# *Listeria monocytogenes* isolates from ready to eat plant produce are diverse and have virulence potential

Alva Smith[1*], Jack Hearn[2], Clare Taylor[1], Nick Wheelhouse[1], Maciej Kaczmarek[1], Edwin Moorhouse[3], Ian Singleton[1]

*1 School of Applied Sciences, Edinburgh Napier University*

*2 Department of Vector Biology, Liverpool School of Tropical Medicine*

*3 Agri-Food Solutions, Cranbrook, UK*

*Corresponding Author – E: alva.smith@napier.ac.uk A: 6/1 Arneil Place, Edinburgh, EH5 2GP.

## Abstract

*Listeria monocytogenes* is sporadically detected on a range of ready to eat fresh produce lines, such as spinach and rocket, and is a threat to public health. However, little is known about the diversity of *L. monocytogenes* present on fresh produce and their potential pathogenicity. In this work, fifteen *Listeria monocytogenes* isolates from the UK fresh produce supply chain were characterised using whole genome sequencing (WGS). Additionally, isolates were characterised based on their ability to form biofilm. Whole genome sequencing data was used to determine the sequence type of isolates based on multi-locus sequence typing (MLST), construct a core single nucleotide polymorphism (SNP) phylogeny and determine the presence of virulence and resistance associated genes. MLST revealed 9 distinct sequence types (STs) spanning 2 lineages (I & II) with one isolate belonging to the ST6 subtype, strains from which have been recently implicated in two large, food-associated *L. monocytogenes* outbreaks in South Africa and across Europe. Although most of the 15 isolates were different, comparison of core genome SNPs showed 4 pairs of 'indistinguishable' strains (< 5 SNPs difference). Virulence profiling revealed that some isolates completely lacked the Listeria pathogenicity island-3 (LIPI-3) amongst other virulence factors. Investigation of the *inlA* gene showed that no strains in this study contained a premature stop codon (PMSC), an indicator of

attenuated virulence. Assessment of biofilm production showed that isolates found in the fresh produce supply chain differ in their ability to form biofilm. This trait is considered important for *L. monocytogenes* to persist in environments associated with food production and processing. Overall the work indicates that a genetically diverse range of *L. monocytogenes* strains is present in the UK fresh produce supply chain and the virulence profiles found suggests that at least some of the strains are capable of causing human illness. Interestingly, the presence of some genetically indistinguishable isolates within the 15 isolates examined suggests that cross-contamination in the fresh produce environment does occur. These findings have useful implications in terms of food safety and for informing microbial surveillance programmes in the UK fresh produce supply chain.

Keywords: Listeria, Contamination, Fresh Produce, Whole Genome Sequencing, Virulence, Food Microbiology

## 1. Introduction

*Listeria monocytogenes* is a Gram-positive, facultative anaerobic bacterial pathogen that is ubiquitous in the environment. It is the causative agent of listeriosis, an intracellular disease which predominantly affects the elderly, immunosuppressed and pregnant women along with their unborn or new-born babies. Although incidence of the disease is low compared to other foodborne pathogens (Hernandez-Milian and Payeras-Cifre, 2014), the disease outcome is often more serious, making it a priority pathogen in many countries. Foods which have been previously implicated in *L. monocytogenes* outbreaks include milk, soft cheeses, delicatessen meats and fresh produce (both fruit and vegetables). Fresh fruit and vegetables account for a significant proportion of the UK market and consumer attitudes are increasingly leaning towards healthy, convenient options such as ready to eat bagged salads and prepared fruit and vegetables (Keynote, 2015). The UK fresh produce supply chain is complex and there are many potential sources of *L. monocytogenes* contamination. These include; soil splash, contaminated irrigation water, wild animal faecal contamination and cross contamination from surfaces and personnel in the processing environment (Smith et al., 2018).

Microbial testing for *L. monocytogenes* and other microbial pathogens in the supply chain of a variety of products, such as spinach, rocket and kale, is obligatory through regulation. According to current practice, discovery of *L. monocytogenes* in the supply chain results in increased sampling and a review of the risk assessments to determine possible sources of the bacterium and demonstrate that there is no on-going risk to other consignments (Monaghan et al., 2009).

The advent of sub-typing techniques, such as pulse-gel field electrophoresis (PGFE) and whole genome sequencing (WGS), has enabled source tracking of *L. monocytogenes* during outbreak investigations, but these technologies are not yet used for general surveillance in food supply chains because of their cost, complexity of analysis and the expertise required to interpret such data. Investigating bodies employ these technologies to identify the source of a *L. monocytogenes* outbreak through identifying related cases of listeriosis, identifying the outbreak strain/strains in one or more common food items, and finally (through targeted sampling and traceability of foodstuffs in the supply chain) identifying the offending stage/location in a supply chain (Pouillot et al., 2016). Once the source of the bacteria has been identified, regulatory authorities can take necessary steps to eradicate the source and prevent further contamination. In addition to source tracking, subtyping of *L. monocytogenes* strains using WGS can reveal pathogenic profiles and infer relative risk to the consumer.

Determining the virulence potential of isolates is important in terms of public health as differences in virulence between *L. monocytogenes* strains may influence infection and clinical outcome. It is worth nothing that all strains of *L. monocytogenes* are currently treated equally for regulatory purposes (Fravalo et al., 2017), however, some strains are highly pathogenic and are more often associated with epidemics of disease, whereas others can be less virulent and are rarely associated with epidemics (Velge and Roche, 2010). For example, the majority of clinical *L. monocytogenes* strains fall into phylogenetic lineage I, and hypervirulent strains belonging to sequence type-6 (ST6) have been implicated recently in two large outbreaks of foodborne listeriosis which caused

widespread illness and mortality (European Food Safety Authority, 2018; National Listeria Incident Management Team, 2018). Further, it is widely acknowledged that premature stop codons (PMSCs) in the *inlA* gene which result in a truncated and expressed internalin-A and attenuated virulence exist within the species (Gelbíčová et al., 2015). Observation of this characteristic in strains isolated from food and environmental sources compared to clinical sources has been recognised (Fravalo et al., 2017). However, a recent study from Ireland demonstrated that five out of the six *L. monocytogenes* strains isolated from food did not contain PMSCs in *inlA* (Hilliard et al., 2018), showing that this characteristic from environmental strains requires further investigation. In addition to virulence potential, it is important to determine the ability of *L. monocytogenes* isolates to form biofilm, as this trait is potentially important for survival in many parts of the fresh produce supply chain including the processing environment, where biofilms can be formed on many different surfaces and serve as a subsequent source of contamination (Colagiorgi et al., 2017; Smith et al., 2018).

To determine the relatedness of strains and establish a virulence profile, this study characterised the subtype of various *L. monocytogenes* isolates found in the UK fresh produce supply chain. This was based on their sequence type and the development of a phylogenetic framework using core genome SNPs. In addition, genomes were examined for the presence of resistance genes, virulence-associated genes and evidence for PMSCs in *inlA*. Furthermore, we assessed the ability of isolates to form biofilm.

## 2. Materials and methods

### 2.1. Bacterial isolates

In total, 15 *L. monocytogenes* strains were sent to Edinburgh Napier University (ENU) from a commercial food testing laboratory. These strains were isolated from various parts of the UK fresh produce supply chain (see Table 1. for details) as part of routine food testing procedures carried out by food production and processing companies between May 2016 and April 2017. *L. monocytogenes*

strains were either isolated directly from product (in the case of the environmental strain, from a swab) using ISO 11290-2: 2017. Bacterial species identification was carried out by a commercial food microbial testing lab and confirmed using biochemical tests (API Listeria, bioMerieux/Microbact Listeria, Thermo Scientific) or using MALDI-TOF technology (Vitek MS, bioMerieux). *L. monocytogenes* was streaked on a nutrient agar slope and sent to Edinburgh Napier University, Sighthill Campus before being sub cultured, twice from a single colony at 37°C for 24h, on OXFORD agar with modified *Listeria* selective supplement (Oxoid). To make long term stocks, a single colony was used to inoculate 10ml BHI broth (Oxoid) and shaking cultures were incubated at 200rpm, 37°C overnight (~16h) before 1ml of this culture was centrifuged at 12,000rpm for 10 min. Resulting bacterial pellets were subsequently washed 3 times in PBS and resuspended in 500µl fresh BHI broth + 500µl of 50% glycerol. Long term stocks were frozen at -80°C until preparation for whole genome sequencing.

*Table 1.* **Characteristics, sample information and sequence type of 15 *L. monocytogenes* isolates from the UK fresh produce supply chain**

| Internal Reference | Sample type | Sample date | Sample location | Stage in supply chain | Lineage | Sequence Type |
|---|---|---|---|---|---|---|
| **NLmo2** | Spinach | May-16 | West Sussex, UK | Raw Product, Unwashed | I | ST-5 |
| **NLmo3** | Spinach | May-16 | West Sussex, UK | Raw Product, Unwashed | I | ST-5 |
| **NLmo4** | Environmental swab (drain) | May-16 | West Sussex, UK | Tray cleaning facility | II | ST-325 |
| **NLmo5** | Spinach | May-16 | West Sussex, UK | Raw Product, Unwashed | II | ST-325 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **NLmo6** | Red leaf lettuce | Jun-16 | Norfolk, UK | Raw Product, Unwashed | I | ST-4 |
| **NLmo7** | Spinach | Jun-16 | West Sussex, UK | Post Cooling, Unwashed | I | ST-1 |
| **NLmo8** | Spinach | Aug-16 | Cambridgeshire, UK | Final Product, Unwashed | I | ST-219 |
| **NLmo9** | Spinach | Jul-16 | Cambridgeshire, UK | Final Product, Unwashed | I | ST-4 |
| **NLmo10** | Spinach | Sep-16 | Cambridgeshire, UK | Post Cooling, Unwashed | I | ST-1 |
| **NLmo13** | Spinach | Sep-16 | Cambridgeshire, UK | Final Product, Unwashed | II | ST-37 |
| **NLmo14** | Beetroot | Oct-16 | Cambridgeshire, UK | Final Product, Washed | I | ST-1 |
| **NLmo15** | Peashoots | Oct-16 | Cambridgeshire, UK | Final Product, Unwashed | II | ST-204 |
| **NLmo16** | Spinach | Nov-16 | Cambridgeshire, UK | Final Product, Unwashed | II | ST-37 |
| **NLmo18** | Baby salad kale | Apr-17 | Cambridgeshire, UK | Post Cooling, Unwashed | II | ST-399 |
| **NLmo20** | Baby salad kale | Apr-17 | Cambridgeshire, UK | Final Product, Unwashed | I | ST-6 |

114

## 2.2. DNA extraction and whole genome sequencing

From long term stocks, a lawn of bacteria was grown on BHI agar (Oxoid) before being transferred to a barcoded cryovial (MicrobesNG) containing beads and preparation broth, then briefly vortexed before being sent to MicrobesNG (University of Birmingham, UK) for DNA extraction and whole genome sequencing. For DNA extraction, three beads were washed with extraction buffer containing lysozyme and RNase A, incubated for 25 min at 37°C. Proteinase K and RNaseA were added and incubated for 5 min at 65°C. Genomic DNA was purified using an equal volume of SPRI beads and resuspended in EB buffer. DNA was quantified in triplicates with the Quantit dsDNA HS assay in an Eppendorf AF2200 plate reader. Genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: two nanograms of DNA instead of one were used as input, and PCR elongation time increased to 1 minute from 30 seconds. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Pooled libraries were quantified using the Kapa Biosystems Library Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. Libraries were sequenced on the Illumina HiSeq using a 250bp paired end protocol, with a target 30-fold depth of coverage.

## 2.3. Bioinformatics methods

### 2.3.1 *De novo* assembly, species identification, multi-locus sequence typing, virulome and resistome

Illumina reads were adapter trimmed using Trimmomatic 0.30 with a sliding window quality cut-off of Q15 (Bolger et al., 2014). Trimmed sequencing reads were then processed using a customised, open source bioinformatics pipeline for the handling of sequence data (https://github.com/tseemann/nullarbor). Within the pipeline, *de novo* assembly was performed on samples using SKESA version 2.1 (Souvorov et al., 2018) with the default parameters. Genome annotation was carried out by Prokka (Seemann, 2014). Species identification was carried out by k-

mer analysis against a known database (MiniKraken 8GB). Assembled genomes were then scanned for sequence type using MLST version 2.11 (Seemann https://github.com/tseemann/mlst). Virulence and resistance genes were detected by Abricate, version 0.8 (Seemann https://github.com/tseemann/abricate), which uses BLAST+ & EMBOSS to screen contigs against databases of known sequences of virulence and resistance genes. Virulence and resistance genes were detected by comparison to the Database for Virulence Factors of Pathogenic Bacteria (VFDB) (L. Chen et al., 2016) and Resfinder (Zankari et al., 2012) databases, respectively. Virulence and resistance genes were considered present when coverage ≥ 95% and identity > 75%, probable when coverage ≥ 36.4% and identity > 75% and missing when undetected. We corroborated Abricate results by manually inspecting Abricate output tables where % coverage, % identity, gene name, accession number and position in a contig were reported. When genes were reported as partial or in two parts, genome annotations (generated by Prokka) were manually inspected, a complete ORF was identified and the annotated nucleotide sequence of the gene was copied and subsequently used in BLAST (Altschul et al., 1990) to ascertain homology with known sequences. When the full-length gene returned coverage ≥ 95% and identity > 75% to *L. monocytogenes*, genes were considered present. If the annotated gene was not found in the genome annotation, as was the case with *ami*, gene nucleotide sequence was copied from the VDFB *Listeria* database and used to BLAST against the contig where it was identified by Abricate. *Ami* was considered present when BLAST results returned ≥ 95% and identity > 75%. Occasionally, virulence genes fell between contigs. In these cases, gene presence was considered probable due to the high % identity of these sequences but low % coverage.

#### *2.3.1.1 Determination of PMSCs in* inlA

To determine whether strains contained a PMSC in the *inlA* gene, Abricate output tables were used to locate the position of full length *inlA* in (2403bp) in SnapGene Viewer. This sequence was then uploaded to MEGAX for all strains and sequences were subsequently aligned to *L. monocytogenes*

EGD-e *inlA* reference using the MUSCLE algorithm. Alignment was then manually inspected for PMSCs in *inlA* based on those which have been previously reported (Gelbíčová et al., 2015).

### 2.3.2. Global core genome alignment and construction of Maximum Likelihood phylogeny based on core genome SNPs

To obtain a global alignment of all isolates included in this study, per sample sequence reads were mapped to *L. monocytogenes* EGD-e (NCBI: AL591824.1, 2,9445,28 bp), a ST35, lineage II, *L. monocytogenes* reference genome. For SNP based analysis, read mapping, and core genome alignment were performed using the Snippy pipeline, version 4.0 (Seemann https://github.com/tseemann/snippy). Sequence reads were aligned to the reference genome and sites that were covered by less than 10 reads were not included in analysis. Any site where at least one of the isolates had a SNP and none of the isolates were absent was considered a core-SNP site. Core-SNPs were used as output to determine the phylogeny of the *L. monocytogenes* population which was inferred by IQ-TREE version 1.6.7 (Nguyen et al., 2015) using the Maximum Likelihood method with model finder (Kalyaanamoorthy et al., 2017) option enabled. Branch support was calculated using ultrafast bootstrap support (Hoang et al., 2017) and the SH-like approximate likelihood ratio test (Guindon et al., 2010), both with 1000 iterations. One *L. ivanovii* isolate was isolated from the fresh produce supply chain and included in sequencing, *de novo* assembly and in core genome SNP analysis. The phylogenetic tree was rooted using *L. ivanovii* as an outgroup. The tree was then modelled and annotated in Figtree (http://tree.bio.ed.ac.uk/software/figtree/).

### 2.3.3. Determination of genetic relatedness between strains

To establish relatedness between *L. monocytogenes* strains from the fresh produce supply chain we considered tree topology, MLST and the pairwise SNP distance matrix computed from the core genome alignment to the reference. When determining SNP based phylogenies and examining bacterial outbreaks, different SNP-based subtyping workflows result in variation in the number of core SNPs predicted within the same data set which means that it is not possible to define a

universal single cut-off value for delineation of outbreaks or in this case, indistinguishable strains (Saltykova et al., 2018). Therefore, for each workflow this threshold should be estimated separately. Whilst different WGS analytical tools, SNP calling algorithms and reference genomes (Kwong et al., 2016) mean that SNP data is not directly comparable between studies, previous investigations on outbreaks have defined 'outbreak strains' that differ by 5, 10, 20 and 28 SNPs (Chen et al., 2017). Additionally, 'plant (i.e. facility) associated' clones have been found to differ by up to 6 SNPs (Morganti et al., 2015). Thus, the current study decided that isolates were 'indistinguishable' i.e. the same strain, when they differed by ≤ 5 SNPs.

## 2.4. Biofilm formation assay

Long term stocks of *L. monocytogenes* were streaked on OXFORD agar plates with modified *Listeria* selective supplement and incubated at 37°C for 24h. 10ml of tryptone soy broth (TSB, Oxoid) was inoculated from 1 isolated colony then incubated overnight in a shaking incubator at 200rpm, 37°C for 16h. After incubation, absorbance was measured at 595nm before back diluting to an OD of 0.05 in 10ml of modified Welshimer's broth (MWB, HiMedia) followed by brief vortexing. 100µl of this diluted culture was then transferred to 6 x wells of a sterile, flat bottomed polystyrene 96-well plate per strain (6 biological replicates from 1 overnight culture). 6 uninoculated MWB wells were included in the plate as a control. Plates were incubated at 20°C for 20h before the turbidity of wells was determined at 595nm using a microtiter plate reader (FLUOstar Omega). MWB was chosen as a medium for biofilm production as this media is a minimal medium and is more likely to be representative of the fresh produce supply chain environment i.e. a more nutrient limited environment, than other, more rich defined media such as brain heart infusion broth.

After measurement of cell turbidity, medium was aspirated from wells using a pipette before wells were washed 3 times with distilled water to remove loosely associated bacteria. Plates were air dried for 45 minutes before staining with 0.1% crystal violet solution in water for 45 minutes. After staining wells were washed again 3 times with distilled water. After washing, plates were left to dry

for a further 45 minutes. For quantitative analysis of biofilm production, 200µl of 95% ethanol was added to de-stain wells and left at room temperature on a shaking platform for 45 minutes. 100µl of this solution was then transferred to a fresh 96-well microtiter plate before the optical density of the de-staining solution was measured on a microtiter plate reader at 595nm. Results were blank corrected by subtracting the average absorbance from control wells (media only) from the average absorbance of test wells. To ensure that differences in biofilm production were not related to differences in growth rate, average absorbance measurements of cell turbidity versus average absorbance measurements from biofilm CV de-stain underwent a Pearson's test for correlation (GraphPad). Differences in biofilm production between strains (n=6) were determined using a one-way ANOVA followed by Tukey's test for multiple comparisons (Graphpad).

# 3. Results

## 3.1. Multi locus sequence typing and SNP based phylogeny using whole genome sequencing

WGS generated various numbers of reads, between 488,048 & 4,673,156. The genomes had average depths of coverage between 31 to 308-fold of the reference. *De novo* assembly resulted in genome assemblies of 9 to 21 contigs and the consensus lengths of genomes ranged from 2.87Mb to 3.06Mb. GC content of isolates ranged from 37.4 - 38.9%. All isolates were identified as *L. monocytogenes* and had 88.87-94.64% similarity to *L. monocytogenes* using the Kraken 8GB database. For isolate specific values of data summarised above, see Appendix A.1. Using data from WGS, MLST determined 9 distinct sequence types from 15 isolates (Table 2). Some isolates shared the same sequence type with the most frequent being ST1 (3/15). No new sequence types were discovered. All isolates belonged to either phylogenetic lineage I or II. Several sets of strains could not be distinguished by MLST analysis; these being NLmo10, NLmo14 & NLmo7 (ST1), NLmo6 & NLmo9 (ST4), NLmo2 & NLmo3 (ST5), NLmo13 & NLmo16 (ST37) and NLmo4 & NLmo5 (ST325).

*Table 2*. **Allelic profile and number of isolates represented by each sequence type of *L. monocytogenes* isolated from the fresh produce supply chain.** The MLST scheme used to

characterise *L. monocytogenes* isolates is determined by the allelic profile of 7 housekeeping genes, these are; ABC transporter (*abcZ*), beta-glucosidase (*blgA)*, catalase (*cat)*, succinyl diaminopimelate (*dapE*), D-amino acid aminotransferase (*dat*), L-lactate dehydrogenase (*ldh*) and histidine kinase (*lhkA*). Clonal complex and lineage information is included.

| Sequence type | Clonal complex | Lineage | Number of isolates |
|---|---|---|---|
| **1** | CC1 | I | 3/15 |
| **4** | CC4 | I | 2/15 |
| **5** | CC5 | I | 2/15 |
| **6** | CC6 | I | 1/15 |
| **37** | CC37 | II | 2/15 |
| **204** | CC204 | II | 1/15 |
| **219** | CC4 | I | 1/15 |
| **325** | CC31 | II | 2/15 |
| **399** | CC14 | II | 1/15 |

A phylogeny of the 15 *L. monocytogenes* strains was obtained using core genome SNPs and showed that when compared, isolates from the fresh produce supply chain differ from 0 up to 40,143 core genome SNPs. SNP analysis allowed differences between isolates to be established in greater detail than MLST. For example, some isolates from the same MLST groups had no SNP differences while others were different by up to 59 SNPs (Appendix A.2). SNP analysis highlighted 4 sets of 'indistinguishable' strains, these were; NLmo2 & NLmo3, NLmo4 & NLmo5, NLmo7 & NLmo14 and NLmo13 & NLmo16.

## 3.2. Prevalence of virulence- and resistance-associated genes

The 15 *L. monocytogenes* isolate genomes were analysed for the presence or absence of 42 key virulence factors (genes) by Abricate version 0.8 using the VFDB database (L. Chen et al., 2016) and

corroborated by manual inspection of genome annotations combined with BLAST of virulence gene sequences against contigs. The presence or absence of the full range of virulence factors can be seen in Appendix A.3. 7 out of the 9 isolates in lineage I had 41 virulence factors present whilst 2 out of 9 (NLmo2 & NLmo3) had 34 virulence factors. Of the 6 strains that were in lineage II, 1 isolate (NLmo18) had 33 virulence factors, 3 isolates (NLmo13, NLmo15, NLmo16) had 32 virulence factors and 2 isolates (NLmo4 & NLmo5) had 31 virulence factors present. Isolates which had ≤ 41 virulence factors present were all missing the LIPI-3 gene cluster (*llsY, llsX, llsP, llsH, llsG, llsD, llsB, llsA*) which encodes to produce Listeriolysin S. Except from NLmo2 & NLmo3, isolates from lineage I were missing the *ami* gene which codes for an autolysin amidase protein. All isolates contained an intact and full length *inlA* gene, apart from NLmo20 which had a 9-nucleotide deletion in position 2212-2220bp (797 aa's), this version of Interalin A is predicted to be fully functional and isolates with this variant show similar invasion ability compared with strains full length Internalin A (Toledo et al., 2018).

Genomes were also analysed for the presence of resistance genes by scanning contigs using Abricate against the ResFinder database. All isolates carried the Fosfomycin resistance thiol transferase (*fosX)* and lincomycin resistance ABC-F type ribosomal protection protein (*lin)* genes whilst 2 isolates (NLmo4 & NLmo5) carried 2 additional, plasmid derived, resistance genes for a quaternary ammonium compound efflux transporter (*bcrB & bcrC*). Figure 1. shows the relationship between *L. monocytogenes* strains and the presence of virulence- and resistance-associated genes.

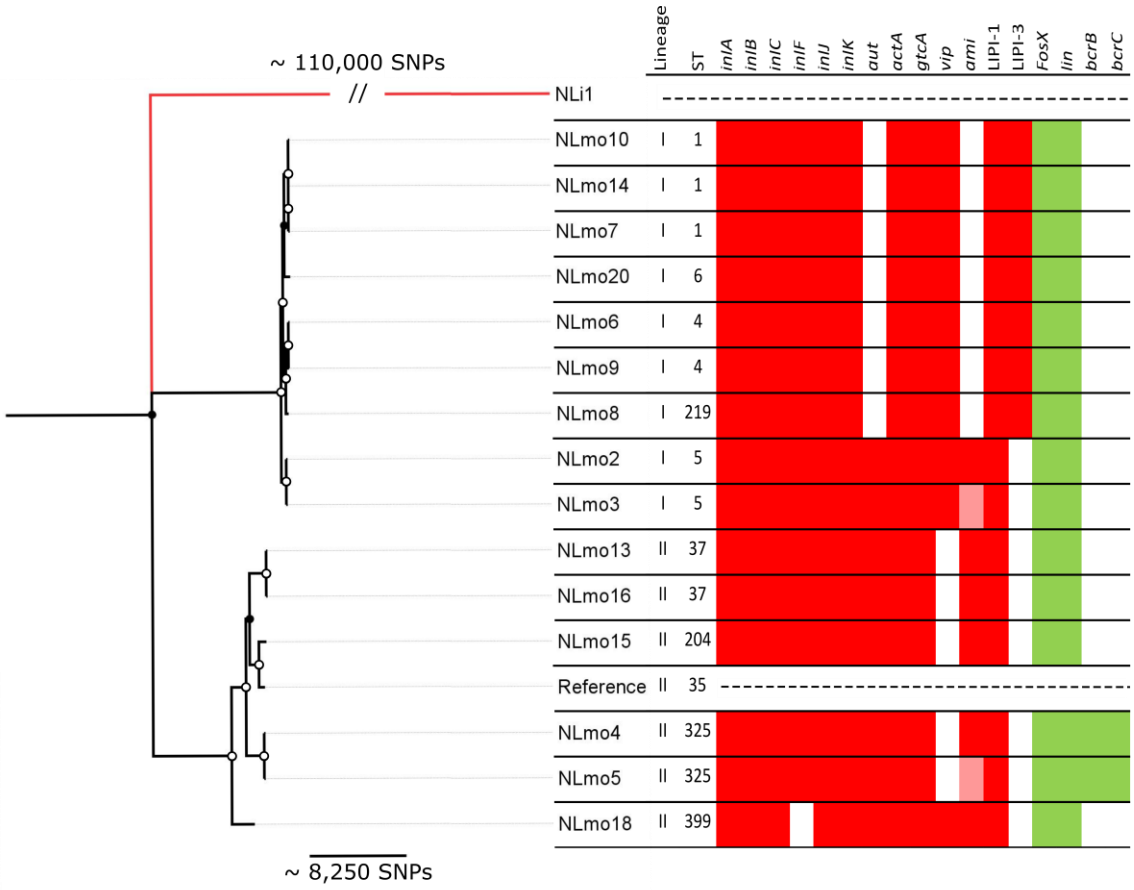



*Figure 1.* **Maximum Likelihood phylogeny of 15 *L. monocytogenes* isolates and 1 *L. ivanovii* isolate from the UK fresh produce supply chain in relation to the presence of virulence and resistance genes.** White circles indicate nodes with ≥ 95% ultrafast bootstrap support and ≥ 80% SH-like approximate likelihood ratio test support. The break (//) in the root branch represents a comparatively long evolutionary distance to the outgroup, *Listeria ivanovii* (NLi1), which is highlighted in red. Approximate branch distance (in SNPs) of this branch is indicated above the break. Columns right of the tree indicate presence (red) probable presence (pink) or absence (white) of *L. monocytogenes* virulence factors and presence (green) or absence (white) of resistance genes (see Appendices A.3 & A.4 for full lists of genes). Evolutionary distances were computed using the Maximum likelihood method and are in units of SNPs. The analyses involved 17 nucleotide sequences. Evolutionary analyses were conducted using IQTree version 1.6.7. tree was generated using FigTree. Dashed lines indicate strains for which no data on the presence of virulence and resistance genes was collected. ST = MLST sequence type.

### 3.3. Biofilm production of *L. monocytogenes* isolates

A Pearson test for correlation between the average values of culture turbidity and biofilm production quantified by CV staining indicated that no correlation existed between these two variables ($R^2 = 0.10$, $P = 0.24$). A one-way ANOVA with Tukeys test for multiple comparisons showed that biofilm production differs significantly ($P < 0.0001$) between strains isolated from the fresh produce supply chain. Figure 2. shows that some isolates produced a comparatively small (Average $OD_{595} \leq 0.053$) amount of biofilm (NLmo4, NLmo5, NLmo7 & NLmo10), whilst others produced a moderate ($0.053 >$ Average $OD_{595} \leq 0.152$) amount of biofilm (NLmo2, NLmo3, NLmo6, NLmo13, NLmo14, NLmo15, NLmo16, NLmo20) and three isolates produced a comparatively high (Average $OD_{595} > 0.152$) amount of biofilm (NLmo8, NLmo9, NLmo18). Biofilm production was inconsistent between pairs of genetically indistinguishable strains.

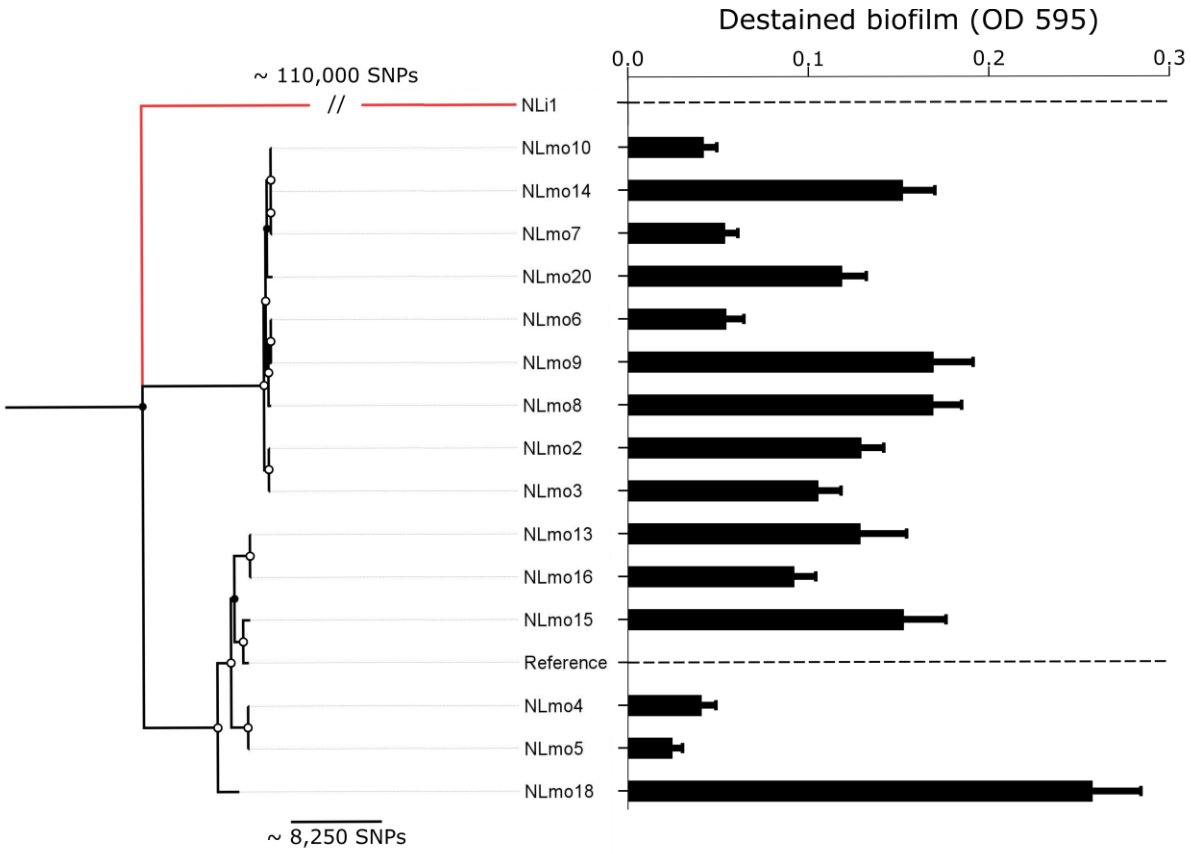



*Figure 2*. **Maximum Likelihood phylogeny of 15 *L. monocytogenes* isolates and 1 *L. ivanovii* isolate from the UK fresh produce supply chain trees inferred using a core-SNP based phylogeny in**

**relation to biofilm production.** Biofilm production assessed by staining with crystal violet, destaining and measuring absorbance at 595nm. White circles indicate nodes with ≥ 95% ultrafast bootstrap support and ≥ 80% SH-like approximate likelihood ratio test support. The break (//) in the root branch represents a comparatively long evolutionary distance to the outgroup, *Listeria ivanovii* (NLi1), which is highlighted in red. The approximate branch distance (in SNPs) of this branch is indicated above the break. Dashed lines indicate strains for which no data on biofilm production was collected. Error bars represent SEM of 6 replicates.

## 4. Discussion

*L. monocytogenes* remains an important foodborne pathogen and is a significant threat to public health in the food supply chain as illustrated by recent outbreaks (European Food Safety Authority, 2018; National Listeria Incident Management Team, 2018). *L. monocytogenes* has been previously identified in UK foodstuffs (Little et al., 2009), but the virulence and relatedness of strains actually found in UK foods has not been characterised. This information is important for inferring the potential risk that isolates pose to consumers and in determining the source of isolates/highlighting persistent strains. Furthermore, the ability of these isolates to form biofilm has not been reported. Understanding these aspects of *L. monocytogenes* ecology in food supply chains can help regulators and operators to design more effective microbial surveillance and prevention strategies. The *L. monocytogenes* isolates from this study were obtained through standard sampling of produce and surfaces in the UK fresh produce supply chain. While it is difficult to draw general conclusions about the whole UK fresh produce supply chain due to the limited number of strains used in this study, our data provide a preliminary insight into the diversity of the *L. monocytogenes* population and virulence potential of *L. monocytogenes* directly isolated from the UK fresh produce supply chain.

MLST of isolates revealed a range of sequence types belonging to lineages I & II (Table 2). MLST is a useful tool for characterising the subtype of *L. monocytogenes* isolates and inferring virulence potential but it lacks the discriminatory power to distinguish amongst closely related strains of

the bacterium which is essential for source tracking in clinical and food environments (Lomonaco and Nucera, 2012). This observation is corroborated with these results which also demonstrate that phylogenies constructed using concatenated MLST sequences result in a tree with comparably low confidence and low discriminatory power (Appendix B). Thus, for inter-strain comparison, deeper resolution is achieved with SNP based phylogenies versus those constructed using MLST.

In this study, fifteen *L. monocytogenes* strains spanning two lineages were isolated from different products and associated environments in the fresh produce supply chain. According to the presence of SNPs, this study identified isolates that were defined as indistinguishable (<5 SNPs), which also fell into the same sequence type and identified isolates that fell into the same sequence type but differed by up to 59 SNPs. With regards to specific cases, indistinguishable strains NLmo2 & NLmo3 were isolated from raw product from the same location (single farm) a week apart. Likewise, isolates NLmo13 & NLmo16 were isolated from raw product in a different common location (holding/packing facility) around 3 weeks apart. These findings could suggest local cross contamination within a given space in the supply chain. In contrast, indistinguishable isolates NLmo7 & NLmo14 came from different locations in the supply chain where product was sampled months apart and isolates NLmo4 & NLmo5 were isolated within a week of each other but were produce was sampled at different points in the supply chain. The complexity of the fresh produce supply chain makes it difficult to draw conclusions about these indistinguishable isolates, for example, whether cross contamination has occurred (and in which direction) or whether strains were spread to these locations from a separate, but common source.

While it can be inferred that indistinguishable strains in the supply chain either came from a common source or were the result of cross contamination, it is impossible to confidently elucidate either scenario without the presence of a robust, targeted sampling plan which characterises *L. monocytogenes* using WGS. As well as a robust sampling plan that uses WGS as a subtyping method,

352 investigations of this kind require strong context knowledge and epidemiological data to elucidate
353 contamination mechanisms (Stasiewicz et al., 2015). If comparisons of strains based on the core
354 genome do not provide sufficient resolution for discrimination of strains, the accessory genome of *L.*
355 *monocytogenes* has been shown to be highly variable between closely related isolates (Casey et al.,
356 2016). Inclusion of the accessory genome during analysis has been used previously to give sufficient
357 discriminatory resolution between isolates for determination of outbreak strains (Y. Chen et al.,
358 2016).

359 Routine commercial sampling of fresh produce and surrounding environments is often infrequent
360 and source tracking in a processing environment remains difficult due to the risk of recontamination,
361 that is, if an indistinguishable strain is found in a processing environment and on a raw material it
362 does not prove that the contamination came from the raw material or vice versa. The current study
363 demonstrates the power of WGS technology in terms of establishing the virulence potential of
364 isolates and discriminating between closely related strains of *L. monocytogenes* and should serve
365 towards designing and implementing a more robust sampling plan for detecting *L. monocytogenes* in
366 the UK fresh produce supply chain. To elucidate a specific source of *L. monocytogenes* in this
367 context, stakeholders would have to embark on a regular sampling regime of both the processing
368 environment and raw (incoming) and processed (outgoing) product combined with WGS and SNP-
369 based analysis. This approach may be able to differentiate sporadic vs. persistent strains and
370 highlight modes of *L. monocytogenes* transmission from the growth environment to inside the
371 processing environment. Within the processing environment, targeted sampling of product contact
372 sites as well as potential harbourage sites may reveal sources of the bacteria. Sampling of processing
373 environments is particularly important as these facilities are 'bottle neck' spaces in food supply
374 chains which encounter all processed produce and thus are potential sites of cross-contamination.
375 Sampling of processed (outgoing) produce is equally important as indistinguishable strains isolated
376 over time, from produce of different origins (growth environments), but passing through the same
377 processing environment would be indicative of cross-contamination and persistence in the

processing environment. Overall, more regular sampling and genome-wide strain characterisation is needed but may currently be beyond the scope of the fresh produce supply chain in terms of time and financial investment. For food production and processing companies, where the financial cost and negative reputational impact of a listeriosis outbreak is potentially large (McCollum et al., 2013), this approach should be more attractive because of the additional information it provides over current standard commercial characterisation techniques which only identify bacteria down to the species level.

Whole genome sequences of isolates revealed that all strains contained the *hly* (listeriolsyin O) gene and all strains had the internalin family of genes present (*inlA, inlB, inlC, inlF, inlJ, inlK*) except strain NLmo18, where *inlF* was missing. This gene codes for a protein that mediates invasion of the brain of the host by binding with vimentin (Ghosh et al., 2018) and suggest that NLmo18 has reduced virulence in the host based on the lack of *inlF*. A subset of isolates from lineage I (47% of total isolates) contained the *Listeria* pathogenicity island LIPI-3, a virulence factor which has been implicated in severe disease (Kim et al., 2018). LIPI-3 is confined to lineage I strains of *L. monocytogenes* and genes from this pathogenicity island encode Listeriolysin S (LLS), a protein which has been shown to display bactericidal activity and has the ability to modify host gut microbiota in mouse models (Quereda et al., 2017). This protein plays a crucial role in the infection cycle of *L. monocytogenes* and is present in epidemic strains. Importantly, whilst other authors have found that some strains isolated from food and environmental samples express a truncated Internalin-A due to PMSCs in *inlA* (Nightingale et al., 2005) all strains in this study, except NLmo20, contained a full and intact *inlA*. Since the short *inlA* variant that NLmo20 contains is predicted to be fully functional, potential virulence attenuation due to truncated Internalin A (Fravalo et al., 2017) can be ruled out for the strains featured in this study. Whilst all strains had two resistance genes present, NLmo4 and NLmo5 had two additional resistance genes, *bcrB & bcrC*, which encode a quaternary ammonium compound efflux SMR transporter. The presence of the virulence factors discussed above in strains isolated from the fresh produce

supply chain coupled with cross referencing of their subtype indicates that they have the necessary genomic prerequisites to cause disease and have many features in common with strains that have previously caused outbreaks of disease. This information is of clear interest to regulators and stakeholders in the UK fresh produce supply chain.

*Listeria monocytogenes* biofilms can be formed on many different surfaces in food processing operations and provide a protective environment for bacterial survival thereby increasing the risk of subsequent contamination (Colagiorgi et al. 2017). While all *L. monocytogenes* in this study formed biofilm, some formed significantly more biofilm than others on polystyrene 96-well plates when grown in MWB, a minimally defined media. This result suggests that some strains may have a competitive advantage over others in the fresh produce supply chain based on their ability to form biofilm, but previous research has generated mixed results when determining whether strong biofilm formation is an indicator of persistence in processing environments (Magalhães et al., 2017; Nowak et al., 2017). Furthermore, it is recognised that many environmental factors contribute to biofilm production in *L. monocytogenes* including substrate, surface material and temperature. In keeping with previous research, biofilm formation of *L. monocytogenes* showed strong strain to strain variation (Colagiorgi et al., 2017). In this study, isolates which were determined to be the same strain (by SNP analysis) did not necessarily produce equal amounts of biofilm, which could suggest that the method for quantifying biofilm in this study was not optimal. On the other hand, changes in biofilm production within *L. monocytogenes* strains may well be affected by phenotypic heterogeneity (Ackermann, 2015) as a strategy to cope with dynamic environments (such as those found in the fresh produce supply chain). Furthermore, the accessory genome (which was not investigated in this study) of all *L. monocytogenes* lineages is enriched for cell surface-related genes (den Bakker et al., 2013), the products of which are likely to be involved in the EPS matrix of biofilms (Colagiorgi et al., 2016). Whilst a correlation between lineage and biofilm production (lineage II isolates producing more biofilm has been reported (Borucki et al., 2003)) our results indicate that phylogenetic lineage does not affect biofilm

production (Di Bonaventura et al., 2008). More research is needed to highlight the intrinsic factors which determine whether strains are strong or weak biofilm formers in the presence of the environmental variables that are representative of environments found in the fresh produce supply chain.

Overall, the *L. monocytogenes* population in the UK fresh produce supply chain is diverse, in line with food isolates other countries such as Ireland (Hilliard et al., 2018), France (Moura et al., 2017), and Australia (Kwong et al., 2015). While some isolates have the necessary genomic components to cause disease and are closely related to outbreak strains, others are distantly related and are of less concern. All isolates formed biofilm but the intrinsic and extrinsic factors affecting this characteristic requires further investigation. In terms of the fresh produce supply chain, a more targeted sampling plan is needed to determine whether potentially virulent strains are sporadic (i.e. infrequently isolated) or persistent in a given environment. This is particularly important in processing environments that have the potential to contaminate produce on a broad scale. Therefore, in terms of policy, fresh produce supply chains should begin to phase in sampling regimes which implement WGS as standard. Implementing this technology may give customers (retail) and consumers added confidence that growers and processors are informed of the *L. monocytogenes* risk in their supply chain and demonstrates a precautionary, rather than reactionary approach to consumer safety.

## 5. Conclusion

In the UK fresh produce supply chain, a variety *L. monocytogenes* isolates were found which are highly related to strains previously implicated in outbreaks of disease. These strains retain the necessary genomic components to cause disease. However, *L. monocytogenes* is isolated relatively infrequently during routine sampling in the fresh produce supply chain and the strains that are isolated come from a range of environments and stages in the supply chain. Overall, *L. monocytogenes* contamination in the fresh produce supply chain may be difficult to prevent because of the range of potential sources of the bacterium in the growing and processing

environments and this is reflected in the diversity of strains isolated. SNP analysis of isolates revealed a mixture of closely and distantly related individuals and suggests that contamination in the supply chain is a combination of sporadic contamination and contamination arising from common sources/cross contamination. This is the first study of its kind in the UK and demonstrates the power of WGS as a subtyping tool for *L. monocytogenes* isolates. More efforts are needed to implement this technology as routine surveillance in the UK.

## 6. Acknowledgments

This work was supported by a joint grant from Edinburgh Napier University, Nature's Way Foods and G's Fresh Ltd. The funders had no role in the decision to publish or in the preparation of the manuscript, study design or data analysis. The views expressed in this report are those of the authors and do not necessarily represent positions or policies of Nature's Way Foods Ltd., G's Fresh Ltd., or Edinburgh Napier University.

## 7. References

Ackermann, M., 2015. A functional perspective on phenotypic heterogeneity in microorganisms. Nat. Rev. Microbiol. 13, 497–508. https://doi.org/10.1038/nrmicro3491

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., Pennsylvania, T., Park, U., 1990. Basic Local Alignment Search Tool. J Mol Biol 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borucki, M.K., Peppin, J.D., White, D., Loge, F., Call, D.R., 2003. Variation in biofilm formation among strains of *Listeria monocytogenes*. Appl. Environ. Microbiol. 69, 7336–7342. https://doi.org/10.1128/AEM.69.12.7336

Casey, A., Jordan, K., Coffey, A., Fox, E.M., McAuliffe, O., 2016. Comparative Genomic Analysis of Two Serotype 1/2b *Listeria monocytogenes* Isolates from Analogous Environmental Niches Demonstrates the Influence of Hypervariable Hotspots in Defining Pathogenesis. Front. Nutr. 3, 54. https://doi.org/10.3389/fnut.2016.00054

Chen, L., Zheng, D., Liu, B., Yang, J., Jin, Q., 2016. VFDB 2016: Hierarchical and refined dataset for big data analysis - 10 years on. Nucleic Acids Res. 44, D694–D697. https://doi.org/10.1093/nar/gkv1239

Chen, Y., Gonzalez-Escalona, N., Hammack, T.S., Allard, M.W., Strain, E.A., Brown, E.W., 2016. Core Genome Multilocus Sequence Typing for Identification of Globally Distributed Clonal Groups and Differentiation of Outbreak Strains of *Listeria monocytogenes*. Appl. Environ. Microbiol. 82, 6258–6272. https://doi.org/10.1128/AEM.01532-16

Chen, Y., Luo, Y., Curry, P., Timme, R., Melka, D., Doyle, M., Parish, M., Hammack, T.S., Allard, M.W., Brown, E.W., Strain, E.A., 2017. Assessing the genome level diversity of *Listeria monocytogenes* from contaminated ice cream and environmental samples linked to a listeriosis outbreak in the United States. PLoS One 12, 1–19. https://doi.org/10.1371/journal.pone.0171389

Colagiorgi, A., Bruini, I., Di Ciccio, P.A., Zanardi, E., Ghidini, S., Ianieri, A., 2017. *Listeria monocytogenes* Biofilms in the Wonderland of Food Industry. Pathogens 6, 41. https://doi.org/10.3390/pathogens6030041

Colagiorgi, A., Di Ciccio, P., Zanardi, E., Ghidini, S., Ianieri, A., 2016. A Look inside the *Listeria monocytogenes* Biofilms Extracellular Matrix. Microorganisms 4, 22. https://doi.org/10.3390/microorganisms4030022

den Bakker, H.C., Desjardins, C.A., Griggs, A.D., Peters, J.E., Zeng, Q., Young, S.K., Kodira, C.D., Yandava, C., Hepburn, T.A., Haas, B.J., Birren, B.W., Wiedmann, M., 2013. Evolutionary Dynamics of the Accessory Genome of *Listeria monocytogenes*. PLoS One 8. https://doi.org/10.1371/journal.pone.0067511

Di Bonaventura, G., Piccolomini, R., Paludi, D., D'Orio, V., Vergara, A., Conter, M., Ianieri, A., 2008. Influence of temperature on biofilm formation by Listeria monocytogenes on various food-contact surfaces: Relationship with motility and cell surface hydrophobicity. J. Appl. Microbiol. 104, 1552–1561. https://doi.org/10.1111/j.1365-2672.2007.03688.x

European Food Safety Authority, 2018. Multi-country outbreak of Listeria monocytogenes serogroup IVb, multi-locus sequence type 6, infections linked to frozen corn and possibly to other frozen vegetables – first update. EFSA Support. Publ. 15. https://doi.org/10.2903/sp.efsa.2018.EN-1448

Fravalo, P., Cherifi, T., Feliciano, K.D.N., Letellier, A., Fairbrother, J.H., Bekal, S., 2017. Characterisation of *InlA* truncation in *Listeria monocytogenes* isolates from farm animals and

human cases in the province of Quebec. Vet. Rec. Open 4, 1–6. https://doi.org/10.1136/vetreco-2016-000199

Gelbíčová, T., Koláčková, I., Pantůček, R., Karpíšková, R., 2015. A novel mutation leading to a premature stop codon in *inlA* of *Listeria monocytogenes* isolated from neonatal listeriosis. New Microbiol. 38, 293–296.

Ghosh, P., Halvorsen, E.M., Ammendolia, D.A., Mor-Vaknin, N., O'Riordan, M.X.D., Brumell, J.H., Markovitz, D.M., Higgins, D.E., 2018. Invasion of the brain by Listeria monocytogenes is mediated by InlF and host cell vimentin. MBio 9, 1–11. https://doi.org/10.1128/mBio.00160-18

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. 59, 307–321. https://doi.org/10.1093/sysbio/syq010

Hernandez-Milian, A., Payeras-Cifre, A., 2014. What is new in Listeriosis? Biomed Res. Int. 2014. https://doi.org/10.1155/2014/358051

Hilliard, A., Leong, D., Callaghan, A.O., Culligan, E.P., Morgan, C.A., Delappe, N., Hill, C., Id, K.J., Cormican, M., Gahan, C.G.M., 2018. Genomic Characterization of Listeria monocytogenes Isolates Associated with Clinical Listeriosis and the Food Production Environment in Ireland. Genes (Basel). https://doi.org/10.3390/genes9030171

Hoang, D.T., Chernomor, O., von Haeseler, A., Quang Minh, B., Sy Vinh, L., 2017. Supplementary Data: Ufboot2: Improving The Ultrafast Bootstrap Approximation 35, 518–522. https://doi.org/10.5281/zenodo.854445

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285

Keynote, 2015. Fruit & Vegetables Market Reports 2015.

Kim, S.W., Haendiges, J., Keller, E.N., Myers, R., Kim, A., Lombard, J.E., Karns, J.S., Van Kessel, J.A.S., Haley, B.J., 2018. Genetic diversity and virulence profiles of Listeria monocytogenes recovered from bulk tank milk, milk filters, and milking equipment from dairies in the United States (2002 to 2014). PLoS One 13, 1–17. https://doi.org/10.1371/journal.pone.0197053

Kwong, J.C., Mccallum, N., Sintchenko, V., Howden, B.P., 2015. Whole genome sequencing in clinical and public health microbiology. Pathology 47, 199–210. https://doi.org/10.1097/PAT.0000000000000235

Kwong, J.C., Mercoulia, K., Tomita, T., Easton, M., Li, H.Y., Bulach, D.M., Stinear, T.P., Seemann, T., Howden, B.P., 2016. Prospective Whole-Genome Sequencing Enhances National Surveillance of *Listeria monocytogenes* . J. Clin. Microbiol. 54, 333–42. https://doi.org/10.1128/JCM.02344-15

Little, C.L., Sagoo, S.K., Gillespie, I.A., Grant, K., McLaughlin, J., 2009. Prevalence and Level of *Listeria monocytogenes* and Other *Listeria* Species in Selected Retail Ready-to-Eat Foods in the United Kingdom. J. Food Prot. 72, 1869–1877. https://doi.org/10.4315/0362-028X-72.9.1869

Lomonaco, S., Nucera, D., 2012. Molecular Subtyping Methods for Listeria monocytogenes: Tools for Tracking and Control, in: DNA Methods in Food Safety: Molecular Typing of Foodborne and Waterborne Bacterial Pathogens.

Magalhães, R., Ferreira, V., Biscottini, G., Brandão, T.R.S., Almeida, G., Teixeira, P., 2017. Biofilm formation by persistent and non-persistent *Listeria monocytogenes* strains on abiotic surfaces. Acta Aliment. 46, 43–50. https://doi.org/10.1556/066.2017.46.1.6

McCollum, J.T., Cronquist, A.B., Silk, B.J., Jackson, K.A., O'Connor, K.A., Cosgrove, S., Gossack, J.P., Parachini, S.S., Jain, N.S., Ettestad, P., Ibraheem, M., Cantu, V., Joshi, M., DuVernoy, T., Fogg, N.W., Gorny, J.R., Mogen, K.M., Spires, C., Teitell, P., Joseph, L.A., Tarr, C.L., Imanishi, M., Neil, K.P., Tauxe, R. V, Mahon, B.E., 2013. Multistate outbreak of listeriosis associated with cantaloupe. N. Engl. J. Med. 369, 944–53. https://doi.org/10.1056/NEJMoa1215837

Monaghan, J., Adams, H., Hutchison, M., 2009. Monitoring microbial food safety of fresh produce.

Morganti, M., Scaltriti, E., Cozzolino, P., Bolzoni, L., Casadei, G., Pierantoni, M., Foni, E., Pongolini, S., 2015. Processing-Dependent and Clonal Contamination Patterns of *Listeria monocytogenes* in the Cured Ham Food Chain Revealed by Genetic Analysis. Appl. Environ. Microbiol. 82, 822–831. https://doi.org/10.1128/AEM.03103-15

Moura, A., Tourdjman, M., Leclercq, A., Hamelin, E., Laurent, E., Fredriksen, N., Van Cauteren, D., Bracq-Dieye, H., Thouvenot, P., Vales, G., Tessaud-Rita, N., Maury, M.M., Alexandru, A., Criscuolo, A., Quevillon, E., Donguy, M.-P., Enouf, V., de Valk, H., Brisse, S., Lecuit, M., 2017. Real-Time Whole-Genome Sequencing for Surveillance of *Listeria monocytogenes* , France. Emerg. Infect. Dis. 23. https://doi.org/10.3201/eid2309.170336

National Listeria Incident Management Team, 2018. Situation Report.

Nguyen, L.-T., Schimidt, H.A., Haeseler, A. von, Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300

Nightingale, K.K., Nightingale, K.K., Windham, K., Windham, K., Martin, K.E., Martin, K.E., Yeung, M., Yeung, M., Wiedmann, M., Wiedmann, M., 2005. Select *Listeria monocytogenes* Subtypes Commonly Found in Foods Carry Disctinct Donsense Mutations in *inlA*, Leading to Expression of Truncated and Secreted Internalin A, and Are Associated with a Reduced Invasion Phenotype for Human Intestina. Appl. Environ. Microbiol. 71, 8764–8772. https://doi.org/10.1128/AEM.71.12.8764

Nowak, J., Cruz, C.D., Tempelaars, M., Abee, T., van Vliet, A.H.M., Fletcher, G.C., Hedderley, D., Palmer, J., Flint, S., 2017. Persistent *Listeria monocytogenes* strains isolated from mussel production facilities form more biofilm but are not linked to specific genetic markers. Int. J. Food Microbiol. 256, 45–53. https://doi.org/10.1016/j.ijfoodmicro.2017.05.024

Pouillot, R., Klontz, K.C., Chen, Y., Burall, L.S., Macarisin, D., Doyle, M., Bally, K.M., Strain, E., Datta, A.R., Hammack, T.S., Van Doren, J.M., 2016. Infectious Dose of *Listeria monocytogenes* in Outbreak Linked to Ice Cream, United States, 2015. Emerg. Infect. Dis. 22, 2113–2119. https://doi.org/10.3201/eid2212.160165

Quereda, J.J., Meza-Torres, J., Cossart, P., Pizarro-Cerdá, J., 2017. Listeriolysin S: A bacteriocin from epidemic *Listeria monocytogenes* strains that targets the gut microbiota. Gut Microbes 8, 1–8. https://doi.org/10.1080/19490976.2017.1290759

Saltykova, A., Wuyts, V., Mattheus, W., Bertrand, S., Roosens, N.H.C., Marchal, K., De Keersmaecker, S.C.J., 2018. Comparison of SNP-based subtyping workflows for bacterial isolates using WGS data, applied to *Salmonella enterica* serotype Typhimurium and serotype 1,4,[5],12:i:-. PLoS One 13, 1–23. https://doi.org/10.1371/journal.pone.0192504

Seemann, T., 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seemann, T., n.d. mlst [WWW Document]. URL https://github.com/tseemann/mlst (accessed 05.08.18).

Seemann, T., n.d. ABRicate [WWW Document]. URL https://github.com/tseemann/abricate (accessed 01.08.18).

Seemann, T., n.d. snippy [WWW Document]. URL https://github.com/tseemann/snippy (accessed 02.08.18).

Smith, A., Moorhouse, E., Monaghan, J., Taylor, C., Singleton, I., 2018. Sources and survival of Listeria monocytogenes on fresh, leafy produce. J. Appl. Microbiol. 1–13.

https://doi.org/10.1111/jam.14025

Souvorov, A., Agarwala, R., Lipman, D.J., 2018. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol. 19, 153. https://doi.org/10.1186/s13059-018-1540-z

Stasiewicz, M.J., Oliver, H.F., Wiedmann, M., den Bakker, H.C., 2015. Whole-genome sequencing allows for improved identification of persistent *Listeria monocytogenes* in food-associated environments. Appl. Environ. Microbiol. 81, 6024–6037. https://doi.org/10.1128/AEM.01049-15

The University of Edinburgh, n.d. FigTree [WWW Document]. URL http://tree.bio.ed.ac.uk/software/figtree/ (accessed 09.11.18).

Toledo, V., Den Bakker, H.C., Hormazábal, J.C., González-Rocha, G., Bello-Toledo, H., Toro, M., Moreno-Switt, A.I., 2018. Genomic Diversity of *Listeria monocytogenes* Isolated from Clinical and Non-Clinical Samples in Chile. Genes (Basel). 9, 1–12. https://doi.org/10.3390/genes9080396

Velge, P., Roche, S.M., 2010. Variability of *Listeria monocytogenes* virulence: a result of the evolution between saprophytism and virulence? Future Microbiol. 5, 1799–1821.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644. https://doi.org/10.1093/jac/dks261

# 8. Appendices

## Appendix A

*Appendix Table A.1* **Assembly and annotation data associated with *L. monocytogenes* whole genomes from Illumina paired-end sequencing data**

| Isolate | No. of reads | Depth of coverage | No. of contigs | Genome size (Mb) | GC content (%) | #1 Identified match (Kraken 8GB database) | % Similarity to #1 match | N50 | rRNA copy number | tRNA |
|---|---|---|---|---|---|---|---|---|---|---|
| **NLmo2** | 1871906 | 119 | 16 | 3022108 | 38.5 | *Listeria monocytogenes* | 89.62 | 537374 | 4 | 54 |
| **NLmo3** | 864832 | 59 | 19 | 3021207 | 38.4 | *Listeria monocytogenes* | 92.34 | 537190 | 4 | 63 |
| **NLmo4** | 909848 | 65 | 19 | 3065932 | 38.4 | *Listeria monocytogenes* | 91.76 | 432370 | 3 | 58 |
| **NLmo5** | 1508824 | 101 | 18 | 3065467 | 37.9 | *Listeria monocytogenes* | 89.49 | 432886 | 4 | 49 |
| **NLmo6** | 4673156 | 308 | 13 | 2878779 | 38.8 | *Listeria monocytogenes* | 92.08 | 477702 | 3 | 50 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **NLmo7** | 3859388 | 262 | 9 | 2919242 | 38.6 | *Listeria monocytogenes* | 91.87 | 596199 | 5 | 50 |
| **NLmo8** | 705258 | 48 | 16 | 2920871 | 38.4 | *Listeria monocytogenes* | 93.06 | 477606 | 6 | 49 |
| **NLmo9** | 612548 | 44 | 18 | 2920275 | 38.4 | *Listeria monocytogenes* | 94.64 | 302065 | 2 | 54 |
| **NLmo10** | 1941944 | 129 | 11 | 2917149 | 38.6 | *Listeria monocytogenes* | 91.47 | 476852 | 5 | 50 |
| **NLmo13** | 1000374 | 67 | 13 | 2929305 | 38.4 | *Listeria monocytogenes* | 92.86 | 1497074 | 3 | 63 |
| **NLmo14** | 1530718 | 103 | 9 | 2918988 | 38.3 | *Listeria monocytogenes* | 92.05 | 596189 | 5 | 50 |
| **NLmo15** | 2202970 | 158 | 16 | 2968413 | 37.9 | *Listeria monocytogenes* | 94.53 | 398348 | 2 | 61 |
| **NLmo16** | 850148 | 57 | 13 | 2927284 | 38.4 | *Listeria monocytogenes* | 93.48 | 1497064 | 3 | 54 |
| **NLmo18** | 488048 | 31 | 21 | 2968827 | 38.4 | *Listeria monocytogenes* | 88.87 | 455612 | 4 | 49 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **NLmo20** | 2052146 | 138 | 10 | 2915683 | 38.9 | *Listeria monocytogenes* | 91.8 | 556332 | 3 | 48 |

*Appendix Table A.2* **Pairwise differences in number of single nucleotide polymorphisms (SNPs) between *L. monocytogenes* isolates from the UK fresh produce supply chain**.

| | **NLmo10** | **NLmo13** | **NLmo14** | **NLmo15** | **NLmo16** | **NLmo18** | **NLmo2** | **NLmo20** | **NLmo3** | **NLmo4** | **NLmo5** | **NLmo6** | **NLmo7** | **NLmo8** | **NLmo9** | **Ref.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NLmo10** | 0 | 39261 | 56 | 39512 | 39262 | 39965 | 3301 | 2738 | 3300 | 40017 | 40017 | 2687 | 56 | 2425 | 2708 | 39752 |
| **NLmo13** | 39261 | 0 | 39260 | 8008 | 1 | 12519 | 39294 | 39353 | 39294 | 8999 | 8999 | 39280 | 39260 | 39334 | 39272 | 7744 |
| **NLmo14** | 56 | 39260 | 0 | 39513 | 39261 | 39965 | 3305 | 2742 | 3304 | 40018 | 40018 | 2677 | 0 | 2415 | 2698 | 39753 |
| **NLmo15** | 39512 | 8008 | 39513 | 0 | 8009 | 12344 | 39571 | 39630 | 39571 | 8513 | 8513 | 39508 | 39513 | 39598 | 39496 | 3541 |
| **NLmo16** | 39262 | 1 | 39261 | 8009 | 0 | 12520 | 39295 | 39354 | 39295 | 9000 | 9000 | 39281 | 39261 | 39335 | 39273 | 7745 |
| **NLmo18** | 39965 | 12519 | 39965 | 12344 | 12520 | 0 | 40016 | 40066 | 40017 | 12815 | 12815 | 39985 | 39965 | 40038 | 39973 | 12125 |
| **NLmo2** | 3301 | 39294 | 3305 | 39571 | 39295 | 40016 | 0 | 3405 | 1 | 40069 | 40069 | 3176 | 3305 | 3218 | 3191 | 39831 |
| **NLmo20** | 2738 | 39353 | 2742 | 39630 | 39354 | 40066 | 3405 | 0 | 3404 | 40133 | 40133 | 2927 | 2742 | 3032 | 2948 | 39872 |
| **NLmo3** | 3300 | 39294 | 3304 | 39571 | 39295 | 40017 | 1 | 3404 | 0 | 40070 | 40070 | 3176 | 3304 | 3218 | 3191 | 39832 |
| **NLmo4** | 40017 | 8999 | 40018 | 8513 | 9000 | 12815 | 40069 | 40133 | 40070 | 0 | 2 | 40042 | 40018 | 40078 | 40030 | 8323 |
| **NLmo5** | 40017 | 8999 | 40018 | 8513 | 9000 | 12815 | 40069 | 40133 | 40070 | 2 | 0 | 40042 | 40018 | 40078 | 40030 | 8323 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NLmo6** | 2687 | 39280 | 2677 | 39508 | 39281 | 39985 | 3176 | 2927 | 3176 | 40042 | 40042 | 0 | 2677 | 1672 | 59 | 39781 |
| **NLmo7** | 56 | 39260 | 0 | 39513 | 39261 | 39965 | 3305 | 2742 | 3304 | 40018 | 40018 | 2677 | 0 | 2415 | 2698 | 39753 |
| **NLmo8** | 2425 | 39334 | 2415 | 39598 | 39335 | 40038 | 3218 | 3032 | 3218 | 40078 | 40078 | 1672 | 2415 | 0 | 1693 | 39825 |
| **NLmo9** | 2708 | 39272 | 2698 | 39496 | 39273 | 39973 | 3191 | 2948 | 3191 | 40030 | 40030 | 59 | 2698 | 1693 | 0 | 39769 |
| **Ref.** | 39752 | 7744 | 39753 | 3541 | 7745 | 12125 | 39831 | 39872 | 39832 | 8323 | 8323 | 39781 | 39753 | 39825 | 39769 | 0 |

*Appendix Table A.3* **Presence/absence of key virulence factors in *L. monocytogenes* strains isolated from the fresh produce supply chain as determined by Abricate.** The accession number for each gene is indicated below gene name. Legend Yes = ≥95% coverage & >75% identity, probable = ≥ 36.4% coverage & >75% identity - = absent.

| **Isolate** | **NLmo2** | **NLmo3** | **NLmo4** | **NLmo5** | **NLmo6** | **NLmo7** | **NLmo8** | **NLmo9** | **NLmo10** | **NLmo13** | **NLmo14** | **NLmo15** | **NLmo16** | **NLmo18** | **NLmo20** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **no. of virulence factors** | 34 | 34 | 31 | 31 | 41 | 41 | 41 | 41 | 41 | 32 | 41 | 32 | 32 | 33 | 41 |
| ***actA* NP_463735** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***ami* NP_466081** | Yes | Probable | Yes | Probable | - | - | - | - | - | Yes | - | Yes | Yes | Yes | - |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***aut*** **NP_466081** | Yes | Yes | Yes | Yes | - | - | - | - | - | Yes | - | Yes | Yes | Yes | - |
| ***bsh*** **NP_465591** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***clpC*** **NP_463763** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***clpE*** **NP_464522** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***clpP*** **NP_465991** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***cdsA*** **NP_464841.1** | Yes | Yes | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | Yes | Yes |
| ***essC*** **NP_645079** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***fbpA*** **NP_465354** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***gtcA*** **NP_466072** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***hly*** **NP_463733** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***hpt*** **NP_464364** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***iap/cwhA*** **NP_464110** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***inlA*** **NP_463962** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***inlB*** **NP_463963** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***inlC*** **NP_465311** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***inlF*** **NP_463939** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | - | Yes |
| ***inlJ*** **NP_466343** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***inlK*** **NP_464815** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***lap*** **NP_465159** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***lapB*** **NP_465191** | Yes | Yes | - | - | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***llsA*** **AHK25016** | - | - | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | - | Yes |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***llsB*** **AHK25020** | - | - | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | - | Yes |
| ***llsD*** **AHK25022** | - | - | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | - | Yes |
| ***llsG*** **AHK25017** | - | - | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | - | Yes |
| ***llsH*** **AHK25018** | - | - | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | - | Yes |
| ***llsP*** **AHK25023** | - | - | - | - | Yes | Yes | Yes | Probable | Yes | - | Yes | - | - | - | Yes |
| ***llsX*** **AHK25019** | - | - | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | - | Yes |
| ***llsY*** **AHK25021** | - | - | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | - | Yes |
| ***lntA*** **NP_463967** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***lpeA*** **NP_465372** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***lplA1*** **NP_464456** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***lspA*** **NP_465369** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

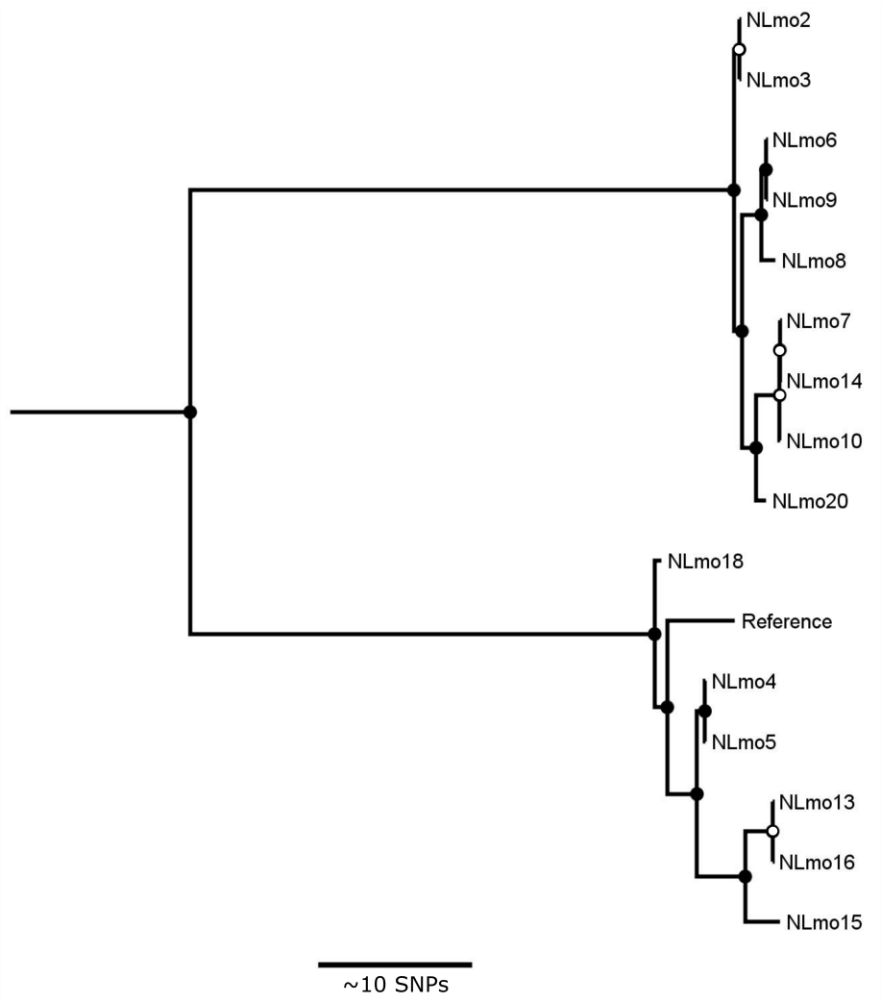
| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***mpl*** **NP_463734** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***oatA*** **NP_464816** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***pdgA*** **NP_463944** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***plcA*** **NP_463732** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***plcB*** **NP_463736** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***prfA*** **NP_463731** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***prsA2*** **NP_465743** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***vip*** **NP_463850** | Yes | Yes | - | - | Yes | Yes | Yes | Yes | Yes | - | Yes | - | - | Yes | Yes |

*Appendix Table A.4* **Presence/absence of key resistance genes in *L. monocytogenes* strains isolated from the fresh produce supply chain.** The accession number for each gene is indicated below gene name. Legend Yes = ≥95% coverage & >75% identity, - = absent.

| Isolate | NLmo2 | NLmo3 | NLmo4 | NLmo5 | NLmo6 | NLmo7 | NLmo8 | NLmo9 | NLmo10 | NLmo13 | NLmo14 | NLmo15 | NLmo16 | NLmo18 | NLmo20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***bcrB*** **A7J11_05169** | - | - | Yes | Yes | - | - | - | - | - | - | - | - | - | - | - |
| ***bcrC*** **A7J11_05170** | - | - | Yes | Yes | - | - | - | - | - | - | - | - | - | - | - |
| ***fosX*** **A7J11_01023** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ***lin*** **A7J11_00340** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

## Appendix B




*Appendix Figure B.1* **Maximum Likelihood phylogeny of 15 *L. monocytogenes* isolates isolate from the UK fresh produce supply chain. Tree was inferred inferred using a core-SNP based phylogeny from a 3288bp concatenated sequence of 7 housekeeping MLST genes in *L. monocytogenes* (IQtree version 1.6.7).** Tree is rooted through the midpoint. White circles indicate nodes with ≥ 95% ultrafast bootstrap support and ≥ 80% SH-like approximate likelihood ratio test support.