-0024 UTK S2-0025
UTK S2-0062	<i>Listeria monocytogenes</i>	1500	14	2925490	37.82	99.45	0	—	
UTK S2-0067	<i>Listeria cossartiae</i> subsp. <i>cayugensis</i>	2500	9	2733526	38.72	99.17	0.55	UTK S2-0030 UTK S2-0032 UTK S2-0033	UTK S2-0065 UTK S2-0066
UTK S2-0072	<i>Listeria cossartiae</i> subsp. <i>cayugensis</i>	4500	9	2745156	38.71	99.17	0.55	—	

17 *The *Listeria* isolates selected as representatives for each group of identical isolates were based on the BLASTN results obtained from Geneious
18 Prime. If a genome of an isolate could be aligned by metagenomic reads, it was selected as the representative isolate for a group of identical isolates.
19 If there was no isolate that could be aligned from metagenomic reads, the isolate with the smallest number was selected.

20 ^a These two groups of isolates were identical (> 99.99% ANI and hqSNP distance of 0), and constitute a single unique strain based on genomics.
21 However, two were selected as representative, as they were isolated from different samples.

Table S2. Summary of raw and processed sequences of metagenomes isolated from soil samples at different altitudes in the Nantahala National Forest

Sample	Sampling altitude (ft)	Total number of bases	Number of sequences	Mean GC content	DRISEE error	QC passed reads (% of total reads)	Predicted protein features	Predicted rRNA features
1500A	1500	4,458,724,630	30,693,922	63 ± 8%	2.853%	79.27%	21,346,196	100,226
1500B	1500	5,077,763,204	35,210,268	64 ± 8%	0.652%	76.20%	23,338,988	124,544
1500C	1500	4,206,087,509	29,403,578	60 ± 11%	4.733%	77.46%	19,182,845	107,779
2500A	2500	4,717,239,687	32,708,202	62 ± 8%	3.332%	75.85%	21,246,860	107,930
2500B	2500	5,061,958,651	35,292,010	64 ± 9%	4.053%	76.10%	23,228,981	127,058
2500C	2500	5,326,361,201	36,714,080	61 ± 9%	0.723%	75.41%	23,736,050	124,146
3500A	3500	4,706,311,427	32,518,686	61 ± 9%	3.621%	77.08%	21,353,053	99,706
3500B	3500	5,529,929,864	38,526,466	61 ± 10%	2.967%	74.24%	24,030,143	142,041
3500C	3500	5,203,513,449	36,082,306	61 ± 9%	0.808%	75.03%	23,203,152	124,986
4500A	4500	4,963,430,243	33,330,456	61 ± 8%	2.630%	75.48%	22,199,297	106,899
4500B	4500	5,554,118,891	38,303,780	62 ± 9%	2.853%	76.08%	24,868,873	132,947
4500C	4500	3,957,001,803	27,565,480	62 ± 9%	2.539%	77.23%	17,994,636	106,278

Table S3. Summary of soil physicochemical properties at four different altitudes in the Nantahala National Forest. Values presented were the average of triplicates (mean $\pm$ standard error, n=3)

Altitude (ft)	Organic matter (%)	Total soil phosphorus (ppm)	Exchangeable cations			pH	Cation exchange capacity (meq/100g)	Total organic carbon (%)	Gravimetric moisture content (%)
			K (mg/kg)	Mg (mg/kg)	Ca (mg/kg)				
1500	19.0 $\pm$ 8.8	18.0 $\pm$ 7.6	82.7 $\pm$ 26.0	272.7 $\pm$ 111.5	2163.3 $\pm$ 808.8	6.3 $\pm$ 0.5	14.0 $\pm$ 4.5	12.5 $\pm$ 6.3	27.6 $\pm$ 14.0
2500	6.2 $\pm$ 2.3	8.0 $\pm$ 2.5	57.0 $\pm$ 4.4	79.3 $\pm$ 21.9	738.0 $\pm$ 521.7	5.6 $\pm$ 0.9	5.9 $\pm$ 2.1	3.6 $\pm$ 1.6	23.6 $\pm$ 5.3
3500	57.6 $\pm$ 14.5	11.3 $\pm$ 3.5	104.7 $\pm$ 10.7	60.7 $\pm$ 3.2	217.0 $\pm$ 31.6	4.4 $\pm$ 0.1	4.9 $\pm$ 0.2	35.4 $\pm$ 9.4	167.3 $\pm$ 48.2
4500	44.9 $\pm$ 16.9	31.7 $\pm$ 9.0	91.3 $\pm$ 19.6	41.7 $\pm$ 4.7	158.0 $\pm$ 24.3	4.2 $\pm$ 0.2	4.2 $\pm$ 0.2	28.0 $\pm$ 9.3	96.2 $\pm$ 16.7

33 **Table S4.** Pearson’s correlation coefficients between soil physicochemical properties. The significance between
 34 factors at Bonferroni corrected level is indicated by * ($p < 0.00139$)

	Organic matter	Total soil phosphorus	K	Mg	Ca	pH	Cation exchange capacity	Total organic carbon	Moisture content
Organic matter	1.000								
Total soil phosphorus	0.004	1.000							
K	0.788	0.204	1.000						
Mg	-0.110	0.214	0.342	1.000					
Ca	-0.236	0.159	0.220	0.944*	1.000				
pH	-0.405	-0.144	-0.057	0.697	0.875*	1.000			
Cation exchange capacity	-0.144	0.235	0.321	0.973*	0.985*	0.784	1.000		
Total organic carbon	0.997*	0.049	0.802	-0.069	-0.199	-0.383	-0.104	1.000	
Moisture content	0.887*	-0.063	0.530	-0.260	-0.376	-0.479	-0.292	0.885*	1.000

35

36

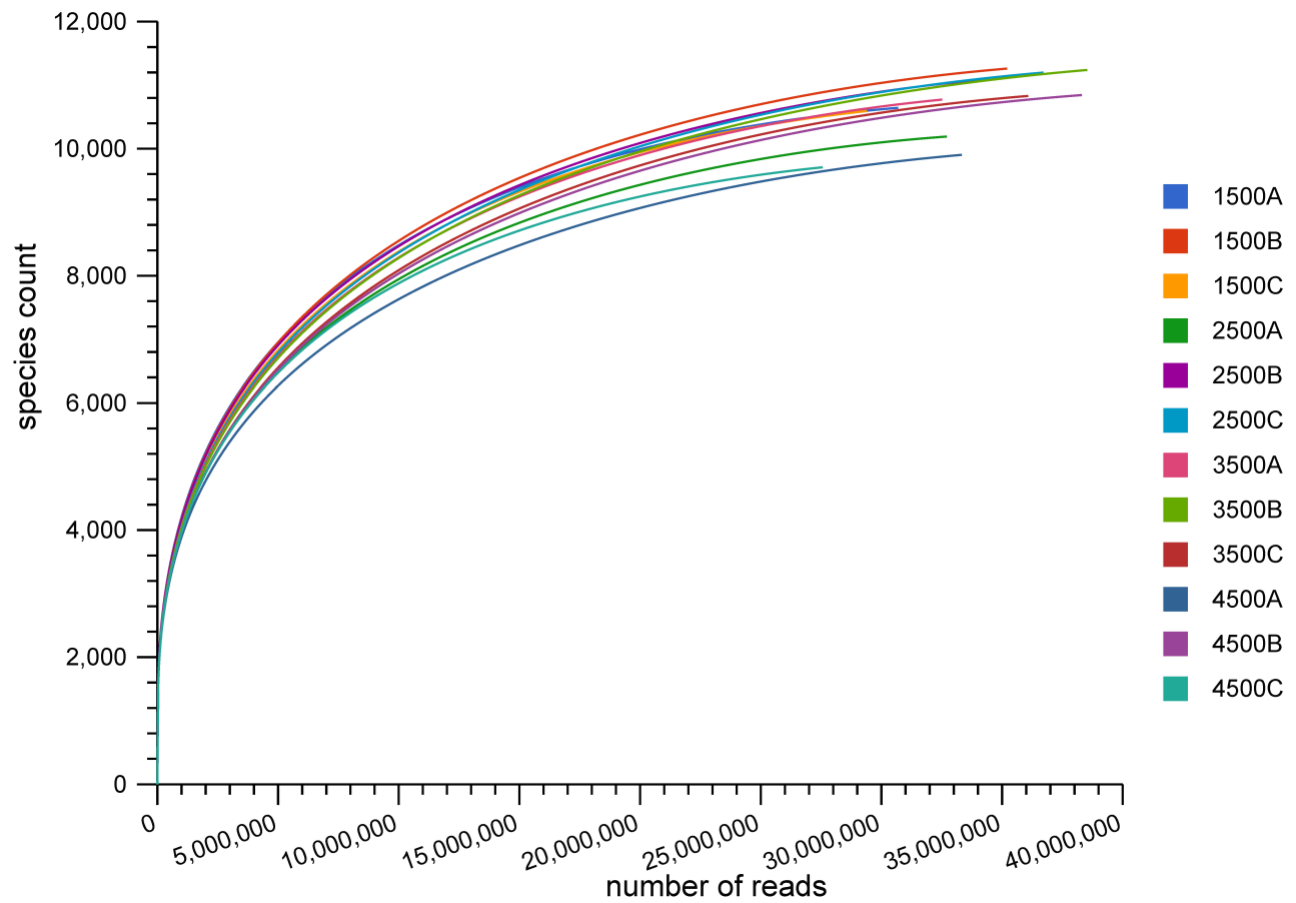


Figure S1. Rarefaction curves for metagenome sequences of bacterial communities in soil samples from the Nantahala National Forest at different altitudes. The alphanumeric labels represent the altitude and replicate of each sample, respectively. The sampling site elevations are indicated by the number: 1,500 ft (1500), 2,500 ft (2500), 3,500 ft (3500), 4,500 ft (4500). The letter following each elevation number represents parallel samples collected at the same elevation.