Genome-wide-analyses of Listeria monocytogenes from food-processing plants reveals clonal diversity and dates the emergence of persisting sequence types

Knudsen, Gitte Maegaard; Nielsen, Jesper Boye; Marvig, Rasmus Lykke; Ng, Yin; Worning, Peder; Westh, Henrik; Gram, Lone

Published in:
Environmental Microbiology Reports

Link to article, DOI:
10.1111/1758-2229.12552

Publication date:
2017

Document Version
Peer reviewed version

Genome-wide-analyses of *Listeria monocytogenes* from food-processing plants reveals clonal diversity and dates the emergence of persisting sequence types

Gitte M. Knudsen[1,5], Jesper Boye Nielsen[2], Rasmus L. Marvig[3], Yin Ng[1], Peder Worning[2], Henrik Westh[2,4] and Lone Gram[1]*

[1] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark

[2] MRSA KnowledgeCenter, Department of Clinical Microbiology, Hvidovre Hospital, Hvidovre, Denmark

[3] Center for Genomic Medicine, Rigshospitalet, Copenhagen, Denmark

[4] Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[5] Current address: Chr. Hansen A/S, Bøge Alle, Hørsholm, Denmark

Running title: Emergence and diversity of *L. monocytogenes*

Keywords: *Listeria monocytogenes*, MLST typing, genomics/functional genomics/comparative genomics, persistence

* Corresponding author

Lone Gram, Department of Biotechnology and Biomedicine, Matematiktorvet, Building 301, room 220, DK-2800 Kgs. Lyngby, Denmark

email: gram@bio.dtu.dk

Phone: +45 23 68 82 95

Fax: +45 45 93 28 09

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1758-2229.12552

ORIGINALITY-SIGNIFICANCE STATEMENT

The ability of some types of the food-borne pathogen *Listeria monocytogenes* to persist in the food-processing environment for years or decades remains unexplained. Persisting types are often the original source of food contamination and the cause of human listeriosis, which has a mortality of 20-30%. By comparative genomic analysis, we found a low genomic diversity of each sequence types (STs) and a low evolutionary rate. Independent analyses of evolutionary rates in three different persisting STs estimated each of these STs to have emerged around year 1910, potentially caused by acquisition of ST-specific genes that enable adaptation to and survival in the industrial food-processing environment. Whole genome sequencing is becoming the epidemiological tool for tracing food-borne disease outbreaks; however, the reliability of this method requires a detailed knowledge of genomic variation within the pathogen investigated.

SUMMARY

Whole genome sequencing is increasing used in epidemiology, e.g. for tracing outbreaks of food-borne diseases. This requires in-depth understanding of pathogen emergence, persistence, and genomic diversity along the food production chain including in food processing plants. We sequenced the genomes of 80 isolates of *Listeria monocytogenes* sampled from Danish food processing plants over a time-period of 20 years, and analyzed the sequences together with 10 public available reference genomes to advance our understanding of inter- and intra-plant genomic diversity of *L. monocytogenes*. Except for three persisting sequence types (ST) based on Multi Locus Sequence Typing (MLST) being ST7, ST8 and ST121, long-term persistence of clonal groups was limited, and new clones were introduced continuously, potentially from raw materials. No particular gene could be linked to the persistence phenotype. Using time-based phylogenetic analyses of the persistent STs, we estimate the *L. monocytogenes* evolutionary rate to be 0.18-0.35 SNPs/year, suggesting that the persistent STs emerged approximately 100 years ago, which correlates with the onset of industrialization and globalization of the food market.

170 words

INTRODUCTION

The ability of a microorganism to survive and grow for longer periods in a specific niche is referred to as persistence and is observed by repeated isolation over time of clonal or close-to-clonal strains (Carpentier and Cerf, 2011; Ferreira et al., 2014; Larsen et al., 2014). However, it is often not known if persistence is clonal ('true persistence') or the result of re-introduction and re-isolation of the same type (Ferreira et al., 2014). The term 'persistence' is not well defined and is used both for long-term survival of specific molecular types in an environment, as it is used in this study (Ferreira et al., 2014) but also for persistence in patients as may be caused by formation of persister cells (Lewis, 2008) or by particular phenotypes such as biofilm formation (Costerton et al., 1999; Romling and Balsalobre, 2012). Several foodborne pathogens can persist at farm level and/or in the food processing environment for years (Ferreira et al., 2014; Larsen et al., 2014; Martin et al., 2014; Orsi et al., 2008a; Wulff et al., 2006). The molecular mechanisms underlying persistence are not well understood but biofilm formation or tolerance to desiccation or biocides may be involved (Cossart, 2011; Ferreira et al., 2014; Larsen et al., 2014).

The food-borne pathogen *L. monocytogenes* is a prime example of a bacterium that can persist in food processing environments, where specific molecular types are isolated repeatedly (Hein et al., 2011; Malley et al., 2015; Orsi et al., 2008a; Ortiz et al., 2014; Vogel et al., 2001; Wulff et al., 2006). Listeriosis is caused by consumption of food contaminated with *L. monocytogenes* and mainly affects elderly and immuno-compromised patients with a mortality of 20-30% (Cossart, 2011). *L. monocytogenes* is divided into four Lineages, of which lineage I isolates are often associated with human outbreaks, whereas Lineage II isolates are often associated with food and food processing environments (Maury et al., 2016; Valderrama and Cutter, 2013). Sequence type (ST) 121 (ST121) identified by Multi Locus Sequence Typing (MLST) is a persistent ST, and has been found in fish processing plants (Wulff et al., 2006), meat processing plants (Martin et al., 2014; Morganti et al., 2015) and in many other food processing plants (Chiara et al., 2014; Ciolacu et al., 2014; Hein et al., 2011; Schmitz-Esser et al., 2015). ST121 is the most common ST isolated from food samples; however, it is rarely isolated from infected patients (Althaus et al., 2014; Ebner et al., 2015; Maury et al., 2016).

Typing, e.g. by MLST, can provide information on persistence; however, it does not allow analyses of clonality due to low discriminatory power. We sequenced the genomes of two ST121 isolates isolated six years apart in two different smoked fish productions and found a remarkable genomic stability, as they only differed by 18 single nucleotide polymorphisms (SNPs; Holch et al., 2013). Similarly, isolates of ST11 that persisted in an American production plant and caused listeriosis in 1988 and 2000, differed by only one SNP in the core genome (Orsi et al., 2008a). This low diversity might be a result of that *L. monocytogenes* has a conserved genome with a high level of core genes and low recombination rates (den Bakker et al., 2008; 2010; 2013; Hain et al., 2012; Moura et al., 2016).

Whole genome sequencing (WGS) is increasingly being used in epidemiology (den Bakker et al., 2014; Kvistholm Jensen et al., 2016; Kwong et al., 2016; Le and Diep, 2013; Moura et al., 2016), and WGS analysis was essential for unraveling a Danish *L. monocytogenes* ST224 outbreak in 2014 (Kvistholm Jensen et al., 2016). However, the genomic stability of *L. monocytogenes* could cause clonal or near-clonal isolates to be isolated from different processing plants and regions, thereby

obscuring the true source. The purpose of the present study was to investigate the genomic diversity of *L. monocytogenes* strains isolated over a 20-year period as food processing plant contaminants and to address if specific genetic features could be associated with the persistent capability.

RESULTS

Persisting and non-persisting *L. monocytogenes* isolates spanning 20 years. In 2013-2014, 233 *L. monocytogenes* strains were isolated from raw material, processing surfaces and product (Table S1 and S2) from processing plants 3 and 5 but no systematic seasonal variation was observed (Table S3). 170 strains were MLST typed (Table S1; Ragon et al., 2008) and 98.8% of the isolates belonged to Lineage II and only two strains belonged to Lineage I (ST1). The strains divided into thirteen different STs, and ST7, ST8, ST121 and ST398 were isolated from both processing plants. Nine STs were detected multiple times (≥3; Table S4), indicating that some STs persist or are continuously introduced, and that others occur sporadically.

From previous studies (Ben Embarek and Huss, 1993; Larsen et al., 2002; Vogel et al., 2001; Wulff et al., 2006), 37 strains were selected and included in the study to broaden the range of food processing plants (*n*=11) and the isolation period (1995-2004). MLST-typing of these isolates resulted in 10 different STs (Table S4) and ST7, ST8 and ST121 were isolated regularly during both sampling periods. We therefore in this study defined them as persisting STs.

Limited clonal persistence in persistent *L. monocytogenes* sequence types. Ninety strains were whole-genome sequenced and SNPs analyzed (Table S5) including 38 isolates 2013-14, 27 isolates 1995-2004 and 10 isolates of non-smoke houses origin (Fig. 1). Thirteen ST121 isolates isolated outside Denmark (Chiara et al., 2014; Fox et al., 2011; Lopez-Alonso et al., 2015; Stessl et al., 2014) were included as this appeared to be the dominant persistent ST type world-wide. Finally, two reference strains F2365 (Nelson et al., 2004) and EGDe (Glaser et al., 2001) were included to represent Lineage I and II, respectively. SNP analysis was performed based on alignment to the genome of the reference strain EGDe (21), and the SNP-based phylogenetic relationship based the core genome was in concordance with lineage and previously identified MLSTs (Fig. 2a).

SNP differences in the 47 ST121 isolates varied from zero to 203 (Fig. 2b and Table S6 and S7) and the ST121 isolates displaying the largest SNP variation (Y16 and Y20) were isolated from the same processing plant (Plant 5) and in the same year (2014). Using an arbitrary and conservative limit of ≤4 core genome SNPs to define a clonal group of isolates, we detected seven clonal groups of ST121 isolates (Fig. 2b and Table S7 and S8). Clonal groups with ≤4 SNPs were all specific to the same processing plant, indicating that clonal isolates originated from one single source of contamination and isolates within plant-specific clonal groups were isolated over a time span from zero to 16 months (Table S8). Plant-specific clonal groups were also found in ST6, ST7, ST101 and ST204 (Table S7 and S8). ST193 was repeatedly isolated from processing plant 3 (45 isolates in total over 16 months) and was not observed in the other production plant or among the earlier isolates. Six of these 45 isolates were genome sequenced (clonal group *ST193_M*). The SNP variation was 7 to 16, which is a high number of SNPs compared to the number observed within the other STs.

SNP differences could not directly related to time or space differences. Thus, clonal group *ST121_A* included only two isolates, M40-1 and M103-1, from plant M isolated one month apart. A third isolate M165-1 was isolated from the same processor, but differed by 8-9 SNPs to clonal group *ST121_A*. A similar distance was found to a fourth isolate X10 (nine SNPs) which was isolated 12 years later than M40-1 at another processing plant.

Evolutionary rate of *L. monocytogenes* in the food-processing environment. The low number of SNPs identified in the core genomes between isolates separated by time and location (Table S7 and S8) is indicative of a common source. In line with the low SNP variation for M40-1 and X10, three ST8 isolates (Y14, R479a and 2R167a) were isolated 15-16 years apart and differed by 17-18 SNPs. By time-based phylogenetic analysis using BEAST (Drummond et al., 2012), we estimated the evolutionary rate of *L. monocytogenes* within the STs ST7, ST8 and ST121 and found rates of 0.18, 0.30, and 0.36 SNPs/year, respectively (Table 1). These rates are among the lowest evolutionary rates determined for any bacterial pathogen. The estimated time of common ancestor of the three persisting STs ST7, ST8 and ST121 was year 1916, 1905, and 1908, respectively, indicating that they have emerged independently in the same time period in the beginning of the twentieth century (Table 1; Fig. S1a, S1b and S2). This correlates with peak time of the industrialization and creation of new niches in food processing plants leading to the emergence of human infection and long-distance dispersal via the food chain.

The *Listeria monocytogenes* core and pan genome. The 90 isolates that represented 15 STs shared a core genome of 2,381 genes and a pan genome of 5,454 genes. The largest variation in gene content was found between isolates belonging to different STs which on average differed by 447 genes (range 219-735 genes), whereas isolates within the same ST on average differed by 146 genes (range 0-356 genes). Isolates from clonal groups *ST121_B*, *ST121_C*, *ST121_F* and *ST193_M* (Table S8) had a high variation in gene content (maximum number of differences was 104, 39, 95, and 139 genes, respectively) due to the presence of different prophages. In contrast, less than five genes differentiated isolates within the other nine clonal groups.

Genes potentially related to persistence. It is not understood why some molecular types of *L. monocytogenes* persist in the food-processing environment; however, it could be related to the ability to withstand stress (Ferreira et al., 2014; Larsen et al., 2014). Among other, we investigated if differences in the alternative sigma factor SigB is involved in stress response could be linked to persistence ability however, only one amino acid change (Y226F) was identified between Lineage I and II isolates in the protein sequences of SigB. The five-gene islet Stress Survival Islet-1 (SSI-1) (Hein et al., 2011; Ryan et al., 2010) and the 12.5 kbp insertion with *LM6179_0173* are suspected to be involved in persistence (Schmitz-Esser et al., 2015). However, we did not find that these presumed persistence genes were limited to the three persisting STs ST7, ST8 and ST121 (Table S9).

Benzalkonium chloride (BAC)-resistant *L. monocytogenes*-isolates have been isolated repeatedly from a meat production plant indicating a potential link to persistence (Ortiz et al., 2014) and some

obscuring the true source. The purpose of the present study was to investigate the genomic diversity of *L. monocytogenes* strains isolated over a 20-year period as food processing plant contaminants and to address if specific genetic features could be associated with the persistent capability.

RESULTS

Persisting and non-persisting *L. monocytogenes* isolates spanning 20 years. In 2013-2014, 233 *L. monocytogenes* strains were isolated from raw material, processing surfaces and product (Table S1 and S2) from processing plants 3 and 5 but no systematic seasonal variation was observed (Table S3). 170 strains were MLST typed (Table S1; Ragon et al., 2008) and 98.8% of the isolates belonged to Lineage II and only two strains belonged to Lineage I (ST1). The strains divided into thirteen different STs, and ST7, ST8, ST121 and ST398 were isolated from both processing plants. Nine STs were detected multiple times (≥3; Table S4), indicating that some STs persist or are continuously introduced, and that others occur sporadically.

From previous studies (Ben Embarek and Huss, 1993; Larsen et al., 2002; Vogel et al., 2001; Wulff et al., 2006), 37 strains were selected and included in the study to broaden the range of food processing plants (*n*=11) and the isolation period (1995-2004). MLST-typing of these isolates resulted in 10 different STs (Table S4) and ST7, ST8 and ST121 were isolated regularly during both sampling periods. We therefore in this study defined them as persisting STs.

Limited clonal persistence in persistent *L. monocytogenes* sequence types. Ninety strains were whole-genome sequenced and SNPs analyzed (Table S5) including 38 isolates 2013-14, 27 isolates 1995-2004 and 10 isolates of non-smoke houses origin (Fig. 1). Thirteen ST121 isolates isolated outside Denmark (Chiara et al., 2014; Fox et al., 2011; Lopez-Alonso et al., 2015; Stessl et al., 2014) were included as this appeared to be the dominant persistent ST type world-wide. Finally, two reference strains F2365 (Nelson et al., 2004) and EGDe (Glaser et al., 2001) were included to represent Lineage I and II, respectively. SNP analysis was performed based on alignment to the genome of the reference strain EGDe (21), and the SNP-based phylogenetic relationship based the core genome was in concordance with lineage and previously identified MLSTs (Fig. 2a).

SNP differences in the 47 ST121 isolates varied from zero to 203 (Fig. 2b and Table S6 and S7) and the ST121 isolates displaying the largest SNP variation (Y16 and Y20) were isolated from the same processing plant (Plant 5) and in the same year (2014). Using an arbitrary and conservative limit of ≤4 core genome SNPs to define a clonal group of isolates, we detected seven clonal groups of ST121 isolates (Fig. 2b and Table S7 and S8). Clonal groups with ≤4 SNPs were all specific to the same processing plant, indicating that clonal isolates originated from one single source of contamination and isolates within plant-specific clonal groups were isolated over a time span from zero to 16 months (Table S8). Plant-specific clonal groups were also found in ST6, ST7, ST101 and ST204 (Table S7 and S8). ST193 was repeatedly isolated from processing plant 3 (45 isolates in total over 16 months) and was not observed in the other production plant or among the earlier isolates. Six of these 45 isolates were genome sequenced (clonal group *ST193_M*). The SNP variation was 7 to 16, which is a high number of SNPs compared to the number observed within the other STs.

The ST-specific genes found in the three persistent ST7, ST8 and ST121 were either involved in resistance towards phages and disinfectant, which potentially could be linked to a survival advantage in the food-processing environment.

DISCUSSION

Molecular typing methods have been extensively used to trace bacterial disease outbreaks or determine persistence of pathogenic bacteria such as *L. monocytogenes* (Hein et al., 2011; Malley et al., 2015; Orsi et al., 2008a; Ortiz et al., 2014; Vogel et al., 2001; Wulff et al., 2006). However, these methods cannot determine if a group of bacterial isolates is 'truly persisting' (clonal) or if a non-related isolate of the same type is repeatedly introduced into the environment. Here, we considered *L. monocytogenes* ST7, ST8 and ST121 as persisting STs under the assumption that repeated isolation was equal to persistence. However, the SNP analysis demonstrated that some isolates were likely 'truly' persisting clonal groups, but also that new *L. monocytogenes* clones of the same ST were introduced continuously probably by raw materials. Raw materials were repeatedly contaminated and on several occasions, isolates belonging to the same clonal groups were sampled from the environment and from products. We defined 'true persistence' based on Ferreira *et al.* (2014) and used an arbitrary limit of ≤4 SNPs for clonal groups combined with isolation over a longer period of time (>2 months). This categorized ST6, ST7, ST101 and ST121 as 'true persisting' STs. To the best of our knowledge, this is the first longitudinal study using WGS to determine the 'true' persistence of a bacterial pathogen in processing plants. However, persistence of *L. monocytogenes* was previously studied using WGS (Morganti et al., 2015; Stasiewicz et al., 2015) and in line with Stasiewicz *et al.* (2015), we identified ST6 as a 'true persisting' ST. Kwong *et al.* (2016), analyzed 423 *L. monocytogenes* strains from foods or patients and observed over 200 SNPs within PFGE groups, consistent with the large number of SNPs in the MLST groups in this study. PFGE groups that linked by epidemiology had less than 10 SNPs, and in patients typically had less than five SNPs. However, we and Morganti *et al.* (2015) show that close-to-clonal groups (4 < SNPs < 20) can be found in different processing plants and raise caution when using a higher clonality limit than 4 SNPs for outbreak investigations unless there are epidemiological data as found by Wang *et al.* (2015). Our study highlights that outbreak investigations of *L. monocytogenes* must consider the low recombination rates (Stasiewicz et al., 2015) or the isolation of near-clonal isolates in different processing plants (Morganti et al., 2015). This is also supported by *Salmonella* data where as few as 15-22 SNPs may differentiate outbreak and non-outbreak isolates (Leekitcharoenphon et al., 2014). Collectively these and our study underline the importance of epidemiological data to support WGS for outbreak investigations.

All of the three persisting STs (ST7, ST8 and ST121) had low evolutionary rates in the range of 0.18 to 0.36 SNPs/year. This is at the same level as estimated by Moura *et al*. (2016) finding an evolutionary rate of 0.41 and 0.38 substitutions per 1.58 MB per year for SL1 (mainly ST1) and SL9 (mainly ST9), respectively. We believe this is the lowest evolutionary rate reported for a bacteria and it is even lower that the rate of *Mycobacterium tuberculosis* estimated to 0.3–0.5 SNPs/year (Ford et al., 2013). The low evolutionary rate and the large number of core genes (2,381 genes) is consistent with a stable genome as found in other studies (den Bakker et al., 2008; 2010; 2013; Hain et al., 2012; Kuenne et al., 2013; Moura et al., 2016; Orsi et al., 2008b).

It is not known, if the ability to be 'truly persisting' is an isolate or ST-specific phenotype. However, based on the low number of SNPs within one ST and high genomic stability within ST, we hypotheses that it is ST dependent and not isolate specific. Thus, all isolates within a ST have the potential to be 'truly persisting' under optimal condition, which is consistent with a recent study of persisting ST8 (Fagerlund et al., 2016). Secondly, this is also in line with many years of research (Hein et al., 2011; Malley et al., 2015; Orsi et al., 2008a; Ortiz et al., 2014; Vogel et al., 2001; Wulff et al., 2006) that have used molecular typing methods, albeit with lower discriminatory power, which defined isolates of the same molecular type as persisting. Finally, it is consistent with the stability of the gene pool within a ST and supported by the study of Stasiewicz *et al*. (2015) that could not identify persisting genes by enrichment of persisting and non-persisting isolates from the same ST. The ST dependency is also found for putative persistence genes, virulence genes and a biocide resistance genes (Maury et al., 2016; Moura et al., 2016; Roche et al., 2009). This high level of ST dependency allowed us to identify ST-specific genes in ST7, ST8 and ST121 and also Fagerlund *et al.* (2016) found ST8 specific genes. While it is possible that ST specific genes increases the likelihood of STs to persist in the processing environment, as also speculated by others (Schmitz-Esser et al., 2015); functional studies and better annotation of the relevant genes is necessary to help build evidence for this.

Prophages were likely a major source of diversity between isolates of the same ST and also the finding of identical prophages in different STs indicated a very high recombination in the prophages as has been found in other studies (Fagerlund et al., 2016; Orsi et al., 2008a; Stasiewicz et al., 2015). Interestingly, several of the ST specific genes in ST7, ST8 and ST121 were involved in phage resistance. Wang *et al.* (2015) recently suggested that 'It is not all about SNP' as they used the presence of rare mobile element to support that environmental samples from a processing plant having 19-20 SNPs to clinical isolates were the cause of an outbreak. To fully understand the diversity of *L. monocytogenes* prophages, further studies are needed to characterize each prophage variation and their potential role in persistence.

L. monocytogenes is not only a food-borne human pathogen but also a ubiquitous saprophyte (Cossart, 2011; Freitag et al., 2009), and it is therefore naturally present on many food raw materials. We found as Fagerlund *et al.* (2016) that different STs and clones were found in raw material. Fagerlund *et al.* (2016) found that three ST8 isolates (MF3949, MF4077 and MF4245) from Norwegian salmon procession plant only had 8-16 SNPs to isolate R479a from a Danish fish smokehouse. R479a has 18 SNPs to Y14, which is isolated from raw material of Norwegian Salmon at plant 5, but Y14 has 14-24 SNPs to the three Norwegian isolates. Thus, the two ST8 isolates from the early isolation (R479a and 2R167a) are more closely related to the Norwegian isolates than our recent Y14 isolate. This highlights that the food producers must have a high focus on eliminating contaminations from raw material in the food processing.

The estimated time of the common ancestor of the three persisting STs, ST7, ST8 and ST121 is 1905-1916, which correlates with the onset of industrialization in Denmark and globalization of the food market (Anonymous, 2016). This timing is consistent with the root of major sublineages of *L. monocytogenes* as determined by Moura *et al.* (2016) based on 1,696 isolates. Although these estimates should be interpreted with caution, the correlation could indicate that *L. monocytogenes* have evolved with the specialized environments in the dairies and slaughterhouses that provided a

new niche. This, in combination with the cooling chains may have caused specific STs to evolve characteristics that improved survival in these environments.

EXPERIMENTAL PROCEDURES

Bacterial strains and growth conditions. Bacterial strains sequenced in this study are listed in Table S5. *L. monocytogenes*-positive samples were received in 2013-2014 from two smoked fish processing plants and isolated strains on *Listeria* selective media. Bacterial stock cultures were stored at -80°C and strains grown on Brain Heart Infusion (BHI) agar or in BHI at 37°C with shaking (250 rpm).

Multi-Locus Sequence Typing (MLST). Genomic DNA was extracted with Dynal Dynabeads DNA Direct System (Invitrogen) and amplified using TEMPase Hot Start 2x Master Mix Blue II (Ampliqon). Sequencing was performed by GATC Biotech AG (Köln, Germany) and Listeria Sequence Typing at Institut Pasteur MLST was used for allele analysis (Institute Pasteur, 2014), Ragon *et al*. (2008)) (http://bigsdb.web.pasteur.fr/listeria/listeria.html).

Whole genome sequencing (WGS). Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen) according to the manufacture's protocol except that cells were pre-lysed with lysozyme for 30 min at 37°C and the proteinase K treatment was extended to 30 min. DNA concentrations were measured and normalized using a Qubit flourometer (Invitrogen, UK) and libraries were prepared using Illumina Nextera XT DNA preparation kit. Sequencing of libraries was performed on an Illumina MiSeq instrument (Illumina, USA). Samples were multiplexed to 24 samples per sequencing reaction and sequenced using 2x150bp paired-end reads. Quality control and de-multiplexing of reads was performed automatically by MiSeq software. The average coverage was 72 (ranging from 22 to 191). Genomes were assembled *de novo* using Velvet version 1.0.11 (Zerbino and Birney, 2008) and VelvetOptimiser version 2.1.7 (Zerbino, 2010) . The assemblies were optimized to maximize N50 and VelvetOptimiser used every odd kmer value between 51 and 137. The raw reads have been deposited to European Nucleotide Archive (ENA) under accession number PRJEB14063.

SNP prediction and phylogenetic analysis using core genome. Ten reference genomes were downloaded from NCBI and 100 bp paired-end reads were generated by selecting random positions in the genome. The 100 bp upstream of each position was used as the first read of a pair and the reverse complement of the 100 bp downstream was used as the second read. The fastq files were constructed to give coverage of 100 and the Phred score at each position is a random value between 34 and 39. To identify the core genome of strains EGDe and 6179, all positions in the reference genomes that were covered by less than five reads in any of the 90 samples were defined as accessory positions and these positions were taken out of the analysis when the distance matrices were calculated. The accessory positions accounted for 191,940 bp out of 2,944,528 bp in the EDGe and 229,610 bp out of 3,010,620 bp in the 6179 genome. For the SNP calling, reads were mapped against *L. monocytogenes* reference genomes EGDe (NC003210) and *L. monocytogenes* 6179

(HG813249) using stampy version 1.0.22 (Lunter and Goodson, 2011) and samtools version 0.1.13 with options –M0 –Q30 –o40 –e20 –h100 –m2 –D –S [62]. Maximum-likelihood phylogenetic trees were inferred by RAxML version 8.2.4 (default settings) (Stamatakis, 2014) based on an alignment consisting of concatenated nucleotide calls from the 58,801 polymorphic positions in the core genome. RAxML was run with default settings using a general time reversible model of nucleotide substitution (option -m GTRCAT).

Evolutionary rate analysis of ST7, ST8 and ST121. TempEst (Rambaut et al., 2016) was used to visualize the temporal signal in maximum-likelihood phylogenies (Fig. S4), and Bayesian analysis of evolutionary rates was performed using BEAST version 1.8 (Drummond et al., 2012). BEAST analyses were run with default settings using a strict molecular clock model and a HKY substitution model, which distinguishes between the rate of transitions and transversions and allows unequal base frequencies. Evolutionary rates were calculated from chains of 10 million steps, sampled every 1,000 steps. The first 1 million steps of each chain were discarded as a burn-in. The effective sample size of all parameters were >200 as calculated by Tracer version 1.5 (available from http://beast.bio.ed.ac.uk/Tracer), which was also used to calculate the 95% highest posterior density (HPD) confidence intervals of the evolutionary rate (*i.e.* an interval within which the modeled parameter resides with 95% probability).

Analysis of core genome, pan genome and accessory genome. All genomes were annotated using Prokka version 1.10 (Seemann, 2014), and gene nucleotide sequences were clustered using CD-HIT version 4.6 (Li et al., 2001) to obtain non-redundant sets of genes (i. e. the pan-genome) with a local sequence identify of at least 90% in an alignment covering at least 80% of the shorter sequence (CD-HIT-EST package with settings -aS 0.8 -c 0.9'). Hereby, we obtained a set of 5,454 genes with a minimum length of 100 nt. Genes were aligned against each of the 90 assembled genomes using BLAST version 2.2.30+ (Shiryev et al., 2007), and a gene was defined as being present in the genome if an alignment with a sequence identity of >90% covering >25% of the gene was obtained. In pairwise comparisons of gene content, genes were only considered absent if no alignment covering ≥25% of the gene was obtained.

Genome mining for presumptive persistence genes. Sequence analysis of SigB protein was performed in CLC Main Workbench 7 (Aarhus, Denmark) using BLAST, alignments and Mega6 (Tamura et al., 2013) to create Maximum-likelihood phylogenetic trees of DNA and protein sequences. Biocide resistance genes were identified using a protein sequence database of 325 confirmed Biocide Resistance Genes from the BacMet database (Pal et al., 2014) and BioEdit as search tool (Hall, 1999). The presence of phage genes was investigated by PhiSpy (Akhter et al., 2012) and PHAST (Zhou et al., 2011) was used for verification of the results. The presence of antibiotic resistance genes was analyzed using ResFinder (Zankari et al., 2012). ST-specific genes were validated against NCBI database for the presence in other sequence types.

ACKNOWLEDGEMENTS

We would like to thank our collaborative companies for the collection of *L. monocytogenes* isolates. This research was supported by Grønt Udviklings- og Demonstrations Program (GUDP) from the Ministry of Food, Agriculture and Fisheries (Grant 34009-12-05-06).

ABBREVIATIONS

ST: Sequence Type

SNP: Single Nucleotide Polymorphism

MLST: MultiLocus Sequence Typing

WGS: Whole genome sequencing

REFERENCES

Akhter,S., Aziz,R.K., and Edwards,R.A. (2012) PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res* **40:** e126.

Althaus,D., Lehner,A., Brisse,S., Maury,M., Tasara,T., and Stephan,R. (2014) Characterization of *Listeria monocytogenes* strains isolated during 2011-2013 from human infections in Switzerland. *Foodborne Pathog Dis* **11:** 753-758.

Anonymous (2016). Danish cooperative movement [WWW document]. URL https://en.wikipedia.org/wiki/Danish_cooperative_movement

Ben Embarek,P.K. and Huss,H.H. (1993) Heat resistance of *Listeria monocytogenes* in vacuum packaged pasteurized fish fillets. *Int J Food Microbiol* **20:** 85-95.

Carpentier,B. and Cerf,O. (2011) Review - Persistence of *Listeria monocytogenes* in food industry equipment and premises. *Int J Food Microbiol* **145:** 1-8.

Chiara,M., D'Erchia,A.M., Manzari,C., Minotto,A., Montagna,C., Addante,N. et al. (2014) Draft Genome Sequences of Six *Listeria monocytogenes* Strains Isolated from Dairy Products from a Processing Plant in Southern Italy. *Genome Announc* **2**.

Ciolacu,L., Nicolau,A.I., Wagner,M., and Rychli,K. (2014) *Listeria monocytogenes* isolated from food samples from a Romanian black market show distinct virulence profiles. *Int J Food Microbiol*.

Cossart,P. (2011) Illuminating the landscape of host-pathogen interactions with the bacterium *Listeria monocytogenes*. *Proc Natl Acad Sci U S A* **108:** 19484-19491.

Costerton,J.W., Stewart,P.S., and Greenberg,E.P. (1999) Bacterial biofilms: a common cause of persistent infections. *Science* **284:** 1318-1322.

den Bakker,H.C., Allard,M.W., Bopp,D., Brown,E.W., Fontana,J., Iqbal,Z. et al. (2014) Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis. *Emerg Infect Dis* **20:** 1306-1314.

den Bakker,H.C., Cummings,C.A., Ferreira,V., Vatta,P., Orsi,R.H., Degoricija,L. et al. (2010) Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC Genomics* **11:** 688.

den Bakker,H.C., Desjardins,C.A., Griggs,A.D., Peters,J.E., Zeng,Q., Young,S.K. et al. (2013) Evolutionary dynamics of the accessory genome of *Listeria monocytogenes*. *PLoS One* **8:** e67511.

den Bakker,H.C., Didelot,X., Fortes,E.D., Nightingale,K.K., and Wiedmann,M. (2008) Lineage specific recombination rates and microevolution in *Listeria monocytogenes*. *BMC Evol Biol* **8:** 277.

Drummond,A.J., Suchard,M.A., Xie,D., and Rambaut,A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29:** 1969-1973.

Ebner,R., Stephan,R., Althaus,D., Brisse,S., Maury,M., and Tasara,T. (2015) Phenotypic and genotypic characteristics of *Listeria monocytogenes* strains isolated during 2011-2014 from different food matrices in Switzerland. *Food Control* **57:** 321-326.

Fagerlund,A., Langsrud,S., Schirmer,B.C., Møretrø,T., and Heir,E. (2016) Genome Analysis of *Listeria monocytogenes* Sequence Type 8 Strains Persisting in Salmon and Poultry Processing Environments and Comparison with Related Strains. *PLoS One* **11:** e0151117.

Ferreira,V., Wiedmann,M., Teixeira,P., and Stasiewicz,M.J. (2014) *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. *J Food Prot* **77:** 150-170.

Ford,C.B., Shah,R.R., Maeda,M.K., Gagneux,S., Murray,M.B., Cohen,T. et al. (2013) *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet* **45:** 784-790.

Fox,E., Hunt,K., O'Brien,M., and Jordan,K. (2011) *Listeria monocytogenes* in Irish Farmhouse cheese processing environments. *Int J Food Microbiol* **145 Suppl 1:** S39-S45.

Freitag,N.E., Port,G.C., and Miner,M.D. (2009) *Listeria monocytogenes* - from saprophyte to intracellular pathogen. *Nat Rev Microbiol* **7:** 623-628.

Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A., Baquero,F. et al. (2001) Comparative genomics of *Listeria* species. *Science* **294:** 849-852.

Hain,T., Ghai,R., Billion,A., Kuenne,C.T., Steinweg,C., Izar,B. et al. (2012) Comparative genomics and transcriptomics of lineages I, II, and III strains of *Listeria monocytogenes*. *BMC Genomics* **13:** 144.

Hall,T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series 41, 95-98. 1999.

Ref Type: Conference Proceeding

Hein,I., Klinger,S., Dooms,M., Flekna,G., Stessl,B., Leclercq,A. et al. (2011) Stress survival islet 1 (SSI-1) survey in *Listeria monocytogenes* reveals an insert common to *Listeria innocua* in sequence type 121 *L. monocytogenes* strains. *Appl Environ Microbiol* **77:** 2169-2173.

Holch,A., Webb,K., Lukjancenko,O., Ussery,D., Rosenthal,B.M., and Gram,L. (2013) Genome sequencing identifies two nearly unchanged strains of persistent *Listeria monocytogenes* isolated at two different fish processing plants sampled 6 years apart. *Appl Environ Microbiol* **79:** 2944-2951.

Institute Pasteur (2014). *Listeria monocytogenes* MLST Database [WWW document]. URL http://www.pasteur.fr/recherche/genopole/PF8/mlst/Lmono.html

Kuenne,C., Billion,A., Mraheil,M.A., Strittmatter,A., Daniel,R., Goesmann,A. et al. (2013) Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Genomics* **14:** 47.

Kvistholm Jensen,A., Nielsen,E.M., Björkman,J.T., Jensen,T., Müller,L., Persson,S. et al. (2016) Whole-genome sequencing used to investigate a nationwide outbreak of listeriosis caused by ready-to-eat delicatessen meat, Denmark, 2014. *Clin Infect Dis* **DOI: 10.1093/cid/ciw192**.

Kwong,J.C., Mercoulia,K., Tomita,T., Easton,M., Li,H.Y., Bulach,D.M. et al. (2016) Prospective Whole-Genome Sequencing Enhances National Surveillance of *Listeria monocytogenes*. *J Clin Microbiol* **54:** 333-342.

Larsen,C.N., Nørrung,B., Sommer,H.M., and Jakobsen,M. (2002) In vitro and in vivo invasiveness of different pulsed-field gel electrophoresis types of *Listeria monocytogenes*. *Appl Environ Microbiol* **68:** 5698-5703.

Larsen,M.H., Dalmasso,M., Ingmer,H., Langsrud,S., Malakauskas,M., Mader,A. et al. (2014) Persistence of foodborne pathogens and their control in primary and secondary food production chains. *Food Control* **44:** 92-109.

Le,V.T. and Diep,B.A. (2013) Selected insights from application of whole-genome sequencing for outbreak investigations. *Curr Opin Crit Care* **19:** 432-439.

Lee,S., Ward,T.J., Siletzky,R.M., and Kathariou,S. (2012) Two novel type II restriction-modification systems occupying genomically equivalent locations on the chromosomes of *Listeria monocytogenes* strains. *Appl Environ Microbiol* **78:** 2623-2630.

Leekitcharoenphon,P., Nielsen,E.M., Kaas,R.S., Lund,O., and Aarestrup,F.M. (2014) Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. *PLoS One* **9:** e87991.

Lewis,K. (2008) Multidrug tolerance of biofilms and persister cells. *Curr Top Microbiol Immunol* **322:** 107-131.

Li,W., Jaroszewski,L., and Godzik,A. (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17:** 282-283.

Lopez-Alonso,V., Ortiz,S., and Martinez-Suarez,J.V. (2015) Genome Sequences of Five Disinfectant-Resistant *Listeria monocytogenes* Strains from Two Iberian Pork-Processing Plants. *Genome Announc* **3:** e00077-15.

Lunter,G. and Goodson,M. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* **21:** 936-939.

Malley,T.J., Butts,J., and Wiedmann,M. (2015) Seek and Destroy Process: *Listeria monocytogenes* Process Controls in the Ready-to-Eat Meat and Poultry Industry. *J Food Prot* **78:** 436-445.

Martin,B., Perich,A., Gomez,D., Yanguela,J., Rodriguez,A., Garriga,M., and Aymerich,T. (2014) Diversity and distribution of *Listeria monocytogenes* in meat processing plants. *Food Microbiol* **44:** 119-127.

Maury,M.M., Tsai,Y.H., Charlier,C., Touchon,M., Chenal-Francisque,V., Leclercq,A. et al. (2016) Uncovering *Listeria monocytogenes* hypervirulence by harnessing its biodiversity. *Nat Genet* **48:** 308-313.

Morganti,M., Scaltriti,E., Cozzolino,P., Bolzoni,L., Casadei,G., Pierantoni,M. et al. (2015) Genetic analysis reveals the processing dependent and clonal contamination pattern of *Listeria monocytogenes* in cured ham food-chain. *Appl Environ Microbiol* **82:** 822-31.

Moura,A., Criscuolo,A., Pouseele,H., Maury,M.M., Leclercq,A., Tarr,C. et al. (2016) Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nat Microbiol* **2:** 16185.

Mûller,A., Rychli,K., Muhterem-Uyar,M., Zaiser,A., Stessl,B., Guinane,C.M. et al. (2013) Tn6188 - a novel transposon in *Listeria monocytogenes* responsible for tolerance to benzalkonium chloride. *PLoS One* **8:** e76835.

Nelson,K.E., Fouts,D.E., Mongodin,E.F., Ravel,J., DeBoy,R.T., Kolonay,J.F. et al. (2004) Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species. *Nucleic Acids Res* **32:** 2386-2395.

Orsi,R.H., Borowsky,M.L., Lauer,P., Young,S.K., Nusbaum,C., Galagan,J.E. et al. (2008a) Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment. *BMC Genomics* **9:** 539.

Orsi,R.H., Sun,Q., and Wiedmann,M. (2008b) Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of *Listeria monocytogenes*. *BMC Evol Biol* **8:** 233.

Ortiz,S., Lopez,V., and Martinez-Suarez,J.V. (2014) Control of *Listeria monocytogenes* contamination in an Iberian pork processing plant and selection of benzalkonium chloride-resistant strains. *Food Microbiol* **39:** 81-88.

Ortiz,S., Lopez-Alonso,V., Rodriguez,P., and Martinez-Suarez,J.V. (2015) The Connection between Persistent, Disinfectant-Resistant *Listeria monocytogenes* Strains from Two Geographically Separate Iberian Pork Processing Plants: Evidence from Comparative Genome Analysis. *Appl Environ Microbiol* **82:** 308-317.

Pal,C., Bengtsson-Palme,J., Rensing,C., Kristiansson,E., and Larsson,D.G. (2014) BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res* **42:** D737-D743.

Ragon,M., Wirth,T., Hollandt,F., Lavenir,R., Lecuit,M., Le,M.A., and Brisse,S. (2008) A new perspective on *Listeria monocytogenes* evolution. *PLoS Pathog* **4:** e1000146.

Rambaut,A., Lam,T.T., Max,C.L., and Pybus,O.G. (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2:** vew007.

Ramirez,M.S. and Tolmasky,M.E. (2010) Aminoglycoside modifying enzymes. *Drug Resist Updat* **13:** 151-171.

Roche,S.M., Gracieux,P., and Velge,P. (2009) Poor detection of low-virulence field strains of *L. monocytogenes* is related to selective agents in selective media and is unrelated to PrfA. *Food Microbiol* **26:** 21-26.

Romling,U. and Balsalobre,C. (2012) Biofilm infections, their resilience to therapy and innovative treatment strategies. *J Intern Med* **272:** 541-561.

Ryan,S., Begley,M., Hill,C., and Gahan,C.G. (2010) A five-gene stress survival islet (SSI-1) that contributes to the growth of *Listeria monocytogenes* in suboptimal conditions. *J Appl Microbiol* **109:** 984-995.

Schmitz-Esser,S., Muller,A., Stessl,B., and Wagner,M. (2015) Genomes of sequence type 121 *Listeria monocytogenes* strains harbor highly conserved plasmids and prophages. *Front Microbiol* **6:** 380.

Seemann,T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068-2069.

Shiryev,S.A., Papadopoulos,J.S., Schaffer,A.A., and Agarwala,R. (2007) Improved BLAST searches using longer words for protein seeding. *Bioinformatics* **23:** 2949-2951.

Stamatakis,A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312-1313.

Stasiewicz,M.J., Oliver,H.F., Wiedmann,M., and 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.

Stessl,B., Fricker,M., Fox,E., Karpiskova,R., Demnerova,K., Jordan,K. et al. (2014) Collaborative survey on the colonization of different types of cheese-processing facilities with *Listeria monocytogenes*. *Foodborne Pathog Dis* **11:** 8-14.

Tamura,K., Stecher,G., Peterson,D., Filipski,A., and Kumar,S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30:** 2725-2729.

Valderrama,W.B. and Cutter,C.N. (2013) An ecological perspective of *Listeria monocytogenes* biofilms in food processing facilities. *Crit Rev Food Sci Nutr* **53:** 801-817.

Vogel,B.F., Jorgensen,L.V., Ojeniyi,B., Huss,H.H., and Gram,L. (2001) Diversity of *Listeria monocytogenes* isolates from cold-smoked salmon produced in different smokehouses as assessed by Random Amplified Polymorphic DNA analyses. *Int J Food Microbiol* **65:** 83-92.

Wang,Q., Holmes,N., Martinez,E., Howard,P., Hill-Cawthorne,G., and Sintchenko,V. (2015) It Is Not All about Single Nucleotide Polymorphisms: Comparison of Mobile Genetic Elements and Deletions in *Listeria monocytogenes* Genomes Links Cases of Hospital-Acquired Listeriosis to the Environmental Source. *J Clin Microbiol* **53:** 3492-3500.

Wulff,G., Gram,L., Ahrens,P., and Vogel,B.F. (2006) One group of genetically similar *Listeria monocytogenes* strains frequently dominates and persists in several fish slaughter- and smokehouses. *Appl Environ Microbiol* **72:** 4313-4322.

Xu,D., Nie,Q., Wang,W., Shi,L., and Yan,H. (2015) Characterization of a transferable *bcrABC* and *cadAC* genes-harboring plasmid in *Listeria monocytogenes* strain isolated from food products of animal origin. *Int J Food Microbiol* **217:** 117-122.

Zankari,E., Hasman,H., Cosentino,S., Vestergaard,M., Rasmussen,S., Lund,O. et al. (2012) Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* **67:** 2640-2644.

Zerbino,D.R. (2010) Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics* **Chapter 11:** Unit.

Zerbino,D.R. and Birney,E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18:** 821-829.

Zhou,Y., Liang,Y., Lynch,K.H., Dennis,J.J., and Wishart,D.S. (2011) PHAST: a fast phage search tool. *Nucleic Acids Res* **39:** W347-W352.

FIGURES AND TABLES

Fig. 1. Time of isolation of *Listeria monocytogenes* isolates. Timeline of isolation for of the 90 *Listeria monocytogenes* included in the whole genome sequencing analysis based on the ST. Asterisk indicates the number of genomes that was downloaded from NCBI.

Fig. 2. Phylogeny of *Listeria monocytogenes* isolates based on whole genome sequencing. A) Maximum-likelihood phylogenetic relationship of all 90 *Listeria monocytogenes* isolates. Isolates are colored according to origin with red are sampled in year 2013-14. Isolates marked in green are of non-Danish origin. B) Maximum-likelihood phylogenetic relationship of the 47 ST121 isolates and isolated are colored according to from which processing plant they are isolated.

Table 1. The evolutionary rate of the three persisting STs ST7, ST8 and ST121 of *Listeria monocytogenes.* The evolutionary rate and estimated time of a common ancestor were calculated by BEAST using Bayesian evolutionary analysis with a HKY substitution model.

	Evolutionary rates						Estimated time of ancestor	
ST	SNPs/year/ site	95% HPD confidence interval		SNPs/year	95% HPD confidence interval		Year	95% HPD confidence interval
ST7	1.27×10^{-7}	2.2×10^{-11}	2.7×10^{-7}	0.35	0.00006	0.72913	1916.1	1615-1989
ST8	6.47×10^{-8}	5.3×10^{-11}	1.5×10^{-7}	0.18	0.00015	0.40396	1905.2	1350-1982
ST121	1.08×10^{-7}	6.69×10^{-8}	1.5×10^{-7}	0.30	0.18405	0.41631	1907.8	1852-1948

Table 2. MLST-dependency of persistence genes, biocide genes and monocin given as percent isolates carrying the gene.

		Presence of persistence genes (Percent)				Presence of biocide resistance genes (Percent)						
		SSI versions[1)]										
MLST ST	# isolates	SSI1	SSI-Li	SSI-F2365	LM6179_0173[2)]	*Tn*6188	*Tn*5422	*lde*	*mdrL*	*cadA2*	pLM80	Presence of full monocin
1	3			100				33	33			Yes
3	2	100					100					Yes
6	3			100								Yes
7	7	100						100	100			Yes
8	7	100					71	100	100			Yes
9	1	100						100	100			Yes
14	3			100	100			100	100			Yes
31	2	100					100	100	100			*lmoB-lmo0127* deletion
35	1	100						100	100			Yes
101	2			100			100	100	100			Yes
121	47		100		100	100	96	100	100			Yes
193	6		100				100	100	100			*lmoB-lmo0127* deletion
204	4	100						100	100	100	100	Yes
372	1	100						100	100			Yes

398	1	100	100	100	*lmoB-lmo0127* deletion)

1) Three version of the SSI island is observed i.e. the five-islet Stress Survival Islet-1 (SSI1) associated with acid and salt stress and assumed to be involved in persistance (Hein et al., 2011; Ryan et al., 2010). In same location in other MLST types is either observed the *L. innocua* version encoded by the lin0464-5, here referred to as SSI-Li, or the one-gene islet identical to the one islet observed in F2365 encoded by LMOf2365_0481, here referred to as SSI-F2365.
2) Schmitz-Esser *et al.* (2015) showed that nine genomes of ST121 all contained a 12.5 kbp insertion with *LM6179_0173* encoding a protein hypothesized to be related to persistence.

Table 3. Genes specific to STs ST7, ST8 and ST121 isolates relative to the 90 *Listeria monocytogenes* isolates, respectively. ST specific genes *Listeria monocytogenes* ST7, ST8 and ST121 were identified and these genes belong to the core genome of the respective ST, but are not found in any other of the 90 isolates. Annotation in previously published genomes 10304S for ST7, R479a for ST8 and 6179 for ST121 are given as and references for previously characterized genes.

ST	Locus tag	Systematic gene name	Description of gene	Observed in other ST	Reference of published genes
ST7	LMRG_0005		hypothetical protein with *motB* domain	CC7 + other ST	
	LMRG_0006		hypothetical protein with ParB-like and HNH nuclease domains	CC7 + other ST	
	LMRG_2573		transcription activator of glutamate synthase operon GltC	CC7 + CC131	
	LMRG_2574		Acetyltransferase (GNAT) family protein	CC7 + CC131	
	LMRG_2575	*hsdR*	RM type I subunit R DEAD/DEAD box helicase	CC7	
	LMRG_2576	*hsdM*	RM type I subunit M - N-6 DNA methylase	CC7	
	LMRG_2577	*hsdS*	RM type I subunit S - enzym specificity subunit S	CC7	
	LMRG_2891		hypothetical protein with no domain	CC7	
	LMRG_2892		hypothetical protein with MTH538 TIR-like domain	CC7	
	LMRG_2933		hypothetical protein with domain of Magnesium-transporting ATPase (P-type)	CC7 + CC415	
	LMRG_2934		hypothetical protein (*ccrB* like recombinase gene)	CC7	
	LMRG_2935 (alternative location)		Integrase with pfam00239 domain	CC7	

	Not found in 10304S, located downstream of LMRG_02935		hypothetical protein with no domain	CC7	
ST8	LMR479a_0530		RM type I subunit S - enzyme specificity subunit S	CC8 + other ST	(Fagerlund et al., 2016)
	LMR479a_0532		RM type I subunit S - enzyme specificity subunit S	CC8	(Fagerlund et al., 2016)
	LMR479a_0811		Hypothetical protein with COG4640 domain.	CC8	(Fagerlund et al., 2016)
	LMR479a_1125		RM type III methylation unit	CC8	(Fagerlund et al., 2016)
	LMR479a_1126		RM type III restriction endonuclease	CC8	(Fagerlund et al., 2016)
	LMR479a_1132		ATP/GTP-binding protein	CC8	(Fagerlund et al., 2016)
	Not found in R479a, located between LMR479a_1132 and LMR479a_1133		ATP/GTP-binding protein - Tn916 transposon	CC8	
	LMR479a_1133		AIPR protein	CC8	(Fagerlund et al., 2016)
	LMR479a_2950		hypothetical protein with DUF4303 domain	CC8	
ST121	LM6179_RS01725		hypothetical protein	ST121 + other ST	
	LM6179_RS01730		Polymorphic toxin systems		
	LM6179_RS02965	*lmoJ2M*	Modification methylase BspRI	ST121 + J2479	(Lee et al., 2012)

LM6179_RS02970	*lmoJ2R*	AlwI restriction endonuclease	ST121 + J2479	(Lee et al., 2012)
LM6179_RS03050		cell surface protein	ST121 + ST11	
LM6179_RS03705		membrane associated lipoprotein	ST121 + other ST	
LM6179_RS05470		hypothetical protein	ST121	
LM6179_RS05815		hypothetical protein	ST121 + other ST	
LM6179_RS11180	*tetR*	type transcriptional regulator BetI	ST121	(Mûller et al., 2013)
LM6179_RS11185	*qacH*	SugE protein	ST121	(Mûller et al., 2013)
LM6179_RS11190	*tnpC*	hypothetical protein	ST121	(Mûller et al., 2013)
LM6179_RS11195	*tnpB*	integrase/recombinase	ST121	(Mûller et al., 2013)
LM6179_RS11200	*tnpA*	integrase/recombinase	ST121	(Mûller et al., 2013)
Not annotated in 6179		CRISPR sequence	ST121	

1) Previously published locus tag in 10304S for ST7, R479a for ST8 and 6179 for ST121

Supporting information Captions

Fig. S1.Bayesian phylogenetic reconstruction and divergence date estimates of the *Listeria monocytogenes* A) ST7 and B) ST8 isolates. Bayesian statistics were used to estimate the divergence times of predicted ancestors. The tree was based on a) 85 SNPs and b) 54 SNPs identified from whole-genome sequencing.

Fig. S2. Bayesian phylogenetic reconstruction and divergence date estimates of the *Listeria monocytogenes* ST121 isolates. Bayesian statistics were used to estimate the divergence times of predicted ancestors. The tree was based on 223 SNPs identified from whole-genome sequencing.

Fig. S3. Biocide Minimal Inhibitory Concentration (MIC) with a selection of *Listeria monocytogenes* isolates. MIC of biocides used in the processing plants included in the 2013-14 sampling performed in TSB with 1% glucose at 20°C with selection of *Listeria monocytogenes* isolates (Table S5).

Fig. S4. Plot of linear regression analysis of root-to-tip distances against sampling time using TempEst (Rambaut et al, 2016).

Table S1. Presumptive *Listeria monocytogenes*-samples included in 2013-14 sampling period. Number of presumptive *Listeria monocytogenes*-samples received at private laboratory from the two processing plants during the 2013-14 sampling period and number of samples that were MLST typed.

Table S2. Origin of *Listeria monocytogenes* positive samples from each plant. Origins of samples that are MLST typed and of the most abundant MLST STs in each processing plant.

Table S3. Time line of *Listeria monocytogenes*-positive samples. Number of *L. monocytogenes* samples that were MLST typed each month over the 2013-14 sampling period showing the season variation both for all MLST sequence typed isolates and the most abundant MLSTs.

Table S4. MLST distribution of *Listeria monocytogenes* isolates included in the study. MLST distribution within 2013-14 sampling period and the collection of old and reference strains showing

the number of isolates of each MLST and how many times this MLST was received at the private laboratory. Further showing the number of isolates selected for genome sequencing and the source of the isolates from the collection of old and reference strains.

Table S5: List of the 90 *Listeria monocytogenes* isolates included in the comparative genome analysis. List of all isolates included in the genome analysis including place and year of isolated. For isolates that are previously published or the genomes that are public available are reference and/or accession number given

Table S6. Characteristics of the 90 *Listeria monocytogenes* dependent on ST. Characteristics of the genome sequenced isolated including number of isolates of each MLST, minimum and maximum number of core genome Single Nucleotide Polymorphism (SNP) between any two isolates with subtype.

Table S7. Distance matrix of Single Nucleotide Polymorphisms (SNPs) of the 90 *Listeria monocytogenes.* Distance matrix of Single Nucleotide Polymorphisms (SNPs) between any two isolates based on core genome of EGDe where 229,610 bp of accessory positions have been removed. The matrix of within same ST SNPs are marked with blue. Yellow marking is the lowest and highest number of SNPs between Lineage I and II isolates. Orange marking is the lowest number of SNPs to another ST.

Table S8. Characteristic of the *Listeria monocytogenes* clonal and close-to-clonal groups. Characteristic of possible Clonal groups divided by the strict clonal group definition (≤4 SNP) or groups of close-to-clonal isolates that are plant-specific (>4 SNPs).

Table S9. Genome mining of presumptive persistence genes, biocide genes and phage genes of the 90 *Listeria monocytogenes* isolates. Contigs or assembled genomes of each isolate were analyzed by blastn against presumptive *Listeria monocytogenes* persistence and biocide genes identified by literature search. If gene was detected in isolate bits score is given. The BacMet database of 325 verified biocide protein was used for a blastx analysis to identify homologies in the sequenced genome. PhySpy was used to analyze for presumptive phage genes and number of presumptive prophages detected is given.

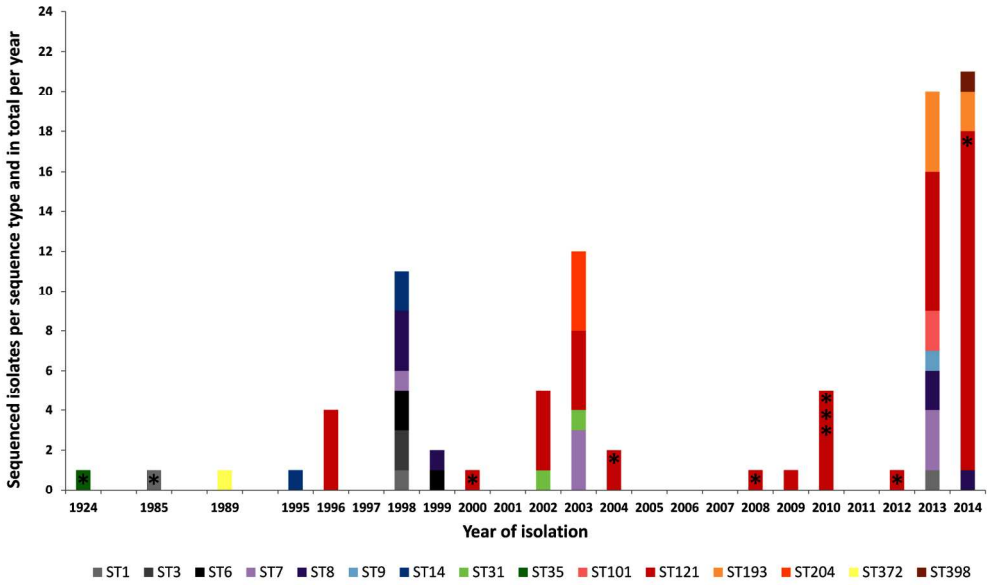

Fig. 1. Time of isolation of *Listeria monocytogenes* isolates. Timeline of isolation for of the 90 *Listeria monocytogenes* included in the whole genome sequencing analysis based on the ST. Asterisk indicates the number of genomes that was downloaded from NCBI.

171x102mm (300 x 300 DPI)

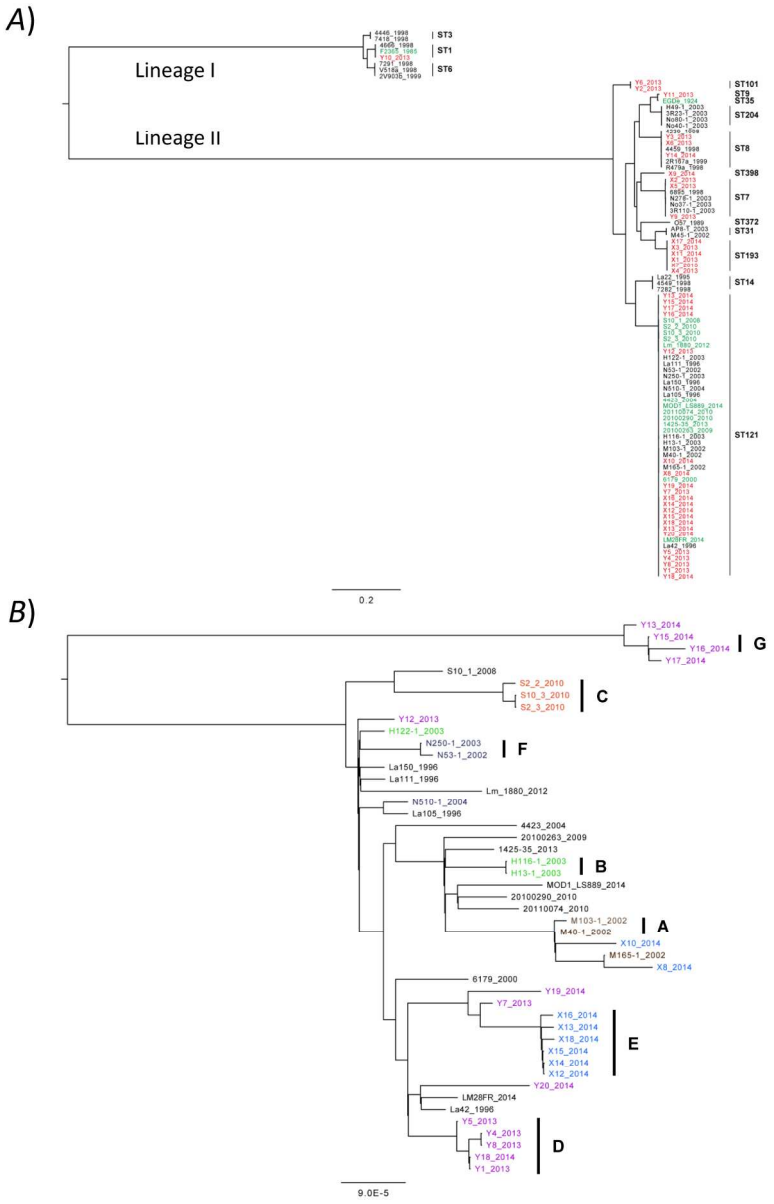

Fig. 2. Phylogeny of *Listeria monocytogenes* isolates based on whole genome sequencing. A) Maximum-likelihood phylogenetic relationship of all 90 *Listeria monocytogenes* isolates. Isolates are colored according to origin with red are sampled in year 2013-14. Isolates marked in green are of non-Danish origin. B) Maximum-likelihood phylogenetic relationship of the 47 ST121 isolates and isolated are colored according to from which processing plant they are isolated.

149x226mm (300 x 300 DPI)