Genomic insights into persistence of *Listeria* species in the food processing environment

**L. Palaiodimou**, S. Fanning and E.M. Fox

1 Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, UK
2 UCD-Centre for Food Safety, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland
3 Institute for Global Food Security, Queen’s University Belfast, Belfast, UK

**Keywords**
disinfection, food safety, *Listeria*, plasmids, virulence.

**Correspondence**
Edward M. Fox, Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, UK.
E-mail: edward.fox@northumbria.ac.uk

2021/0082: received 15 January 2021, revised 18 February 2021 and accepted 21 March 2021
doi:10.1111/jam.15089

**Abstract**

**Aims:** *Listeria* species may colonize and persist in food processing facilities for prolonged periods of time, despite hygiene interventions in place. To understand the genetic factors contributing to persistence of *Listeria* strains, this study undertook a comparative analysis of seven persistent and six presumed non-persistent strains, isolated from a single food processing environment, to identify genetic markers correlating to promoting persistence of *Listeria* strains, through whole genome sequence analysis.

**Methods and Results:** A diverse pool of genetic markers relevant to hygiene tolerance was identified, including disinfectant resistance markers *qacH*, *emrC* and the efflux cassette *bcrABC*. Both persistent and presumed non-persistent cohorts encoded a range of stress resistance markers, including heavy metal resistance, oxidative and pH stress, although trends were associated with each cohort (e.g., *qacH* and *cadA1C* resistance was more frequently found in persistent isolates). Persistent isolates were more likely to contain mutations associated with attenuated virulence, including a truncated InlA. Plasmids and transposons were widespread between cohorts.

**Conclusions:** Results suggest that no single genetic marker identified was universally responsible for a strain’s ability to persist. Persistent strains were more likely to harbour mutation associated with hypovirulence.

**Significance and Impact of the Study:** This study provides additional insights into the distribution of genetic elements relevant to persistence across *Listeria* species, as well as strain virulence potential.

**Introduction**

*Listeria* species comprise an expanding genus of bacteria, which to date includes 21 recorded species, many of which are relatively recently described (Leclercq et al. 2019; Quereda et al. 2020). Of these species, *Listeria monocytogenes* is of primary concern to public health, although *Listeria ivanovii* is an important pathogen of animals (Orsi and Wiedmann 2016). These bacteria can be found ubiquitously in the environment and may contaminate foods including ready-to-eat foods, vegetable, seafood, meat, eggs and dairy products; thus, incidence of disease is mainly linked to infections via foodborne transmission (Fugett et al. 2007; Scallan et al. 2011; McAuley et al. 2014). Both pathogenic and non-pathogenic species are known to share common niches, and as such, non-pathogenic *Listeria* species, such as *Listeria innocua* and *Listeria welshimeri*, may be used as index organisms for potential contamination and/or colonization of *L. monocytogenes* in food processing environments (FPEs). Hygiene regimes, which include cleaning, sanitizing and disinfection cycles, are among the primary interventions applied in FPEs to control *Listeria* species. This includes the use of antimicrobial agents such as quaternary ammonium compounds (QACs), as well as other antimicrobial formulations.

The persistence of *L. monocytogenes* has frequently been reported in FPEs, with several studies from various
locations reporting re-isolation of the same clone over extended periods up to and exceeding 10 years (Wulff et al. 2006; Chambel et al. 2007; Lomonaco et al. 2009; Fox et al. 2011b). This presents an important challenge to food producers, as these persistent contaminants are associated with an increased likelihood of cross-contamination of food products produced, because they are not eliminated from the FPE. Persistence may be caused, or at least contributed to, by several factors such as poor hygiene practice and/or ineffective sanitizers; harbourage sites in the FPE, such as damaged equipment of surfaces; the presence of genetic markers providing a competitive advantage to persistent strains; the efficient production of biofilms by persistent strains; or interactions with native microbiota (Fox et al. 2014; Fox et al. 2015; Schmitz-Esser et al. 2015; Harter et al. 2017; Rodríguez-Campos et al. 2019). Although a number of genetic elements have been purported to play a role in colonization and persistence dynamics, the nature of the role of these mechanisms to the persistence phenomenon remains poorly understood. This may, in part, be due to the different environmental conditions across sometimes disparate FPEs, which may vary in aspects such as hygiene systems, temperature conditions and resident microbiota, among other factors. A particular challenge to studying the persistence phenotype is the difficulty in replicating experimental conditions amenable to its study under laboratory conditions. Similarly, introduction of pathogenic L. monocytogenes to an FPE to examine colonization dynamics represents an unacceptable food safety risk. To further elucidate the potential genetic factors that may promote persistence of Listeria strains and to understand other relevant aspects of interest such as pathogenic potential, this study aimed to characterize seven persistent and six presumed non-persistent strains isolated from the same FPE, over the same time frame, to correlate genetic traits across persistent and/or non-persistent cohorts. This would facilitate the comparison of strains experiencing comparable environmental selection, such as antimicrobial agents in use, temperatures, resident microbiota and surface materials. This study included persistent and presumed non-persistent strains from three species (L. monocytogenes, L. welshimeri and L. innocua), to examine pangenome markers across multiple members of the genus Listeria.

Materials and methods

Bacterial isolates in this study

This study characterized 13 Listeria strains collected over a 2-year period from a meat processing facility in Ireland. Seven of them (L. monocytogenes: UCDL011, UCDL016, UCDL019 and UCDL187; L. innocua: UCDL146; and L. welshimeri: UCDL122) were characterized as ‘persistent’ contaminants, representing pulsed-field gel electrophoresis (PFGE) pulsotypes isolated multiple times over 6 months or more. All isolates were subjected to the PulseNet Standard PFGE method for subtyping L. monocytogenes (PulseNet 2013), utilizing two restriction enzymes (viz., AscI and Apal). Isolates were classified as the same strain based on an indistinguishable PFGE pulsotype considering fingerprints of both enzymes (i.e., 100% similarity score). The other six isolates (L. monocytogenes: UCDL037, UCDL133, UCDL150 and UCDL175; L. innocua: UCDL085; and L. welshimeri: UCDL063) were designated as ‘presumed non-persistent’ contaminants, comprising genotypes only identified a single time over the 2-year surveillance period.

Genome assembly and annotation

Genomic DNA was extracted using the QIAGEN DNeasy kit (Qiagen, Hilden, Germany). Sample quality was confirmed using a NanoDrop instrument (Thermo Fisher Scientific, Waltham, MA) to confirm 260:280-nm and 260:230-nm ratios between 1.8 and 2.0. Library preparation using genomic DNA of isolates was performed using the Nextera XT library prep kit (Illumina, San Diego, CA). Raw read sequences were then generated using 250-bp paired end sequencing on the MiSeq platform (Illumina). The raw read quality was assessed with FastQC (ver. 0.11.8). These raw reads were subsequently processed to remove adapter sequences and low-quality reads using Trimmomatic software ver. 0.22 (Bolger et al. 2014). Draft genomes were assembled using SPAdes (Species Prediction and Diversity Estimation) software ver. 2.5.1 based on an algorithm that employs multisized De Bruijn graphs with \( k \)-mer values of ‘21, 33, 55, 77’ to construct the contigous sequences (Bankevich et al. 2012). All draft genomes were annotated using the RAST online platform tool and using Prokka algorithms (Aziz et al. 2008; Seemann 2014).

Molecular subtyping of isolates

The serotype of L. monocytogenes isolates was identified using previously described in silico schemes (Doumith et al. 2004). Strain MLST type was derived using the seven housekeeping gene targets previously described (Ragon et al. 2008) and referencing the Institut Pasteur BIGSdb-Lm database (https://bigsdb.pasteur.fr/listeria). Novel alleles were submitted to the Institut Pasteur BIGSdb-Lm database for assignment of novel sequence types (STs).
Persistence of Listeria species

L. Palaiodimou et al.

Genome screening for molecular markers and comparative visualization of sequence data

A strain BLAST database was created using the Geneious Prime software platform (Kearse et al., 2012). Additional databases were created comprising genes of interest relating to virulence, stress resistance or other features such as mobile genetic elements, as detailed in Table S1. Sequence alignments were performed using MAFFT program (Katoh et al., 2002). EasyFig software was utilized to visualize sequence alignment similarities, including transposon and phage alignments (Sullivan et al., 2011). The BLAST ring image generator (BRIG) platform was used to visualize BLAST comparisons using constructed pangenome references (Alikhan et al., 2011).

Pangenome analysis, core SNP analysis and phylogenetic tree construction

Pangenome analysis was performed utilizing the Roary pipeline (Page et al., 2015), and maximum likelihood phylogenetic trees were constructed using RAXML (Stamatakis et al., 2014). Phylogenetic analysis of core genome alignments was performed using FastTree, with the GTR+CAT model (Price et al., 2009). Plasmid searches were performed using PlasmidFinder 2.1, interrogating against the Gram-positive database (Carattoli et al., 2014), coupled with BLAST searches of draft genomes utilizing published plasmid sequences pL1100, pN1-011A and p6179 (NCBI accession numbers NC_003383, NC_022045 and NZ_HG813250, respectively).

Draft sequence archiving

Draft genome sequences for strains from this project have been deposited in the NCBI genome database, BioProject PRJNA692370, under the following accessions: UCDL011, SAMN17321004; UCDL016, SAMN17321005; UCDL019, SAMN17321006; UCDL037, SAMN17321007; UCDL063, SAMN17321008; UCDL085, SAMN17321009; UCDL122, SAMN17321010; UCDL133, SAMN17321011; UCDL146, SAMN17321012; UCDL150, SAMN17321013; UCDL162, SAMN17321014; UCDL175, LXRD00000000; and UCDL187, SAMN17321015.

Results

Overview of the genomes

An overview of the genetic subtypes and genome characteristics of isolates in this study is presented in Table 1. The L. monocytogenes strains’ genomes ranged from 2 913 782 to 3 080 560 bp, the L. innocua from 2 991 782 to 3 026 780 bp and the L. welshimeri from 2 856 944 to 2 946 539 bp. GC content was lowest among L. welshimeri (36-1-36-3%), followed by L. innocua (37-3%), and highest among L. monocytogenes strains (37-8-37-9%). The total pangenome size was 7669 CDS sequences, of which 1314 were core to all strains (17%; Fig. 1). Listeria monocytogenes shared a greater number of genes exclusively with L. innocua (n = 477), relative to genes exclusive with L. welshimeri (n = 178); this is supported by the likelihood of these species being closer in evolutionary terms (Orsi and Wiedmann, 2016). Among the L. monocytogenes strains, clonal complex 9 (CC9) strains were most common (50%, 4/8); this

<table>
<thead>
<tr>
<th>Strain and molecular subtype</th>
<th>Genome Size (bp)</th>
<th>GC%</th>
<th>CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCDL011 L. monocytogenes ST9 CC9 1/2c, 3c II 3</td>
<td>3 073 215</td>
<td>37-8</td>
<td>3058</td>
</tr>
<tr>
<td>UCDL016 L. monocytogenes ST9 CC9 1/2c, 3c II 3</td>
<td>3 063 336</td>
<td>37-8</td>
<td>3035</td>
</tr>
<tr>
<td>UCDL019 L. monocytogenes ST121 CC121 1/2a, 3a II 3</td>
<td>3 057 649</td>
<td>37-8</td>
<td>3018</td>
</tr>
<tr>
<td>UCDL187 L. monocytogenes ST31 CC31 1/2a, 3a II 3</td>
<td>3 080 560</td>
<td>37-8</td>
<td>3094</td>
</tr>
<tr>
<td>UCDL122 L. welshimeri ST168 CC168 ND – 2</td>
<td>2 856 944</td>
<td>36-3</td>
<td>2787</td>
</tr>
<tr>
<td>UCDL162 L. welshimeri ST2688 – ND – 2</td>
<td>2 832 429</td>
<td>36-1</td>
<td>2817</td>
</tr>
<tr>
<td>UCDL146 L. innocua ST602 – ND – 2</td>
<td>3 026 780</td>
<td>37-3</td>
<td>3054</td>
</tr>
<tr>
<td>UCDL037 L. monocytogenes ST1 CC1 4b, 4e, 4d I 3</td>
<td>2 913 758</td>
<td>37-9</td>
<td>2836</td>
</tr>
<tr>
<td>UCDL133 L. monocytogenes ST9 CC9 1/2c, 3c II 3</td>
<td>3 009 680</td>
<td>37-8</td>
<td>2977</td>
</tr>
<tr>
<td>UCDL150 L. monocytogenes ST622 CC9 1/2c, 3c II 3</td>
<td>3 030 065</td>
<td>37-9</td>
<td>3007</td>
</tr>
<tr>
<td>UCDL175 L. monocytogenes ST204 CC204 1/2a, 3a II 3</td>
<td>3 020 537</td>
<td>37-8</td>
<td>2951</td>
</tr>
<tr>
<td>UCDL063 L. welshimeri ST1084 CC1084 ND – 2</td>
<td>2 946 539</td>
<td>36-3</td>
<td>2999</td>
</tr>
<tr>
<td>UCDL085 L. innocua ST1008 – ND – 2</td>
<td>2 991 782</td>
<td>37-3</td>
<td>3002</td>
</tr>
</tbody>
</table>

-, not defined; ND, not determined.

TABLE 1 Summary of the genetic subtypes and genomes of strains in this study.
included three ST9 isolates and a single allele variant (ST622). Other STs identified included ST1, ST121 and ST204. Among the *L. welshimeri* strains was ST168 and ST1084 isolates, along with a novel ST not previously described, designated as ST2688 in the Institut Pasteur BIGSdb-*Lm* database. For *L. innocua*, ST602 and ST1008 were observed.

**Stress resistance markers**

Three disinfectant resistance markers were identified among strains in this study: *bcrABC*, *emrC* and *qacH*. The most common of these genetic markers found was *bcrABC*, present in four strains, followed by *emrC* in three strains and *qacH* in two strains (Fig. 2). Interestingly, strains only harboured one of these disinfectant resistance markers. When considering persistent and presumed non-persistent cohorts, 5/7 persisters harboured disinfectant resistance markers (71%), compared with 4/6 non-persisters (67%). The *qacH* marker was only identified in persistent strains (2/7), whereas the other two markers were present in both cohort groups.

Cadmium resistance cassettes were prevalent among both persistent (86%) and presumed non-persistent (83%) cohorts; this included the *cadA1*, *cadA2* and *cadA3* variants of this resistance system (Fig. 2). Of these, *cadA1* was more frequent among persisters compared with non-persisters (57% vs. 17%), whereas *cadA4* was only identified in non-persisters.

Listeria Genomic island 2 (LGI2) encodes a large arsenic resistance operon (*arsD1A1R1D2R2A2B1B2*) and...
was identified in two presumed non-persistent isolates: UCDL037 and UCDL175. Another known arsenic cassette carrying element, a Tn544 transposon, was found in six isolates (four persistent and two presumed non-persistent). An additional screen of 100 isolates previously described by Hurley et al. (2019) was then analysed for the presence of arsA1, arsA2 and the Tn544 resistance cassette, with an overall prevalence of 12, 2 and 1% identified, respectively.

The *L. monocytogenes* stress survival islets (SSIs) encode genetic mechanisms for resistance to stress conditions such as temperature, pH and osmotic stress (Ryan et al. 2010; Harter et al. 2017). Of the previously described SSIs, SSI-1 was the most common; it was found in eight isolates, including 5/7 persisters (71%) and 3/6 non-persisters (50%; Fig. 2 and Fig. S1). SSI-2 was identified in three isolates (UCDL019, UCDL085 and UCDL146).

**Virulence markers**

The distribution of a number of important virulence markers is shown in Fig. 3. None of the virulence genes in Fig. 3 were identified outside of *L. monocytogenes* in this study. All of the 12 internalins, however, were identified among the *L. monocytogenes* strains; of these, *inlA, inlB, inlC, inlE, inlf, inll, inlf* and *inlk* were present in all strains. The *inlg* gene was absent in two persistent strains and a presumed non-persistent strain, whereas *inlc2, inld* and *inlh* were absent from four strains (two persistent and two presumed non-persistent). Premature stop codons were identified in *inla* sequences in five *L. monocytogenes* strains (all persisters and one non-persister); in addition, UCDL187 harboured PMSCs in *inlc*, in *ascB* and a hypothetical protein in its *inlc2de* locus and in the *prfa* virulence regulator (Figs S2 and S3).

The *Listeria* Pathogenicity Islands (LIPIs) 3 and 4 contribute to the pathogenesis of strains and are associated with increased virulence in mammalian host infections (Cotter et al. 2008; Maury et al. 2016). LIPI-3 was identified in a single presumed non-persistent strain, the *L. monocytogenes* strain UCDL037; LIPI-4 was absent in all isolates in this study.

**Mobile genetic elements**

Plasmids were identified in 10 of the strains in this study (77%), with three of these containing two plasmids (Fig. 4 and Fig. S4). Both *L. welshimeri* UCDL063 and UCDL122 contained identical plasmids (pUCDL063-1 and pUCDL122-1); two ST9 strains (UCDL016 and UCDL133) contained similar plasmids (pUCDL016-1 and pUCDL133-1), and a smaller plasmid of 4265 bp was present in three isolates in this study (UCDL011, UCDL016 and UCDL133).

The *cosK* phage insert was identified in seven strains in this study (54%); this included four persistent (57%) and three presumed non-persistent isolates (50%; Fig. S5). This phage was identified in both *L. monocytogenes* and *L. innocua* species, but not in *L. welshimeri*.

The maximum likelihood tree generated from the pan-genome analysis of all coding sequences among the strains in this study identified three clades, one representing each of the three species (Fig. S6). The *L. monocytogenes*-containing clade included a CC9 subclade, supporting their genetic similarity relative to other STs identified. The CC9 subclade was investigated through a core SNP analysis and SNP frequencies supporting
diverse strains (ranging from 198 to 1236 SNPs between isolates) (Fig. S7).

**Discussion**

This study characterized persistent and presumed non-persistent strains of three *Listeria* species. While a diverse population was noted (Table 1 and Fig. S6), for the *L. monocytogenes* strains, CC9 was the most common. This CC was represented among both persistent and presumed non-persistent isolates, suggesting that it is associated with food-related niches. This finding supports previous similar studies from Ireland and Europe, which found a high incidence of CC9 among food-related niches.
sources (Ebner et al. 2015; Henri et al. 2016; Hurley et al. 2019). One each of ST1, ST121 and ST204 was also identified, the latter two STs also being among the more common food-related clonal subgroups (Schmitz-Esser et al. 2015; Jennison et al. 2017). Lineage II was the most common genetic lineage observed among the L. monocytogenes strains, which supports the assertion that this lineage has adapted to food-associated niches, compared with lineage I, which is more frequently associated with clinical incidence of disease (Maury et al. 2016; Jennison et al. 2017).

The ability of strains to colonize and persist in FPEs is thought to be a multifaceted phenomenon. Strains must colonize an environmental niche, where efficient biofilm production is likely to be important (Norwood and Gilmour 1999; Rodríguez-Campos et al. 2019), although a direct correlation with persistence has not always been identified (Djordjevic et al. 2002). Subsequently, these bacteria must tolerate a plethora of environmental stressful conditions, many which may be unfavourable or antagonistic to their survival. Exposure to disinfectants imposes a continual stress on bacterial species colonizing FPEs, as these are central to hygiene efforts and food safety. Previous studies have suggested that disinfectant resistance may be a feature of strains encountered, or persisting, in FPEs; this is thought to contribute to the dominance of ST121 and ST204 clonally related strains (Schmitz-Esser et al. 2015; Fox et al. 2016); however, a positive association with persistence of specific strains in FPEs varies (Kastbjerg and Gram 2009; Fox et al. 2011a; Rodríguez-Campos et al. 2019). A variety of disinfectant resistance markers have been identified in Listeria species, typically comprising efflux pump systems; these include the bcrABC cassette, emrC, emrE, qacA, qacC, qacH and qacEA1 determinants. These systems are typically associated with resistance to QACs, with this class of disinfectant in use at the food processing facility where these strains were isolated. Comparative analysis of the aforementioned genetic markers between persistent and presumed non-persistent isolates did not suggest a correlation with either phenotype group. Although three different markers were identified among persisters (bcrABC, emrC and qacH) relative to two among non-persisters (bcrABC and emrC), the overall prevalence was similar (Fig. 2); 5/7 persisters harboured disinfectant resistance markers (71%), compared with 4/6 non-persisters (67%). This suggests that the presence of known disinfectant resistant markers was not the sole causative mechanism for persistence. It does, however, support the previous associations of Tn6188 with strains associated with food environments and their propensity to be associated with survival and/or persistent contamination dynamics, because this transposon, including the qacH gene, was only identified among persistent strains in this study (Muller et al. 2013; Ortiz et al. 2015; Hurley et al. 2019).

Heavy metal resistance has been frequently observed in Listeria species, principally to cadmium and arsenic; the associated genetic resistance markers are among the more commonly found stress resistance markers associated with mobile genetic elements across the genus (Parsons et al. 2018). Cadmium resistance is generally mediated through the cadAC cassette system in Listeria species (Lebrun et al. 1994a, 1994b), with six cadA variants (cadA1–A6) described to date (Chmielowska et al. 2020). Of these variants, cadA4 is thought to provide the lowest relative tolerance to cadmium, permitting growth up to approximately 50 µg ml⁻¹ (Parsons et al. 2017); cadA1 and cadA2, however, facilitate growth at concentrations >140 µg ml⁻¹. Although a direct link to persistence and cadmium resistance has not been demonstrated, there is growing evidence that the prevalence of cadmium resistance is higher among clones showing recurrent contamination patterns in FPEs compared with their sporadically contaminating counterparts (Harvey and Gilmour 2001; Parsons et al. 2020). Results of this study suggest that high frequencies of known cadmium resistance cassettes were present among both persistent (86%) and presumed non-persistent (83%) cohorts. Although found at high incidence among strains in this study, results suggest that cadA1 is more common in persisters, whereas cadA4, which provides lower tolerance than cadA1, was only carried in non-persisters.

Arsenic resistance is typically associated with higher prevalence among serotype 4b strains of L. monocytogenes (McLauchlin et al. 1997; Mullapudi et al. 2008); this study only included a single 4b isolate, UCDL037, harbouring LGI2, which carries a large arsenic resistance operon (arsD1A1R1D2R2A2B1B2). LGI2 was also present in UCDL175, and this operon encodes both the arsA1 and arsA2 ATP transporters, as well as the membrane transporters arsB1 and arsB2. Interestingly in this study, lineage II L. monocytogenes had a relatively high rate of carriage of arsenic resistance determinants, with 86% (6/7) of these strains encoding an arsenic transporter. This was primarily due to the presence of a Tn544 resistance transposon containing an arsCDABR cassette (Kuenne et al. 2013). This prevalence is higher than previously noted by McLauchlin et al. (1997) or Mullapudi et al. (2008), who both reported lineage II resistance rates of 3%. An extended wider analysis of 100 isolates was conducted to further investigate if the higher arsenic resistance prevalence from the facility in this study was also observed in other facilities in Ireland. A set of 100 isolates previously described by Hurley et al. (2019) was analysed for carriage of arsenic resistance markers (arsA1, arsA2 and
the Tn544 resistance cassette); although the carriage rate among the lineage II isolates was higher than the previously mentioned studies at 14%, it was still lower than the prevalence observed in this study. In both cases, the Tn544 cassette was the most common resistant determinant. Interestingly, Pasquali et al. (2018) noted a high carriage rate of LGII-associated ars operon among ST14 isolates; however, this was absent in ST121 isolates collected from the same environment. The reason for the higher prevalence of arsenic resistance in the present study is not clear but may elude to introduction of resistant isolates from ingredient suppliers and/or horizontal gene transfer (HGT) events at the facility. The high carriage rate of Tn544-mediated resistance among persisters (50%), coupled with carriage across different species, supports the likelihood of HGT dynamics.

A number of broad spectrum SSIs of *L. monocytogenes* and/or *L. innocua*, denoted as SSIs (SSI-1 and SSI-2), have been described; these provide benefits to growth and/or survival under suboptimal or stress conditions, such as low pH (SSI-1), alkaline pH (SSI-2) or oxidative stress conditions (both islets) (Ryan et al. 2010; Harter et al. 2017). Their carriage is typically overrepresented among food isolates and has been implicated in persistence of clonally related groups, such as ST121; serotype 4b isolates often lack SSI-1 or SSI-2 but instead harbour a 549-bp hypothetical protein CDS (referred to as the variation ‘SSI-V’ islet in this study). In line with previous studies, UCDL019 (an ST121 strain) and both *L. innocua* isolates harboured SSI-2; the other lineage II isolates harboured SSI-1 (UCD011, UCDL016, UCDL133, UCDL150, UCDL175 and UCDL187), whereas the ST1 isolate UCDL037 contained SSI-V (Fig. 2 and Fig. S1). Ryan et al. (2010) noted that *L. welshimeri* strain SLCC5334 lacked any genes in the SSI insertion hotspots; interestingly, in our study, both persistent *L. welshimeri* isolates harboured SSI-1, whereas the non-persistent isolates had an absence of any insert in the SSI locus. This may allude to the possible contribution of the SSI inserts to persistence of *Listeria* strains in FPEs and should be further investigated among other persistent and presumed non-persistent clones to provide additional insights.

The internalin family of proteins comprise 25 members with characteristic leucine-rich repeat domains and have demonstrated roles in virulence and host pathogen interactions (Radoshevich and Cossart 2017). The most well characterized of these, InlA, mediates entry to host cells through binding of the E-cadherin host cell receptor (Gaillard et al. 1991; Mengaud et al. 1996). A number of mutations have been reported in the coding gene, *inlA*, which lead to production of truncated InlA variants (Van Stelten et al. 2010). These variants typically lack the LPXTG motif at the C-terminal end and are not bound to the bacterial cell wall by the sortase enzyme. Associated strains of *L. monocytogenes* lacking *inlA* are generally attenuated in their pathogenicity (Olier et al. 2005). In this study, of the eight *L. monocytogenes* isolates (each harbouring *inlA*), five contained mutations leading to a PMSC in the gene sequence, which would produce a truncated InlA lacking the LPXTG sequence motif. This included all the persistent strains (4/4; 100%) and one of the presumed non-persistent strains (1/4; 25%). This suggests that all the persistent isolates included in this study would be associated with reduced virulence in vivo. Persistent isolate UCDL185 also contained a single nucleotide insertion in a polyA region at the N-terminal end of *inlC* (nucleotide positions 7–15), causing a frameshift mutation and leading to a downstream PMSC (Fig. S2). This mutation may also have a negative impact on the virulence of this strain.

The *ascB-dapE* internalin cluster includes variable combinations of *inlC2, inlD, inlE, inlG* and *inlH* and has been suggested as of potential use as a marker for sublineage classification (Chen et al. 2012). Sublineages IIA, IIB and IIC were noted in this study. However, strain UCDL019 included another sublineage II variant, similar to that described by Dramsi et al. (1997). Another interesting feature of this locus was noted in UCDL187, where the flanking *ascB* β-glucosidase gene, as well as one of the hypothetical proteins in the locus, was identified as pseudogenes (Fig. S3).

The expression of most key *L. monocytogenes* virulence factors identified to date is under the control of *prfA*, the main virulence regulator; this regulator is responsible for the switch to in vivo pathogenesis, when the bacterium enters its mammalian host (Chakraborty et al. 1992; Freitag et al. 2009). One persistent isolate in this study, UCDL185, harboured a seven-nucleotide insertion in *prfA*, causing a downstream PMSC at amino acid position 185 (A185*). This mutation has been associated with attenuated virulence in vivo (Roche et al. 2005; López et al. 2013).

Apart for LIPI-1 and LIPI-2 (the former encoding the main virulence gene locus in *L. monocytogenes* and the latter encoding virulence factors in *L. ivanovii*), two additional pathogenicity islands of note have been described: LIPI-3 and LIPI-4. The LIPI-3 pathogenicity island encodes listeriolysin S, which is associated with increased strain virulence. This has been associated with functionality as a bacteriocin when expressed in the intestinal microenvironment, positively contributing to strains’ capacity to colonize this niche, and with a role as an alternative haemolysin/cytolysin (Gotter et al. 2008; Querda et al. 2017). In this study, only a single isolate harboured LIPI-3: UCDL037, a presumed non-persistent strain. The LIPI-4 pathogenicity island is associated with
Persistence of *Listeria* species

L. Palaiodimou *et al.*

hypervirulence in a subset of *L. monocytogenes* genetic clones, encoding a putative phosphotransferase system (Maury *et al.* 2016). No isolates in this study harboured LIPI-4. Overall, these results suggest that these additional pathogenicity islands are not common among food isolates and do not correlate persistence to increased/hypervirulence.

Taken together, our results suggest a lower virulence potential of persistent isolates in this study, due to the widespread prevalence of truncated InlA among the persistent *L. monocytogenes* strains, the lack of additional virulence factors such as LIPI-3 and LIPI-4 and the other notable mutations such as that of prfA in persistent strain UCDL187. Because persistence of pathogenic bacteria in FPEs can be associated with an increased risk to public health, due to an ongoing risk of cross-contamination of food products associated with the colonized environment, the attenuated virulence observed among persistent isolates in this study is positive from a food safety perspective. These results also suggest that in the facility studied, persistent strains were likely to be less virulent than other transient strains found in the same environment. None of the non-*monocytogenes* species in this study contained homologues of any of the virulence genes shown in Fig. 3.

The presence of mobile genetic elements typically gives rise to diverse functional variation in the *L. monocytogenes* accessory genome, although species-specific differences have also been noted (Glaser *et al.* 2001; Hain *et al.* 2006; Fox *et al.* 2016). Plasmids were found to contribute to variation across strains of the same species, as well as interspecies, with at least one plasmid present in 10 of the isolates included in this study (77%); of these, three isolates contained two plasmids (Fig. S4); These plasmids encoded a number of genetic markers related to stress resistance, as illustrated in Fig. 4. This included determinants related to disinfectant resistance (bcrABC and emrC), heavy metal resistance systems (including cadmium and copper resistance) and other stress resistance markers with roles in oxidative and temperature stress, such as clpB, clpL and NADH peroxidase. Interestingly, homologues of the same plasmid were found in multiple strains: for example, both *L. welshimeri* UCDL063 and UCDL122 strains contained identical plasmids (pUCDL063-1 and pUCDL122-1; Fig. S4); similarly, two ST9 strains (UCDL016 and UCDL133) contained similar plasmids (pUCDL016-1 and pUCDL133-1), and a smaller plasmid previously described in ST6 strains (Kropac *et al.* 2019), encoding *emrC*, was present in three isolates in this study (UCDL011, UCDL016 and UCDL133). Interestingly, all strains harbouring this small 4265-bp plasmid also harboured larger plasmids, but these larger plasmids did not encode bcrABC, nor did these strains carry the qacH-containing transposon Tn6188. The carriage of extrachromosomal plasmid DNA confers an associated fitness cost to strains; whether the presence of either disinfectant resistance plasmid leads to exclusion or unstable carriage of the other requires further investigation.

The *comK* gene is a known phage insertion hotspot in *L. monocytogenes* and may contain the variant of A118, or other, phage (Loessner *et al.* 2000; Orsi *et al.* 2008; Fox *et al.* 2016). Interestingly, the presence of a *comK* phage insertion (*φcomK*) has been suggested to play a role in colonization and persistence by functioning as a rapid adaptation island through recombination events (Vergheze *et al.* 2011). Analysis of isolates in this study found *φcomK* variants among both persistent (57%) and presumed non-persistent isolates (50%), suggesting that the presence of an insert does not predispose a strain to persistence. The *φ comK* genotypes also suggested that multiple recombination events had occurred, and no clear genotype responsible for persistence was apparent (Fig. S5). Interestingly among the isolates in this study, no *L. welshimeri* had a *φ comK* insert. To further investigate this, we examined the phage *attP* attachment site and the corresponding *attB* bacterial site. The *comK* phage *attP* site is unusual in that the core insert sequence is just three nucleotides in length (−GGA−). When comparing the insert site sequence across isolates in this study, all three *L. welshimeri* isolates contained an SNP in their *attB* (G/G; Fig. 5). To further elaborate this, we compared this region in four additional *L. welshimeri* strains (Fig. S8); all of these also harboured the ‘GGT’ variant. Taken together, these results may indicate that the *comK* gene in *L. welshimeri* may not be an insertion hotspot due to *attPlattB* sequence variation. This could be further investigated as more *L. welshimeri* genomes become available.

To investigate if similar genotypes were shared by persistent isolates, we conducted a pangenome analysis and constructed an associated maximum likelihood phylogenetic analysis of the strains in this study (Fig. S6). No clear segregation was noted based on pangenome genotype; similarly, considering plasmid carriage and *φcomK* status, no clear clustering was observed. In addition, further analysis of the CC9 subclade was undertaken, comparing core SNPs across the four associated strains. This subclade included both persistent and presumed non-persistent isolates, and the SNP differences supported a diverse strain cohort, with no clear segregation between persistent and presumed non-persistent strains, again reinforcing the observations of the pangenome phylogeny.

This study sought to further investigate persistence of *Listeria* species in FPEs by comparing cohorts of persistent and presumed non-persistent isolates collected from the same environment. This facilitated evaluation of related molecular mechanisms, in the context of an environment exerting similar selective pressures on associated
strains of *Listeria*. Taken together, the insights provided in this study do not point to a single genetic mechanism driving the persistence of *Listeria* strains in the FPE of their isolation. Persistent strains of *L. monocytogenes* were more likely to harbour mutations associated with hypovirulence and less invasive disease. Although disinfectant resistance markers were found in both persistent and presumed non-persistent strains, *qacH* was only identified among the persistent cohort. Persistent *L. welshimeri* strains harboured SSI-1, and strains of this species may be less prone to *comK* phage insertion due to *attB* site mutation.

**Acknowledgements**

The authors wish to sincerely thank the food business involved in this study. We thank the Institut Pasteur teams for the curation and maintenance of BIGSdb-Pasteur databases at http://bigsdb.pasteur.fr/.

**Conflict of Interest**

No conflict of interest declared.

**References**


Chambel, L., Sol, M., Fernandes, I., Barbosa, M., Zilhao, I., Barata, B., Jordan, S., Perni, S. *et al.* (2007) Occurrence and persistence of *Listeria* spp. in the environment of ewe and cow’s milk cheese dairies in Portugal unveiled by an integrated analysis of identification, typing and spatial-


Persistence of *Listeria* species


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** List of genes used for BLAST analysis, and associated NCBI accession numbers.

**Figure S2.** Comparative analysis of the *inlC* gene showing the wild type sequence (UCDL019), and the UCDL187 sequence harbouring a single nucleotide insertion in the poly(A) tract from nucleotide positions 7–15.

**Figure S3.** Comparative analysis of the *ascB-dapE* internalin cluster among strains in this study.

**Figure S4.** BLAST Ring comparison of plasmids identified among isolates in this study, showing sequence homology between different plasmids.

**Figure S5.** Comparative analysis of the *comK* gene and associated *comK* phage inserts.

**Figure S6.** Maximum likelihood analysis of strains in this study, based on a comparative pangenome analysis.

**Figure S7.** Core SNP analysis of the CC9 strains in this study.

**Figure S8.** Alignment of the *comK* gene sequence form *L. welshimeri* strains in this study, as well as an additional 4 strains taken from the NCBI genome database (strains CDPFHDLB, F4083, NCTC 11857, and SLCC5334).

**Table S1.** List of genes used for *BLAST* analysis, and associated NCBI accession numbers.