


# Lineage-specific evolution and gene flow in *Listeria monocytogenes* are independent of bacteriophages

Roxana Zamudio,<sup>1</sup> Richard D. Haigh,<sup>2</sup> Joseph D. Ralph,<sup>1</sup> Megan De Ste Croix,<sup>1</sup> Taurai Tasara,<sup>3</sup> Katrin Zurfluh,<sup>3</sup> Min Jung Kwun,<sup>4</sup> Andrew D. Millard,<sup>1</sup> Stephen D. Bentley,<sup>5</sup> Nicholas J. Croucher,<sup>4</sup> Roger Stephan<sup>3</sup> and Marco R. Oggioni <sup>1\*</sup>

<sup>1</sup>Department of Genetics and Genome Biology, University of Leicester, Leicester, UK.

<sup>2</sup>Department of Respiratory Sciences, University of Leicester, Leicester, UK.

<sup>3</sup>Institute for Food Safety and Hygiene, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland.

<sup>4</sup>MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London, UK.

<sup>5</sup>Parasites and Microbes, Wellcome Trust Sanger Institute, Hinxton, UK.

## Summary

*Listeria monocytogenes* is a foodborne pathogen causing systemic infection with high mortality. To allow efficient tracking of outbreaks a clear definition of the genomic signature of a cluster of related isolates is required, but lineage-specific characteristics call for a more detailed understanding of evolution. In our work, we used core genome MLST (cgMLST) to identify new outbreaks combined to core genome SNP analysis to characterize the population structure and gene flow between lineages. Whilst analysing differences between the four lineages of *L. monocytogenes* we have detected differences in the recombination rate, and interestingly also divergence in the SNP differences between sub-lineages. In addition, the exchange of core genome variation between the lineages exhibited a distinct pattern, with lineage III being the best donor for horizontal gene transfer. Whilst attempting to link bacteriophage-mediated transduction to observed gene transfer, we found an inverse correlation between phage presence in a

lineage and the extent of recombination. Irrespective of the profound differences in recombination rates observed between sub-lineages and lineages, we found that the previously proposed cut-off of 10 allelic differences in cgMLST can be still considered valid for the definition of a foodborne outbreak cluster of *L. monocytogenes*.

## Introduction

*Listeria monocytogenes*, one of the most important foodborne pathogens, is the causative agent of Listeriosis, an infection that particularly affects the elderly, immuno-compromised individuals, and pregnant women (Radoshevich and Cossart, 2017; Lamond and Freitag, 2018). *Listeria monocytogenes* infections can progress to life threatening conditions including septicaemia, meningitis, encephalitis and spontaneous abortion, resulting in a high mortality rate (EFSA and ECDC, 2017). In recent years, a number of large listeriosis outbreaks have been seen worldwide, including the latest one reported in 2017–2018 in South Africa, which was caused by the consumption of contaminated ready-to-eat processed meat products from one company (WHO, 2018; Smith *et al.*, 2019). This means that surveillance of outbreaks is a major concern for national and international authorities. With the advent of low-cost whole-genome sequencing in the past few years, outbreak analysis has shifted from the previous gold standard of PFGE analysis to core genome multilocus sequencing (cgMLST) (Lüth *et al.*, 2018). cgMLST is based upon just the genes present in the core genome of a species, or at least those present in the panel of genomes available for analysis (Ruppitsch *et al.*, 2015). The 1701-gene cgMLST scheme (Ruppitsch *et al.*, 2015) has a much higher capacity for resolution of isolates differences than the classical 7-locus MLST (Salcedo *et al.*, 2003; Ragon *et al.*, 2008). As such cgMLST is not only able to provide insight into overall population structure but is also able to contribute to short-term nomenclature and the analysis of outbreaks. In the specific case of *L. monocytogenes*, the cut-off value for outbreak detection was fixed to  $\leq 10$  cgMLST alleles difference (Ruppitsch *et al.*, 2015). More recently after analysis of isolates with a known epidemiological link, it was proposed that

Received 26 February, 2020; revised 26 May, 2020; accepted 30 May, 2020. \*For correspondence. E-mail mro5@leicester.ac.uk; Tel. +44 (0)116 252 2261.

this cut-off value should be reduced to  $\leq 7$  cgMLST allele differences (Moura *et al.*, 2016).

Reference laboratories favour cgMLST analysis over other genomic analyses due to the easier standardization of the cgMLST scheme; however, analysis of population structure both at the outbreak and at the lineage scales also benefits from the analysis of how SNPs and recombination contribute to bacterial evolution. This is of particular interest as *L. monocytogenes* is not a homogeneous species but contains four separate deep-branching lineages as is evident from both MLST work and more recent global genomic analyses (Haase *et al.*, 2014; Moura *et al.*, 2016; Lees *et al.*, 2019). When testing for point mutations and recombination using the 7-locus MLST data, it was found that the occurrence of recombination versus point mutations was about six times more frequent in lineage II than in lineage I (den Bakker *et al.*, 2008). However, the knowledge is limited on population structure within sub-lineages (SL) and the gene flow between lineages at the finer scale.

Horizontal gene transfer mechanisms alter the core and accessory genome and are an important driving force in shaping bacterial population structures. With transformation not being documented for *L. monocytogenes* (Borezee *et al.*, 2000), transduction, which is well documented in many bacterial species (Popa *et al.*, 2017), automatically gains relevance as a candidate mechanism for gene transfer. In this species interest in phages also has practical aspects due to the differences in lineage-specific (teichoic acids) and genus-specific (cell wall) phage adsorption targets (Dunne *et al.*, 2018). Furthermore, there is significant research in the utilization of phages as a biocontrol in the food industry (Carlton *et al.*, 2005; Hagens and Loessner, 2014). With respect to the actual contribution of bacteriophages to gene flow within *Listeria*, there is conflicting evidence. In addition to the relevance of phages in the shaping of the accessory genome of *L. monocytogenes* (Klump and Loessner, 2013; Kuenne *et al.*, 2013), there is clear evidence that phages are responsible for the transfer of virulence-associated genes from other species to *Listeria* (Chen and Novick, 2009). With this, phages have been implicated with facilitating core-genome gene transfer by generalized transduction possibly with distinct impact in distinct lineages (Hodgson, 2000; Orsi *et al.*, 2011). Contrary to this, the short size of homologous recombination events detected does not support transduction as the main transfer mechanism (den Bakker *et al.*, 2008).

In the present work, we identify new outbreak isolates from a panel of genome sequences from Swiss *L. monocytogenes* clinical and food isolates by using cgMLST scheme, and then we used core SNPs to characterize differences in the population structure and the gene flow at a high resolution of the lineages I, II and III.

## Methods

### Bacterial isolates

The draft genome sequences of 158 *L. monocytogenes* isolates identified during surveillance by the Institute for Food Safety and Hygiene, the Swiss National Reference Laboratory, at the University of Zurich in the Swiss region of Europe mainly from 2011 to 2014 are presented in this work (Table S1). The addition of two lineage III isolates and six further outbreak isolates (Weinmaier *et al.*, 2013; Tasara *et al.*, 2015, 2016) creates a total database of 166 Swiss isolates. Ninety-six of the isolates were from human clinical cases (Althaus *et al.*, 2014) and 66 isolates were isolated from a variety of food products (Ebner *et al.*, 2015), one was from the environment and three were of unknown origin. Metadata for the isolates, including geographic location, collection data and source of the isolation, are listed in Table S1. For the comparative genomic analysis, we included 414 isolates from Germany (Halbedel *et al.*, 2018) and 128 Dutch isolates (Kremer *et al.*, 2017; Lees *et al.*, 2019) (Table S2) making a total of 708 genomes. We used the German isolates because it was our reference study for the outbreaks analysis and we included the Dutch isolates because it was our reference study for the PopPUNK analysis. In addition, for the population structure and gene flow analysis, we included 36 lineage III and four lineage IV isolates from public databases (Table S3).

Serotyping of the Swiss isolates was performed using the commercial set of *Listeria* O-factor and H-factor antisera from Denka Seiken (Pharma Consulting, Burgdorf, Switzerland) according to the manufacturer's instructions.

### Genome sequencing

DNA extraction for the 160 Swiss isolates was performed from pure cultures of each isolate grown in Brain Heart Infusion. In brief, 1.5 ml of overnight culture was pelleted, washed with 100% acetone and re-suspended in 400  $\mu$ l TE (10 mM Tris-HCl, 10 mM EDTA, pH 7), 10  $\mu$ l lysozyme (10 mg/ml, Sigma UK) and 10  $\mu$ l RNase A (10 mg/ml, Sigma UK) and incubated at 37°C for 30 min. Then, 75  $\mu$ l of 10% SDS (Sigma UK) and 10  $\mu$ l proteinase K (10 mg/ml, Sigma UK) were added and incubated at 65°C for 30 min. DNA was then purified using the Genomic Clean and Concentrator kit (Zymo Research), using the manufacturer's instructions. 150 bp paired-end DNA libraries, with an average insert size of 451 bp were produced using the NEB Ultra II custom kit and sequenced in pools on the HiSeq X10 platform; this resulted in an average of 2 666 434 reads per sample, corresponding to an average of 100 $\times$  coverage per genome. Adapter and poor quality sequences were removed using Trimmomatic (Ver. 0.36) (Bolger *et al.*, 2014) and the genomes were assembled using SPAdes (Ver. 3.9.0, default parameters) (Bankevich *et al.*, 2012) and the assembly

statistics for each isolate are found in Table S1. For the downstream analysis assemblies were only included if they had a minimum depth coverage of 10-fold. Raw sequence files and draft genomes in contig level for all 160 Swiss isolates were submitted to the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>) under study accession number PRJNA486730 (Table S1). In silico, MLST was performed using a BLAST-based tool (<https://github.com/tseemann/mlst>) on de novo genome assemblies (Jolley and Maiden, 2010; Seemann, 2018).

#### cgMLST and SNP analysis

A set of 1701 genes well defined for cgMLST analysis (Ruppitsch *et al.*, 2015; Halbedel *et al.*, 2018) were used for the identification of cluster type, SL and lineages in a total of 708 isolates (Tables S1 and S2). The alleles for the 1701-loci cgMLST were downloaded from the database <https://www.cgmlst.org/ncs/schema/690488/> and used as a query in BLASTn in order to identify and extract the target gene sequence. The parameters used in BLASTn were 100% of query coverage required, word size 11, mismatch penalty -3, match reward 1, gap open costs 5, and gap extension costs 2 and the hit with the highest % identity were kept. The sequences for each target gene per genome were extracted using bedtools v2.25.0 (<https://bedtools.readthedocs.io>), which creates a new FASTA file for each extracted sequence. Using a custom-made bash script, a multi-FASTA file was obtained for each target gene, and then the genes were aligned using Muscle v3.8.31 (Edgar, 2004).

The evaluation of the cgMLST scheme was done using standard pairwise allelic mismatches and SNPs pairwise. The matrix of pairwise allelic mismatches per target gene was obtained from the previously aligned sequences using snp-dists v0.6 (<https://github.com/tseemann/snp-dists>), where any differences (SNPs and/or indels) was scored as 1 and no differences was scored as 0. Using a python custom script, the matrices of allele pairwise differences were combined obtaining the overall allele differences in 1701-loci for each pairwise comparison. Gene absence was scored as a missing value. A total of 1596 (93.8%) target genes were present in all 708 isolates and this set of genes was used for the SNP pairwise analysis. A total of 1596 core genes from the gene-by-gene alignment were concatenated according to the order of the reference genome EGD-e (NC\_003210.1) and then the matrix of SNPs pairwise differences was obtained using the snp-dists v0.6. The maximum-likelihood core genome phylogenetic tree was inferred from the concatenated alignment core genes using the GTR model in RAXML (Randomized Axelerated Maximum Likelihood) (version 8.2.12) (Stamatakis, 2014). The visualization, manipulation and annotation of the phylogenetic tree were done using the ggtree R package v1.15.6 (Yu *et al.*, 2017).

#### PopPUNK analysis

Clustering analysis of 708 isolates was performed using the PopPUNK tool, which uses a K-mer approach to find genetic distances between isolates and then identify clusters of closely related isolates (Lees *et al.*, 2019). PopPUNK was run on the genome assemblies using the default settings (lower k-mer length was 13). Through PopPUNK we identified clusters of isolates, which were then compared with the clustering obtained from the phylogenetic analysis that was based on SNPs.

#### Chromosome painting and fineSTRUCTURE analysis

In order to detect genome-wide transfer of DNA sequence chunks through homologous recombination events between lineages, we expanded our initial dataset (708 isolates) with additional genomes from lineages III (36 isolates) and IV (four isolates) taken from public databases (Table S2) making a total of 748 isolates. We inferred population structure at a finer scale among the isolates from the core-genome SNPs data by using ChromoPainter (version 0.04) and fineSTRUCTURE (version 2.1.1). (Lawson *et al.*, 2012). Before executing the analysis, we removed isolates belonging to the same CC to generate a dataset with a single randomly selected genome for each CC. The genome collection for this analysis now contained 81 isolates with 15 isolates belonging to lineage I, 33 to lineage II, 29 to lineage III and four to lineage IV. In this genome collection, the genetic distance was more than 3000 SNPs between the single core-genomes (1569 core genes). The genome sequences of the isolates were mapped against the complete reference genome N2306 (CP011004) through the SKA tool (Harris, 2018), thus the alignment of the whole genome was obtained. ChromoPainter infers where DNA 'chunks' have been donated from a donor to a recipient and reconstructs the chromosome haplotype of 'recipient' individuals as a series of chunks from the other 'donor' individuals in the sample. A 'chunk' refers to a set of neighbouring/linked SNPs copied from a donor to a recipient. The output from ChromoPainter is a 'co-ancestry matrix', which summarizes the expected number of chunks of DNA imported from a donor to a recipient genome. fineSTRUCTURE uses the output of ChromoPainter for the clustering of individuals based on the co-ancestry matrix (Lawson *et al.*, 2012). ChromoPainter and fineSTRUCTURE were used according to the procedure described in <http://www.paintmychromosomes.com>.

#### Recombination rate

The recombination rate in the core genome alignment was assessed with Gubbins Version: 2.3.1 (Croucher *et al.*, 2015). Gubbins was run independently using

separate datasets for lineage I (371 isolates), lineage II (335 isolates) and lineage III (38 isolates). The whole-genome alignment was done by using the SKA tool (Harris, 2018) and complete reference genomes were used for the alignment; LL195 (HF558398) from Lineage I, Lm3136 (CP013723) from lineage II and M7 (CP002816) from lineage III. We did not estimate the recombination for lineage IV due to the small sample size of genomes deposited (four isolates). Reference-based alignments of the core genome were analysed using the default parameter of Gubbins, which identifies regions in the alignment with an atypically high density of SNPs, and that infers hypothetical recombination events. Two measures of the recombination rate were estimated:  $\rho/\theta$  estimate the relative frequency of occurrence of recombination and mutation in the history of the lineage (Milkman and Bridges, 1990), and  $r/m$  assess the relative impact of recombination and mutation in the genetic diversification of the lineage (Guttman and Dykhuizen, 1994).

#### *Transduction in silico analysis*

*Listeria monocytogenes* prophages were identified in the set of 81 isolates (see above) using the PHASTER tool, which provides the phage sequence annotation as either intact, questionable or incomplete (Zhou et al., 2011; Arndt et al., 2016). The incomplete prophages were removed for further analysis. The intact and questionable prophages were validated against a well-characterized set of known prophages evaluated using MASH v2.1 (Ondov et al., 2016). This set drew on an additional 45 *L. monocytogenes* prophages available from a public database (<http://millardlab.org/bioinformatics/bacteriophage-genomes>). Through the Mash analysis, we obtained a distance matrix to produce a heatmap of similarity between the prophages, and to determine whether questionable prophage was prophage or host chromosomal sequence. The same analysis allowed us to identify those prophages from a public database with no similarity with the prophage in our dataset. The mapping of the prophages to the set of 81 isolates was done by BLASTn. The identity threshold was 72% and the coverage value was recorded and plotted as a heatmap to quantify the length of the prophage region. To test *in silico* the effect of phages on competence, we mapped the known insertion of phages into the *comK* (lmo2270-lmo2233) in the set of 81 isolates by using BLASTn.

## Results

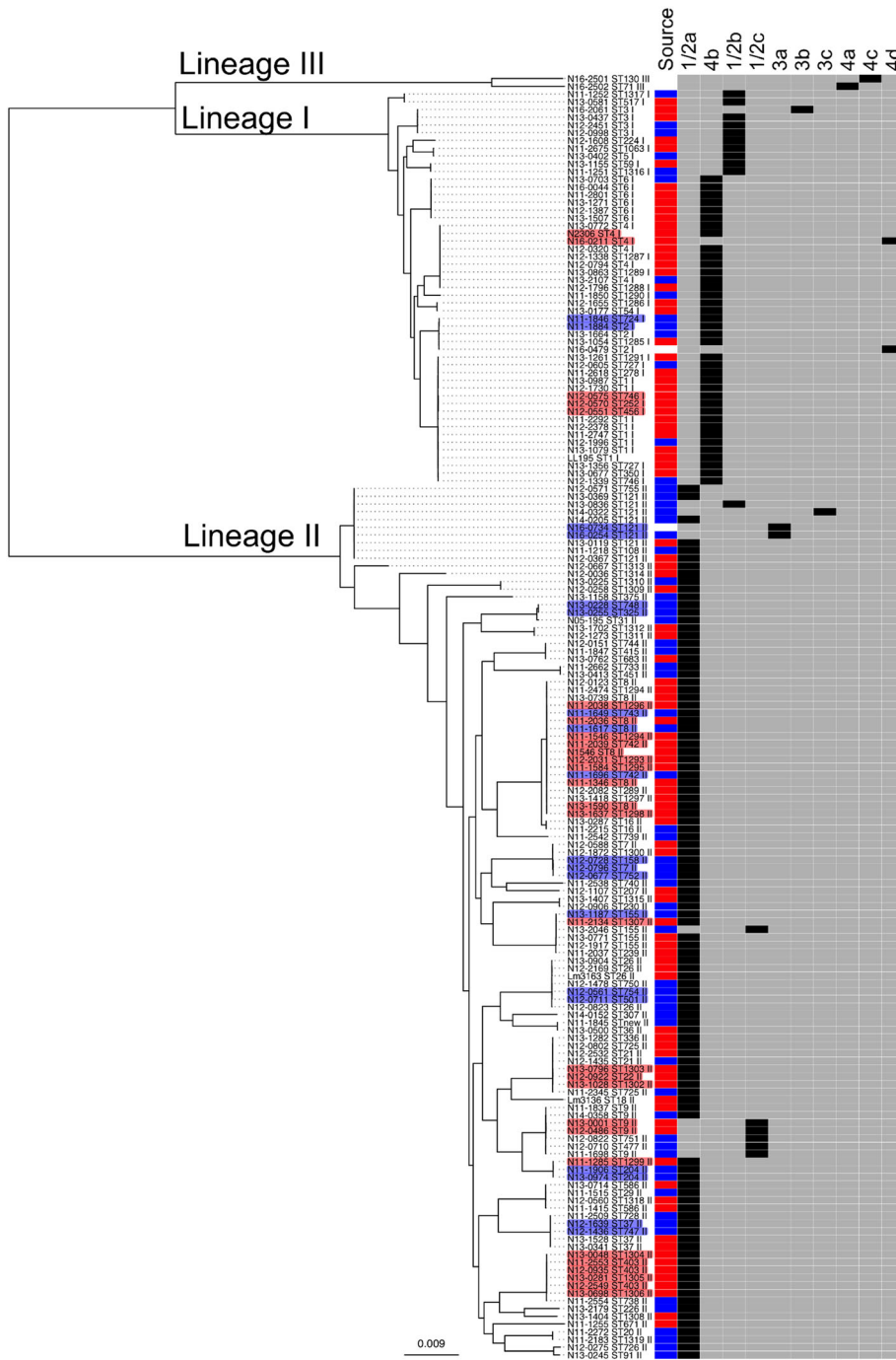
### *The genome sequences of the 166 Swiss L. monocytogenes isolates*

The species *Listeria monocytogenes* is subdivided into four phylogenetic lineages (I–IV) (Ward et al., 2008;

Orsi et al., 2011). Here, we present the draft genome sequences of *L. monocytogenes* isolates identified during surveillance by the Institute for Food Safety and Hygiene, the Swiss National Reference Laboratory, at the University of Zurich in the Swiss region of Europe mainly from 2011 to 2014 (Table S1). The mean length of the draft genomes was 2 948 113 bp with a mean GC% content of 37.8%. The summary data of the assembly for each genome are reported in Table S1. Core genome phylogenetic analysis of the 166 Swiss *L. monocytogenes* isolates found 51 isolates belonged to lineage I and 113 to lineage II (Fig. 1; Table S1). No significant association of the two lineages was found with respect to the origin (clinical or food) of the isolates (clinical isolates: lineage I 36 isolates, lineage II 60 isolates; food isolates: lineage I 14 isolates and lineage II 52 isolates) (Fig. 1; Table S1). Phenotypic testing was used to confirm the serotypes (1/2a, 1/2b, 1/2c, 3a, 3b, 3c, 4a, 4b, 4c and 4d) across the isolates. The two most prevalent serotypes were the lineage II serotype 1/2a (103 isolates) and the lineage I serotype 4b (38 isolates) (Fig. 1). MLST sequence type (ST) information showed ST1 (eight isolates) to be the most prevalent ST in lineage I and ST8 and ST121 (with seven and eight isolates respectively) in lineage II (Fig. 1; Table S1).

### *Genome evolution in lineages I and II*

In recent years, outbreak analysis has shifted from PFGE to cgMLST and a difference of  $\leq 10$  cgMLST alleles was identified as the most appropriate cut off (Ruppitsch et al., 2015). To verify if the proposed cut off for outbreaks of  $\leq 10$  cgMLST differences was appropriate in our dataset, we included in the analysis recently published isolates from Germany (414 isolates) (Halbedel et al., 2018) and the Netherlands (128 isolates) (Kremer et al., 2017; Lees et al., 2019). To visualize the distribution and frequency of differences, we constructed a pairwise matrix of either the core genome SNPs or the cgMLST allele differences and plotted the data (Fig. 2). In the overall pairwise plots (Fig. 2E), we clearly identified differences that mapped to outbreaks, SL and lineages. When limiting the analysis to groups of isolates distinguished by fewer than 150 differences in cgMLST or SNPs (Fig. 2A and B) outbreaks were clearly identifiable on the plots both by analysing cgMLST allele difference and core genome SNPs. When choosing SNPs or cgMLST differences with the lowest frequency both approaches yielded the same cut-off of  $\leq 12$  differences separating outbreak isolates but if being more stringent cut-offs of  $\leq 8$  alleles and  $\leq 7$  SNP differences could also be considered (Fig. 2A and B). The pairwise analysis is plotted separately for diversity within sub-lineage (Fig. 2A and B), between sub-lineage within a lineage (Fig. 2C and D) and between lineages (Fig. 2E).

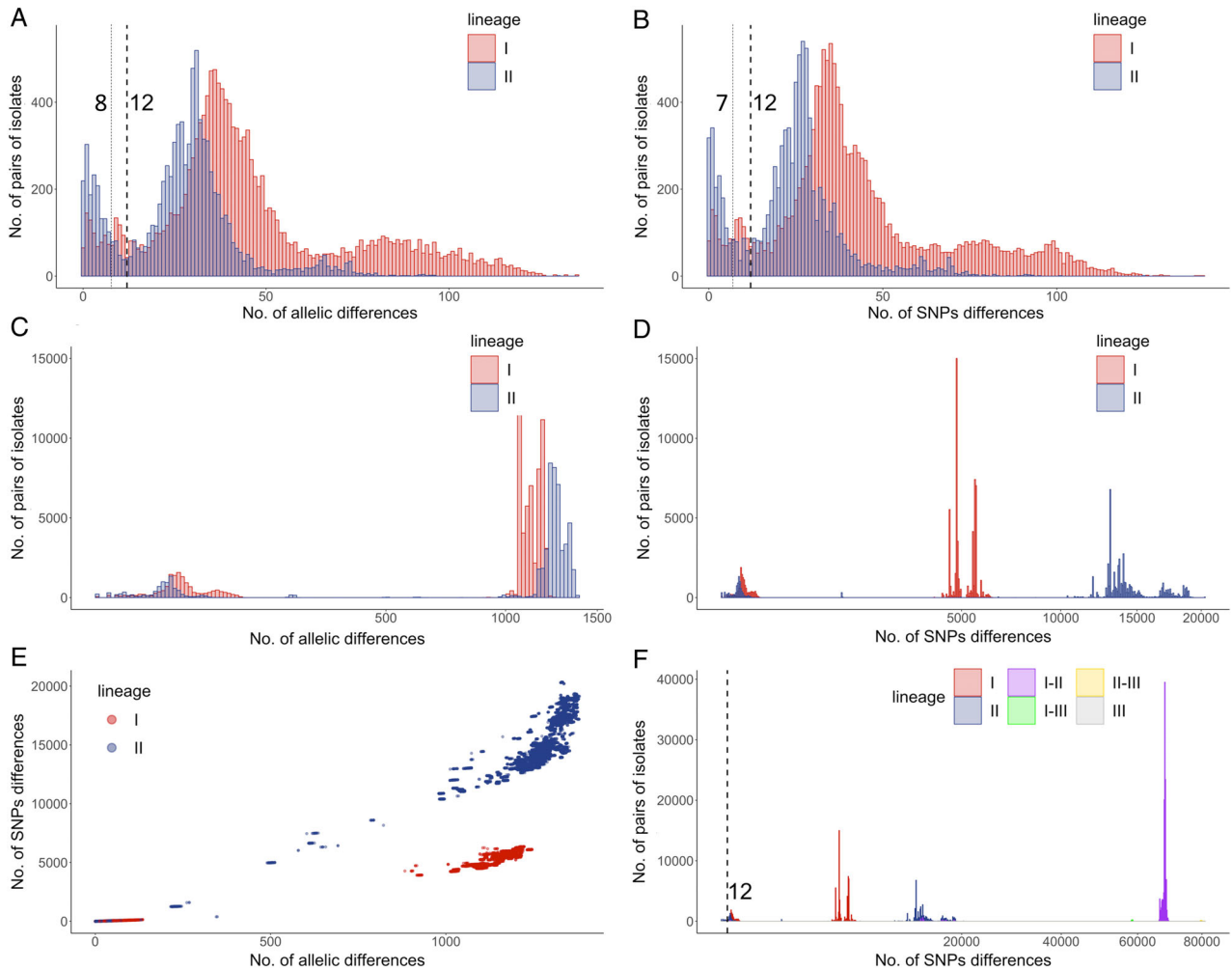


**Fig. 1.** Swiss *L. monocytogenes* core-genome phylogenetic tree. A maximum-likelihood phylogenetic SNP tree was constructed using 1596 core genes with the genome sequences of 166 Swiss isolates, which are associated with the lineages I, II and III. In the heatmap to the right of the tree the source of the isolates is shown (clinical in red, food in blue and other in white), and the serotype indicated (in black). Outbreak isolates are shaded in red and food clusters in blue.

When we plotted the data separately for lineage I and lineage II isolates (red and blue respectively in Fig. 2) we noted differences in their genome diversity. Lineage I isolates were more diverse within sub-lineages than lineage II (Kruskal–Wallis test  $p$ -value  $< 0.01$ ), with a broader distribution of pairwise distances, and a less clear peak corresponding to putative outbreaks. In both lineages the outbreak cut off was comparable, but the sub-lineages had different population structures with lineage I having

peaks at 36 and at 80 cgMLST differences while lineage II showed tighter distribution with peaks at 31 and 65 cgMLST differences (Fig. 2A). A comparable distribution was given by plotting SNP differences where lineage I had more SNPs within sub-lineages than lineage II (Fig. 2B).

When looking instead at the differences between SL within a lineage (cut off 1500 cgMLST alleles or 22 000 core genome SNPs) we observed the opposite arrangement with



**Fig. 2.** Lineage-specific evolution of *L. monocytogenes*. Pairwise cgMLST (A, B) and SNP (C, D, F) differences were analysed for 708 isolates from Switzerland, Germany and the Netherlands. Pairwise analyses are plotted separately for diversity within sub-lineage (A–B; up to 150 differences), between sub-lineage and within a lineage (C–D; up to 1500 cgMLST differences and 20 000 SNPs) and between lineages (F). The correlation between SNP and cgMLST differences is plotted in panel E. The pairwise SNP analysis considering the whole population (F) includes data for within lineage I differences (red), within lineage II differences (blue), within lineage III differences (grey) and the differences between lineages I and II (purple), between lineages I and III (green), and between lineages II and III (yellow).

lineage I showing less genome diversity than lineage II (Kruskal–Wallis test  $p$ -value  $<0.01$ ) (Fig. 2C and D). cgMLST did not resolve this difference effectively as almost all the alleles were different; however, SNP differences showed dramatic diversity with lineage I between-sub-lineages differences of about 5000 SNPs and lineage II of about 15 000 SNPs differences. These differences became more evident when plotting SNPs versus a cgMLST differences (Fig. 2E). In conclusion, this suggests lineage II is composed of sub-lineages that are more highly divergent from one another than those within lineage I, but each lineage I sub-lineage contains slightly more diversity than its lineage II equivalent, which might complicate the inference of outbreaks in lineage I.

#### Genome diversity within sub-lineages

To better understand this potentially confounding within-lineage population structure, the genetic diversity of the collection was analysed using MLST and PopPUNK. This revealed a close correspondence between the cgMLST-identified sub-lineages and both MLST-defined clonal complexes (CCs) and PopPUNK clusters. The differences within sub-lineages were clearly defined by both cgMLST difference and SNPs but also mapped perfectly to clusters defined by PopPUNK (Lees *et al.*, 2019) (Fig. S1). To explore the rationale for the multiple peaks in the cgMLST and SNP analysis within SL, we plotted the values separately for the different CC/PopPUNK

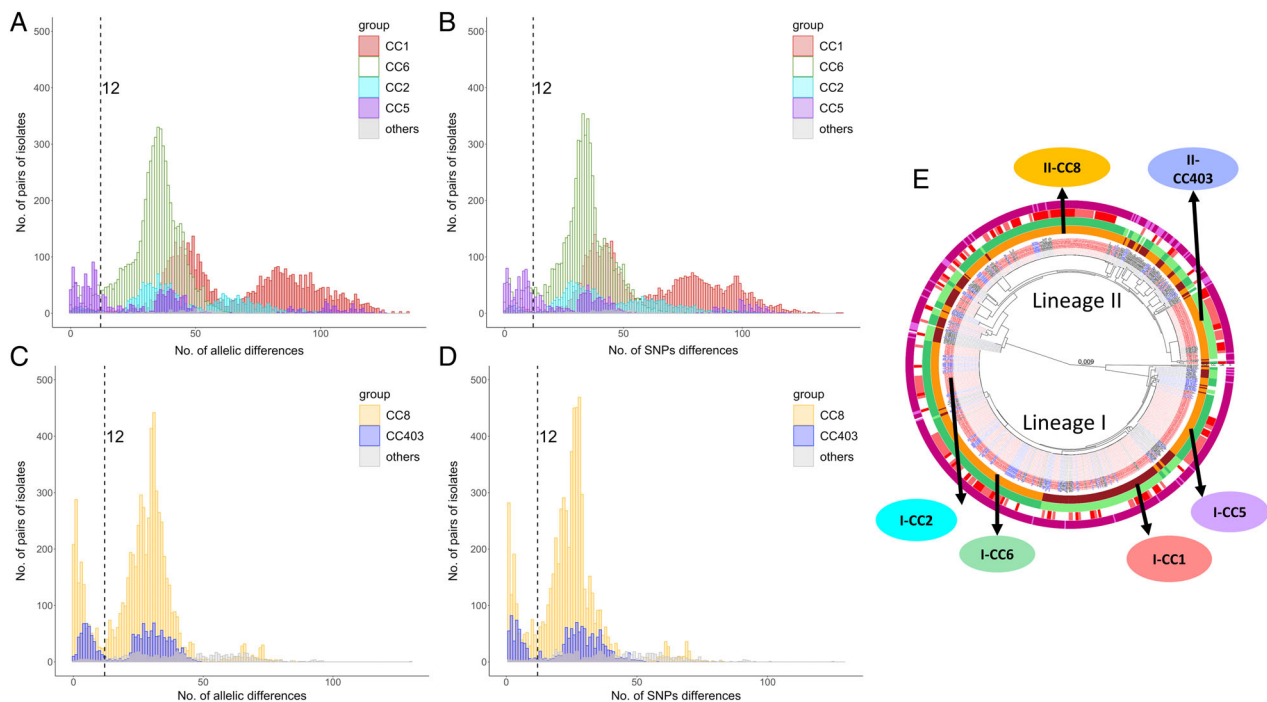


clusters/sub-lineages (Fig. 3). Data show that the main sub-lineages and PopPUNK clusters match to CC. In lineage I, CC1 included 100 isolates (89 isolates of which were ST1), CC2 included 60 isolates (58 of which were ST2), CC5 included 56 isolates (55 of which were ST5 and one ST1063) and CC6 included 111 isolates (110 of which were ST6) (Fig. 3). Irrespective of the number of outbreak isolates (those with less than 12 cgMLST allele differences), there were different levels of SNP diversity within these CC/PopPUNK clusters/sub-lineages (Fig. 3A and B). For example, CC6 had one peak with cgMLST difference of 35 alleles and a comparable number of SNPs (Fig. 3A and B). In contrast, both CC1 and CC2 showed two clearly distinct peaks, and CC5 had three peaks (Fig. 3A and B) indicating population structure at a finer level than sub-lineages. In CC1 there were several groups of isolates that differed by about 40–50 cgMLST alleles within the groups and which differed by 80–100 cgMLST alleles between the groups. Similarly, in CC2 there were a number of groups of isolates that differed by about 40 cgMLST alleles within the groups and differed by about 60 cgMLST alleles between the groups. The few isolates in clusters not belonging to either ST1 or ST2 could not account for this peculiar within the ST population structure. In lineage II the two main two clusters

were CC403, which included 54 isolates (50 of which ST403) and a large CC8 cluster with 118 isolates (102 of which ST8) (Fig. 3E). The two lineage II CC analysed (Fig. 3C and D) showed a similar population structure with a median cgMLST difference of 30 alleles for CC8 and 31 alleles for CC403. When analysing the k-mer distance using PopPUNK, data on pairwise distance could be generated in addition to the core genome, also for the accessory genome (Fig. S2A,B), with both sets of data showing quite a good overlap.

### Swiss outbreaks

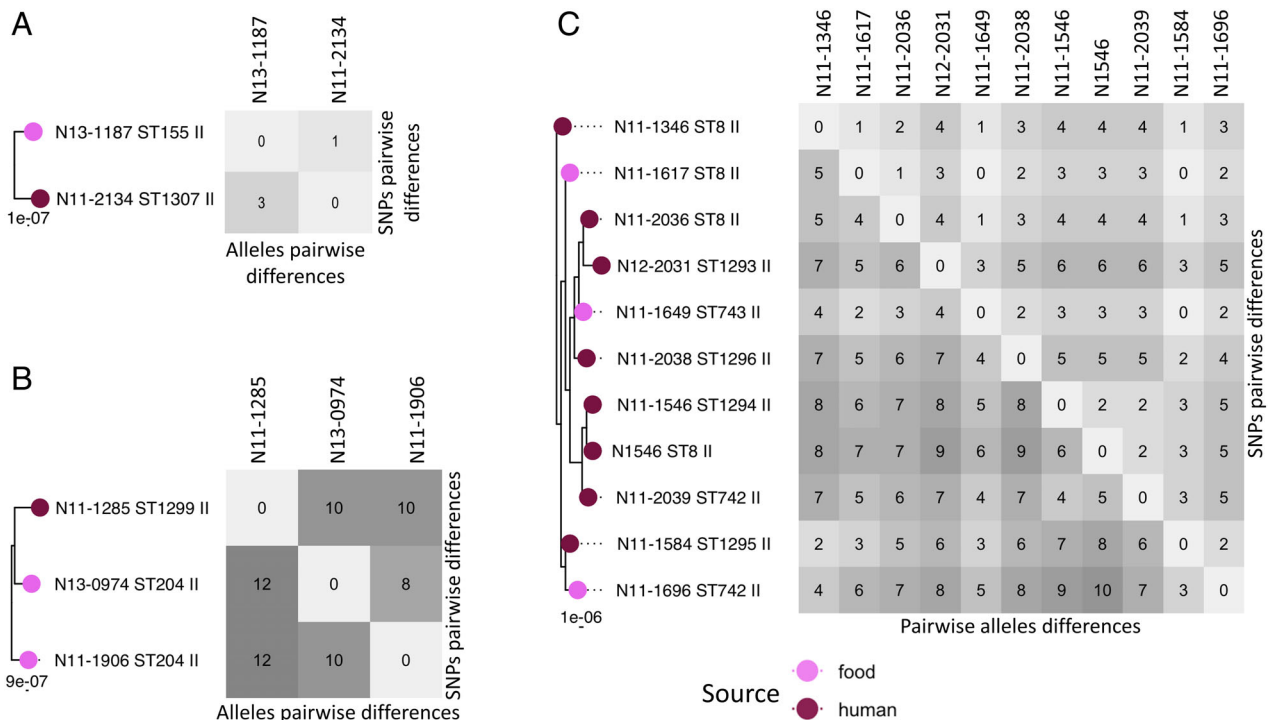
In the 166 Swiss isolates, from clinical and food sources, we could identify 15 clusters of isolates with  $\leq 12$  cgMLST allele differences (Table 1). Nine cases were known outbreaks and six were clusters from food (Table 1). SNP differences correlated well to cgMLST data but often yielded fewer differences as our SNP data did not include indels and were calculated on just 1594 core genes, while cgMLST was performed on all 1701 genes. Two small clusters of two and three isolates respectively were from lineage I while all others mapped to lineage II with three clusters in CC8. In three instances isolates from human and food sources are placed in the same cluster



**Fig. 3.** Genome diversity within clonal complexes within lineages I and II. Pairwise cgMLST and SNP differences were analysed for the major clonal complexes for each lineage. For lineage I the distribution of the pairwise cgMLST (A) and SNP (B) differences are shown for clonal complexes CC1 (red), CC2 (cyan), CC5 (purple) and CC6 (green). For lineage II the distribution of the pairwise cgMLST (C) and SNP (D) differences are shown for CC8 (yellow) and CC403 (blue). The CCs shown in panels A–D are indicated on the core genome phylogenetic tree and the arrows indicate the respective PopPUNK clusters (orange and brown) plotted on the first ring (E). The complete legend and high-resolution image of this circular tree are in Fig. S1.

**Table 1.** Swiss *L. monocytogenes* outbreaks and food clusters.

Outbreak	Lineage	CC (ST)	No. of isolates	cgMLST differences	SNP differences	Comment
<i>Outbreak</i>						
1	I	1 (252, 456, 746)	3	4–6	3–4	Human isolates
2	II	8 (8, 742, 743, 1293–1296)	11	2–10	1–6	Food and human isolates
3	II	8 (289)	1	2	1	Human isolates, international
4	II	8 (8, 1298)	2	2	1	Human isolates, international
5	II	9 (9)	2	8	6	Human isolates
6	II	21 (22, 1302, 1303)	3	5–7	4–6	Human isolates
7	II	155 (155, 1307)	2	3	1	Food and human isolate
8	II	204 (204, 1299)	3	10–12	8–10	No conclusive epidemiological link
9	II	403 (403, 1304–1306)	6	2–11	1–10	Human isolates
<i>Food cluster</i>						
1	I	2 (2, 724)	2	3	2	Food isolates
2	II	7 (7, 158, 752)	3	3–5	1–3	Food isolates
3	II	26 (501, 754)	2	2	0	Food isolates
4	II	31 (325, 748)	2	7	7	Food isolates
5	II	37 (37, 747)	2	1	0	Food isolates
6	II	121 (121)	2	3	1	Food and environment isolates

**Fig. 4.** Swiss outbreaks with mix source origin. A subset of three phylogenetic trees for three cases of Swiss outbreaks, which included both human (dark purple) and food isolates (light purple). To the right of the trees is the respective matrix of the pairwise cgMLST and SNPs differences.

(Fig. 4A–C). When we included the German and Dutch isolates, we observed that three ‘Swiss’ outbreak clusters actually also contained isolates from Switzerland and/or Germany and/or the Netherlands.

### Recombination

The recombination rate of *L. monocytogenes* was analysed separately for lineages I, II and III using Gubbins

on 371, 336 and 38 isolates respectively. The mean of the number of polymorphisms introduced via recombination with respect to mutation ( $r/m$ ) and the number of recombination events with respect to polymorphisms introduced by mutation ( $\rho/\theta$ ) are reported for internal and terminal branches of the lineages (Table 2). There are differences in the mean recombination ratio between the internal and external branch for lineage I and III, but there are considerably fewer for lineage II, with recombination hotspots

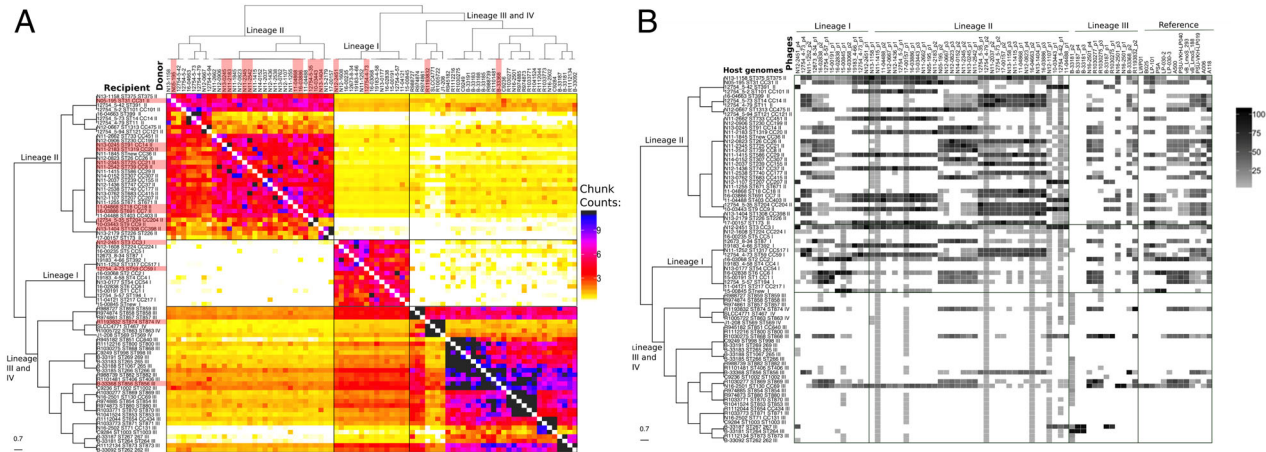


**Table 2.** Recombination statistics among lineages.

Lineages	Mean $r/m^a$			Mean $\rho/\theta^b$		
	Internal	Terminal	Total	Internal	Terminal	Total
Lineage I ( $n = 371$ )	0.0046	0.0091	0.0137	0.0002	0.0013	0.0015
Lineage II ( $n = 336$ )	0.3146	0.2324	0.5470	0.0149	0.0075	0.0224
Lineage III ( $n = 39$ )	0.1100	0.3460	0.4559	0.0038	0.0114	0.0151

<sup>a</sup> $r/m$  is the ratio of base substitutions predicted to have been imported through recombination compared with those occurring through point mutations.

<sup>b</sup> $\rho/\theta$  (rho/theta) is the ratio of the number of recombination events to point mutations, a measure of the relative rates of recombination and point mutation.



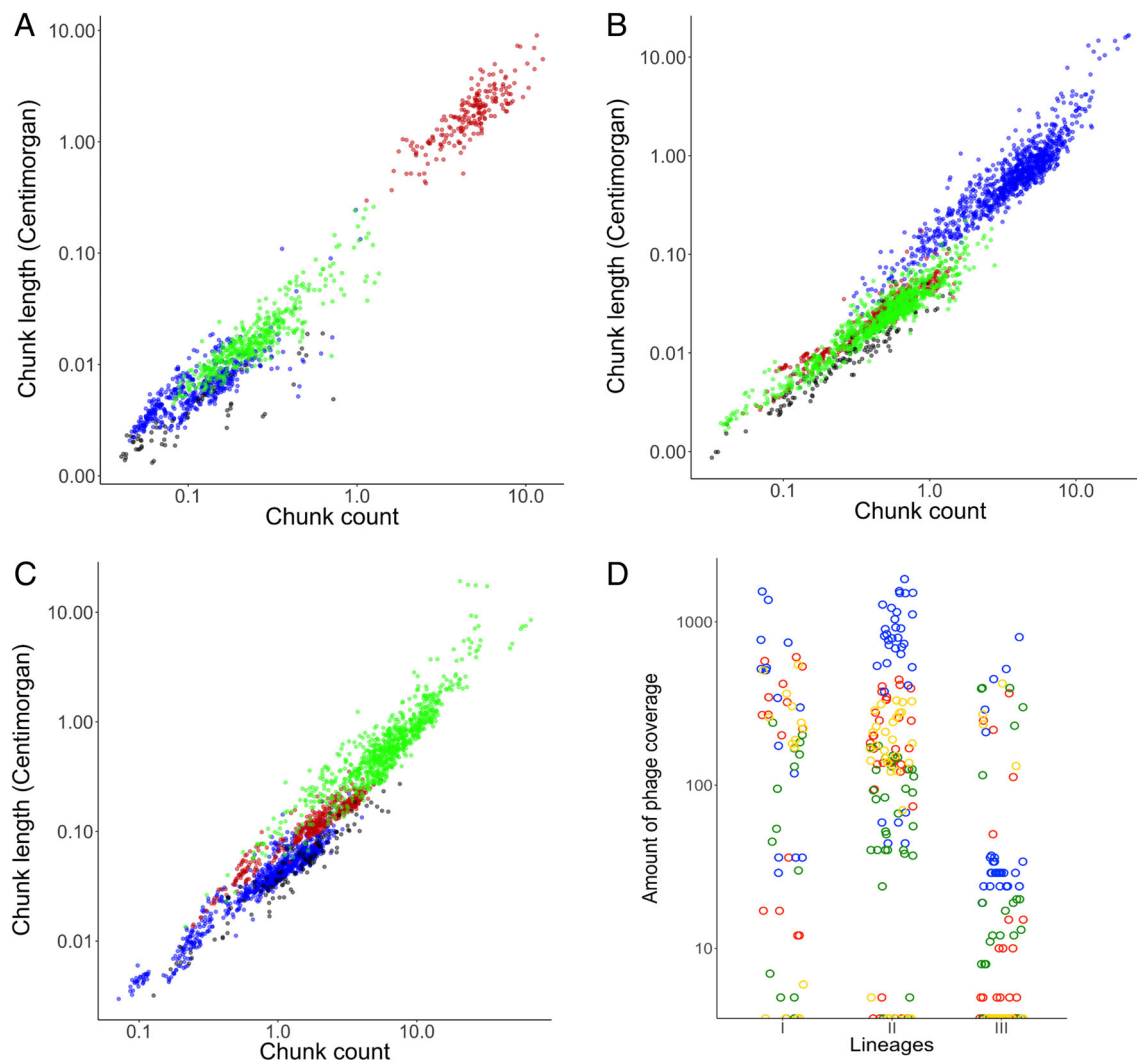
**Fig. 5.** The heatmap of ChromoPainter's co-ancestry matrix and occurrence of the prophages. A. The genome collection of 81 isolates for this analysis contained only a single isolate for each clonal complex including 15 isolates belonging to lineage I, 33 to lineage II, 29 to lineage III and 4 to lineage IV. The clustering of the genomes seen above the heatmap was done using FineStructure. Each row of the heatmap represents the genomes of recipient isolates, with the number of chunks copied from each donor individuals as columns. Dark colouring denotes a high chunk count and light colour a low chunk count. The colour bar on the right indicates the number of chunks. B. Heatmap of phage alignment coverage in a set of 81 isolates. Each row of the heatmap represents the genomes of the host (bacteria) with the percentage of phage alignment coverage from each phage as columns. A square with dark colouring denotes high coverage and with light colouring low coverage.

(Kuenne *et al.*, 2013) readily detectable in the core genome (Figs S3–S5). In agreement with our data above, recombination has occurred more frequently in lineages II and III where the  $r/m$  rate was 40–33 times higher when compared with the lineage I (Table 2). Graphical representations of the blocks of recombination clearly underline the higher frequency of recombination and horizontal gene transfer in lineages II and III (Figs S3–S5).

#### Sequence exchange between lineages

We characterized patterns of gene flow and population structure by using ChromoPainter and fineSTRUCTURE. To perform this analysis, we selected a panel of 81 isolates, with one isolate representing each CC, from all four lineages of *L. monocytogenes*. The co-ancestry matrix was visualized as a heatmap with each cell indicating the DNA 'chunk' counts that a recipient received from each donor (Fig. 5A). We used all isolates as potential donors in order to characterize the gene flow among the isolates. The

boxes in Fig. 5A indicate evident signs of 'gene flow'; we use the term 'gene flow' when there is a horizontal transfer of DNA chunks between lineages. The co-ancestry matrix revealed asymmetries in the gene flow between lineages; i.e., the four lineages imported higher numbers of chunks from other members of their own lineage. Both lineage I ( $p$ -value < 0.05) and lineage II ( $p$ -value = 0.018) imported more chunk counts from lineage III than from lineages II and I respectively (Fig. S6A). There is a high genetic exchange from both lineages I and II to lineage III, with more chunk counts imported from lineage I ( $p$ -value < 0.05) (Fig. S6A). Interestingly, despite the high frequency of exchanges observed overall, the analysis found little evidence for sequence exchange between lineages I and II (Fig. 5A, Fig. S6A). Lineages I and II donated more chunk counts to lineage III and lineage III donated more chunk counts to lineage II ( $p$ -value < 0.05) (Fig. S6B). When we considered the total chunk counts of transferred DNA per isolate, we observed again that lineage I had fewer chunk counts from external lineage population, and that lineages



**Fig. 6.** Distribution of the total chunk count for lineages and total occurrence of the prophages. The total chunk count was obtained for each genome from the co-ancestry matrix obtained from ChromoPainter and the distribution of this total count is shown in the dot plot. Each circle represents one single genome for the lineage I (red), lineage II (blue) and lineage III (green). Data for lineage I are in panel A, for lineage II in panel B and for lineage III in panel C. Panel D shows the distribution of the total occurrence of prophages (the total of the numerical values for the coverage by lineage is in Fig. 5B) plotted as a dot plot for lineage I (red), lineage II (blue), lineage III (green) and prophages from public database (yellow). In these graphs, we have not included the results for lineage IV due to the small sample size.

II and III had more chunk numbers, which matched well with the lower and higher respective rates of recombination observed with Gubbins (Fig. 6A). Importantly the chunk counts also showed that each lineage predominantly imported DNA from the same lineage and that lineage III was the second 'best' donor in both lineages I and II (Fig. 6A). Here, we did not analyse lineage IV due to the small sample size available.

#### *Horizontal gene transfers by transduction and transformation*

To test if bacteriophage were responsible for the horizontal gene transfer in our dataset, we mapped the prophages in

our set of genomes, inferring that detection of similar phages or phage-remnants through the genomes of different CCs could be taken as a proxy for the transfer of DNA by transduction. We identified 41 intact and 47 questionable prophages across the 81 *L. monocytogenes* isolates (one representative for each CC) using PHASTER. After Mash analysis, through the heatmap of similarity, we identified 28 of the questionable prophages as chromosomal sequence and thus they were removed, also 31 lytic phages from the public database (Dorscht *et al.*, 2009; Denes *et al.*, 2014) were removed as none had similarity with our genomes (Table S4). We then used the remaining 60 prophages plus 13 temperate phages from the public database as a query in a BLASTn analysis with all

81 isolates as the subject in order to predict sequence exchange between lineages mediated by prophage transduction. From BLASTn we obtained a phage coverage matrix, which was plotted as a heatmap and linked to the phylogenetic tree (Fig. 5B). Data showed that there is a high occurrence of prophages in both lineages I and II and a lower occurrence of prophages in lineage III. Some genomes, particularly in lineage II, showed extensive homology to many phages or phage-remnants present in lineages I and II. The prophages identified in lineage III had greater similarity to prophage of lineages I and II than with others in lineage III. A similar pattern was observed with the prophages from the public databases (Fig. 5B); however, for these, we had the limitation that we did not have access to the lineage data from the isolates from which they were originally sourced. When we quantified the total coverage of an individual phage throughout all isolates of a given lineage, we observed that lineage II phages scored much higher than lineage I ( $p$ -value  $<0.05$ ), followed by lineage III (Fig. 6D, Fig. S7) indicating that phages from lineage II show much higher levels of sequence identity to the phages present in the other lineages. To check the impact of phage on competence for genetic transformation, we mapped the incidence of insertion of prophage into the competence regulator gene *comK* in the 81 isolates (Table S5). We detected a phage-interrupted *comK* gene in 2/15 isolates in lineage I, 10/33 isolates in lineage II, 1/29 in lineage III and 1/4 in lineage IV, however, none of these genomes showed any evidence of a reduction in DNA exchange (Fig. 5A).

When testing the gene flow between lineages using Chromopainter the data clearly showed a correlation between the number of chunks of exchanged DNA and their size (Fig. 6A–C). In each of the three lineages incoming DNA originating from the same lineage is detected more frequent (higher chunk counts) and larger (chunk size) (Fig. 6A–C). Of the three lineages, lineage III shows by far the highest numbers of recombination events (Fig. 6C; note different X-axis scale). Interestingly, when we compare incoming DNA from other lineages the data show that lineage III is also the main donor to both lineages I and II (Fig. 6A and B, Fig. S6A). Finally, through Pearson correlation, we confirm an inverse correlation ( $p = 0.023$ ) between the prophage coverage and the chunk counts donated from lineages II and III to lineage I, and from I and III to lineage II.

## Discussion

The cgMLST scheme set up for *L. monocytogenes* typing is currently the most widely accepted approach for the identification of outbreaks (Ruppitsch *et al.*, 2015; Moura *et al.*, 2016; Halbedel *et al.*, 2018). This is because cgMLST has a greater discriminatory power

when compared with MLST or PFGE (Lüth *et al.*, 2018). In this study, we explored the cut-off for outbreak identification by performing a pairwise comparison with a 1701-locus cgMLST scheme (Ruppitsch *et al.*, 2015) against 708 *de novo* assembled genomes using BLASTN. Our work analysed 158 genomes of Swiss human and food *L. monocytogenes* isolates (this study) to which we added, for comparison, six known Swiss outbreak isolates (Weinmaier *et al.*, 2013; Tasara *et al.*, 2015, 2016), 414 German human isolates (Halbedel *et al.*, 2018) and 128 Dutch human isolates (Kremer *et al.*, 2017; Lees *et al.*, 2019). Our results show a cut-off for the definition of outbreaks between 7 and 12 allelic or SNPs differences. Using the same set of cgMLST target genes the cut-off for the determination of outbreak isolates was previously identified as  $\leq 10$  alleles differences by analysing 42 and 424 genomes respectively (Ruppitsch *et al.*, 2015; Halbedel *et al.*, 2018). In contrast, a lower threshold of  $\leq 7$  alleles was identified by Moura *et al.* (2016) using 1696 genomes with a 1748 gene cgMLST scheme. The use of different sets of cgMLST has not been found by others to be critical for the general identification of clonal groups or outbreak isolates (Chen *et al.*, 2016). While we agree with this cut-off, especially when plotting all isolates on the same graph, it should be noted that there are also more subtle differences to be seen when mapping pairwise differences of different lineages and sub-lineages/clonal-complexes. Our data clearly show that dependent upon the sub-lineage analysed, the most likely cut-off could be anywhere between 7 and 12 cgMLST differences when using the 1701-locus cgMLST scheme (Fig. 3). Plotting core genome SNP differences shows the same pattern of variation when the different lineages and sub-lineages are plotted separately. Our pairwise data analysis, in common with previous observations (Chen *et al.*, 2016), does not allow us to propose defining a clonal-complex specific cut-off, however, we would underline the necessity to consider the possibility of occurrence of outbreaks with a slightly higher cgMLST or SNP diversity. SNP analysis in core-genomes was used in previous work for the characterization of the genome diversity within SL but this was limited to CC1 and CC9 only (Moura *et al.*, 2016). In this study, we have greatly expanded this and characterized the genome diversity within lineages for CC1, CC2, CC5 and CC6 for lineage I and CC8 and CC403 for lineage II.

While the web-based identification of CC is straight forward, the definition of SL is not as obvious. To identify a computationally more approachable method, we tested the PopPUNK program for clustering clonal isolates (Lees *et al.*, 2019). Our data show that PopPUNK clustering matches perfectly to isolates grouped into CC. Importantly, this overlaps perfectly with the SL identified by cgMLST that are defined by having less than  $<150$  alleles or SNPs differences (Moura *et al.*, 2016). One important aspect is

that PopPUNK is much faster, more flexible and computationally less heavy for the definition of SL, which confirms it as an ideal tool for the analysis of large datasets (Lees *et al.*, 2019).

Nine cases of listeriosis were recorded in Switzerland in 2011 as due to imported cooked ham (Hächler *et al.*, 2013), and in this study, we have sequenced the genomes for the isolates from these cases. A follow-up study reported a genome sequence for the isolate N1546, which was associated with this listeriosis outbreak in 2011 (Tasara *et al.*, 2015). In this study, we have identified a total of 11 isolates belonging to this outbreak; these include the seven isolates previously associated with the outbreak, in addition, four isolates newly reported in this study, three of which were from a food source (meat). Another isolate N2306, which is associated with the listeriosis outbreak in 2013–2014 (Tasara *et al.*, 2016), also fell into a cluster with one of the newly sequenced isolates. Overall, using the cgMLST scheme and SNP approach, we have identified nine Swiss outbreaks of which, three included isolates with food and human sources. The fact that we have identified three outbreaks, which include isolates from Switzerland and Germany, shows that *a posteriori* genome analysis can highlight cross-border outbreaks, which may fail to be noted by national surveillance programs. In this case, inclusion of the large collection of French genomes (Moura *et al.*, 2017) may well have shown other cross-border outbreaks.

The population structure in *L. monocytogenes* had been previously evaluated using the STRUCTURE program to analyse seven house-keeping genes. These studies on MLST loci found that the genetic diversity between lineages was due to mutation rather than recombination (Ragon *et al.*, 2008; Haase *et al.*, 2014). In contrast, previous work, which involved the examination of two virulence genes, had found that the rate of recombination was six times higher than the mutation rate (den Bakker *et al.*, 2008). In our work, which used 1569 core genes, we have identified a higher recombination rate in lineages II and III, 40 and 33 times higher respectively, than in lineage I; these figures match well to the data of den Bakker (den Bakker *et al.*, 2008). Our work is the first study to demonstrate that horizontal gene transfer occurs predominantly within a lineage, and that lineage III shows the highest rate of within-lineage exchange. In addition, lineage III was also observed to be the most proficient donor of DNA chunks. The surprisingly low gene flow between lineages I and II indicates a, so far, an unexplained barrier to horizontal gene transfer between these lineages. As restriction-modification systems in *Listeria* are specific to STs (Chen *et al.*, 2017; Croix *et al.*, 2017) these barriers to horizontal gene transfer cannot account for the observed lineage-specific DNA

exchange. Bacteriophages have also been proposed to be responsible for DNA transfer in *Listeria* (Orsi *et al.*, 2011). The preferential within-lineage transfer of DNA we have observed could have matched well to the lineage-specific serovars of wall teichoic acids to which *Listeria* phages adsorb and which therefore determine their serovar-dependent host range (Dunne *et al.*, 2018). In contrast, our results from the prophages mapping showed the opposite findings, with lineage III harbouring almost no bacteriophages whilst being the most proficient acceptor and donor of core genome chunks. Given that the detection of phage sequences was inversely proportional to the amount of recombination the hypothesis that horizontal gene transfer in *Listeria* is mainly due to transduction had to be excluded. The observation of the preferential gene flow to and from lineage III in absence of phage detection would exclude for *Listeria* phage-related mechanisms as major contributors to core genome evolutions, as also for example in *E. coli* (Tree *et al.*, 2014). This observation would also exclude other phage-related mechanisms of genome evolution in *Listeria* as for example the genome hypermobility through lateral gene transfer by transduction, as recently reported for the related species *Staphylococcus aureus* (Chen *et al.*, 2018) for which phage-mediated gene transfer to *Listeria* had been documented (Chen and Novick, 2009). The potential contribution of the conjugation of plasmids and genomic islands to the horizontal gene flow was not evaluated here as their contribution to homologous recombination within the core genome is considered generally less significant. The other hypothesis for horizontal gene transfer, i.e. that it most likely occurs through transformation, has been examined previously by analysis of MLST locus recombination based on the relatively short size of recombining DNA (den Bakker *et al.*, 2008). The well-described insertion of phages into the competence regulator *comK* (Loessner *et al.*, 2000; Kuenne *et al.*, 2013) was detected in a third of lineage II isolates and few other isolates however, this did not correlate with an appreciable reduction in horizontal gene transfer in these isolates. It is important to underline that the phage insertion into *comK* in the isolates tested for gene flow was generally representative of all isolates (as they were selected randomly) of those specific CCs, thus representing a stable trait that could have a potential impact on CC evolution. This does not rule out transformation as the mechanism for gene transfer in the core genome because the *comK*-integrated phages have been shown to excise efficiently during stress allowing *comK* transcription and ComK-mediated gene regulation in conditions like intracellular infection of macrophages (Molloy, 2012; Rabinovich *et al.*, 2012). The observation that the gene flow shows clearly that within lineage gene exchange is favoured over the transfer of DNA to or from other lineages is consistent with

homologous recombination, the main mechanism for horizontal gene transfer in the core genome, being facilitated by the lower diversity between isolates in the same lineage. While we have no direct evidence for the mechanism of core genome evolution, the fact that phage presence inversely correlates to recombination, indicates that transformation, even if it has never been experimentally tested in *L. monocytogenes*, should be viewed as the most likely mechanism.

In conclusion, our data show important differences in the evolution of the four lineages of *L. monocytogenes*, with differences in recombination rate and flow of core genome genes. Furthermore, the finding that phage presence inversely correlates with core genome gene flow indicates that transformation and not transduction is the most likely mechanism for the evolution of the core genome. From an epidemiological point of view, where the aim is for a clear definition of a cut-off for an outbreak, our work demonstrates that the proposed cut-offs of 10 cgMLST differences can still be considered valid irrespective of the different population structures found in the different *L. monocytogenes* lineages.

### Acknowledgements

The authors wish to acknowledge funding from BBSRC Grant BB/N002903/1 to MRO, SDB and NJC. NJC is also supported by a Sir Henry Dale Fellowship, jointly funded by Wellcome and the Royal Society (Grant No. 104169/Z/14/Z). JDR is funded by a KTN CASE BBSRC studentship BB/P504737/1. The authors thank the work of the management team of the ALICE High Performance Computing Facility at the University of Leicester.

### Data availability

The genome sequences were deposited as reads at the European Nucleotide Archive ENA with Study accession PRJEB17888 and accession 4421STDY6601709 to 4421STDY6601879 and as FASTA files at GenBank as Bioproject PRJNA486730 with the accession numbers QYDA00000000 to RDTB00000000.

### References

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.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**: W16–W21.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Borezee, E., Msadek, T., Durant, L., and Berche, P. (2000) Identification in *Listeria monocytogenes* of MecA, a homologue of the bacillus subtilis competence regulatory protein. *J Bacteriol* **182**: 5931–5934.

Carlton, R.M., Noordman, W.H., Biswas, B., de Meester, E. D., and Loessner, M.J. (2005) Bacteriophage P100 for control of *Listeria monocytogenes* in foods: genome sequence, Bioinformatic analyses, oral toxicity study, and application. *Regul Toxicol Pharmacol* **43**: 301–312.

Chen, J., and Novick, R.P. (2009) Phage-mediated intergeneric transfer of toxin genes. *Science* **323**: 139–141.

Chen, J., Quiles-Puchalt, N., Chiang, Y.N., Bacigalupe, R., Fillol-Salom, A., Chee, M.S.J., *et al.* (2018) Genome hypermobility by lateral transduction. *Science* **362**: 207–212.

Chen, P., den Bakker, H.C., Koriach, J., Kong, N., Storey, D. B., Paxinos, E.E., *et al.* (2017) Comparative genomics reveals the diversity of restriction-modification systems and DNA methylation sites in *Listeria monocytogenes*. *Appl Environ Microbiol* **83**: e02091-16.

Chen, Y., Gonzalez-Escalona, N., Hammack, T.S., Allard, M. W., Strain, E.A., and Brown, E.W. (2016) Core genome multilocus sequence typing for identification of globally distributed clonal groups and differentiation of outbreak strains of *Listeria monocytogenes*. *Appl Environ Microbiol* **82**: 6258–6272.

Croix, D.S., Vacca, I., Kwun, M.J., Ralph, J.D., Bentley, S. D., Haigh, R., *et al.* (2017) Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. *FEMS Microbiol Rev* **41**: S3–S15.

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., *et al.* (2015) Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**: e15.

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

Denes, T., Vongkamjan, K., Ackermann, H.-W., Moreno Switt, A.I., Wiedmann, M., and den Bakker, H.C. (2014) Comparative genomic and morphological analyses of *Listeria* phages isolated from farm environments. *Appl Environ Microbiol* **80**: 4616–4625.

Dorscht, J., Klumpp, J., Biemann, R., Schmelcher, M., Born, Y., Zimmer, M., *et al.* (2009) Comparative genome analysis of *Listeria* bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site. *J Bacteriol* **191**: 7206–7215.

Dunne, M., Hupfeld, M., Klumpp, J., and Loessner, M.J. (2018) Molecular basis of bacterial host interactions by Gram-positive targeting bacteriophages. *Viruses* **10**: 397.

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.
- Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf* **5**: 113.
- EFSA and ECDC. (2017) The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. *EFSA J* **16**: 5500.
- Guttman, D., and Dykhuizen, D. (1994) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266**: 1380–1383.
- Haase, J.K., Didelot, X., Lecuit, M., Korkeala, H., L. monocytogenes MLST Study Group, and Achtman, M. (2014) The ubiquitous nature of *Listeria monocytogenes* clones: a large-scale multilocus sequence typing study. *Environ Microbiol* **16**: 405–416.
- Hächler, H., Marti, G., Giannini, P., Lehner, A., Jost, M., Beck, J., et al. (2013) Outbreak of Listeriosis due to imported cooked ham, Switzerland 2011. *Eurosurveillance* **18**: 20469.
- Hagens, S., and Loessner, M.J. (2014) Phages of *Listeria* offer novel tools for diagnostics and biocontrol. *Front Microbiol* **5**: 1–6.
- Halbedel, S., Prager, R., Fuchs, S., Trost, E., Werner, G., and Flieger, A. (2018) Whole-genome sequencing of recent *Listeria monocytogenes* isolates from Germany reveals population structure and disease clusters. *J Clin Microbiol* **56**: e00119-18.
- Harris, S.R. (2018) SKA: Split kmer analysis toolkit for bacterial genomic epidemiology. *BioRxiv*, 453142. <https://doi.org/10.1101/453142>.
- Hodgson, D.A. (2000) Generalized transduction of serotype 1/2 and serotype 4b strains of *Listeria monocytogenes*. *Mol Microbiol* **35**: 312–323.
- Jolley, K.A., and Maiden, M.C.J. (2010) BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinf* **11**: 595.
- Klumpp, J., and Loessner, M.J. (2013) *Listeria* phages. *Bacteriophage* **3**: e26861.
- Kremer, P.H.C., Lees, J.A., Koopmans, M.M., Ferwerda, B., Arends, A.W.M., Feller, M.M., et al. (2017) Benzalkonium tolerance genes and outcome in *Listeria monocytogenes* meningitis. *Clin Microbiol Infect* **23**: 265.e1–265.e7.
- 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.
- Lamond, N., and Freitag, N. (2018) Vertical transmission of *Listeria monocytogenes*: probing the balance between protection from pathogens and fetal tolerance. *Pathogens* **7**: 52.
- Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012) Inference of Population Structure Using Dense Haplotype Data. *PLoS Genet* **8**: e1002453.
- Lees, J.A., Harris, S.R., Tonkin-Hill, G., Gladstone, R.A., Lo, S.W., Weiser, J.N., et al. (2019) Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* **29**: 304–316.
- Loessner, M.J., Inman, R.B., Lauer, P., and Calendar, R. (2000) Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. *Mol Microbiol* **35**: 324–340.
- Lüth, S., Kleta, S., and Al Dahouk, S. (2018) Whole genome sequencing as a typing tool for foodborne pathogens like *Listeria monocytogenes* – the way towards global harmonisation and data exchange. *Trends Food Sci Technol* **73**: 67–75.
- Milkman, R., and Bridges, M.M. (1990) Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. *Genetics* **126**: 505–5017.
- Molloy, S. (2012) Bacterial pathogenicity: a competent escape for *Listeria*. *Nat Rev Microbiol* **10**: 670.
- Moura, A., Tourdjman, M., Leclercq, A., Hamelin, E., Laurent, E., Fredriksen, N., et al. (2017) Real-time whole-genome sequencing for surveillance of *Listeria monocytogenes*, France. *Emerg Infect Dis* **23**: 1462–1470.
- 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.
- Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and Phillippy, A.M. (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* **17**: 132.
- Orsi, R.H., den Bakker, H.C., and Wiedmann, M. (2011) *Listeria monocytogenes* lineages: genomics, evolution, ecology, and phenotypic characteristics. *Int J Med Microbiol* **301**: 79–96.
- Popa, O., Landan, G., and Dagan, T. (2017) Phylogenomic networks reveal limited phylogenetic range of lateral gene transfer by transduction. *ISME J* **11**: 543–554.
- Rabinovich, L., Sigal, N., Borovok, I., Nir-Paz, R., and Herskovits, A.A. (2012) Prophage excision activates listeria competence genes that promote Phagosomal escape and virulence. *Cell* **150**: 792–802.
- Radoshevich, L., and Cossart, P. (2017) *Listeria monocytogenes*: towards a complete picture of its physiology and pathogenesis. *Nat Rev Microbiol* **16**: 32–46.
- Ragon, M., Wirth, T., Hollandt, F., Lavenir, R., Lecuit, M., Le Monnier, A., and Brisse, S. (2008) A new perspective on *Listeria monocytogenes* evolution. *PLoS Pathog* **4**: e1000146.
- Ruppitsch, W., Pietzka, A., Prior, K., Bletz, S., Fernandez, H. L., Allerberger, F., et al. (2015) Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*. *J Clin Microbiol* **53**: 2869–2876.
- Salcedo, C., Arreaza, L., Alcalá, B., de la Fuente, L., and Vazquez, J.A. (2003) Development of a multilocus sequence typing method for analysis of *Listeria monocytogenes* clones. *J Clin Microbiol* **41**: 757–762.
- Seemann, T. (2018) Scan contig files against PubMLST typing schemes: tseemann/mlst. <https://github.com/tseemann/mlst>.
- Smith, A.M., Tau, N.P., Smouse, S.L., Allam, M., Ismail, A., Ramalwa, N.R., et al. (2019) Outbreak of *Listeria monocytogenes* in South Africa, 2017–2018: laboratory activities and experiences associated with whole-genome



- sequencing analysis of isolates. *Foodborne Pathog Dis* **16**: 524–530.
- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Tasara, T., Ebner, R., Klumpp, J., and Stephan, R. (2015) Complete genome sequence of *Listeria monocytogenes* N2306, a strain associated with the 2013-2014 listeriosis outbreak in Switzerland. *Genome Announc* **3**: e00553-15.
- Tasara, T., Klumpp, J., Bille, J., and Stephan, R. (2016) Genome sequences of *Listeria monocytogenes* strains responsible for cheese- and cooked ham product-associated Swiss Listeriosis outbreaks in 2005 and 2011. *Genome Announc* **4**: e00106-16.
- Tree, J.J., Granneman, S., McAteer, S.P., Tollervey, D., and Gally, D.L. (2014) Identification of bacteriophage-encoded anti-SRNAs in pathogenic *Escherichia coli*. *Mol Cell* **55**: 199–213.
- Ward, T.J., Ducey, T.F., Usgaard, T., Dunn, K.A., and Bielawski, J.P. (2008) Multilocus genotyping assays for single nucleotide polymorphism-based subtyping of *Listeria monocytogenes* isolates. *Appl Environ Microbiol* **74**: 7629–7642.
- Weinmaier, T., Riesing, M., Rattei, T., Bille, J., Arguedas-Villa, C., Stephan, R., and Tasara, T. (2013) Complete genome sequence of *Listeria monocytogenes* LL195, a serotype 4b Strain from the 1983-1987 Listeriosis epidemic in Switzerland. *Genome Announc* **1**: e00152-12.
- WHO. (2018) *Listeriosis – South Africa*. World Health Organization. URL <https://www.who.int/csr/don/28-march-2018-listeriosis-south-africa/en/>.
- Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017) Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**: 28–36.
- 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.

### Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

- Table S1.** Swiss *L. monocytogenes* isolates.
- Table S2.** Germany and The Netherlands *L. monocytogenes* isolates.
- Table S3.** *L. monocytogenes* lineage III and IV isolates.
- Table S4.** *Listeria monocytogenes* bacteriophages.
- Appendix S1.** Lineage specific evolution and gene flow in *Listeria monocytogenes* are independent of bacteriophages.