



Population Structure of *Listeria monocytogenes* in Emilia-Romagna (Italy) and Implications on Whole Genome Sequencing Surveillance of Listeriosis

Erika Scaltriti¹, Luca Bolzoni^{1*}, Caterina Vocale², Marina Morganti¹, Ilaria Menozzi¹, Maria Carla Re² and Stefano Pongolini¹

¹ Risk Analysis and Genomic Epidemiology Unit, Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia-Romagna, Parma, Italy, ² Operating Unit of Clinical Microbiology, Regional Reference Center for Microbiological Emergencies, St. Orsola-Malpighi Polyclinic, Bologna, Italy

OPEN ACCESS

Edited by:

Vitali Sintchenko,
The University of Sydney, Australia

Reviewed by:

Alexandre Leclercq,
Institut Pasteur, France
Heather A. Carleton,
Centers for Disease Control and
Prevention (CDC), United States

*Correspondence:

Luca Bolzoni
luca.bolzoni@izsler.it

Specialty section:

This article was submitted to
Infectious Diseases - Surveillance,
Prevention and Treatment,
a section of the journal
Frontiers in Public Health

Received: 11 December 2019

Accepted: 13 August 2020

Published: 18 September 2020

Citation:

Scaltriti E, Bolzoni L, Vocale C,
Morganti M, Menozzi I, Re MC and
Pongolini S (2020) Population
Structure of *Listeria monocytogenes*
in Emilia-Romagna (Italy) and
Implications on Whole Genome
Sequencing Surveillance of Listeriosis.
Front. Public Health 8:519293.
doi: 10.3389/fpubh.2020.519293

The population structure of human isolates of *Listeria monocytogenes* in Emilia-Romagna, Italy, from 2012 to 2018 was investigated with the aim of evaluating the presence of genomic clusters indicative of possible outbreaks, the proportion of cluster-associated vs. sporadic isolates and different methods and metrics of genomic analysis for use in routine surveillance. In the 2012–2018 period the notification rate of confirmed invasive cases in Emilia-Romagna was 0.91 per 100,000 population per year, more than twice the average rate of EU countries. Out of the total 283 cases, 268 (about 95%) isolates were typed through whole genome sequencing (WGS) for cluster detection with methods based on core-genome multi-locus sequence typing and single nucleotide polymorphisms. Between 66 and 72% of listeriosis cases belonged to genomic clusters which included up to 27 cases and lasted up to 5 years. This proportion of cluster-associated cases is higher than previously estimated in other European studies. Rarefaction analysis, performed by reducing both the number of consecutive years of surveillance considered and the proportion of isolates included in the analysis, suggested that the observed high proportion of cluster-associated cases can be ascribed to the long surveillance duration (7 years) and the high notification and typing rates of this study. Our findings show that a long temporal perspective and high surveillance intensity, intended as both exhaustiveness of the system to report cases and high WGS-typing rate, are critical for sensitive detection of possible outbreaks within a WGS-based surveillance of listeriosis. Furthermore, the power and complexity of WGS interpretation emerged from the integration of genomic and epidemiological information in the investigation of few past outbreaks within the study, indicating that the use of multiple approaches, including the analysis of the accessory genome, is needed to accurately elucidate the population dynamics of *Listeria monocytogenes*.

Keywords: *Listeria monocytogenes*, genomic epidemiology, surveillance, whole genome sequencing, cgMLST, SNPs, relatedness threshold, interpretation

INTRODUCTION

Listeria monocytogenes is the causative agent of listeriosis, a severe food-borne disease mainly observed among elderly people, immunocompromised individuals, pregnant women, and newborn. The disease is relatively rare, but its high hospitalization and fatality rates and increasing incidence make it one of the most important food-borne infections in Europe and worldwide (1).

The confirmed identification of outbreaks and their sources of contamination is generally hampered by the long incubation of listeriosis and the common finding of *L. monocytogenes* in the food-chain and the environment. Consequently, this is only possible when laboratory surveillance is implemented and high-resolution typing methods, able to pinpoint similar isolates in the population and in the sources of infection, are used. For this reason, whole genome sequencing (WGS), a highly discriminatory method, is being progressively introduced in routine surveillance of listeriosis and in food-safety monitoring (2). However, the complexity of WGS data not only brings the potential of high-resolution molecular epidemiology, but also entails the need for robust and shared methods and standards for the generation and interpretation of typing results. In recent years different methods and interpretation criteria for WGS surveillance of *L. monocytogenes* have been developed from collections of isolates of various origin (3–8). This study analyzed the use of WGS for epidemiological surveillance of *L. monocytogenes* in the real-life conditions of the actual surveillance system of a specific administrative territory and explored its implications on surveillance strategies. The study territory is the Italian administrative Region of Emilia-Romagna (ER) having a population of about 4.5 million and a surface of about 22,450 Km². Since 2012, the Region has implemented laboratory surveillance of food-borne diseases integrated with official microbiological monitoring of foodstuff and represents a typical European example of regional health administration. Emilia-Romagna adopted PCR-serotyping as the routine typing method for *L. monocytogenes* isolates until 2017, while adopted WGS as the routine typing method starting from 2018. In the 2012–2018 period the notification rate of confirmed invasive cases, here defined as the number of confirmed invasive cases per 100,000 population [as used in (9, 10)], was 0.91 per 100,000 per year, more than twice the average rate of EU countries (9, 10). Despite the high notification rate, unexpectedly, no listeriosis outbreaks were detected in Emilia-Romagna before of the implementation of WGS surveillance. In this work, a retrospective (2012–2017) and prospective (2018) study of the population structure of human *L. monocytogenes* from ER was conducted to support the adoption of functional interpretation criteria for WGS surveillance. The aims of the study were: (i) to assess the presence of clusters of isolates similar enough to represent possible outbreaks, considering that before WGS typing was introduced all notified cases had been classified as sporadic and no outbreaks had been reported in ER; (ii) to determine the structure of the population of *L. monocytogenes*, namely the proportion of isolates belonging to clusters vs. the proportion of sporadic isolates, as an indicator of the

prevalent mode of infection (outbreak-associated vs. sporadic) and the duration and geographical extension of clusters; (iii) to create a background of genomes of *L. monocytogenes* from the Regional territory to allow for accurate assessment of genomic relationships among the incoming isolates in the new surveillance; and (iv) to evaluate methods and metrics of genomic analysis for their use in the surveillance.

MATERIALS AND METHODS

Bacterial Isolates, Whole Genome Sequencing, and Archive Genomes

The study included all 283 *L. monocytogenes* isolates notified to the laboratory surveillance of Emilia-Romagna (Italy) in the 2012–2018 period. All isolates originated from invasive cases and were checked to avoid any duplication (i.e., multiple isolates for single patient). The average notification rate for confirmed cases in Emilia-Romagna in the study period was 0.91 per 100,000 population per year, more than three times the overall national rate of 0.25 per 100,000 population per year (9, 10) and higher than those estimated in all but two EU countries in the same period, i.e., Finland and Denmark (9, 10). A total of 268 of the isolates (about 95%) underwent WGS, for the others, WGS was not possible (isolate not viable) or associated metadata were missing. Additionally, four environmental isolates that originated from outbreak investigations performed during the perspective surveillance underwent WGS. Genomic DNA was extracted with DNeasy Blood and Tissue kit (Qiagen), sequencing libraries were prepared using Nextera DNA Flex Library Prep Kit (Illumina) and run with a Miseq sequencer (Illumina) in pair-end mode (2 × 250 bp). The raw reads of the 272 newly sequenced isolates of the study were deposited at EBI under Project number PRJEB34036 (see **Table S1**). The raw reads of all available Italian genomes of human isolates, isolated in the years 2012–2018, were recovered from Sequence Reads Archive of NCBI [see (11–14)]. **Table S2** provides a complete list of the accession numbers of the NCBI raw reads used.

Core-Genome Multi-Locus Sequence Typing and SNP Analysis

Raw reads were checked for quality and length with FastQC (15) and for species confirmation using the miniKraken database (16), then filtered with Trimmomatic ver. 0.38 (17) according to (8), i.e., removal of (a) any adaptor sequences, (b) leading bases with PHRED < 25, (c) trailing bases with PHRED < 25, and (d) the entire read if length < 36 bases, clipping of the remainder of the read when a sliding window of 20 bases has average PHRED < 25. *De-novo* assembly was done with SPAdes Assembler ver. 3.9.0 (18) as in Van Walle et al. (8) and evaluated by QUAST ver. 4.2 (19) retaining high quality assemblies, i.e., post-trimming coverage > 40X, N50 > 20,000 Kb, (4), minimum contig length of 300 nucleotides (8), and contig number < 50. Based on assemblies, phylogenetic lineage, clonal complex (CC), seven-loci Multi-Locus Sequence Types (ST) (20) were assigned through Bacterial Isolate Genome Sequence Database (BIGSdb), and core-genome Multi-Locus Sequence Types (cgMLST) were

assigned through Bionumerics Software ver. 7.6.3 (Applied-Maths, Biomerieux) according to the Pasteur cgMLST scheme (4). Isolates with more than one locus showing multiple alleles and/or <95% of core genome coverage were re-processed, in accordance with Van Wall et al. (8). Cluster analysis based on cgMLST was done with single-linkage clustering. The metadata (Accession and sample numbers, source of origin, isolation material, date of isolation, age) and the main genomic clustering data (lineage, CC, ST) on the newly sequenced isolates were summarized in **Table S1**.

Matrices of Single Nucleotide Polymorphisms (SNP) were generated by CFSAN Pipeline for SNP analysis ver. 0.8.2 (FDA, USA) (21) and used to infer phylogeny within each ST through Maximum Likelihood Algorithm in RaxML software (22). The presence in ST7 isolates of the 31 kb prophage reported in 2015TE24968 genome (13) was checked by reads mapping using Geneious software (Biomatters, Ltd).

Criteria for Cluster Definition

Clusters of similar isolates representing putative outbreaks of infection were defined according to different internationally proposed thresholds and criteria. We defined clusters as groups of at least two isolates similar according to specific allelic/SNPs thresholds [as in (8, 23)] and we classified as sporadic the isolates not belonging to any cluster. In the study, clusters of similar isolates were only considered as indicators of possible outbreaks while a confirmed outbreak is such only following identification of a source of infection common to the involved cases. Two cluster-defining thresholds were represented by different values of the maximum acceptable number of differing cgMLST loci between any pairs of genomes in a cluster. This number is commonly referred to as allelic distance (AD). The threshold values for AD adopted in this study were 7 and 4, hereafter referred to as AD7 and AD4, as already proposed by Moura et al. (4) and Van Walle et al. (8), respectively. In other words, isolates were considered part of the same cluster when their pairwise AD was ≤ 7 or ≤ 4 in cgMLST. The third threshold considered was the value of 5 as the maximum acceptable number of pairwise SNPs between any pairs of genomes in the same cluster, proposed by Dallman et al. (24), hereafter referred to as SNP5. The fourth approach adopted for cluster definition was that proposed by Pightling et al. (6), consisting in the combination of the following three criteria: (i) pairwise SNP distance <21; (ii) bootstrap support > 0.89 for the cluster clade in the SNP-based phylogenetic tree; (iii) monophyletic topology of the cluster isolates in a SNP-based phylogenetic tree, hereafter referred to as SBT (i.e., SNPs, Bootstrap, Topology).

Rarefaction Analysis

The variation of the overall proportion of isolates attributed to clusters as a function of (a) the number of consecutive years of surveillance considered in the analysis and of (b) different fractions of isolates included in the analysis was estimated (rarefaction analysis). Cluster detection in rarefaction analysis was based on cgMLST with thresholds AD7 and AD4. Different rarefaction curves were generated

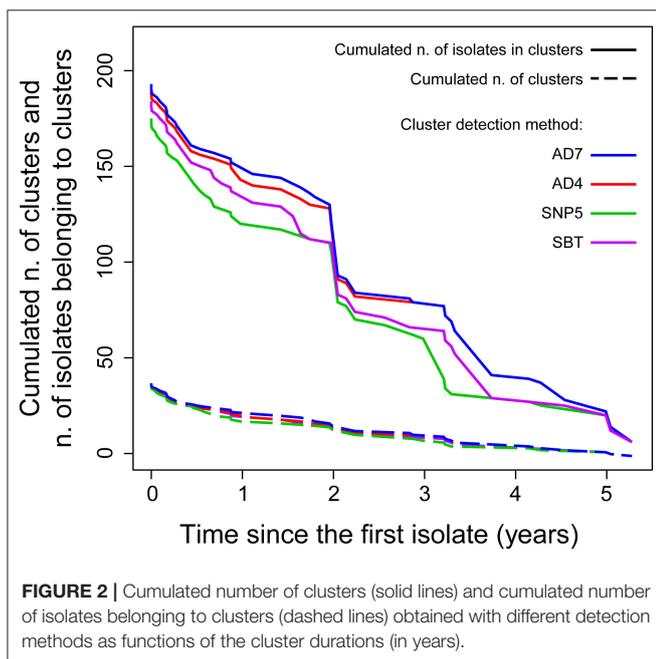
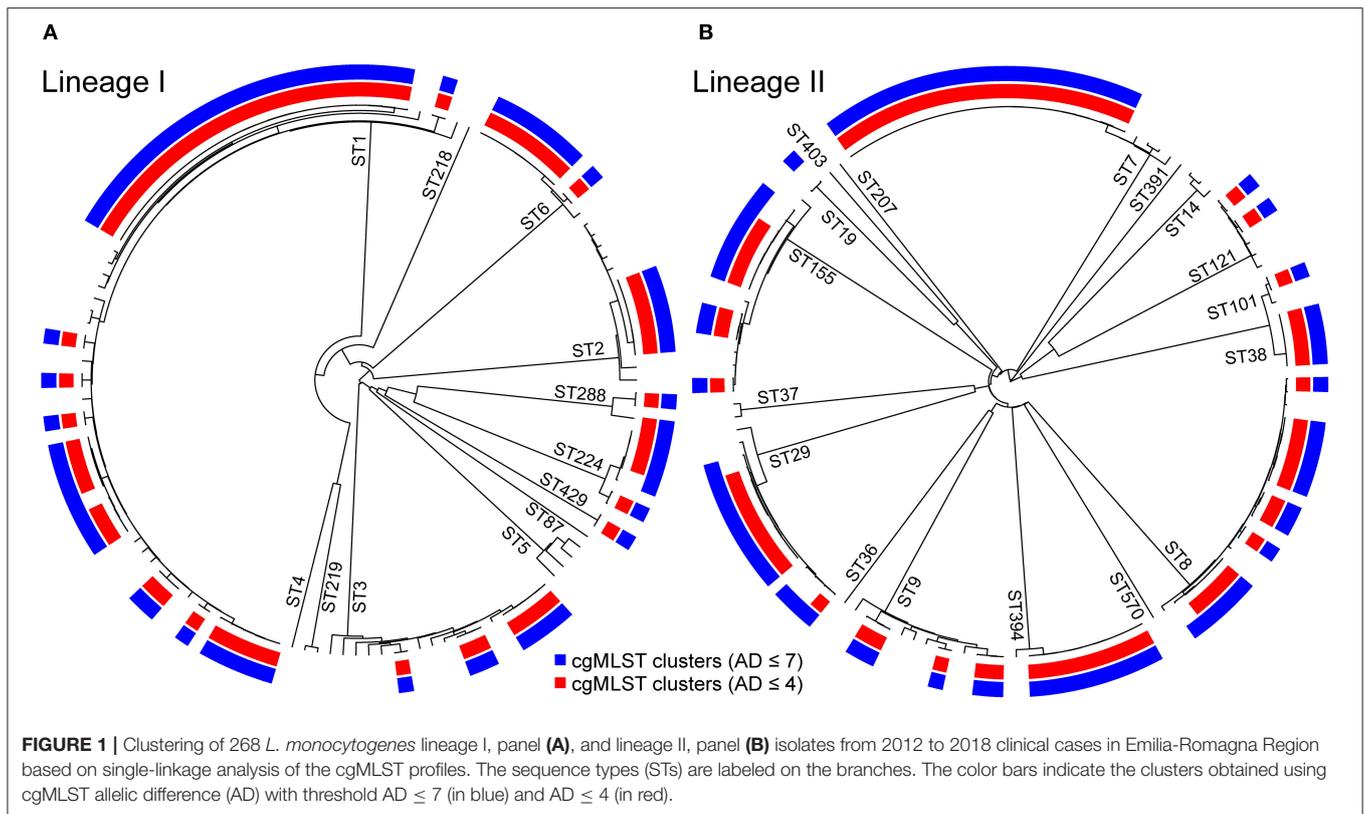
corresponding to proportions of included isolates of 25, 50, and 100%. Estimations for all rarefaction curves were performed with 1,000 random samples of the isolates dataset per point.

RESULTS

Highly Clustered Population of *Listeria monocytogenes* Disclosed by WGS Analysis

The isolates were distributed across phylogenetic lineages as follows: 135/268 were lineage I, while 133/268 were lineage II. The population structure of phylogenetic lineage I, panel (A), and II, panel (B), of *L. monocytogenes* in Emilia-Romagna is shown in **Figure 1**. Twenty-seven CCs (11 in lineage I and 16 in lineage II) and 29 STs (12 in lineage I and 17 in lineage II) were identified by seven-loci MLST, showing that the population is divided into clearly distinct phylogenetic groups. The distribution of isolates across the STs is skew, with the six most prevalent STs (ST1, ST7, ST8, ST3, ST155, ST29) accounting for about 64% of total isolates. Overall cgMLST- and SNP-based analyses show the presence of several clusters in the population, 38–39 depending on the detection method (**Figure 2**), with 7–11 of the clusters including more than 5 isolates, and the largest consisting of 26–27 isolates. The proportion of cases belonging to clusters ranges from 69 to 72% (when estimated through cgMLST criteria) and it ranges from 66 to 71% (when estimated through SNP-based criteria)—indicating that most cases of listeriosis in ER could be attributed to possible outbreaks—with 18–23 clusters lasting more than 1 year (up to five) and 117–146 (43–54%) of isolates belong to multi-year clusters. The performance of methods is compared in **Figure 2**. The maximum number of clusters and the largest proportion of cluster-associated cases were identified by AD7-cgMLST while SNP-based detection with cut-off ≤ 5 SNPs (SNP5) identified the minimum, providing maximum discrimination. The results of the rarefaction analysis, using AD7 criterium for cluster definition, are summarized in **Figure 3** (see **Figure S1** for the analysis with AD4). The analyses showed that the proportion of isolates belonging to clusters crucially depends—as expected—on the fraction of isolates included, but also on the number of consecutive years of surveillance considered. In particular, **Figure 3** shows that this proportion decreases for datasets shorter than 4 years, with a drop of more than 10% for datasets of 2 years (or less) compared with datasets of four (or more) years, regardless of the fraction of isolates included.

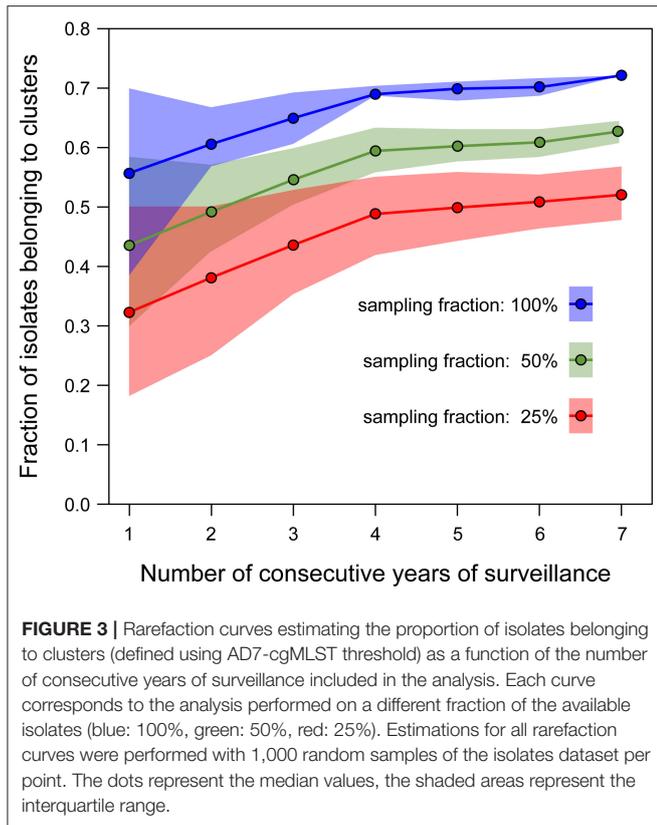
All 108 genomes of *L. monocytogenes* publicly available from the other Italian regions collected in the 2012–2018 period, representing about 10% of the 1,061 confirmed invasive listeriosis cases in Italy (9, 10) were jointly analyzed with the ER isolates through SNP-based procedures with the aim of: (i) investigating the possible extension of the clusters beyond the borders of ER; and (ii) verifying whether a larger geographical scale leads to the detection of additional clusters that include ER cases. Actually, nine of the clusters detected within ER included isolates from



neighboring regions and two additional small clusters (with two and four isolates, respectively) were detected upon extension of the geographical scale. (See the SNP-based phylogenetic trees in Figures S2–S12).

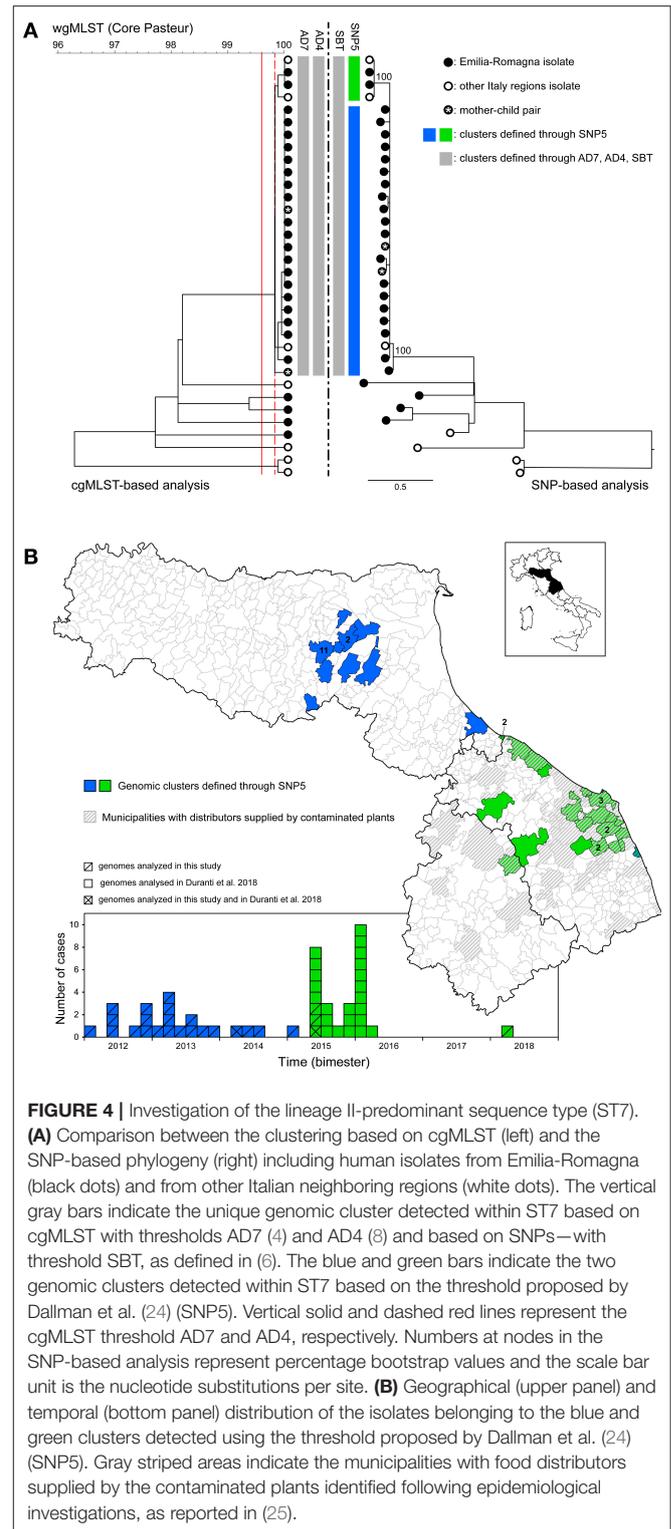
Complexity of WGS Interpretation Illustrated by Analysis of Lineage II-Predominant ST7

About 20% of ER lineage II isolates ($n = 27$) belonged to ST7 (Figure 1), featuring a main cgMLST cluster of 23 cases (Figure 4). The cluster included, among the others, a mother-child pair of isolates and three additional genomes from a neighboring region. This mother-child pair of isolates clearly represents a duplicate of the same strain and as such the pair should be included as a single isolate in the analysis for cluster detection in order not to artificially generate false clusters. Nevertheless, in this study both isolates of the pair were included with the specific aim of testing the ability of the different approaches to assign the pair to the same cluster, as expected. Actually, all cgMLST and SNP-based methods placed the mother-child pair in the same cluster (see Figure 4A). In the specific context of the study, the inclusion of this pair of isolates did not result in an artificial increase of the number of detected clusters as the pair was part of the main ST7 cluster together with several other isolates. With regard to this cluster, while the SNP-based method SBT defined by Pighling et al. (6) provided the same clustering as cgMLST methods, the SNP-based method derived from (24) divided the same isolates in two distinct clusters (blue and green in Figure 4A). Actually, available epidemiological evidence does support the existence of two distinct clusters with independent origin. Firstly, the two extra-ER genomes of the green cluster were part of a confirmed outbreak of 25 cases affecting two neighboring regions caused

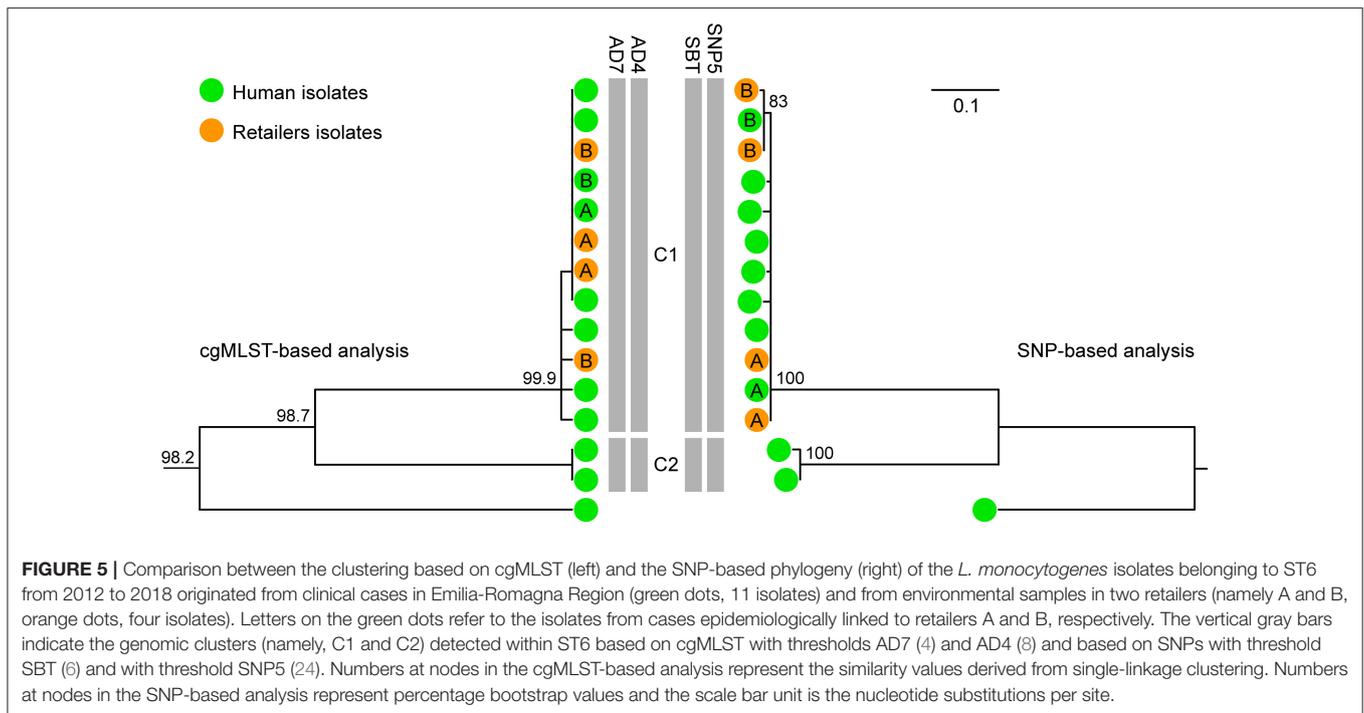


by the consumption of pork products as investigated by Duranti et al. (25). Therefore, the two ER cases of the green cluster very likely belonged to that outbreak also considering their closeness to the outbreak area (Figure 4B). Concomitantly, during the investigation of the outbreak, those authors also detected cases belonging to the blue cluster on their territory but could not demonstrate epidemiological association of those cases with the food producers implicated in the outbreak and eventually classified the cases as not belonging to the outbreak. Secondly, the two clusters are geographically and temporally segregated (Figure 4B). Thirdly, the trade area of the responsible food overlaps the outbreak territory (green cases) but not the area of the blue cluster (Figure 4B). The independence of the two clusters is further confirmed by the presence of a 31 kb prophage integrated in the green cluster genomes of ER and extra-ER origin as well as in the isolates from the pork products, but missing in the blue cluster genomes (13).

Altogether, these findings would suggest the existence of independent sources for the green and blue clusters; nevertheless, further understanding of the relationship between the two clusters emerges from a more in-depth analysis of the genomic information available. Their overall genetic similarity is considerable, with pairwise difference of only 9–12 SNPs (min-max); moreover, the phylogenetic tree topology shows that the green cluster originates from the blue cluster as a sub-clone (Figure 4A) and, consistently with this observation, the green cases occurred after the blue ones. Overall, these data indicate the possible existence of a recent shared origin for the two clusters represented by a source of contamination responsible at first



of the blue cases and later causing the green cases through the expansion of a sub-clone of *L. monocytogenes* from the original clone. This sub-clone was evidently distributed through the pork products associated with the green outbreak, while the direct vehicle of infection of the blue cases remained undisclosed.



Use of WGS in Prospective Surveillance

WGS typing was integrated in the prospective surveillance of invasive listeriosis in Emilia-Romagna in 2018. Forty-six cases were notified to the surveillance system in 2018 (1.03 cases per 100,000 population). The cgMLST-based surveillance identified nine new genomic clusters, which included 21 isolates, while another 17 isolates were assigned to six previously detected clusters. Two of these clusters, which included four or more 2018 isolates, underwent further epidemiological and microbiological investigation to identify the source of infection. The choice to further investigate these two clusters was based on the evidence produced by Rounds et al. (26, 27) that genetic clusters of ≥ 4 isolates from foodborne infections were more likely to be epidemiologically confirmed than clusters with less isolates. The investigated clusters belonged to ST224 (with four 2018 cases) and ST6 (with six 2018 cases). No suspected source was identified for ST224 cluster. On the other hand, three cases included in the ST6 cluster were linked, through food questionnaires, to the consumption of pork products sold in large retail shops located in three different provinces of the region. All six 2018 cases of the ST6 cluster were hospitalized and three died. Environmental swabs collected at the retail shops on equipment used for the pork sale were positive for *L. monocytogenes* belonging to the ST6 cluster in two of the three shops (see cluster C1 in **Figure 5**). Specifically, isolates matching the ST6 cluster were sampled on a meat slicer and a sausage knife in one of the shops (retailer A) and on a meat slicer and on the surface of the pork counter in the other (retailer B). **Figure 5** shows that both the cgMLST-based and the SNP-based metrics assigned the 2018 cases to cluster C1 along with the retailers' isolates. However, despite the overall high genetic similarity within cluster C1 (pairwise SNP difference: median = 2, range = 0–5), the SNP-based phylogeny provided

better resolution than cgMLST by differentiating (although with bootstrap value = 83) the isolates linked to retailer B, both clinical and environmental, from the others within cluster C1 (see **Figure 5**). The trace-back to a specific food source shared by the retailers was not conclusive.

In 2018, the regional surveillance system evaluated two ECDC urgent inquiries upon request of the Italian Operational Contact Point for listeriosis, UI-426 (28), and UI-444 (29), concerning multi-country outbreaks in Europe. No sequences fulfilling the matching criteria set in the inquiries were present in the regional database.

DISCUSSION

The results of the study reveal that the population of *Listeria monocytogenes* from humans of ER is predominantly structured in clusters of isolates similar enough to be indicative of several possible listeriosis outbreaks over the 7 years of observation. Indeed, between 66 and 72% of the isolates, depending on the method adopted, belong to genomic clusters. This finding contrasts with the evidence that no outbreaks were detected in ER before the implementation of WGS-based surveillance, essentially due to the scattering in time and space of the cases that were consequently classified as sporadic instead of outbreak-associated. This highlights the critical importance of WGS methods to support *L. monocytogenes* surveillance for outbreak identification. Supportive evidence of this was the first identification of an outbreak by *L. monocytogenes* in ER in the first year of adoption of prospective WGS surveillance in 2018. While this is widely recognized, the cluster-associated proportion of cases that we observed was

higher than those reported through cgMLST methods by Moura et al. (23), 40% (AD7 threshold), and Van Walle et al. (8), 45% (AD4) and 52% (AD7 threshold), who investigated the population structure of *L. monocytogenes* in France and Europe, respectively. As shown by the rarefaction analysis (see **Figure 3**), the observed difference with Moura et al. (23) could be attributed to the wider time-window (seven vs. 2 years) and to the higher rate of notified cases (0.91 vs. 0.59 per 100,000 population per year) in our study compared to the French one. Specifically, the rarefaction analysis predicts (both for AD4 and AD7) a drop of about 13% in the percentage of cluster-associated isolates following a reduction of the time-window from seven to 2 years. The observed difference with Van Walle et al. (8) could be attributed to the higher percentage of confirmed cases undergone WGS included in our study compared to the European one, where <24% of listeriosis cases reported in the European Union over a 6-year time-window were included, as opposed to 95% of the cases reported from Emilia-Romagna. Interestingly, our rarefaction analysis performed on a 6-year time-window including only 25% of our isolates predicts a proportion of cluster-associated isolates strikingly similar to that estimated by Van Walle et al. (8), i.e., 46.5% (with AD4) and 50.8% (with AD7). In agreement with our finding, Van Walle et al. (8) highlighted that the proportion of sporadic cases estimated in their work is likely to be lower for more comprehensive samplings. These findings indicate that a wide time-window and high surveillance intensity (intended as both high notification rate and high WGS-typing rate) represent critical conditions for the sensitive detection of potential listeriosis outbreaks in a WGS-based surveillance. This high rate of WGS-typed isolates can only be achieved through optimal local interoperation of diagnostic laboratories, typing laboratories, and health services which is easier to achieve through distributed systems on the territory than through centralized organizations. At the same time, our results confirmed that outbreaks can extend beyond a single region and even a single country as already observed in Europe (8, 29, 30). This simultaneous need of local-scale interoperation and large-scale cooperation suggests that networks of regional services and laboratories connected to each other in a coordinated organization could represent a proper approach to effective surveillance as a balanced solution to both needs.

Our study highlighted the existence of multi-year clusters, up to 5 year long. This finding reflects the ability of *L. monocytogenes* to persistently colonize food-processing facilities, as already reported (31–33) and has an effect on newly implemented WGS surveillance in its starting stage. In fact, gathering a multi-year retrospective dataset of genomic sequences is essential for reliable cluster detection at the beginning of surveillance when incoming isolates often belong to outbreaks originated in the past.

The performance of four previously validated cgMLST- and SNPs-based methods was evaluated. While generally providing comparable outcomes in cluster detection, cgMLST- and SNPs-based approaches may sometimes differ in assigning some

isolates to clusters and in the overall inclusiveness of isolates and resolution of clusters. In particular AD7-cgMLST determined the largest number of clusters and proportion of cluster-associated cases, while SNP-based detection with cut-off ≤ 5 SNPs provided the highest discrimination between clusters. According to these results, AD7-cgMLST appears the best rapid approach (screening tool) to detect putative outbreaks from WGS data with maximum sensibility, also considering its independence from dataset size and consequent scalability to large numbers of isolates and the immediacy to share information on cgMLST data among laboratories. On the other hand, SNP-based analysis on selected and relatively small sub-sets of data appears necessary to refine the analysis and accurately confirm clusters, as indicated by Schurch et al. (34) and Pightling et al. (6). The importance of adopting a multiple approach has clearly emerged in the investigation of the ST7 cluster, where both SNP-based analysis, including SNP differences and phylogenetic tree topology, and the analysis of the accessory genome (a prophage) were needed to elucidate the likely epidemiological dynamics of the cluster.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the EBI under Project number PRJEB34036.

AUTHOR CONTRIBUTIONS

ES, LB, and SP conceived and design the study. ES, CV, MM, IM, and MR collected the samples and performed the laboratory analysis. ES, LB, and IM analyzed the data. LB and SP drafted the manuscript. All authors revised the manuscript.

FUNDING

This study was supported by Italian Health Ministry grant for ES postdoctoral fellowship and PRC2017_001 for WGS surveillance system development.

ACKNOWLEDGMENTS

The network of clinical microbiology laboratories of Emilia-Romagna is greatly acknowledged for their active participation in the surveillance of listeriosis through the submission of the isolates of *Listeria monocytogenes* for genomic typing. The authors wish to acknowledge two referees, whose several suggestions helped us to significantly improve the quality of the paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2020.519293/full#supplementary-material>

REFERENCES

- World Health Organization (2015). *WHO Estimates of the Global Burden of Foodborne Diseases: Foodborne Disease Burden Epidemiology Reference Group 2007-2015*. WHO Library Cataloguing-in-Publication Data. p. 255.
- Jackson BR, Tarr C, Strain E, Jackson KA, Conrad A, Carleton H, et al. Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clin Infect Dis*. (2016) 63:380–6. doi: 10.1093/cid/ciw242
- Chen Y, Luo Y, Carleton H, Timme R, Melka D, Muruvanda T, et al. Whole genome and core genome multilocus sequence typing and single nucleotide polymorphism analyses of *Listeria monocytogenes* isolates associated with an outbreak linked to Cheese, United States, 2013. *Appl Environ Microbiol*. (2017) 83:e00633–e00617. doi: 10.1128/AEM.00633-17
- Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, et al. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nat Microbiol*. (2016) 2:16185. doi: 10.1038/nmicrobiol.2016.185
- Nielsen EM, Björkman JT, Kiil K, Grant K, Dallman T, Painset A, et al. Closing Gaps for Performing a Risk Assessment on *Listeria monocytogenes* in Ready-to-Eat (RTE) Foods: Activity 3, the Comparison of Isolates From Different Compartments Along the Food Chain, and From Humans Using Whole Genome Sequencing (WGS) Analysis. EFSA supporting publication 2017 (2017). p. 170. doi: 10.2903/sp.efsa.2017.EN-1151
- Pightling AW, Pettengill JB, Luo Y, Baugher JD, Rand H, Strain E. Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory applications and outbreak investigations. *Front Microbiol*. (2018) 9:1482. doi: 10.3389/fmicb.2018.01482
- Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, et al. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*. *J Clin Microbiol*. (2015) 53:2869–76. doi: 10.1128/JCM.01193-15
- Van Walle I, Bjorkman JT, Cormican M, Dallman T, Mossong J, Moura A, et al. Retrospective validation of whole genome sequencing-enhanced surveillance of listeriosis in Europe, 2010 to 2015. *Euro Surveill*. (2018) 23:1700798. doi: 10.2807/1560-7917.ES.2018.23.33.1700798
- EFSA/ECDC. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. *EFSA J*. (2017) 15:5077. doi: 10.2903/j.efsa.2017.5077
- EFSA/ECDC. Scientific report on the European Union one health 2018. Zoonoses report. *EFSA J*. (2019) 17:5926. doi: 10.2903/j.efsa.2019.5926
- Comandatore F, Corbella M, Andreoli G, Scaltriti E, Aguzzi M, Gaiarsa S, et al. Genomic characterization helps dissecting an outbreak of listeriosis in Northern Italy. *PLoS Curr*. (2017) 9:ecurrents.outbreaks.633fd8994e9f06f31b3494567c7e504c. doi: 10.1371/currents.outbreaks.633fd8994e9f06f31b3494567c7e504c
- Lomonaco S, Gallina S, Filipello V, Sanchez Leon M, Kastanis GJ, Allard M, et al. Draft genome sequences of *Listeria monocytogenes* strains from food isolates and human listeriosis cases from northern Italy. *Genome Announc*. (2018) 6:e01276–e01217. doi: 10.1128/genomeA.01276-17
- Orsini M, Cornacchia A, Patavino C, Torresi M, Centorame P, Acciari VA, et al. Whole-genome sequences of two *Listeria monocytogenes* serovar 1/2a strains responsible for a severe listeriosis outbreak in central Italy. *Genome Announc*. (2018) 6:e00236–e00218. doi: 10.1128/genomeA.00236-18
- Orsini M, Torresi M, Patavino C, Centorame P, Rinaldi A, Acciari VA, et al. Whole-genome sequence of a reemerging *Listeria monocytogenes* serovar 1/2a strain in Central Italy. *Microbiol Resour Announc*. (2018) 7:e01069–e01018. doi: 10.1128/MRA.01069-18
- Babraham B. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. (2010). Available online at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. GPL v3 or later.
- Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. (2014) 15:R46-2014-15-3-r46. doi: 10.1186/gb-2014-15-3-r46
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. (2012) 19:455–77. doi: 10.1089/cmb.2012.0021
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics*. (2013) 29:1072–5. doi: 10.1093/bioinformatics/btt086
- Stessl B, Ruckerl I, Wagner M. Multilocus sequence typing (MLST) of *Listeria monocytogenes*. *Methods Mol Biol*. (2014) 1157:73–83. doi: 10.1007/978-1-4939-0703-8_6
- Pettengill JB, Luo Y, Davis S, Chen Y, Gonzalez-Escalona N, Ottesen A, et al. An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with Salmonella. *PeerJ*. (2014) 2:e620. doi: 10.7717/peerj.620
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. (2014) 30:1312–3. doi: 10.1093/bioinformatics/btu033
- Moura A, Tourdjman M, Leclercq A, Hamelin E, Laurent E, Fredriksen N, et al. Real-time whole-genome sequencing for surveillance of *Listeria monocytogenes*, France. *Emerging Infect Dis*. (2017) 23:1462–70. doi: 10.3201/eid2309.170336
- Dallman T, Ashton P, Schafer U, Ironkin A, Painset A, Shaaban S, et al. SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. *Bioinformatics*. (2018) 34:3028–9. doi: 10.1093/bioinformatics/bty212
- Duranti A, Sabbatucci M, Blasi G, Acciari VA, Ancora M, Bella A, et al. A severe outbreak of listeriosis in central Italy with a rare pulsotype associated with processed pork products. *J Med Microbiol*. (2018) 67:1351–60. doi: 10.1099/jmm.0.000785
- Rounds JM, Hedberg CW, Meyer S, Boxrud DJ, Smith KE. *Salmonella enterica* pulsed-field gel electrophoresis clusters, Minnesota, USA, 2001–2007. *Emerging Infect Dis*. (2010) 16:1678–85. doi: 10.3201/eid1611.100368
- Rounds JM, Boxrud DJ, Jawahir SL, Smith KE. Dynamics of *Escherichia coli* O157:H7 outbreak detection and investigation, Minnesota 2000–2008. *Epidemiol Infect*. (2012) 140:1430–8. doi: 10.1017/S0950268811002330
- EFSA/ECDC. *Multi-Country Outbreak of Listeria Monocytogenes Sequence Type 8 Infections Linked to Consumption of Salmon Products*. Stockholm; Parma: EFSA supporting publication 2018 (2018). p. 16.
- ECDC/EFSA. *Multi Country Outbreak of Listeria monocytogenes Serogroup IVb, Multi-Locus Sequence Type 6, Infections Probably Linked to Frozen corn*. Joint ECDC – EFSA Rapid Outbreak Assessment. Stockholm; Parma: EFSA supporting publication 2018 (2018). p. 18.
- Schjorring S, Gillesberg Lassen S, Jensen T, Moura A, Kjeldgaard JS, et al. Cross-border outbreak of listeriosis caused by cold-smoked salmon, revealed by integrated surveillance and whole genome sequencing (WGS), Denmark and France, 2015 to 2017. *Euro Surveill*. (2017) 22:17–00762. doi: 10.2807/1560-7917.ES.2017.22.50.17-00762
- Morganti M, Scaltriti E, Cozzolino P, Bolzoni L, Casadei G, Pierantoni M, et al. Processing-dependent and clonal contamination patterns of *Listeria monocytogenes* in the cured ham food chain revealed by genetic analysis. *Appl Environ Microbiol*. (2015) 82:822–31. doi: 10.1128/AEM.03103-15
- Orsi RH, Borowsky ML, Lauer P, Young SK, Nusbaum C, Galagan JE, et al. Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment. *BMC Genomics*. (2008) 9:539. doi: 10.1186/1471-2164-9-539
- Vongkamjan K, Roof S, Stasiewicz MJ, Wiedmann M. Persistent *Listeria monocytogenes* subtypes isolated from a smoked fish processing facility included both phage susceptible and resistant isolates. *Food Microbiol*. (2013) 35:38–48. doi: 10.1016/j.fm.2013.02.012
- Schurch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic

analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect.* (2018) 24:350–4. doi: 10.1016/j.cmi.2017.12.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Scaltriti, Bolzoni, Vocale, Morganti, Menozzi, Re and Pongolini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.