RESEARCH IN PROCESS

Listeria whole genome sequence data reference sets are needed to allow for improved persistence assessment and source tracking

SUMMARY

Routine use of whole genome sequencing (WGS) to "fingerprint" *Listeria monocytogenes* (*LM*) from humans and foods has considerably increased the number of disease outbreaks detected and traced back to specific foods, including produce. WGS also is used to identify instances where a specific type of bacteria appears to survive in a given food processing facility, indicating a particular food safety risk. However, our ability to interpret WGS data is hampered by (i) a lack of WGS data for bacteria from sources other than humans and foods and (ii) the need to better define how likely closely related bacteria can be found in different locations. Thus, collection of *LM* and other *Listeria* spp. from environmental sources and comprehensive genomic comparisons among these bacterial isolates along with isolates from produce associated environments and human cases will facilitate more accurate interpretation of WGS data relevant to produce related food safety issues.

OBJECTIVES

1. Develop a sampling plan for collection, across the US, of soil samples focusing on natural environments, followed by testing for *LM* and *Listeria* spp.
2. Perform WGS of the *LM* and *Listeria* spp. isolates obtained through Obj. 1, and assess associations between WGS sequence type and geographical origin.
3. Perform WGS of *Listeria* spp. isolated from throughout the produce chain; isolates will be obtained from pre-existing isolate collections, and through concurrent sampling efforts that are part of ongoing studies.
4. Perform a comprehensive analysis of *LM* and *Listeria* spp. WGS data to provide information on the number of single nucleotide polymorphisms or allelic differences that provide an appropriate cut-off to identify isolates with a likely epidemiological link.

METHODS

During the first 3 months of this project, the work focused on Objective 1, in which two main methods – (1) soil collection and (2) *Listeria* identification – were employed.

(1) Briefly, a sampling kit was distributed to collectors. Soil samples were collected at 5 sites (each >0.25 miles apart) within one sampling area. Sample sites identified by ArcGIS were between 30–150 feet from roads, or 20–150 feet from trails. At each site, 3 subsamples of topsoil were collected using a 5-oz sterile scoop, yielding 1 pooled sample.

(2) *Listeria* was enriched from soil using buffered *Listeria* enrichment broth. Sample enrichment was streaked onto *Listeria monocytogenes* plating medium agar (LMPM) and Modified Oxford agar (MOX) plates. Blue colonies and white colonies on LMPM, and black colonies on MOX, were sub-streaked onto brain heart infusion agar (BHI) plates. Presumptive *Listeria* colonies selected from BHI plates were confirmed by *sigB* sequencing.

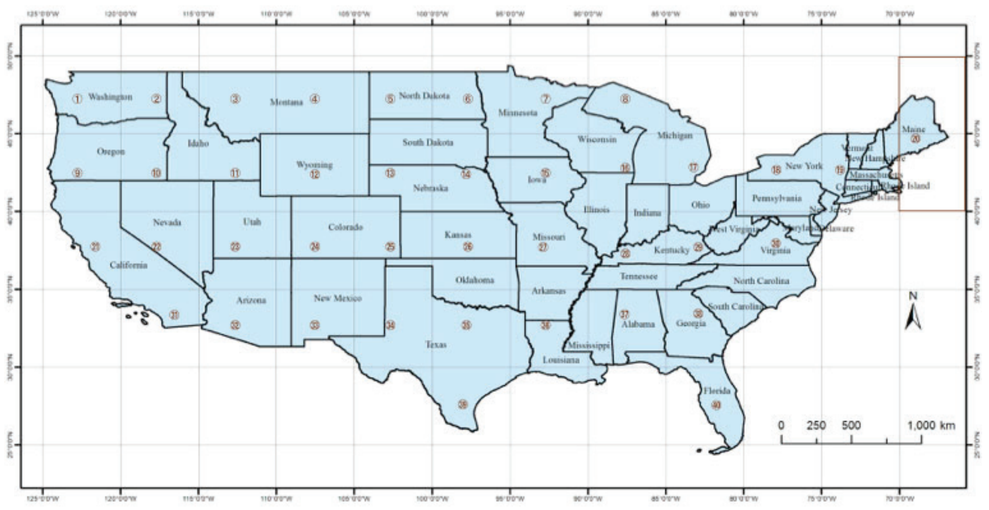

Figure 1. Proposed sampling grids (n=40) across continental US. The longitude and latitude coordinates of each intersection are provided. Dots in this map are the numbers assigned to each sampling grid, ranging from 1 to 40.

RESULTS TO DATE

The sampling plan and sampling protocol have been developed. The continental United States was divided into 40 equisized grids (Figure 1). Five sampling areas within each grid are currently being selected; all sampling areas are at least 20 miles apart and are limited to natural environments. Within each area, 5 sampling sites are randomly selected. A list of 60 volunteer collectors has been compiled and the individuals have been contacted. A total of 30 *Listeria* isolates have been obtained from soil collected at 5 sites in Great Smoky Mountains National Park. Based on the *sigB* tree (Figure 2), 6 isolates from site 1 are *LM*; 4 isolates from site 2 are *L. booriae*; 6 isolates from site 3 are *L. innocua*; 7 isolates and 1 isolate from site 4 are *LM* and *L. booriae*, respectively; 5 isolates and 1 isolate from site 5 are *L. marthii* and *L. booriae*, respectively.

BENEFITS TO THE INDUSTRY

Anticipated benefits include (i) baseline data on the frequency of *LM* and *Listeria* spp. across environmental sources in the US, (ii) data on the effects of geo-spatial, soil, and meteorological parameters on the likelihood of *LM* and *Listeria* spp. detection, (iii) data on the distribution of identical or similar *LM* and *Listeria* spp. WGS sequence types in different locations, and (iv) produce relevant data on number of SNP differences that likely indicate a recent common ancestor. These data will help industry interpret WGS data. For example, regulatory data may suggest *Listeria* persistence based on the isolation of closely related *Listeria* spp. with X SNP differences 9 months apart in a facility; data collected here would help industry to estimate (i) how likely isolation of *Listeria* with X SNPs represent persistence, and (ii) what these SNP differences mean (in terms of "These two isolates likely shared a common ancestor XX years ago").

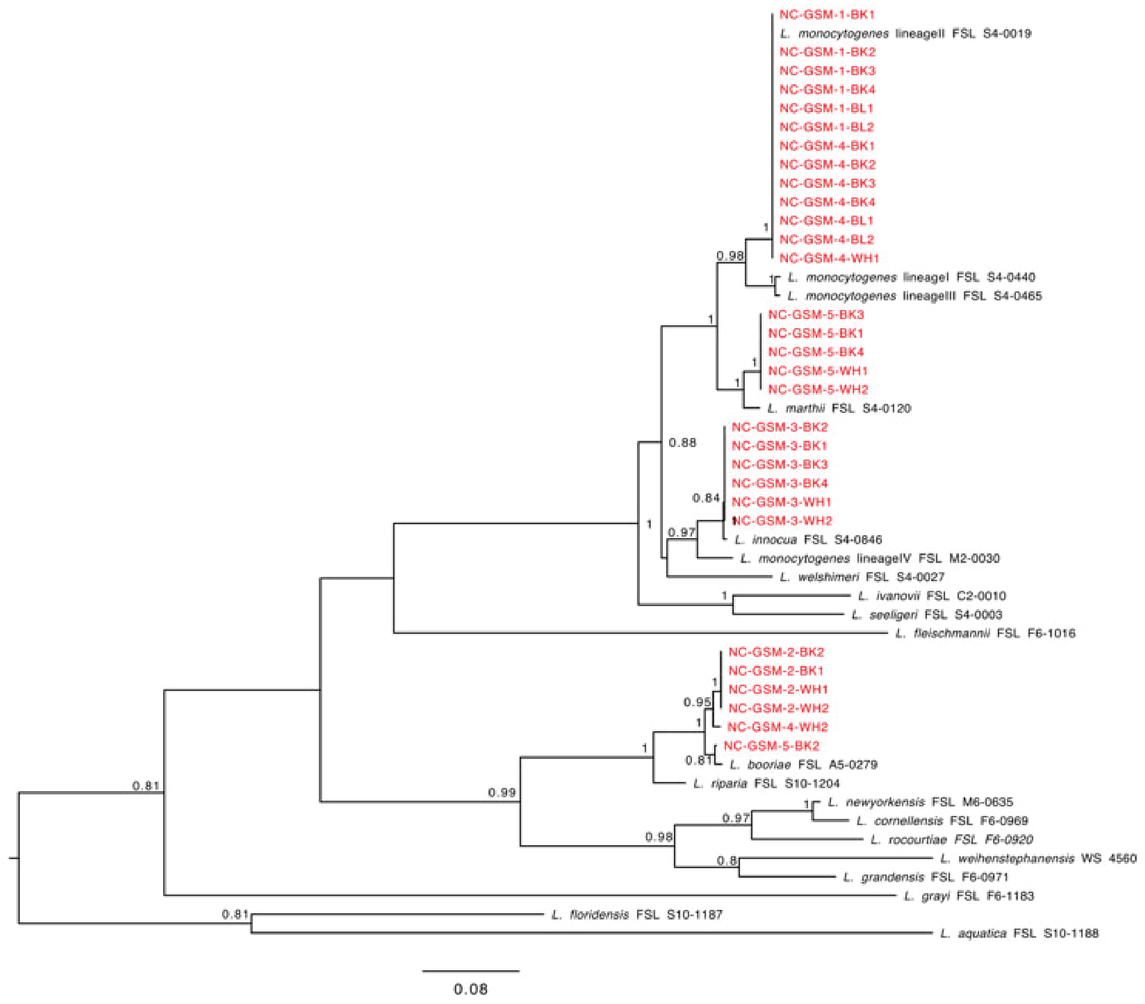

Figure 2. Phylogenetic trees inferred by maximum likelihood method using sigB of 20 Listeria allelic reference types and 30 isolates obtained through this study. The tree was rooted by midpoint. The substitution model T92+I+G was used for constructing the tree with 1000 bootstrap repetitions. Only bootstrap values >70% are presented on the tree. Isolates obtained through this study are in red.

CONTACT

Martin Wiedmann, Dr. med. vet., Ph.D.
Department of Food Science
Cornell University
E: mw16@cornell.edu
T: 607-254-2838

AUTHORS

Jingqiu Liao, Daniel Weller, and Martin Wiedmann

ACKNOWLEDGEMENTS

We are grateful to all volunteer sample collectors who have agreed to participate in this project.

LENGTH OF FUNDING

January 1, 2018 – December 31, 2019

CENTER *for* PRODUCE SAFETY