Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze e tecnologie agrarie, ambientali e alimentari

Ciclo XXX

Settore Concorsuale: 07/H2

Settore Scientifico Disciplinare: VET/04

TITOLO TESI

Genomic methods for enhanced surveillance and persistence investigations of foodborne pathogens

Presentata da: Federica Palma

Coordinatore Dottorato **Supervisore**

Prof. Giovanni Dinelli **Prof. Gerardo Manfreda**

Esame finale anno 2018

Abstract

As result of globalization, the food supply chain became a vast and complex network that leads to an increased risk of spread of known and emerging foodborne pathogens (FBPs). *Listeria monocytogenes* and *Salmonella enterica* serovar Typhimurium variant 4,[5],12:i:- are foodborne pathogens with relevant rate of associated foodborne illness cases reported in Europe. Species-specific genetic bacterial determinants and matrix-specific ecological factors (e.g. physical factors, microbial competition, predation, etc.) can contribute to the persistence of these foodborne pathogens (FBPs) in the food chain. Whole Genome Sequencing (WGS) based analyses have recently revealed an unparalleled potential for multiple investigations as a one-serve-all approach. A valuable potential to enhance surveillance and persistence investigations of FBPs has been shown combining genomic approaches, in particular, to identify specific genetic features to be used as markers for phenotype prediction. In the studies presented in this thesis, genomics have been essential in investigating the persistence, virulence potential and physiological adaptation of *Listeria monocytogenes* in food processing plants (Study I) and the micro-evolution and phylogeography of *Salmonella enterica* serovar 4,[5],12:i:- (Study II).

Study I: Based on a previous study of persistence of *L. monocytogenes* in Italian rabbit meat production plants, two of most represented *ApaI*-PFGE profiles, belonging to subtypes ST14 and ST121, provided an interesting dataset to investigate persistence, virulence potential and physiological adaptation to food-processing environmental stresses. In this study, 27 ST14 and 6 ST121 newly sequenced genomes collected over one year on the same rabbit meat processing plant were investigated in comparison to a selection of publicly available genomes. Newly sequenced genomes were firstly clustered based on a 1748 loci cgMLST analysis which showed higher discriminatory power in comparison to conventional typing methods (PFGE, MLVA) for detecting two singletons not observed by other molecular methods. *In silico* Multi Virulence-Locus Sequence Typing and alignments of a representative database of virulence determinant genes were performed on ST121 and ST14 newly sequenced genomes in comparison to a set of publicly available genomes. ST14 strains were classified as Virulence-type VT107, a closely related subtype to VT1 strains, identified as part of an epidemic clone responsible of outbreaks in US. A full-length and a truncated version of *inlA* gene were identified in ST14 and ST121 respectively, predicting a higher virulence potential for ST14 isolates in comparison to ST121. wgSNPs based phylogenetic reconstructions were inferred on 273 newly sequenced and publicly available ST121 and ST14 draft genomes. Within ST14, a persistent clone was circulating in the Italian rabbit-meat plant isolates along with not persistent strains. Nevertheless, given the limited number of isolates it would be complimentary to investigate on a wider representative diversity of persistent and not persistent ST14. A novel dataset of accurately selected genes was designed to investigate physiological adaptation to food-processing environment of *L. monocytogenes*. A significant enrichment in ST121 genomes concerned genetic features related to adaptation to sanitizing procedures, whereas within ST14, genes conferring enhanced ability to form biofilm and heavy metals resistance was associated. Phenotypic tests performed on *L. monocytogenes* isolates from the Italian rabbit meat plant confirmed that ST121and ST14 were resistant to cadmium chloride 70 mg/ l and 35 mg/ l and that ST121 isolates are characterized by lower biofilm forming ability in comparison to ST14 isolates. Overall, these results suggest that ST14 strain-specific genetic repertoire may represent a selective advantage for this genotype to colonize harbourage sites where sanitizing procedures are difficult to occur. In conclusion, these study highlights that the

incidence and persistence of ST14 *L. monocytogenes* strains within food-processing plant should be taken into greater consideration, given its hypervirulent potential.
Study II: 148 *Salmonella enterica* ser. Typhimurium monophasic variant 4,[5],12:i:- (MVSTm) circulating in human and swine in Italy have been whole genome sequenced and investigated within an extended contest of selected publicly available *S.* Typhimurium/ MVSTm strains collected in Italy and worldwide. In this study, an innovative genome-wide investigative approach was applied. Population structure analysis was performed on a large dataset (~4,000 genomes) including several *Salmonella* serovars and Italian MVST strains revealing that this last belong to a large population of ~1,300 clonal *S.* Typhimurium/ MVSTm isolates (maximum of 2.5% of allele differences), collected from a wide-range of countries in last two decades. cgSNPs based phylogenetic reconstruction performed on these genomes revealed that isolates from the same geographical origin shaped several highly supported monophyletic groups, suggesting discrete geographical segregation. A Genome-Wide Association Study (GWAS) performed on the pangenome matrix of 1326 *S.* Typhimurium/ MVSTm identified a number of genetic markers (e.g. plasmid- and prophage-related genes) statistically associated with Italian origin. Evidence that geographical isolation has had a strong impact on the accessory gene content had particular significance for a large SopE-containing prophage detected in most of the isolates. In conclusion, these data suggest that the expansion of successful epidemic clones harbouring unique gene clusters, that constitute specific biomarkers, is driven by the adaptation of certain lineages to specific hosts or food production systems, with regard to a local geographical scale. Furthermore, in order to improve identification of the source in course of large epidemics, studies focusing on optimal dataset of genomes, representative of all the source of isolation of MVSTm in the food chain, should be performed even taking account of biogeographical genetic markers.

Table of content

1. Food Safety in the Genomic Era 3

1.1 Introduction 3

1.2 *Salmonella enterica* 5

1.2.1 Genome structure and virulence determinants of *Salmonella enterica* 5

Salmonella pathogenicity 6

Salmonella antigens 6

Salmonella antimicrobial resistance (AR) 7

1.2.2 Epidemiology of *Salmonella* infection 9

1.2.2.1 Epidemiology of multi-resistant monophasic *S.* Typhimurium strains variant 4,[5],12:i:- 10

1.3 *Listeria monocytogenes* 10

1.3.1 Genetics determinants of Listeria monocytogenes 11

Pathogenicity and virulence of *Listeria monocytogenes* 11

Persistence of *Listeria monocytogenes* 12

1.3.2 Epidemiology of *Listeria monocytogenes* infection 13

1.4 Molecular and genomic-based typing for enhanced surveillance of foodborne pathogens 14

1.5 Enhanced surveillance: from conventional molecular typing to Whole Genome Sequencing 15

1.5.1 Conventional molecular typing 15

1.5.2 Whole Genome Sequencing applied to foodborne pathogens 20

WGS for outbreak investigation and surveillance purposes 21

WGS for source attribution 21

1.5.2.1 Example of application of WGS on *Listeria monocytogenes* and *Salmonella enterica* 22

Listeria monocytogenes 22

Salmonella enterica 22

1.6 European initiatives for implementation of genomic-based typing for enhancing surveillance of food-borne pathogens 23

2. Bioinformatics for comparative genomics 25

2.1 *De Novo* assembly 26

2.2 Reference based read mapping and variant calling 26

2.3 Structural and functional genome annotation 27

2.4 “Gene-by-gene” approaches 27

2.5 Pangenome analysis and Genome Wide Association Studies 28
2.6 Phylogenetic approaches 29
2.7 *In silico* pheno-genotyping using WGS 29
2.8 Management, analyses and visualisation of WGS data 30

References 32

Aims of the work 50

Study I: Listeria monocytogenes in a rabbit meat processing plant: persistence investigations

Introduction 54
Materials and methods 55
Bacterial isolates 55
DNA extraction, sequencing and assembly 56
Bioinformatics data analyses for comparative genomics of *L. monocytogenes* 56
Genomic MLST analysis 56
Virulotyping of environmental ST14 and ST121 L. monocytogenes 56
Statistical analysis: index of diversity 56
Phylogenetic reconstruction based on SNPs analysis 57
Dataset of putative gene markers of ecophysiology in L. monocytogenes 57
Data availability 57
Phenotypic tests for confirming genotyping results 58
Antimicrobials, cadmium and benzalkonium chloride susceptibility tests 58
Crystal violet staining assay 58
Results 58
De novo assembly 58
Core genome MLST based phylogeny 60
Statistical evaluation of typing methods discriminatory power 61
In silico characterization of virulence potential 62
Phylogenetic trees based on SNPs analysis 62
Screening of putative gene markers of ecophysiology and related phenotypic tests 65
Discussion 66
Conclusions 69
Manuscripts submission and acknowledgements 69

References 70

Study II: Salmonella enterica ser Typhimurium variant 4,[5],12:i:- enhanced surveillance

Introduction 76
Materials and methods 77
Bacterial strains and genome sequencing 77
De novo assembly and *in silico* MLST 77
Salmonella reference genomic dataset 77
Whole genome MLST (wgMLST) schema creation, validation and allele calling 77
Genome annotation, pangenome analyses and genome-wide association studies (GWAS) 78
In silico antimicrobial resistance, plasmid and phage typing 78
Single-nucleotide polymorphism (SNP) analysis 78
Data availability 79
Results 79
Quality of *de novo* assembly 79
Population structure *Salmonella* genomes 79
Pangenome analysis 80
In silico characterization of the isolates 81
Genome-wide association study identified geographical segregated genetic markers in Italian MVSTm 82
Characterization of the prophage region 84
Phylogenomic reconstruction of the MVSTm strains 85
Discussion 88
Conclusions 90
Manuscripts submission and acknowledgements 90
References 91

Useful links 96
Discussion and conclusions 98
Acknowledgments

‘

1. Food Safety in the Genomic Era

1.1 Introduction

As result of globalization the food supply chain became a vast and complex network that leads to an increased risk of spread of known and emerging foodborne pathogens (FBPs). World Health Organization (WHO) has estimated the global impact of food and water-borne diseases (FWDs) in 2010 to be roughly 600 million cases and 420,000 associated deaths per year (1). In particular, the gastrointestinal FBPs *Escherichia coli*, non-typhoidal *Salmonella enterica* and *Campylobacter jejuni/ coli*, were the most common cause of illness compared with invasive infection diseases, with approximately 550 million cases and 230,000 deaths year (WHO, 2015). In EU countries, FWD infections (sporadic incidences and outbreaks) in human, and the presence of zoonotic agents are regularly monitored in food-producing animals and food in accordance with the Directive 2003/99/EC. Data on humans, animals and food are compiled and analysed jointly by the European Food Safety Agency (EFSA) and the European Centre for Disease Prevention and Control (ECDC) and presented yearly in the EU Summary Report on trends and sources of zoonosis, zoonotic agents and foodborne outbreaks. In 2015 at the European level, most of the FWDs infections were caused by FBPs and toxins, leading to 45,874 cases of illness, 3,892 hospitalisations and 17 deaths in 26 European Union member states (2). In comparison with 2014, the total number of human reported cases from the European Surveillance System (TESSy) on the most common FBPs [*Campylobacter jejuni/ coli, Salmonella enterica, Yersinia spp*., verotoxigenic/ shigatoxigenic *E. coli* (VTEC/ STEC, and *Listeria monocytogenes*] in the European Union (EU) showed a similar or increasing trend in 2015 (Figure 1) (2, 3) with different notification rates (Table 1).

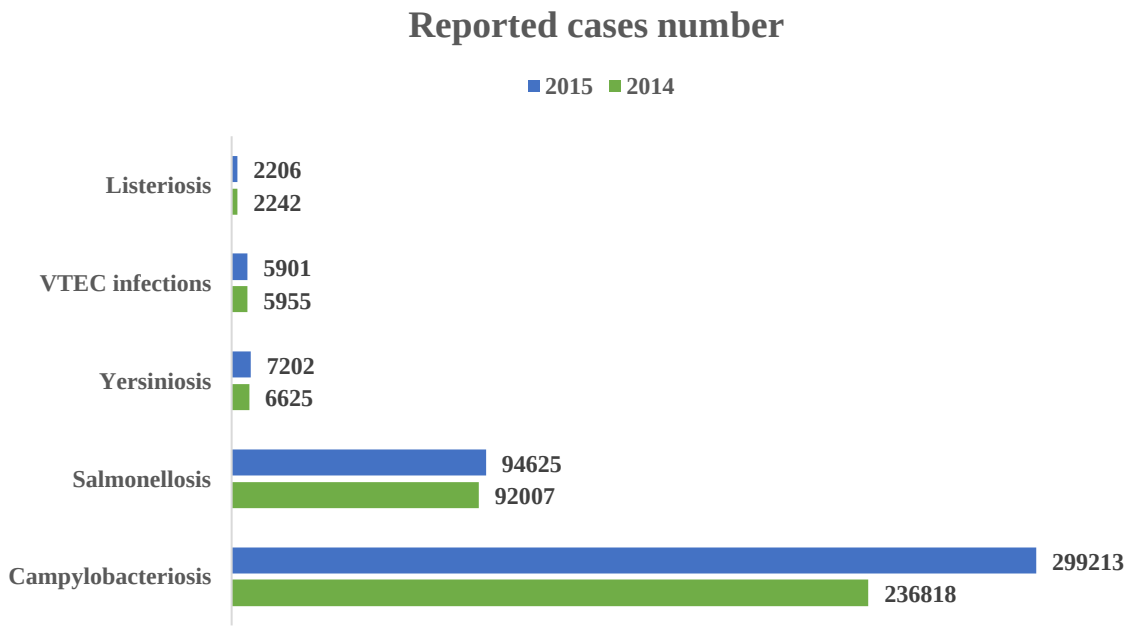

Figure 1. Notified cases of the most common foodborne diseases in EU in 2015 (blue) and in 2014 (green) (2, 3).

Table 1. Major bacterial food-borne bacterial pathogens, related diseases, human notification rate in EU and related foodstuffs and animal sources (2, 4).

Pathogen(s)/ Disease	Illnesses	Notification rate of cases per 100,000 population in EU in 2015	Foodstuffs and animals sources
***Salmonella* Enteritidis and *S.* Typhimurium/ Salmonellosis**	Self-limiting gastroenteritis, 5% of cases develop bacteraemia	21.2	Minced meat and meat (from poultry, turkey, pork, beef), ready-to-eat (RTE) food and eggs, vegetables and dried seeds. Laying hen, broiler, and turkey flocks, pigs and cattle but also animal- and vegetable-derived feed
***Listeria monocytogenes*/ Listeriosis**	Self-limiting gastroenteritis, sepsis, meningitis, encephalitis, abortion	0.46	Milk and dairy products, fishery products (mainly smoked fish), meat, eggs and RTE food. Domestic ruminants (cattle, sheep and goats), pigs and broilers but also pets and wild animals
***Campylobacter jejuni* and *C. coli*/ Campylobacteriosis**	Self-limiting gastroenteritis, ≤ 1% of cases develop Guillain-Barré and Miller-Fisher syndromes	65.5	Mainly poultry meat but also water and RTE food. Broiler, pigs, cattle, wild animals and pets
Verocytotoxigenic/ Shiga toxin–producing *E. coli* (VTEC/STEC)	Intestinal discomfort, haemolytic uremic syndrome, renal disease	1.27	Primarily meat from ruminants (sheep and goat) but also wild ruminants and cattle, Raw milk and dairy products. Very low in fruit and vegetables (sprouted seeds)

Although FWDs lead to significant economic losses for the agri-food sector and contribute substantially to the global border of diseases (5–7) measures to prevent and mitigate the risk of acquiring infections caused by FBPs are a significant challenge. Understanding the ecology of FBPs in different food matrixes is still a key problem in microbial risk assessment. Repeated food contamination due to food safety noncompliance promote the persistence of FBP which relies on species-specific genetic bacterial determinants and on matrix-specific ecological factors (e.g. physical factors, microbial competition, predation, etc.) (7). Moreover, the identification of the point-of-exposure source of foodborne outbreaks and, in general, the reservoir of FBPs are critical tasks in risk management and public health preventions. However, the population dynamic of several FBPs and the complexity of the transmission pathways make these goals hard to be achieved.
The introduction of genomics in the study of FBPs changed completely our understanding of microbial evolution, ecology and population dynamic. It revealed new insights into the persistence of FBP in food. Moreover, genomic based typing of FBPs turned out to be a key tool in retrospective and prospective epidemiological investigations ensuring enhanced surveillance (8–10). In the study presented in this thesis, genomics have been essential in investigating the evolution and phylogeography of *Salmonella enterica* serovar 4,[5],12:i:- and the persistence of *Listeria monocytogenes* in food processing plants. In the following chapters, the biology and epidemiology of both *Salmonella enterica* and *Listeria monocytogenes* as well as the impact of whole genome sequencing in FBP surveillance have been reviewed.

1.2 *Salmonella enterica*

Salmonella enterica, a facultative anaerobe, Gram-negative motile bacillus, was first isolated from pigs by Dr. Salmon and Dr. Smith in 1885 and is one of the leading cause of human gastroenteritis worldwide (11).
Salmonella genus was named in 1900 by Lignières in honor of Dr. Salmon, and originally it included numerous species. However, later it has been discovered that several of those species were indeed the result of antigenic variation of one. Therefore, *Salmonella* species were defined as "a group of related sero-fermentative phage-types" by Kaufmann (1961) after the development and application of serological analysis, firstly conceiving by White (1926): the White-Kauffman-Le Minor scheme. This serotyping scheme is based on the antigenic variation of the *Salmonella* Lipopolysaccharide O-chain and on variation in the major protein of the flagellum (Antigen H) (12, 13). The White-Kauffman-Le Minor serotyping scheme is still the central method for *Salmonella* nomenclature (14) and it is currently maintained and updated by the World Health Organization (WHO) Collaborating Centre for Reference and Research on *Salmonella* at the Pasteur Institute, Paris, France (WHO Collaborating Centre) (15).
Currently the genus is composed by two species, *S. bongori* and *S. enterica* and the latest divided into six subspecies: *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV) and *indica* (VI). *Salmonella enterica* ssp. *enterica* (I) is the most important subspecies accounting for the majority of human and animal infectious. According to the White-Kauffman-Le Minor scheme, more than 2,500 different serotypes are classified within this subspecies (13, 16).
Salmonella enterica serovars are either strictly adapted to one particular host or generalists, meaning they are present in a wide range of hosts. Salmonellosis from human-restricted serovars like Typhi and Paratyphi causing typhoid and paratyphoid fever respectively (17) are transmitted from person to person, without an intermediate host, through faecal contamination of water and food. Non-typhoidal *Salmonella* serovars in humans are usually caused by generalist serovars. Transmission to human is by faecal-oral route, and the majority of cases are foodborne, caused by consumption of contaminated food of animal origin (18). Most of the non-typhoidal serovars within *Salmonella enterica* cause gastroenteritis with *S. enterica ssp. enterica* Typhimurium and Enteritidis the most common pathogenic serovars, responsible of foodborne outbreaks. Low infective doses are sufficient to cause clinical symptoms (19). Non-typhoidal *Salmonella* normally causes gastroenteritis, bacteraemia, and subsequent infection (16) with an estimation of 93.8 million cases of gastroenteritis globally each year, including 155,000 deaths (20). Although most human infections are self-limiting, salmonellosis of new-borns and infants (more susceptible to infections than adults) presents diverse clinical symptoms, from a grave typhoid-like illness with septicaemia to a mild or asymptomatic infection.
Nevertheless, after recovery from a clinical case of salmonellosis, some patients, although asymptomatic, could remain carriers for a prolonged time contributing to the dissemination of diseases. Interestingly, antibiotics used in human therapy are usually ineffective in the treatment of the carrier state (21).

1.2.1 Genome structure and virulence determinants of *Salmonella enterica*

Already in pre-genomic era, the physical map of the chromosome of several strains shows that *Salmonella* serovars share similar genome size (between 4.6 and 4.8 Mb), all have seven *rrn* operons, and the order of genes on chromosome segments is usually the same (22–24). It was already clear that genomic diversity across *Salmonella* strains and serovars is determined by gain/ loss of functions via horizontal gene transfer. Several strains harbour "foreign" mobile genetic elements (MGEs), such as phages and plasmids, that include genes linked to virulence, antimicrobial resistance, antigenic changes of O antigen, and metabolic characteristics commonly used in diagnostic identification (e.g. lactose or sucrose fermentation) (21). The advent of genomic era these features have been extensively

characterized. Below a non-exhaustive summary of the main genomic features related to pathogenesis, antigenic variation and antibiotic resistance.

Salmonella pathogenicity

Several steps are involved in *Salmonella* pathogenicity starting from the invasion of the small bowel mucosa to the rapidly adhesion and invasion of follicle-associated epithelium cells and absorptive enterocytes. Subsequently, *Salmonella* overcome host-specific defence mechanisms which include antibacterial actions of phagocytic cells coupled with the immune response. The interaction between *Salmonella* and the host is a complex mechanism. and depends on several loci. Most of these genes are located in five highly conserved chromosomal *Salmonella* pathogenicity islands (SPI-1-5) and in virulence plasmids (e.g. pSLT), essential for dissemination in the appropriate host, guaranteeing bacterial survival. Gene products of the SPI-1 are key determinants of the intestinal inflammation induced by *Salmonella* (25). Into host cells, phage-mediated horizontal transfer and SPI-1 type 3 secretion system (T3SS-1) translocate effector proteins (SopB, SopE, and SopE2) that can induce inflammation within the gut (26). In particular, it has been shown that *sopE* gene can increases the severity of intestinal inflammation (27), and increase the fitness of *S*. Typhimurium by conferring a nitrate respiration-dependent luminal growth advantage (26). Only a limited number of *Salmonella* serovars within subspecies *enterica* carry a large, low-copy-number plasmid that contains virulence genes. *Salmonella* serovars-specific virulence plasmid, containing a highly conserved 7.8 Kb region, *spv*, can varies in size: 95 kb for serovar Typhimurium, 60 kb for Enteritidis, 80 kb for Dublin (28). Furthermore, the pSLT plasmid also contains *pef*, a fimbrial operon encoding genes involved in colonization of the small intestine (29).

Salmonella antigens

The O-antigen (or somatic antigen), the external component of the lipopolysaccharide, consists of a long linear polysaccharide and is classified using the characteristic O factor (e.g. O:4). The genes involved in the biosynthesis of the O antigen are generally found in the chromosome in the cluster (frequently a single operon) names as *rfb* gene cluster and are divided in three groups: (i) genes involved in the biosynthesis of the precursors of nucleotide sugars; (ii) glycosyltransferases; (iii) gene involved polymerization and in the translocation through the membrane (30). Diverse combination of these three set of genes are at the basis of the antigenic variation of the O-antigen. The H-antigen (or flagella antigen) correspond to the flagellin which is the major component of the flagella responsible of *Salmonella* motility. Many *Salmonella* serovars contain a biphasic character of the flagellar antigen giving the ability to change its composition with a switch for the expression of two loci encoding the major flagellar protein, FliC (phase 1 antigen) or FljB (phase 2 antigen) flagellin, described as "phases". Therefore, the capacity of *Salmonella* to express two distinct flagellar antigens is termed "biphasic" (e.g. *Salmonella* Typhimurium) and results in a "phase variation" with respect to their flagellar antigen. In order to express only one variety of flagellin protein at time (31), a switch mechanism, under the control of the recombinase Hin, regulate the expression of these two loci to facilitate inversion of a promoter element so that it transcribes fljB (which encodes the phase 2 antigen FljB) and fljA (which encodes a repressor of fliC, the gene encoding the phase 1 antigen FliC). If the promoter is located in an orientation that does not allow for transcription of *fljB* and *fljA*, the lack of a repression of *fliC* transcription leads to expression of phase 1 flagellar antigens (Figure 2) (31). Based on flagellar antigen expression ability, monophasic variant (producing only one type of flagellin), and less frequently triphasic or other complex (producing more than three flagellins) variants were identified and detected among *Salmonella* serovars. A small type letter ‘a to z’ indicate the first phase, followed by the second phase of the H-antigens indicated by numbers or letters. If *Salmonella* do not have H-antigens is considered non-motile and does not possess flagella. The Vi-antigen is a surface (capsular) polysaccharide structure and occurs in only three *Salmonella* serovars (*Salmonella* Typhi, *Salmonella* Paratyphi C and *Salmonella* Dublin) (32).

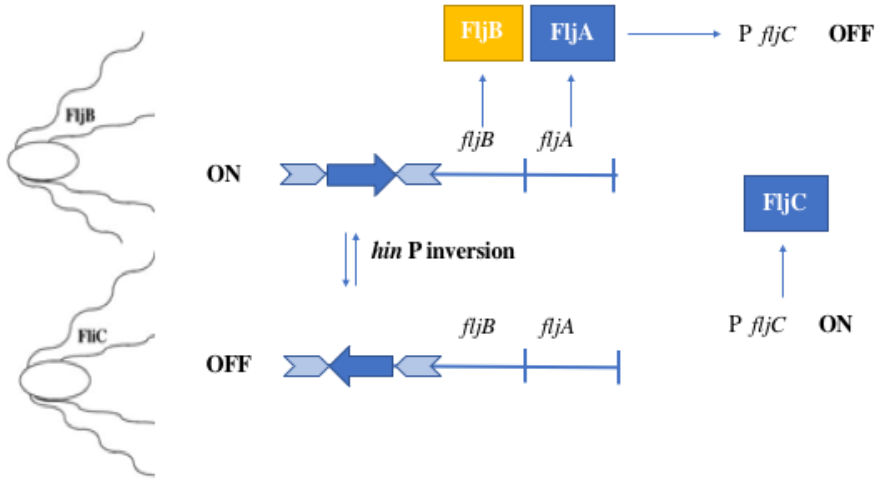

Figure 2. Flagellar phase variation in *S.* enterica. The FljB flagellin is produced when the promoter (P) for the *fljBA* operon is expressed (orientation ON) along with FljA, repressor of *fliC* gene that encodes FliC flagellin. The promoter inversion mediated by the Hin recombinase (orientation OFF) implicate that neither the fljB gene nor the repressor FljA are expressed, allowing the transcription of *fliC* (adapted from Soyer et al., 2009) (31).

Salmonella antimicrobial resistance (AR)

As other bacterial species, *Salmonella* can readily acquire through horizontal transfer MGEs containing genes conferring resistance to a number of antimicrobials. These genes could be located either in chromosomal regions (SGI-1) or in plasmids (33, 34). In addition, point mutations in housekeeping genes can confer resistance to selected antimicrobial agents (e.g., fluoroquinolones and quinolones). Extensive use of antibiotics lead to the selection of strains harbouring multiple resistance genes in MGE which can move between human and animal through the food chain (35). In particular, antibiotic selective pressure events lead to the spread of strains with plasmids-mediated MDR to high priority therapeutic antibiotics (36). Among *Salmonella* serovars with multidrug resistance (MDR) *S.* Typhimurium and monophasic variant 4,[5],12:i:- strains are global health problems and are commonly found in animal species such as poultry, pigs and sheep (37, 38). In particular, *S.* Typhimurium definitive type (DT) 104, monophasic U302 and DT193 variants have been described as largely responsible for the increase of MDR *Salmonella* isolates in Europe (39, 40). Several studies have identified the specific resistance genes and genetic mechanisms associated with antimicrobial drug resistance phenotypes in different *Salmonella* serovars (41), and it is summarized in Table 2.

Table 2. Common resistance genes associated to the antimicrobial compounds class and Salmonella serotypes (adapted from Alcaine et al., 2007) (41).

Associated resistance genes	**Antimicrobial compounds class**	***Salmonella* serotypes**
aac(3)-IV, aac(3)-IVa, aacC2, strA, strB, aph(3)-IIA, aadA1, aadA2, aadB	**Aminoglycosides**	4,5,12:i:-, Agona, Anatum, Blockley, Bredeney, Derby, Give, Hadar, Heidelberg, Kentucky, London, Infantis, Saintpaul, Newport, Typhimurium
blaCMY-2, blaCTX-M9, blaTEM-1, blaTEM-53, blaCARB2, blaOXA-30	**Beta-lactams**	Anatum, Agona, Blockley, Dublin, Enteritidis, Haardt, Muenchen, Newport, Stanley, Typhimurium, Virchow
cat1, cat2, cmlA, floR	**Chloramphenicols**	Albany, Agona, Derby, Enteritidis, Haardy, Kiambo, Newport, Typhimurium
gyrA, gyrB, parC	**Quinolones ***	Enteritidis, Typhimurium
tet(A), tet(B)	**Tetracyclines**	Agona, Anatum, Blockley, Bredeney, Colorado, Derby, Dublin, Enteritidis, Give, Haardt, Hadar, Heidelberg, Infantis, Orion, Senftenberg, Typhimurium
sul1, sul2, sul3	**Sulfonamides**	4,5,12:i:, Agona, Albany, Anatum, Brandenburg, Derby, Djugu, Enteritidis, Hadar, Heidelberg, Orion, Rissen, Typhimurium

* Quinolones resistance is mediated by point mutations in the associated gene

1.2.2 Epidemiology of *Salmonella* infection

In the EU, notification and surveillance of food-borne salmonellosis in humans is mandatory in accordance to European Commission Decision 2000/96/EC and Decision No 2119/98/EC. Since 2007, EU and EEA countries report their public health surveillance data to TESSy at ECDC. Among all the serovars, *S.* Enteritidis (39.5%%-45.7%), *S.* Typhimurium (15.8%-20.2%) and S. Typhimurium monophasic variant 1,4,[5],12:i:- (~8%) were the three most frequently reported in the 2013-2015 period (Figure 3) (2).

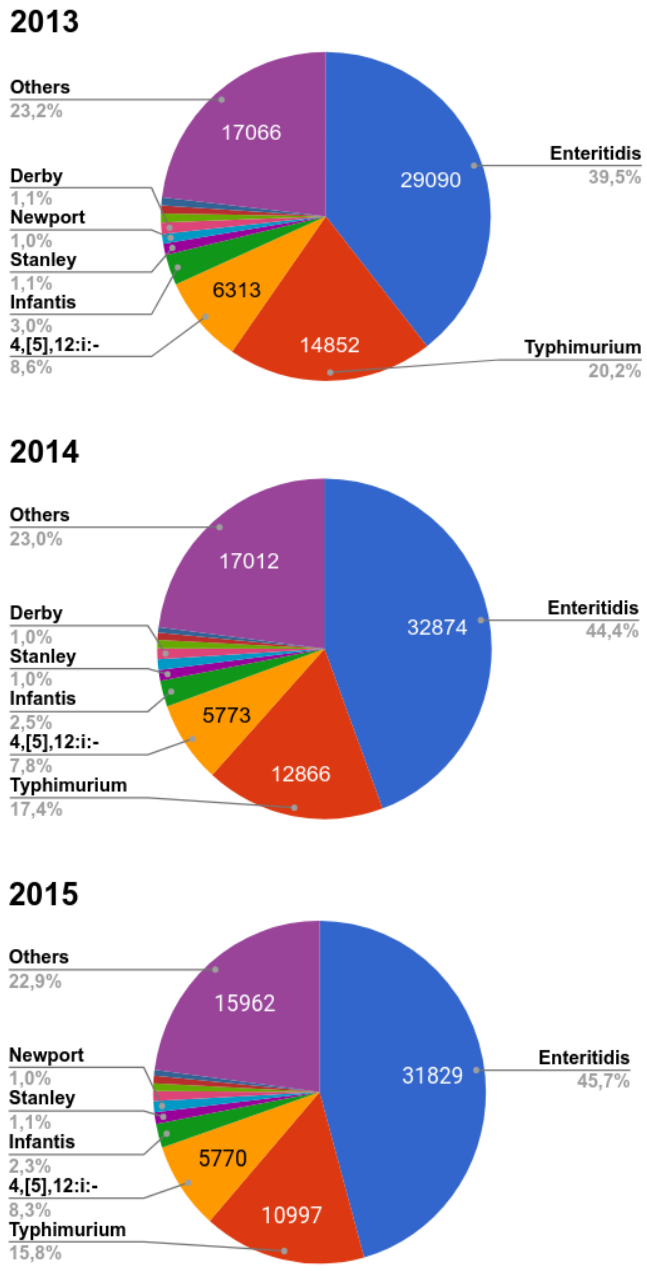

Figure 3. Distribution of the 10 most frequent non-typhoidal *Salmonella* serovars in 2015, 2014, 2013 (2).

Among the implicated food vehicles those of animal origin were dominant: eggs and egg products, pig meat, broiler meat and cheese, followed by fish and fish products, milk and dairy products, bovine meat and crustaceans. Among the reported place of exposure by Member States in 2015, household was by far the most frequent accountable in terms of risk exposure for food-borne outbreaks. A typical factor that contribute to the spread of a particular serovar into the food chain is the import of food. Indeed, besides a decreasing trend in 2015, the highest number of reported food-borne outbreaks caused by *Salmonella*, in particular *Salmonella* Enteritidis was associated with the consumption of imported eggs and egg products.

Overall, the annual percentage of the *Salmonella* Typhimurium isolates progressively decrease over last year's, but an increase was observed in the number of isolates from pigs in EU. Notably, besides a decreasing trend of *S.* Typhimurium in pork meat was observed, the number of isolates of monophasic strains of *S.* Typhimurium 4,[5],12:i:- constantly increased. In particular in Italy, Portugal and United Kingdom (UK), this specific serovar were collected in more than 50% of *Salmonella* serovars isolated from pig meat and were also very frequent among pig isolates in Italy and UK (2). *S. enterica* serovar 4,[5],12:i:-, considered a monophasic variant of Typhimurium, appears to be antigenically similar and genetically closely related to *Salmonella* Typhimurium (31, 42, 43). Moreover, the emergence of this serovar is highlighted by the high level of resistance to antimicrobials and heavy metals for the different clones of monophasic strains of *S.* Typhimurium, especially along the swine chain (44).

1.2.2.1 Epidemiology of multi-resistant monophasic *S.* Typhimurium strains variant 4,[5],12:i:-

In the United Stated, *S.* Typhimurium and its monophasic variant 4,[5],12:i:- were responsible of 6,500 and 1,019 human illness, respectively, from 2002 to 2012 (45). In particular, a significant increase of *Salmonella* 1,4,[5],12:i:- causing human diseases was observed also in Canada from 2003 to 2010, becoming the fifth most common serovar in 2007 (46). In Europe, an increasing emergence of monophasic variant *S.* Typhimurium have been reported already in 2010 (EFSA). Several studies in numerous EU countries confirmed the rapid emergence and dissemination of such strains in food animals, companion animals and humans (32, 47–49). In particular, serotype 1,4,[5],12:i:- has been described as the most frequently identified serotype in swine in Spain (50) as well as the causative agent of at least two *Salmonella* outbreaks in Luxembourg always linked to consumption of contaminated pork (51).
S. Typhimurium and its monophasic variant have similar virulence-related features as well MDR patterns and are generally resistant to a variable number of different drugs (ampicillin (A), chloramphenicol (C), streptomycin (S), sulfamethoxazole (Su), tetracycline (T), trimethoprim (Tp), nalidixic acid (NA) and gentamicin (G). Several studies described monophasic variant 1,4,[5],12:i:- strains as belonging to multiple clones or clonal lines (31, 44, 51–54). Multiple clones of *Salmonella* serotype 1,4,[5],12:i:- with distinct geographical distributions, MDR profiles and deletion patterns have been described as emerged through independent deletion events (31, 44, 55). Within European countries two major clonal lines of monophasic *S.* Typhimurium, the so-called European and Spanish clones, have been described. The first reported one was the 'Spanish clone' in the late 1990s, characterised by a plasmid-mediated resistance against up to seven antimicrobials, ACSuGSTTp, also described as a heptaresistance by Petrovska and colleagues (44, 54, 55). In contrast to the 'Spanish clone', the 'European clone' emerged since 2000 and has been characterised by a predominant ASSuT resistance profile, harbouring the AR genes on a chromosomal region (33, 34). These clones have probably evolved from a traditional biphasic S. Typhimurium resulting in a monophasic variant due to the lack of the second-phase flagellin-encoding gene or to the ability to express it (56).
Within the European clone, the majority of monophasic *S.* Typhimurium strains belong to the phage types DT193 or DT120 (51, 57) whereas the Spanish clone includes strains belonging mostly to the phage type U302 (50). The reasons why these MDR strains colonize successfully are still unknown. The molecular basis for the success of specific epidemic clones of bacterial pathogens has important implications for the surveillance and management of outbreaks related to monophasic S. Typhimurium. Several factors including improved genetic mechanism to survive in the host and acquisition of bacteriophages encoding AR to additional molecules such as heavy metals and detergents affect the fitness, the virulence and adaptive response of monophasic S. Typhimurium (36)

1.3 *Listeria monocytogenes*

Listeria ssp. are Gram positive rods, psychrophile, ubiquitous bacteria and was first isolated in 1924 from livers of laboratory rabbits with bacterial sepsis by Dr. Murray (58). The genus includes six species from which *Listeria monocytogenes* turned out to be by far the most pathogenic one for humans. *L. monocytogenes* is a highly heterogeneous species and can be divided into four major evolutionary lineages (59–61), 13 serotypes (21) and four PCR serogroups (62). Multilocus sequence typing (MLST) further subdivides the above categories into clones, which are geographically and temporally widespread (63, 64). *Listeria monocytogenes* MLST clonal complexes (CC) have been recently defined by Maury and colleagues as described in Supplementary Note available from the study (65). Herein, the authors stated that clonal complexes belonging to Lineage I, which includes serovars 4b, 1/2b and 3b, are more frequent among clinical isolates then other lineages. Clonal complexes belonging to Lineage II, including serovars 1/2a, 1/2c and 3c (Figure 4) are the most

frequently detected in food and environment and relative underrepresented in clinical samples (60, 65).

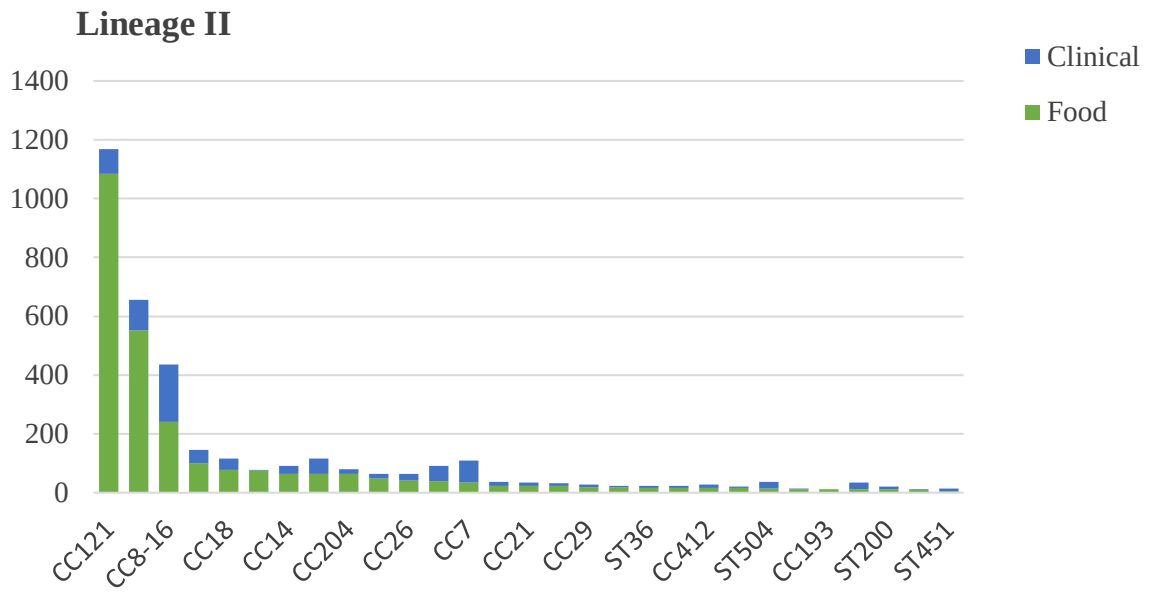

Figure 4. Prevalence of MLST clones belonging to Lineage II and isolated from 2005 and 2013 in food and clinical sources (adapted from Maury et al. 2016) (65).

1.3.1 Genetics determinants of Listeria monocytogenes

Genome evolution in *Listeria* has involved limited gene gain and loss as suggested by relatively high conservation of the predicted pan-genome (66). Conserved genome size ranges between 2.8 and 3.2 Mb, a low C+G content of about 37-38% and a highly synthetically linked genome (67, 68). The genus *Listeria* provides an example of a group of bacteria that appears to evolve through a loss of virulence rather than acquisition of virulence characteristics (66). In fact, the limited gene loss described in *Listeria* includes mainly virulence genes, likely associated with multiple transitions to a saprotrophic lifestyle.

Pathogenicity and virulence of *Listeria monocytogenes*

Listeria monocytogenes is the etiologic agent of listeriosis, an opportunistic, invasive illness which occurs in immunocompromised individuals, such as HIV patients, elderly persons, infants, and pregnant women (69). The potential severity of *L. monocytogenes* infection is linked to the invasive capacity of the bacterium. A non-invasive form of *listeriosis* is known to occur in immunocompetent individuals, and it is characterized by febrile gastroenteritis and flu-like symptoms (70). By far the most common passage for the bacterium into the human body is via the digestive tract. *L. monocytogenes* crosses the intestinal epithelium and disseminate through the body by exploiting the lymph and blood circulation and thereby reaching the liver and the spleen. In particular, *L. monocytogenes* is able to penetrate and adhere into different types of eukaryotic cells, to multiply and spread into the cytoplasmic reaching the adjacent cells. The molecular mechanisms involved in the different stages of infection are relatively well known and involve numerous virulence factors. The main *Listeria monocytogenes* virulence gene cluster, also known as the *prf*A virulence cluster or the *Listeria* pathogenicity island (LiPI), encodes several proteins that are necessary for intracellular survival and motility. Specific functions encoded in this cluster include hemolysin, two phospholipases and a metalloprotease (encoded by *hly*, *plc*A, *plc*B, and *mpl* genes), which all

contribute to escape from host cell vacuoles, an actin polymerizing protein (encoded by *act*A), and a global regulator of virulence gene transcription (encoded by *prf*A). In particular, the latter, PrfA, is the main transcriptional regulator for key virulence factors and (directly or indirectly) controls the expression of 145 other genes, which are encoding transporters, metabolic enzymes, regulators as well as of the members of the regulation of stress response sigma B, a general stress-responsive sigma factor in Gram-positive bacteria (71). Members of the internalin protein family, which are cell wall anchored or secreted proteins that are characterized by the presence of leucine rich repeats, are also associated with virulence in *L. monocytogenes*. While a considerable number of genes encoding internalins have been found in pathogenic and non-pathogenic *Listeria*, clear virulence related functions have been assigned to specific alleles, including *inlA* and *inlB*, which encode proteins required for invasion of different cells types, including human intestinal epithelial cells, and *inlC* (71). Nonetheless, premature stop codons (PMSC) in the key virulence genes *inl*A and *prf*A resulting in the production of truncated InlA and PrfA proteins, were associated to attenuated virulence (72–75).

Persistence of *Listeria monocytogenes*

Listeria monocytogenes is able to persist within food processing environments for long periods of time, due in part to its ability to grow at wide-ranging temperatures and pH (-0.4°C to 45°C, optimum 37°C; pH 4.39 to 9.4, optimum 7.0 (ICMSF, 1996)), to resist to desiccation and to the ability to form biofilms (7). The persistence of a single subtype of *L. monocytogenes* in processing facilities or on equipment has been reported for up to 10 years, resulting in contamination of food at several stages from farm to fork (76). The contamination route (Fig. 5) starts from the direct contamination of raw food material (meat, milk, vegetable and fish) by water, microflora and soil. This last plays a central role in the transfer of the pathogen into farms resulting in the contamination of raw material that is one of the main entries for the pathogen into the food processing environment. A relevant role may also be played by livestock, a reservoir for *L. monocytogenes* that can survive in the environment since contaminated faeces and manure are transferred back to the soil.

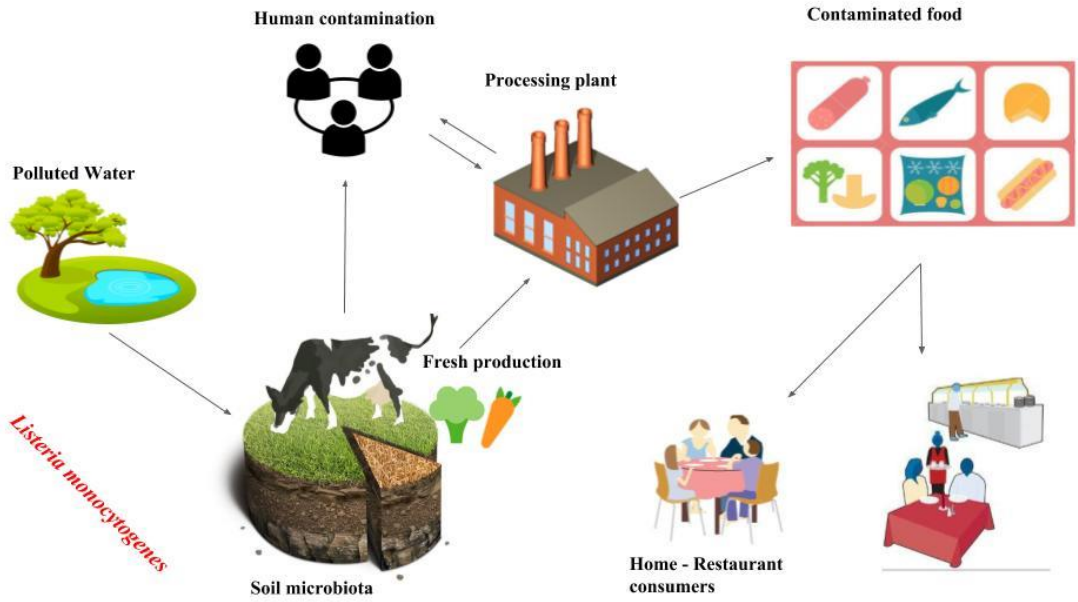

Figure 5. Contamination route of *Listeria monocytogenes* from the natural environment to humans.

Furthermore, the raw material contamination can be transferred to processing and post-processing of food, defined as secondary contamination. The risk of this contamination is for all food sectors and include different environments and tools where operators of the food chain handle the food items during the processing of food as well as the storage and transportation of the product.
There is a gap of knowledge concerning mechanisms of persistence, although several studies focused on investigating the tolerance to environmental stresses such as cold, salt, acid and oxidative stresses as well as resistance to heavy metals and quaternary ammonium compounds (QAC) (77–80) of *L. monocytogenes* isolated from food processing plants. In particular, tolerance to QAC disinfectants and heavy metals have been largely described alone or in combination in *L. monocytogenes* persistent and presumed non-persistent strains in food processing plants (77, 79, 81, 82), showing contradictory results. Elhanafi and colleagues (77) identified a plasmid-based gene cassette that conferred increased resistance to benzalkonium chloride, a widely used QAC disinfectant in the food industry. Several genetic determinants (e.g. *bcr*ABC, *erm*E) were described as associated to benzalkonium chloride resistance and located in *Listeria* genomic island 1 (LG1). In addition, a novel transposon, Tn6188, carrying the transporter QacH, which is responsible for benzalkonium chloride (BC) tolerance, has been described (83–86). Regarding heavy metals, the plasmid-mediated *cadAC* determinant were described as associated to inducible cadmium resistance mediated by efflux mechanism (87). Moreover, a novel cadmium resistance determinant in *L. monocytogenes* has been recently described as *cadA4* providing evidence for roles of this genetic cassette in enhanced cadmium tolerance as well as in virulence and biofilm formation (80). Several studies have investigated attachment/ biofilm formation of persistent and presumed non-persistent strains of *L. monocytogenes* to surfaces, with varying results. For example, a study by Lunden and colleagues showed that persistent strains enhanced attachment over short periods of time, however, some presumed non-persistent strains matched, or in some cases surpassed, the levels of attachment of persistent strains after 72h. A possible explanation for this observation may involve differences in flagella between different strains; flagella have been shown to facilitate early attachment to stainless steel (88). A more recent study described a better adherence of persistent strains from dairy environment than sporadic strains (89) while a higher biofilm formation among persistent compared to non-persistent strains from bulk milk samples was also described (90).
Stress tolerance have been observed in *L. monocytogenes* associated to specific serotypes and clonal complexes (91–93). Gene cluster involved in tolerance to low pH and high salt concentrations in *L. monocytogenes* predominantly belonging to serotypes 1/2c, 3b and 3c have been described so far as Stress Survival Islet 1 (SSI-1) and SSI-2 (93). Further gene cluster involved in alkaline and oxidative stress response in *L. monocytogenes* predominantly belonging to ST121 have been described as well (91, 94). Higher cold tolerance has been recently associated to the biomarker of virulence *inl*A gene, in that strains harbouring a full-length *inl*A gene (92). Other factors, alone or in combination, such as the reintroduction of the "persistent" strain-type from an external habitat, the recontamination events due to not sufficiently sanitised sites, the increased survival and growth capacity at the conditions in particular niches (95) could influence the persistence of *L. monocytogenes* phenomenon at secondary production (7).

1.3.2 Epidemiology of *Listeria monocytogenes* infection

As previously mentioned, *Listeria monocytogenes* is transmitted via three main routes: direct contact with animals, cross-infection of new-born babies in hospital and food-borne common vehicles of infection as packaged ready-to-eat foods. Listeriosis is associated with enormous public health and economic consequences due to the high hospitalization rate (91%) and fatality rate (20-30%) among foodborne illnesses, with most cases reported to be sporadic in nature, and associated with long-term sequelae (96, 97). In Europe, the observed trend of listeriosis changed during 2001-2006 where an increasing in the number of confirmed cases (1,558 human cases of listeriosis were reported in 2007) was observed especially in patient over 60 years old (98). Moreover in 2015, of the 2,206 confirmed

human cases of listeriosis in EU countries 270 deaths occurred showing a statistically significant increase trend over 2008-2015 (2).
Ready-to-eat (RTE) products with a long shelf-life are associated with a direct risk of transmission of *L. monocytogenes* that is detected above the legal safety limit (100 CFU/g). Moreover, non-compliance in samples of fishery products (mainly smoked fish), dairy products (other than cheeses) and meat products collected at processing have been reported. Even if meat products are cooked before consumption, *L. monocytogenes* can survive and grow during storage, posing cross-contamination issues in processing plants as well as at household level. At retail, non-compliance was highest in batches of fishery products and 'soft and semi-soft cheeses' (2). It is recognised that the presence of *L. monocytogenes* in raw foods cannot be completely eliminated, but through the application of effective hygiene measures, it is possible to reduce its occurrence and level. In addition, to ensure the safety of food products, all the processing and food supply systems must be carried out in order to control the growth of *L. monocytogenes* and to prevent its multiplication in the food-processing environment (99). Therefore, it is important to investigate which *L. monocytogenes* strains contaminate food products and persist in processing plants and which are the stress-tolerance and virulence genetic determinants that may influence infection and clinical outcome.

1.4 Molecular and genomic-based typing for enhanced surveillance of foodborne pathogens

Contamination events in the production or distribution of food leading to FWDs may be hard to detect and correlate, and have a big impact on the public health especially if geographically dispersed. To prevent and control FWDs within the European Union, three platforms have been developed for rapid exchange of information between member states (MS) on detected food-borne threats in humans and hazards in food or feed:

1. The Epidemic Intelligence Information System for Food- and Waterborne Diseases (EPIS-FWD): an ECDC-hosted platform for information communication and exchange on emerging clusters and outbreaks as well as unusual increases in human cases detected at the national level. Since it was launched in 2010 more than 300 outbreaks have been assessed from extra EU/EEA countries (100).
2. The Early Warning and Response System (EWRS): a web based official notification system linking the European Commission (EC) and competent Public Health Authorities (PHA) and MS to control communicable diseases event at the EU level, ensuring a rapid and effective response.
3. The Rapid Alert System for Food and Feed (RASFF): an official system for sharing information on hazards found in food and feed and trade of (potentially) contaminated batches between MS, and for tracing these batches back and forward (https://ec.europa.eu/food/safety/rasff_en).

Since 2012, molecular typing data of strains isolated from human *Salmonella*, *L. monocytogenes*, and STEC infections are submitted to ECDC by public health authorities and laboratories of the MS through the common-shared database system TESSy. Non-human typing data collected by EU/ EEA countries on related FBPs isolates from food/feed/animals and environment are reported to EFSA (through the joint EFSA/ ECDC molecular typing data collection system) by the food and veterinary authorities and MS laboratories. These network process, is aimed to perform regular joint analysis of the molecular typing data in order to share comparable typing data in a common repository, referred to as 'the joint database', so that human's data can be linked to similar data from the food chain (101, 102). Moreover, the analysed data are published annually in two European Union Summary Reports: one report on zoonoses, zoonotic agents and foodborne outbreaks, and another on antimicrobial resistance (2, 3).

At the EU level, Directive 99/2003/EC and Decision 1082/2013/EU provides criteria for data collection from humans and food, with a view to an integrated multidisciplinary approach in a context of real-time molecular surveillance of FBPs. For this purpose, ECDC developed a set of specific principles and prerequisites relevant for a routine EU-level surveillance of circulation food and clinical strains and outbreak investigations. External quality assessment (EQA) schemes to support molecular typing methods in order to improve data quality and allow internationally comparable results across laboratories were developed (2, 100, 101). Consequently, standardised Pulsed Field Electrophoresis (PFGE) and Multi Locus Variable tandem repeat Analysis (MLVA)-based methods emerged for the characterisation and comparison of *Salmonella*, *Listeria monocytogenes*, STEC (103, 104) and of *Salmonella* serovars Enteritidis and Typhimurium (105), respectively.
Several other initiatives have been promoted for enhancing surveillance of foodborne pathogens. Local capacity and a robust and harmonised surveillance system are paramount for rapid detection and investigation of outbreaks supported by good communications between clinical laboratories and the public health authorities (5, 106). PulseNet International is a global laboratory network dedicated to bacterial foodborne disease subtype-based surveillance (11). It involved European and other 85 countries in the world, and its aims are the implementation of standards for molecular typing and the sharing information of laboratory surveillance data for the detection and correlation of foodborne outbreaks even if geographically separated.

1.5 Enhanced surveillance: from conventional molecular typing to Whole Genome Sequencing

Molecular typing of FBPs supply the traditional epidemiological surveillance by providing relationship between strains at suitable discriminatory power to support the rapid and early detection of widespread international outbreaks, to detect and investigate transmission pathways and the relatedness of strains. Moreover, tracing the source of an outbreak and identifying new risk factors are also enabled by molecular typing.
Molecular typing methods have been widely and successfully used for FBPs investigation in EU surveillance and outbreak investigation (107), and recently Whole Genome Sequencing (WGS)-based analysis is replacing traditional methodologies becoming a standard approach in national and multinational food-borne outbreaks and epidemiological investigations (108–113). Therefore, the integration of WGS into real-time food safety management and the collection of WGS data at European level have become a central aspect in the European strategy to control foodborne pathogens (114, 115). From the recently published ECDC "Expert Opinions", standards that fulfil the criteria for integration into EU level surveillance networks and systems are going to be established in the next future, enabling the EU-wide use of WGS as the method of choice for typing of microbial pathogens (114, 116).
WGS was revealed to be one of the most promising one-serve all approach. Besides typing, the sequencing data on the whole genome are useful to explore genotypes and predict phenotypes, including antibiotic resistance, virulence, pathogenicity and infectivity. One of the greater advantages of the WGS based approaches is that thanks to their digital nature, WGS data are sharable and storable in dedicated databases (117).

1.5.1 Conventional molecular typing

Several molecular techniques with high phylogenetic resolution have been developed and extensively used for typing foodborne pathogens, differentiating closely related strains and identifying the sources of contamination (118). Until the recent advent of Whole Genome Sequencing, the most common subtyping techniques for outbreak investigations and for source attribution, population structure, epidemiological and evolutionary studies have been phenotypic and genotypic-based

methods. In outbreak investigations and in the development of strategies to prevent further FWDs spread, subtyping schemes with high discriminatory power are required to distinguish all epidemiologically unrelated or sporadic isolates and to discriminate closely related isolates to eventually reveal strain transmission pathways. At the same time, others performance criteria such as reproducibility along with convenience (cost and availability of reagents and equipment; versatility, time and complexity of execution and interpretation of results) are considered (119, 120). Moreover, typing methods validation has to be assessed by determining if highly similar isolates are gathered accordingly to the observed techniques (121). In order to implement a typing method for routine-surveillance in international networks for an efficient control of FWDs, adequate stability over time and full access to results via open source web-based database are essential. In particular, molecular typing methods provided such a level of intra- and inter-laboratory reproducibility that several databases were developed (122, 123). Internationally standardised nomenclature and protocols on different bacterial species are crucial prerequisite for a typing method to be used in surveillance systems in order to allow comparable results across laboratories (124, 125). A summary of traditional typing methodology and corresponding international databases is available in Table 3. Moreover, in Table 4 the most commonly applied methods with technical details and relevant strengths and limitations have been summarized (107, 118, 119, 124–127). Among these methods, Pulse-field Gel Electrophoresis (PFGE) is widely used for surveillance and outbreak investigations of *Salmonella*, VTEC and *Listeria monocytogenes*. PFGE has been considered the "gold standard" typing technique and global protocols have been developed and standardised by the United States (US) Centre for Disease Control and Prevention (CDC) (104, 128). PFGE has been successfully and widely used and has proven to be very accurate and reproducible in epidemiological investigations for almost all bacterial species and in backtracking of foodborne outbreak becoming the only generic method for typing *Salmonella* serovars (using *XbaI* enzyme) and *Listeria* serotype (103, 122). PulseNet-standardized PFGE protocol with *Apa*I/*Asc*II restriction enzymes for *Listeria monocytogenes* typing were developed by CDC and Association of Public Health Laboratories and also used by ECDC as gold standard procedure to track *L. monocytogenes* isolates from food processing facilities, foods, and clinical samples (129, 130).

Table 3. Molecular typing databases (adapted from Carriço et al., 2013) (123).

Typing method	Database	URL
MLST	MLST.net	http://www.mlst.net
	Pubmlst.org	http://www.pubmlst.org
	Institut Pasteur MLST	http://www.pasteur.fr/mlst/
MLVA	Pasteur MLVA:MLVA-NET	http://www.pasteur.fr/mlva
	MLVA.net	http://www.mlva.net
PFGE	PulseNet International	http://www.pulsenetinternational.org
	PulseNet USA	www.cdc.gov/pulsenet/
	Ridom database	http//:spaserver.ridom.de
Ribotyping	Dipnet	http://www.dipnet.org/ribo.public.php

Furthermore, PFGE protocols based on more than one restriction enzyme have been studied as described by Saidijam and colleagues (2003) that used a two-enzymes PFGE (*Sma*I and *Sal*I) to differentiate sporadic-case strains of *C. jejuni* (68). Unfortunately, more than one epidemiological study demonstrated that PFGE alone could not prove an adequate epidemiologic connection for the

identification of the source of an outbreak since PFGE sometimes discriminate isolates of the same outbreak and characterize with an identical pulsotype two samples not epidemiologically linked (131, 132). This suggests that PFGE results must be combined with conventional epidemiologic data in order to establish the right epidemiology (131,132)(131).

Table 4. Classification level, technical details, reproducibility and discriminatory power and relevant strengths and limitations of most commonly used FBPs typing methods.

Typing method	Reproducibility	Discrimination power	Advantages & Disadvantages
Serotyping	Good reproducibility	Poor discriminatory power for many pathogens when compared to other typing methods but good for Salmonella enterica serotypes	Most important developed phenotypic method although the possible high cost of antisera. Time consuming and usually restricted to only a few reference laboratories
Multi Locus Sequence Typing (MLST)	Highly reproducible	Even though MLST became the gold standard for long-term epidemiological surveillance of several species, PFGE remains important for outbreak detection because it often has higher discriminatory power	Internationally standardised nomenclature producing unambiguous and portable results but the selected genes and their number need to be suitable to discriminate among isolates with more recent genetic mutation. MLST is expensive and will require skilled researcher to perform
Multi Locus Variable Number Tandem Repeat (VNRT) Analysis (MLVA)	Highly reproducible	Insufficient for routine use in long-term surveillance although is a very popular tool for epidemiological and evolutionary studies of specific pathogenic serovars (*S.* Typhimurium and Enteritidis) and show higher discriminatory power than most other typing methods	Possible automation and high resolution. The results can be expressed in digital format facilitating inter- and intra-laboratory comparison but not universal method, time consuming and variable cost based on the loci array
Pulsed Field Gel Electrophoresis (PFGE)	Very good reproducibility	Excellent discriminatory power to become the 'Gold standard' typing technique	High epidemiological concordance, excellent typeability and intra-laboratory reproducibility but technically demanding and labour-intensive, time-consuming and relatively expensive
Ribotyping	Good reproducibility	Relatively limited discriminatory power and resolution, not as high as that of PFGE	Automation, 100% typeability and straightforward and quick protocol, enables analysis without prior knowledge of genomic DNA sequence. Resulting profiles are comparable among laboratories and can be used to build databases but requires 3-4 days to complete the analysis

For *Salmonella*, Multi Locus Variable Number Tandem Repeat (VNRT) Analysis (MLVA) is serotype specific, and has a higher discrimination power compared with PFGE for *S.* Typhimurium and *S.* Enteritidis is widely used for surveillance and outbreak investigations (107, 133).. In 2006 Hyytia-Trees and colleagues demonstrated promising epidemiological concordance of MLVA for subtyping *E. coli* O157:H7 (134). Their results showed that MLVA correctly clustered isolates belonging to eight well-characterized outbreaks (134). In addition, MLVA showed high concordance with PFGE (68). An MLVA scheme for subtyping *L. monocytogenes* was described and evaluated by Murphy and colleagues as able to discriminate strains of the same serotype and to be well-correlated with PFGE data (135). Recently, MLVA protocol based on 18 VNTRs was evaluated by Chenal-Francisque and colleagues (63) and considered useful for *L. monocytogenes* characterization and epidemiological investigations as well as for listeriosis surveillance (136, 137). Ribotyping was firstly used by Wiedmann and colleagues for lineage-specific differentiation of *L. monocytogenes* isolates

in 1997 and to also assess their pathogenic potential (61). Later, ribotyping has also been used successfully for *L. monocytogenes* typing from different sources (138–140). Ribotyping have been also applied for *Salmonella* identification and characterization (141, 142), serotype prediction (143) and prevalence study (144).

Recently, clustered regularly interspaced short palindromic repeats (CRISPR) typing have been applied for the analysis of a range of *Salmonella* serovars including *S.* Enteritidis and *S.* Typhimurium (145–147) and for typing *Listeria monocytogenes* strains in evolutionary studies (148).

DNA-based typing has been largely applied for *L. monocytogenes* and *Salmonella* genotyping starting from molecular serotyping through multiplex PCR (polymerase chain reaction) for serotype prediction (12, 62, 149). Multi Locus Sequence Typing (MLST) have been used as gold standard typing methodology for long-term epidemiological surveillance for both Salmonella and Listeria. Achtman and colleagues in 2012 proposed an MLST-based, 'e-burst' method for serovar prediction (Fig. 6), although limitations of this method could arise in serovars that shared a common ancestor resulting in identical MLST types for different serovars and in polyphyletic serovars (150). In a study based on MLST seven housekeeping genes (*acbZ, bglA, cat, dapE, dat, ldh, lhkA*) adapted from the MLST system proposed by Salcedo and colleagues, 360 *L. monocytogenes* representative isolates, from the collections of the French National Reference Centre for *Listeria* and the WHO Collaborative Centre for foodborne listeriosis, were grouped into seven well-demarcated CC within the different *L. monocytogenes* lineages (see above), showed by means of the MST in Figure 7 (adapted from Ragon et al., 2008 (151)).

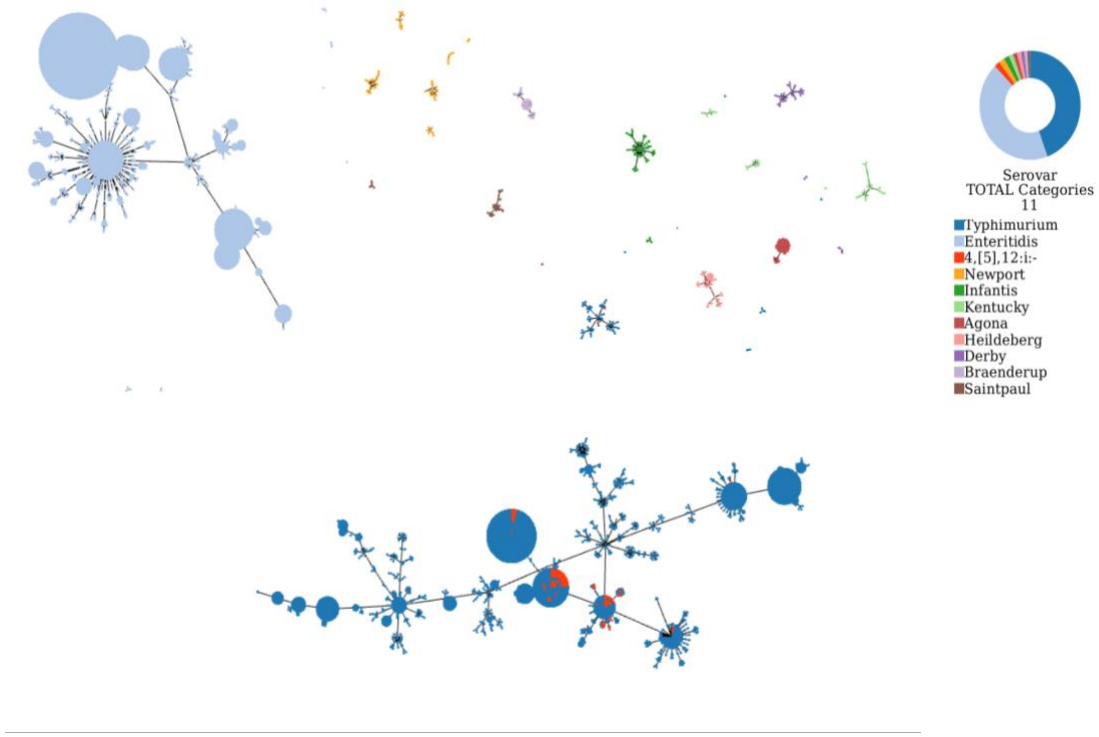

Figure 6. Minimum spanning tree (MST) build on PHYLOViZ Online from MLST data of 3,536 publicly available genomes belonging to 11 most commonly reported serovars of *S. enterica* subsp. *enterica*, downloaded from Enterobase (data from this thesis).

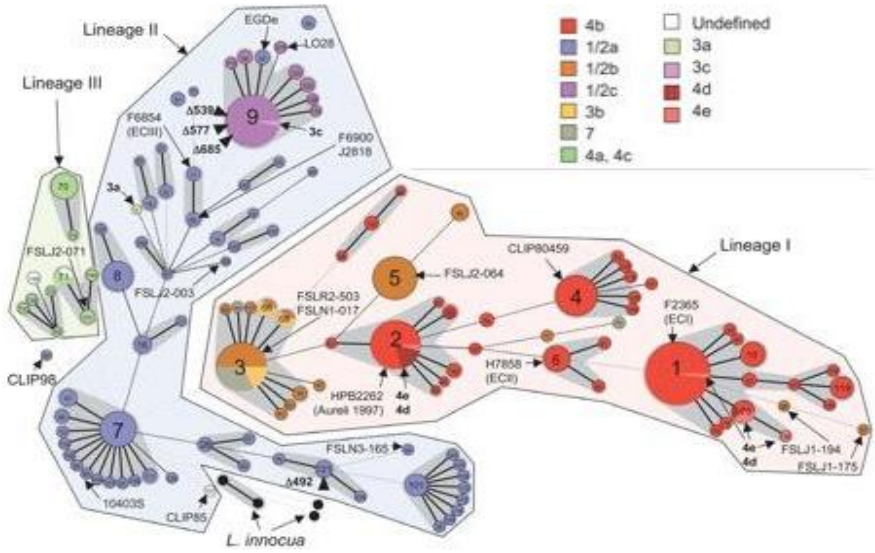

Figure 7. MST analysis of 360 *L. monocytogenes* and four *L. innocua* strains based on MLST data (adapted from Ragon et al., 2008) (151).

1.5.2 Whole Genome Sequencing applied to foodborne pathogens

More than ten years ago, the release of a pyrosequencing-based next generation sequencer (Roche's 454 Genome Sequencer FLX) marked the beginning of high throughput sequencing generating reads of 100 bp up to 500 bp in the successive version (152). In 2006, sequencing-by-synthesis-based Illumina Genome Analyzer was able to produce a Gigabases of sequence data. In the years following the increase in throughput and the drop of costs allowed large bacterial sample collection with 96 samples to be sequenced simultaneously in each lane using the Illumina HiSeq platform. Nowadays, the most popular platforms in microbiology for fast and compact bench-top machines are reported in Table 5. Moreover, several authors have described the application of WGS for genomic and metagenomics studies of FBPs from food processing environments and retail settings as well as across the whole food chain (110, 153, 154).

In recent years, several studies demonstrated that WGS is applicable to all organisms and shows a great potential to perform several *in silico* analysis from a single assay, including rapid characterization, functional annotation and prediction of phenotypic characteristics (antimicrobial resistance) and genome-wide association studies, with an unprecedented level of precision (106, 155–157). Additionally, based on WGS data nowadays public health microbiologist has the right tool to precisely interpret the transmission pathways of infectious diseases, to identify the critical points of contamination along the food chain (from environmental, animal or food sources) and to characterize specific features related to persistent of certain bacterial strains in food and food processing plants. Indeed, compared to conventional typing techniques, WGS offers cost-competitive and labour-efficient performances with similar turnaround times for FBPs typing (116). WGS-based datasets containing all relevant genetic information (genotype and other characteristics like antibiotic or chemical compounds resistance profiles and serovars) on the observed isolates along with related epidemiological data, are stepping stone for an integrated surveillance optimization, not least to enable the establishment of a threshold value for the genetic variation level among epidemiologically related isolates (101).

In addition, the achievement, handling and management of WGS-based data along with efficient bioinformatics analytical workflows standardisation and a comprehensive genomic sequence database are the upcoming objectives to make as effective for an integrated surveillance (114).

Table 5. Next Generation Sequencing platforms and relative year, amplification and sequencing technology and references.

Technology	Year	Amplification method & Sequencing chemistry	Advantages & Disadvantages	References
Illumina™ (Illumina)	2006	Bridge amplification on solid surface and reversible dye termination Sequencing by Synthesis	Highest throughput and best cost-effectiveness but long run time, short read length and substitution errors	(158)
SOLiD (Life Technologies)	2006	Emulsion PCR on beads and Sequencing by nucleotide ligation and detection	Highest throughput but long run and time, very short read length	(159)
Pacific Biosciences	2010	Amplification-free and monitoring of individual DNA polymerase molecules n zero-mode waveguide detectors.	Very long read length and fast run time but high cost and error rate with low throughput	(160)
Ion Torrent (Life Technologies)	2011	Emulsion PCR on beads and Sequencing by Synthesis with detection of H+ ions on silicon chip	Fast run time but errors in homopolymer repeats	(161)
Oxford nanopore tecnology	2014	Discriminate individual nucleotides by measuring the change in electrical conductivity as DNA molecules pass through the pore	Low capital cost, portable DNA sequencer, and can produce data in real-time	(162)

WGS for outbreak investigation and surveillance purposes

WGS has emerged as a powerful tool for outbreak detection and investigation and surveillance studies especially for monomorphic (clonal) microorganisms such as *Salmonella* (4) and *Listeria monocytogenes* (163). As reviewed by Moran-Gilad the routine use of WGS technology is growing among public health microbiology laboratories and governmental agencies along with harmonized, standardized and transferable bioinformatics tools and approaches for analysing WGS data (164). By 2015, ECDC reported that WGS technology has been accessible to public health laboratories in the 68% of the EU/ EEA countries (114).

Nowadays, the harmonisation and standardization of bioinformatics analytical approaches for WGS-based analysis of FBPs across public health and food safety sectors are the main objectives within EU surveillance. In particular, the analytical strategies that can be adopted for comparative genome analysis and genome evolution studies are:

- Phylogenetic analysis to measure of the evolutionary distance among genomes with the determination of the most recent common ancestor (MRCA). The resulting tree or network graph shows the linkage among isolates from different sources (human, food, animal, environment).
- Phenotype prediction of biological properties with clinical and epidemiological relevance such as antimicrobial resistance profile, stress adaptation and virulence profile as well as the genetic determinants included in mobile genetic elements (plasmids, phages, transposons). The assignment of the biological function is provided looking for gene homologies in a reference database.
- Nomenclature assignment through an openly accessible database which return the allele identifiers of a sequence for each FBPs, indispensable for communication of results to public health and food safety professionals in a view of epidemiological surveillance and outbreak rapid response (116, 165).

The high resolution of the outputs of WGS-based analytical strategies results in highly specific and sensitive case definition increasing the detection of the source of infection also in retrospective FWDs outbreak investigation (127, 166). Moreover, an enhanced ability to detect temporal and spatial clusters of genetically related FBPs may enable the detection of a higher number of outbreaks (167–169), the detection of the emergence and monitoring of the evolution/dynamics of MDR pathogen spread (44, 54, 170) hugely impacting on surveillance-specific epidemiological investigations (171). Along with these analytical strategies, a crucial step for meaningful epidemiological surveillance is the integration and combination of WGS data with accessible and sharable contextual (sensitive, clinical and epidemiological) metadata, using a common source attribution for contextual data to be deposited in a publicly available database in accordance with the EU legislation (114,116).

WGS for source attribution

Providing information on the genetic features of strains and their distribution to identify the primary source (animal reservoirs and vehicles, e.g. food) of human infection caused by foodborne pathogen is crucial to control and reduce the burden of human diseases (101). The growing number of genome-based methodological approaches and data analysis comparison tools is significantly contributing to outcome of novel approach for source attribution studies (171, 172). Indeed, NGS technology applied on FBPs demonstrated an unprecedented accuracy in studies for attribution of contamination sources as was the case for the EHEC outbreak in Germany in 2011 (173). Recently, WGS was also applied to compare *Escherichia coli* O157 strains from cattle and sheep isolated in Scotland to demonstrates that geographic variation dominates host association (174). Moreover, WGS-based investigation on recent multi-country outbreak of *Salmonella* Enteritidis successfully provided evidence of linkage between the outbreak and eggs sources produced and exported from Poland (175).

Besides the superior resolution of WGS data can improve the accuracy in the attribution of specific ST to specific reservoirs, there is often a considerable uncertainty in source attribution concerning

particular pathogen serotypes or clonal complexes. Therefore, the analytical methods used in source attribution studies should be discriminatory enough to correlate human isolates with their sources of contamination but, on the other hand, they should not be too discriminatory to avoid the risk that correlated isolates might be distinguished as unrelated (172, 176). Moreover, the optimal level of discrimination is not only influenced by the level of clonality but also by the degree of host association of the observed pathogen (101). Due to the source specificity of certain pathogen subtypes and assuming that the transmission pathways start from sources and end to humans, the relative contribution of each source to human cases can be inferred probabilistically by comparing the human and source subtype distributions (172). Therefore, combining the analysis of epidemiological (case–control) and molecular based source attribution data has also been shown to bridge the gap between attributing human cases at the point of exposure and at the top of the transmission chain (177). Moreover, an integrated surveillance through the collection of isolates extended to all major sources to which human population is exposed to may facilitate source attribution analysis (101).

1.5.2.1 Example of application of WGS on *Listeria monocytogenes* and *Salmonella enterica*

Listeria monocytogenes

In 2013, to evaluate the applicability of WGS for enhanced surveillance of *L. monocytogenes* in United States, isolates from different sources (clinical, food and environment) were sequenced within a nationwide *Listeria* Whole Genome Sequencing Project (4). From this wide project emerged that WGS enable the identification of more clusters than with PFGE, decreasing the cluster size median (167) and effectively attributing the diseases to their sources. These results suggest that outbreak detection may be faster. This was the case described by Angelo and colleagues that confirmed *Listeria monocytogenes* strains from caramel apples originating from a single supplier. Since *L. monocytogenes* isolates have never been associated with this kind of food product without WGS the outbreak source might not have been conclusively identified (178). Other outbreak associated with consumption of packaged salad (179) and stone fruits have been successfully untangled using WGS technology leading to the identification of the source of contamination and the removal of contaminated foods from circulation (180). Genomic comparison using WGS data has also been useful in establishing potential links between food and environmental isolates of *L. monocytogenes* and the limited numbers of epidemiologically linked outbreak strains indicating the genomic relationship between strains (181). In a retrospective study on persistence of *L. monocytogenes* contaminating different delis, WGS approaches were successfully applied revealing that *L. monocytogenes* strains isolated in different geographic areas and in different time had identical or nearly identical genome, improving the subtyping and identification of this persistent pathogens in food associated environments (83, 110, 167, 181–183).

Salmonella enterica

Opposite to traditional typing techniques, subtyping methods based on WGS data have been shown to successfully delineate clonal *S. enterica* serovars. Although the first sequencing platform was developed in 2005, the big potential of WGS technology was shown in 2012 during a *Salmonella* outbreak. A retrospective study was conducted on *Salmonella* isolates which belonged to the same pulsotype of the ongoing outbreak. A collection of clinical, food and historical *Salmonella* were sequenced on the Illumina MiSeq and further analysed distinguishing and correlating all the isolates with high resolution (184). Moreover, in a retrospective and prospective analysis on outbreak-associated *S.* Enteritidis from Connecticut and New York in 2010, 93 isolates previously characterized through PFGE were whole genome sequenced and further analysed demonstrating that whole genome cluster analysis enhance outbreak clusters detection and resolution improving surveillance and tracking of this pathogen (185). Later in 2014 several cases of *S.* Enteritidis phage

type (PT) 4 infection from two geographically separated areas in Belgium were investigated with MLVA. Results did not sufficiently discriminate or assign a common origin of contamination for the two outbreaks. On the contrary, WGS confirmed and/or discriminated food and human isolates between and within outbreaks, separating isolates into two different clades and confirming that two overlapping outbreaks were taking place (186). In a different study, 52 *S.* Enteritidis isolates analysed by MLVA, CRISPR, and PFGE in tandem comparison with WGS. The latter clustered isolates into 16 outbreaks whereas the next most accurate technique, MLVA, correctly grouped only 6 of the 16 outbreaks (187).
Also concerning *S.* Typhimurium, a study conducted in Denmark compared 18 isolates from 6 salmonellosis outbreaks to 16 unrelated strains retrospectively identifying outbreak cases with 100% of accuracy (188). In addition, WGS based approaches were also evaluated for the ability to gather outbreak cases from unrelated isolates within the same PT revealing that highly clonal isolates may be clustered in well-defined different outbreak clades (189).

1.6 EU initiatives to implement genomic-based typing for enhancing surveillance of food-borne pathogens

Several international initiatives and networks in the area of food safety have been underpinned to harmonize and standardize FBPs detection and subtyping. These have been aimed not only for surveillance and outbreak investigations but also for investigating population structure and transmission patterns, source attribution and antimicrobial resistance studies. In particular in 2011 the Global Microbial Identifier (GMI) project, a common platform including a global surveillance of microbial pathogens and infectious diseases identification and diagnostic, was developed with the aim to aggregate, share, mine and use microbial genomic data for public health and clinical purposes. Since 2014, in collaboration with members of GMI initiative, the European COMPARE (COllaborative Management Platform for detection and Analyses of (Re-) emerging and foodborne outbreaks in Europe) project began. The main objectives of the multidisciplinary research network gathered in COMPARE are:

- enabling analytical framework and globally linked data and information sharing platform system for the rapid identification, containment and mitigation of emerging infectious diseases and foodborne outbreaks
- integrate state-of-the-art strategies, tools, technologies and methods for collecting, processing and analyzing sequence-based pathogen data in combination with associated (clinical, epidemiological and other) data, for the generation of actionable information to relevant authorities and other users in the human health, animal health and food safety domains.

Another large European project co-founded by EFSA named INNUENDO was launched to develop a cross-sectorial platform for the integration of genomics surveillance of foodborne pathogens in the context of small countries with limited resources. Not far from the COMPARE project, the main objective of INNUENDO is to establish a “one size fits all” framework to ensure the accessibility to all the WGS-based strategic approaches for surveillance of foodborne diseases also to all the players and end-users in the field of public routine health surveillance and epidemiological and outbreak investigations.
Several other organizations, projects and consortia actively working to overwhelm WGS technology obstacles as well as to establish robust bioinformatics pipelines and develop pilot studies based on WGS typing of foodborne pathogens (116). In particular ECDC have planned WGS-based actions for surveillance of *Listeria monocytogenes* to be conducted during 2017–19. In addition, the WGS cost-efficiency (based on a moderate number of clinical cases) and the typing feasibility of this pathogen thanks to the validated typing and nomenclature schemes along with the progress of MS capacity for WGS application underpin the use of this method to complement and replace the gold standard typing PFGE. A consensus methodology for comparable WGS data production, exchange,

analysis and EU reporting of *L. monocytogenes* is being launched along with an international WGS-based nomenclature for this and others food-borne pathogens (106).
Concerning *Salmonella enterica*, EU-wide WGS-based surveillance is still challenging and standard protocols and WGS-derived nomenclatures are not yet validated at international level. Despite the high priority and health impact of salmonellosis infections in the EU/ EEA, the high volume of case reports and samples to be typed in order to fulfil the main surveillance goal of rapid outbreak detection make the implementation process more complex. Nevertheless, has been envisaged that, within the next years, once WGS approaches for *S. enterica* will be well established it definitely becomes the all-embracing technology for identification and characterisation of this food-borne pathogen across the EU/ EEA countries. In the meanwhile, ECDC will assist MS by supporting ad hoc comparative WGS-based typing for the investigation and joint assessment with EFSA of multistate outbreaks of salmonellosis in Europe (114), as was the case of *S.* Enteritidis contaminating Polish eggs (190).

2. Bioinformatics for comparative genomics

The comparative analysis of multiple genomes of the same species makes possible the high throughput investigation of the micro-evolution and genetic variations of strains as well as the identification of specific genetic markers determining the pan-genome of related microorganisms and exploring the nucleotide substitutions between genomes (165). This approach not only allows to reveal genomic rearrangements, such as inversions or translocations, but also to explore the conserved and shared genes as well as the more variable or accessory genes. Based on genomic data, the evolutionary relationships between pathogenic strains and a number of epidemiological relevant genotypic and phenotypic features (e.g. virulence, stress tolerance, environmental adaptation) can be inferred. Furthermore, the presence and location of particular genomic islands (GIs) and of external genetic elements such as prophage clusters and transposable element may be detected by aligning two or more sequences (4). Nevertheless, the massive rate of genomic data generated through WGS poses significant challenges for comparative genomics, such as speed and complexity of analysis, data quality assessments, and results management, visualization and interpretation.
Currently, multiple bioinformatics approaches (Fig.8) have been developed to overcome these hurdles mainly relying on reference based read mapping, de novo assembly, core and whole genome single nucleotide variant (also called single nucleotide polymorphism - SNP) and the "gene by gene" approach, based on the extension of the original seven genes based MLST concept to the core/whole-genome level (cg/ wgMLST).

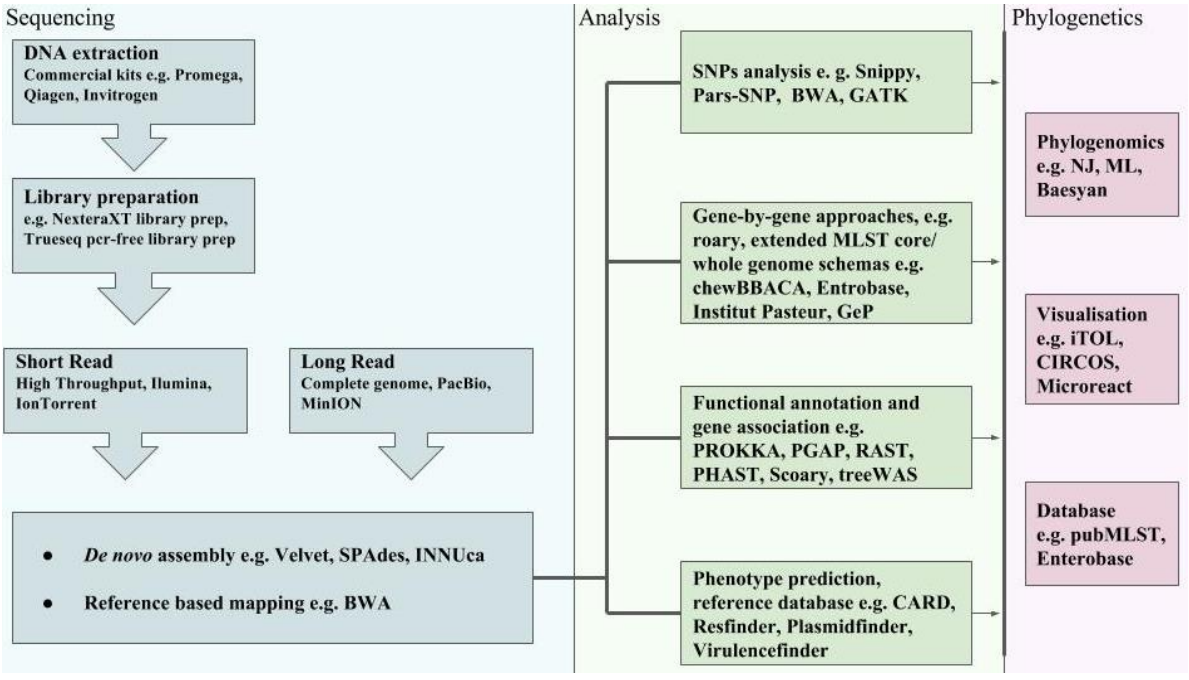

Figure 8. Illustration of genomic approaches for foodborne pathogens outbreak investigation and surveillance purposes.

2.1 *De Novo* assembly

As descripted above, the most spread methodology to generate WGS data is the shotgun sequencing through short read sequencers able to produce many copies of randomly fragmented DNA segments (191). These fragments are termed *reads* and are commonly stored in FASTQ files in association with a *Phred* quality score for the predicted nucleotide. Commonly, these files are inspected for assessing quality of sequence reads and subject to pre-processing step such as trimming sequencer adapters and low-quality reads (192). Pre-processed data are used for the reconstruction of genomes: starting from reads and ending with an assembled draft sequence (193). If this process is performed without a reference genome is referred as *de novo* assembly. The bioinformatics process behind the *de novo* assembly analysis is aimed to reconstruct long contiguous consensus sequences, so-called contigs, through computationally efficient algorithms that overlap reads (194). Additionally, reads are mapped back to the contigs generating longer sequences called scaffolds so that the sequence is polished resulting in high quality draft genomes. Typically, the assembler output is a FASTA file containing the contigs nucleotide sequences. This allows an easier management and storage of the sequence data in dedicated database thanks to reduced file size and standard format that enhance transferability for further comparative genomic analysis (195). Nevertheless, a crucial point in draft genomes generation is the quality of the assembled sequence. Given the relevance of this key bioinformatics process several software are available for performing de Bruijin graph-based *de novo* assembly, such as Velvet (194) or SPAdes (193) and *de novo* quality assembly evaluation such as QUAST (196). Quality check of assemblies are based on several parameters such as the number of contigs, the size of the assembled genome, the contigs lengths and the median contig size (N50). Coverage is also a central aspect for evaluation the quality of final assembly. Coverage refers to the time each nucleotide position in the assembled genome is covered by reads. , not many pipelines have been developed to perform *de novo* assembly based on a consistent quality assessment and quality check (QC/QA). One of these is INNUca (https://github.com/INNUENDOCON/INNUca), an automatic bioinformatics pipeline specialised on systematic QA/QC respecting well-defined criteria and calculating coverage in different analytical steps in order to control the quality of reads, assess the quality of *de novo* assembled contigs and searching for possible contamination.
High quality *de novo* assemblies may be used as reference sequences for SNP variant calling (see below) in comparative phylogenomics approaches especially in intra-cluster comparisons of strains within outbreaks (197) or monophyletic strains (198).

2.2 Reference based read mapping and variant calling

The generation of genome sequences along with the detection of genetic variants can be performed mapping pre-processed reads against a high-quality reference genome, if available (199–201). This is one of the most common strategies applied in SNPs analysis based pipelines such as snippy, a variant calling pipeline by Torsten Seemann (https://github.com/tseemann/snippy). The process behind read-mapping against a reference genome relies on algorithms that index the reference genome, align sequence reads to the reference and index the differences between them assigning confidence levels to each variant position (4, 165). Among the several read mappers software, the most popular in FBPs analysis are Burrows-Wheeler Aligner (BWA) (199) and Bowtie2 (202), both based on Burrows-Wheeler transform based algorithm. These tools generate a Sequence Alignment Map (SAM) file so that the aligned reads can be processed for SNPS detection. A list of the identified polymorphisms is produced applying variant caller algorithms such as Varscan (203), Freebayes (200) and Haplotype Caller as part of Genome Analysis Toolkit, GATK (204, 205).
Ad hoc pipelines for variant calling focused on SNPs analysis have been successfully applied in several outbreak detection and investigations studies and in food-borne pathogens surveillance projects and networks (10, 110, 206–208). Nevertheless, one of the relevant drawback of reference mapping based approach is the selection of the most appropriate reference genome to be queried.

Characteristics like gaps or sequence errors, the genome quality and the relatedness are essential in the reference choice in order to avoid misalignments or variants misidentification. If poorly related genomes are chosen as reference, distant and divergent genomic regions won't likely be considered from the SNPs analysis so that likely variant sites will remain undetected. One of the strategies to overcome this limitation is to perform SNPs analysis selecting a high quality *de novo* assembled draft genome as reference genetically closely related to the observed dataset and filtering that genomic regions responsible of ambiguous read mapping (209).

2.3 Structural and functional genome annotation

Genomes annotation refers to the identification of sequences features and the subsequent association to the biological role (157, 210). Annotation is one the first steps applied for the analyses of draft genome assemblies in order to identify the genetic regions that code for proteins (coding DNA sequences, CDS) as well as non-protein-coding, functional RNA molecules (transfer, tRNA, and ribosomal, rRNA), regulatory genetic regions and "foreign" DNA such as that originated from plasmids or bacteriophages (211, 212). Occasionally, higher order features such as operons, CRISPR elements and specific genomic islands can be identified during the annotation phase (145,147). This process is usually performed through automatic pipelines that use multiple external feature prediction algorithms and often followed by manual curation (212). One of the most widely used algorithm for sequence similarity search against a database of known protein and functional characterization of gene sequences is the basic local alignment search tool (BLAST). Different type of analysis can be performed with this tool including comparing nucleotide query sequences to a nucleotide reference database (BLASTn), comparing protein query sequences to a protein reference database (BLASTp), comparing translated nucleotide query sequences to the protein database (BLASTx), comparing protein query sequences to the translated nucleotide database (tBLASTn) and comparing translated nucleotide query sequences to a translated nucleotide database (tBLASTx) (213). The quality of pairwise sequence alignment performed with BLAST is assessed by a score, which is positive for good matches and negative for mismatches or gapped positions. Among the several freely available online web services, the NCBI prokaryotic genome annotation pipeline (PGAP) (210) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/) is one of the well-known for genome annotation along with the RAST Server (214). The drawback of these tools is that they need from 12-24 hours to few days to get the genome assembly annotated. This feature makes it unsuitable to work with hundreds or thousands of genomes. Nevertheless, a locally performing faster option is to use PROKKA (157), an accurate pipeline that fully annotate a draft genome in few minutes. The output of the annotation process consists in 10 files which include Genbank file (GBK) and GFF v3 file, containing sequences and annotations. These files can be visualised with freely available software such as the Artemis genome viewer in order to explore the genomic features of interest and the Artemis Comparison Tool (ACT) in order to identify and analyse regions of similarity and difference between genomes and to explore conservation of synteny, in the context of the entire sequences and their annotation (215).

2.4 "Gene-by-gene" approaches

Along with SNPs analysis, one of the mayor strategies in WGS-based FBPs subtyping using assembled genomes or reads as input is the "gene-by-gene" analysis (165, 216). This approach reveals the full gene content of a given dataset of bacterial genetic sequences referred as pan-genome. The pan-genome comprises a set of core genes, shared among nearly all the members of the same species, and a set of accessory genes, present only in a variable percentage of strains or specific to a single strain (217, 218). Therefore, several curated MLST scheme may be adopted to explore strain

relationships based on the amount of and diversity in gene content (allelic variation) and on the required resolution level:

- ribosomal MLST (rMLST), targeting 53 ribosomal protein subunit genes that allow for bacterial typing at taxonomic level;
- core genome MLST (cgMLST), targeting genes shared by all members of the observed species;
- whole genome MLST (wgMLST), enabling the indexing of allelic variation in both core and accessory genes.

Approaches based on whole genome MLST scheme allow the identification of the gain/loss process characterizing the members of a bacterial species and providing discriminatory power levels comparable to SNPs analyses especially in phylogenetic studies, outbreak investigations and comparative analyses for routine-surveillance of clonal strains such as *Listeria monocytogenes*, *Salmonella enterica* (106, 165). At the time of writing, two publicly online databases are available, Pasteur Institute database (http://bigsdb.pasteur.fr/listeria/) and EnteroBase (https://enterobase.warwick.ac.uk/species/index/senterica), hosting validated cgMLST schemes for *Listeria monocytogenes* and cg/ wgMLST schemes for *Salmonella enterica*, respectively.
Nevertheless, designing *ad hoc* cg/ wgMLST schemes and performing the analyses locally in a closed system it's also possible through pipelines such as Genome Profiler (GeP) (219) or the recent compressively workflow for schema creation, validation and allele call chewBBACA (220) (https://github.com/B-UMMI/chewBBACA/). ChewBBACA used an allele calling algorithm based on Blast Score Ratio (BSR) and include functions for validating allele variation in the loci (220). In addition, a recent study conducted by McNally and colleagues highlights that core and whole genome gene-by-gene analysis in combination with regulatory gene regions (non-coding DNA sequences) analysis across genomes may reveal with high resolution population evolutionary details of particular bacterial pathogenic lineages (221).

2.5 Pangenome analysis and Genome Wide Association Studies

An open source tool that rapidly build large-scale pan-genomes, without requiring a pre-defined schema and identifying core and accessory genes, is Roary (222). It can analyse datasets with thousands of samples in few hours, comparing the pan-genome of related and annotated draft genomes. Moreover, a novel pipeline emulating Roary but relying on intergenic regions (IGRs) variation analysis is Piggy (223) (https://github.com/harry-thorpe/piggy). In particular, for *Staphylococcus aureus* differences in gene expression have been associated with highly divergent ('switched') IGRs and a multi-locus reference database of IGR alleles (igMLST) has been developed and implemented in BIGSdb (224).
The specific genomic content identified by gene presence/absence analysis in the draft genomes performed by Roary and Piggy can be further analysed with different systems in order to perform Genome Wide Association Studies (GWAS). One easy-to-use and ultra-fast software tool with wide applicability for pan-GWAS is Scoary (155) (https://github.com/AdmiralenOla/Scoary), designed to take the gene presence/ absence file from Roary as well as a traits file created by the user. All the accessory components of the pan-genome are therefore scored and associated with specific genes or different phenotypic traits, accounting for population stratification with minimal assumptions about evolutionary processes (155). Another efficient system for rapidly locating differentially abundant genomic regions based on probabilistic models is Neptune (225). This tool is aimed to identify unique and conserved sequences that are sufficiently common to a group of target sequences (inclusion group) and sufficiently absent from non-targets (exclusion group), returning a genomic signature without requiring genome annotation information. Signatures may reside in genic or intergenic regions and may correspond to genomic islands, phage regions or entire operons. As in Scoary-based GWAS, Neptune might correlate genomic features with phenotypic traits using probabilistic models

instead of heuristic strategies and uniquely delineate groups of organisms associated with particular disease cluster or event, from unrelated sporadic or environmental microbes (225). Methodologies accounting for GWAS can be useful in the identification of loci that are strongly associated with strain virulence, antibiotic resistance and host or niche adaptation. Nevertheless, in association studies methods accounting for scale genetic differences and phenotypic clustering, large volumes of valuable data can be discarded reducing the association detecting power. To overcome this drawback phylogenetic analyses accounting for clonal population structure and genetic recombination can be inferred and combined to GWAS (226).

2.6 Phylogenetic approaches

Phylogenetic tree-based approaches have been largely applied for the reconstruction of the evolutionary history of bacterial population from WGS data in outbreak detection or surveillance studies. Phylogenetic analyses allow for the identification of genetic evolutionary relationships not only for population clusters, but also for subpopulations and individual relationships taking into account clonality and genetic recombination (227).
Depending on the observed species and the study purposes, the inputs for phylogenetic analyses can be different, from SNPs or pan-genome matrix to core SNPs or core/accessory genes alignment files (165, 228). Several algorithms can be adopted to construct and infer trees, including distance-based algorithms: Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (229) and neighbour-joining (NJ) (230); and character-based algorithms: Maximum parsimony (MP), Maximum likelihood (ML) and Bayesian methods. In particular, the character-based algorithms are the most commonly applied in WGS outbreak investigation studies since attempt to find (i) the tree that requires the fewest evolutionary changes, MP method, (ii) the tree that, given a model of molecular evolution, is most likely to have produced the observed data, ML method, and (iii) the tree or parts of it that has the highest posterior probability given the observed data, Bayesian method (4). Otherwise, distance-based methods infer the tree that best represents the observed distances or the tree whose terminal nodes have the same distance from the root assuming that the rate of mutation is constant over time, NJ and UPGMA (229,230). One of the main advantages of using character-based algorithms as tree methods is the high sensitivity and their ability to model the evolution of a whole group of genes or genomes from multiple sequence alignments at once. Nevertheless, the dataset size, the homologous recombination rate and the accuracy of multiple sequence alignments may influence the inference of phylogenetic trees reducing the number of information usable to build evolution models in comparative genomics analyses (231). Some of the most frequently used software to infer phylogeny on WGS data are MEGA6 (232), FastTree (233), PhyML and IQ-Tree (234) for distance based, MP and ML analyses and BEAST (235) for Bayesian analyses. In addition, phylogenetic methods that detect and account genetic recombination of bacterial genomes in ML tree inference, such as ClonalFrameML (https://github.com/xavierdidelot/ClonalFrameML), have been recently developed (236).

2.7 *In silico* pheno-genotyping using WGS

Dedicated bioinformatics online resources for antimicrobial resistance prediction have been developed, including the ResFinder database (156) (https://cge.cbs.dtu.dk/services/ResFinder/) and the Comprehensive Antimicrobial Resistance Database (CARD) (237). An automatic pipeline for mass screening of contigs for antimicrobial resistance and virulence genes has been recently developed and named abricate (https://github.com/tseemann/abricate). This tool can be locally performed and can rely on seven different databases. Among these are the Virulence Factors database (VFDB) (http://www.mgc.ac.cn/VFs/) that enables the search of virulence-related genes from various

bacterial pathogens in nucleotide sequences. Another similar tool for detecting virulence genes is the VirulenceFinder (156) (https://cge.cbs.dtu.dk/services/VirulenceFinder/), optimized for the examination of phage nucleotide sequences. Not all antimicrobial resistance or virulence-related genes that are identified by molecular methods are consistently expressed, in order to enhance the fitness of the organism (238), but identification of these genes can serve as a trigger for further phenotypic investigations.
Publicly available web tool PlasmidFinder (239) (https://cge.cbs.dtu.dk/services/PlasmidFinder/), PHAST (211) (http://phast.wishartlab.com) and MLST server (www.cbs.dtu.dk/services/MLST) have been also developed to identify plasmids and prophage regions in raw reads, contigs or closed plasmid sequencing data and to identify the sequence types (STs) of bacterial pathogens from WGS data, respectively. Additionally, command line based tools that locally scan contigs files against traditional PubMLST typing schemes determining STs is also available (https://github.com/tseemann/mlst) and implemented in the above mentioned automatic INNUca pipeline.

2.8 Management, analyses and visualisation of WGS data

The ability to easily transfer and share sequence data analyses along with efficiently store and define a common language is essential for identification of bacterial clones, supporting and enhancing surveillance as well as tracking and monitoring of FPBs (102,123). Since a massive amount of DNA sequences is currently generated by WGS technologies and since bioinformatics tools have been being more commonly applied, several software systems and freely available databases that store, manage, analyse and visualize the WGS-based "big data" from FBPs surveillance and outbreak investigations have been developed (102). In particular, databases and software systems allow to identify and investigate about clusters of related bacterial profiles, plasmids, bacteriophages, and other mobile elements spread in geographically and temporally distant isolates from human and non-human source (food, feed, environment and animal) in association with linked metadata. Moreover, epidemiologically relevant metadata including year and source of isolation and geographical localisation in association with molecular typing data of released isolates are essential for focusing investigation on strains that might be part of ongoing and recent outbreaks or of particular clones relevant for epidemiological studies (102). Community-oriented databases include National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/genome/), pubMLST database (http:// pubmlst.org/databases/) and nucleotide-sequence repository, such as the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena). Species-specific databases based on genomic data have been developed including BigsDB (216) or EnteroBase (https://enterobase.warwick.ac.uk/), a genome database including more than 100,000 *Salmonella* isolates of different serovars along with their temporal and geographic metadata. This web service allows to analyse and visualise genetic variation among isolates combining a set of modular, open-source and web-based tools integrating bacterial genomics with epidemiological disease patterns. However, the currently available databases show several drawbacks, including the lack of automated process for data and metadata submission and curation protocols along with the multiplicity of data sharing formats. Therefore, published data collection is still a challenging and laborious manual process and needs of integration with data from different databases (123) that could be analysed by locally available software.
A suite of analysis options including quality metrics for the data, reconstruction of MLST profiles with pre-defined schemes, resistance prediction, gene content variation analyses and phylogenetic inferences are provided by WGSA system (http://WGSA.net) which interactively displays data via trees and shows maps of geographical sources of the isolates. Downloadable systems with easy-to-use interfaces, such as PHYLOViZ 2.0 (240), enables to perform locally multiple data analyses (e.g. goeBURST analysis) of molecular epidemiological data, including SNPs and cg/ wgMLST allelic profiles, and visualize results through Minimum Spanning Trees (MST) or hierarchical

clustering along with related metadata. The extended version is PHYLOViZ Online (241), a web-based platform which allows phylogenetic inference, visualization of trees including large dataset along with metadata and data sharing.

References

1. Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, Praet N, Bellinger DC, Silva NR de, Gargouri N, Speybroeck N, Cawthorne A, Mathers C, Stein C, Angulo FJ, Devleesschauwer B, Group on behalf of WHOFDBER. 2015. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLOS Med 12:e1001923.
2. European Food Safety Authority, European Centre for Disease Prevention and Control. 2016. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2015. EFSA J 14:n/a-n/a.
3. European Food Safety Authority, European Centre for Disease Prevention and Control (ECDC). 2015. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2014. EFSA J 13:n/a-n/a.
4. Deng X, Bakker HC den, Hendriksen RS. 2016. Genomic Epidemiology: Whole-Genome-Sequencing–Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens. Annu Rev Food Sci Technol 7:353–374.
5. Tauxe RV, Doyle MP, Kuchenmüller T, Schlundt J, Stein CE. 2010. Evolving public health approaches to the global challenge of foodborne infections. Int J Food Microbiol 139:S16–S28.
6. Hussain MA, Dawson CO. 2013. Economic Impact of Food Safety Outbreaks on Food Businesses. Foods 2:585–589.
7. Larsen MH, Dalmasso M, Ingmer H, Langsrud S, Malakauskas M, Mader A, Møretrø T, Smole Možina S, Rychli K, Wagner M, John Wallace R, Zentek J, Jordan K. 2014. Persistence of foodborne pathogens and their control in primary and secondary food production chains. Food Control 44:92–109.
8. Grad YH, Lipsitch M. 2014. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. Genome Biol 15.
9. Aung MM, Chang YS. 2014. Traceability in a food supply chain: Safety and quality perspectives. Food Control 39:172–184.
10. Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R. 2016. Practical Value of Food Pathogen Traceability through Building a Whole-Genome Sequencing Network and Database. J Clin Microbiol 54:1975–1983.
11. Swaminathan B, Barrett TJ, Fields P. 2006. Surveillance for human Salmonella infections in the United States. J AOAC Int 89:553–559.
12. McQuiston JR, Parrenas R, Ortiz-Rivera M, Gheesling L, Brenner F, Fields PI. 2004. Sequencing and comparative analysis of flagellin genes fliC, fljB, and flpA from Salmonella. J Clin Microbiol 42:1923–1932.
13. Grimont P, Weill F-X. 2007. Antigenic Formulae of the Salmonella serovars, (9th ed.) Paris: WHO Collaborating Centre for Reference and Research on Salmonella.
14. Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B. 2000. Salmonella Nomenclature. J Clin Microbiol 38:2465–2467.

15. Ryan MP, O'Dwyer J, Adley CC. 2017. Evaluation of the Complex Nomenclature of the Clinically and Veterinary Significant Pathogen Salmonella. BioMed Res Int. Research article.
16. Hohmann EL. 2001. Nontyphoidal salmonellosis. Clin Infect Dis Off Publ Infect Dis Soc Am 32:263–269.
17. Holt KE, Thomson NR, Wain J, Langridge GC, Hasan R, Bhutta ZA, Quail MA, Norbertczak H, Walker D, Simmonds M, White B, Bason N, Mungall K, Dougan G, Parkhill J. 2009. Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. BMC Genomics 10:36.
18. Silva C, Calva E, Maloy S. 2014. One Health and Food-Borne Disease: Salmonella Transmission between Humans, Animals, and Plants. Microbiol Spectr 2:OH-0020-2013.
19. Blaser MJ, Newman LS. 1982. A Review of Human Salmonellosis: I. Infective Dose. Rev Infect Dis 4:1096–1106.
20. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, Hoekstra RM, International Collaboration on Enteric Disease "Burden of Illness" Studies. 2010. The global burden of nontyphoidal Salmonella gastroenteritis. Clin Infect Dis Off Publ Infect Dis Soc Am 50:882–889.
21. Bergey's Manual of Systematics of Archaea and Bacteria - Wiley Online Library.
22. Liu SL, Sanderson KE. 1992. A physical map of the Salmonella typhimurium LT2 genome made by using XbaI analysis. J Bacteriol 174:1662–1672.
23. Liu SL, Hessel A, Sanderson KE. 1993. The XbaI-BlnI-CeuI genomic cleavage map of Salmonella typhimurium LT2 determined by double digestion, end labelling, and pulsed-field gel electrophoresis. J Bacteriol 175:4104–4120.
24. Liu SL, Hessel A, Sanderson KE. 1993. Genomic mapping with I-Ceu I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in Salmonella spp., Escherichia coli, and other bacteria. Proc Natl Acad Sci U S A 90:6874–6878.
25. Lee CA, Silva M, Siber AM, Kelly AJ, Galyov E, McCormick BA. 2000. A secreted Salmonella protein induces a proinflammatory response in epithelial cells, which promotes neutrophil migration. Proc Natl Acad Sci U S A 97:12283–12288.
26. Lopez CA, Winter SE, Rivera-Chávez F, Xavier MN, Poon V, Nuccio S-P, Tsolis RM, Bäumler AJ. 2012. Phage-Mediated Acquisition of a Type III Secreted Effector Protein Boosts Growth of Salmonella by Nitrate Respiration. mBio 3.
27. Zhang S, Santos RL, Tsolis RM, Mirold S, Hardt W-D, Adams LG, Bäumler AJ. 2002. Phage mediated horizontal transfer of the sopE1 gene increases enteropathogenicity of Salmonella enterica serotype Typhimurium for calves. FEMS Microbiol Lett 217:243–247.
28. Guiney DG, Fierer J. 2011. The Role of the spv Genes in Salmonella Pathogenesis. Front Microbiol 2.
29. Chessa D, Dorsey CW, Winter M, Baümler AJ. 2008. Binding specificity of Salmonella plasmid-encoded fimbriae assessed by glycomics. J Biol Chem 283:8118–8124.
30. Clark CG, Kropinski AM, Parolis H, Grant CCR, Trout-Yakel KM, Franklin K, Ng L-K, Paramonov NA, Parolis LAS, Rahn K, Tabor H. 2009. Escherichia coli O123 O antigen genes and polysaccharide structure are conserved in some Salmonella enterica serogroups. J Med Microbiol 58:884–894.

31. Soyer Y, Switt AM, Davis MA, Maurer J, McDonough PL, Schoonmaker-Bopp DJ, Dumas NB, Root T, Warnick LD, Gröhn YT, Wiedmann M. 2009. Salmonella enterica Serotype 4,5,12:i:−, an Emerging Salmonella Serotype That Represents Multiple Distinct Clones. J Clin Microbiol 47:3546–3556.
32. EFSA Panel on Biological Hazards (BIOHAZ). 2010. Scientific Opinion on monitoring and assessment of the public health risk of "Salmonella Typhimurium-like" strains. EFSA J 8:n/a-n/a.
33. Lucarelli C, Dionisi AM, Filetici E, Owczarek S, Luzzi I, Villa L. 2012. Nucleotide sequence of the chromosomal region conferring multidrug resistance (R-type ASSuT) in Salmonella Typhimurium and monophasic Salmonella Typhimurium strains. J Antimicrob Chemother 67:111–114.
34. Argüello H, Sørensen G, Carvajal A, Baggesen DL, Rubio P, Pedersen K. 2014. Characterization of the Emerging Salmonella 4,[5],12:i:- in Danish Animal Production. Foodborne Pathog Dis 11:366–372.
35. Anderson ES, Humphreys GO, Willshaw GA. 1975. The molecular relatedness of R factors in enterobacteria of human and animal origin. J Gen Microbiol 91:376–382.
36. Beceiro A, Tomás M, Bou G. 2013. Antimicrobial Resistance and Virulence: a Successful or Deleterious Association in the Bacterial World? Clin Microbiol Rev 26:185–230.
37. Wray C, Davies RH, Unit WHOVPH. 1994. Guidelines on detection and monitoring of salmonella infected poultry flocks with particular reference to salmonella enteritidis.
38. Threlfall EJ. 2000. Epidemic salmonella typhimurium DT 104--a truly international multiresistant clone. J Antimicrob Chemother 46:7–10.
39. Hall RM. 2010. Salmonella genomic islands and antibiotic resistance in Salmonella enterica. Future Microbiol 5:1525–1538.
40. Threlfall E j., Hampton M d., Chart H, Rowe B. 1994. Identification of a conjugative plasmid carrying antibiotic resistance and salmonella plasmid virulence (spv) genes in epidemic strains of Salmonella typhimurium phage type 193. Lett Appl Microbiol 18:82–85.
41. Alcaine SD, Warnick LD, Wiedmann M. 2007. Antimicrobial Resistance in Nontyphoidal Salmonella. J Food Prot 70:780–790.
42. Zamperini K, Soni V, Waltman D, Sanchez S, Theriault EC, Bray J, Maurer JJ. 2007. Molecular characterization reveals Salmonella enterica serovar 4,[5],12:i:- from poultry is a variant Typhimurium serovar. Avian Dis 51:958–964.
43. Mandilara G, Lambiri M, Polemis M, Passiotou M, Vatopoulos A. 2013. Phenotypic and molecular characterisation of multiresistant monophasic Salmonella Typhimurium (1,4,[5],12:i:-) in Greece, 2006 to 2011. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 18.
44. Petrovska L, Mather AE, AbuOun M, Branchu P, Harris SR, Connor T, Hopkins KL, Underwood A, Lettini AA, Page A, Bagnall M, Wain J, Parkhill J, Dougan G, Davies R, Kingsley RA. 2016. Microevolution during the emergence of a monophasic Salmonella Typhimurium epidemic in the United Kingdom. Emerg Infect Dis 22.
45. Shah DH, Paul NC, Sischo WC, Crespo R, Guard J. 2017. Population dynamics and antimicrobial resistance of the most prevalent poultry-associated Salmonella serotypes. Poult Sci 96:687–702.

46. Mulvey MR, Finley R, Allen V, Ang L, Bekal S, El Bailey S, Haldane D, Hoang L, Horsman G, Louie M, Robberts L, Wylie J, McCracken M, Langner S, Ahmed R, Tabor H, Gilmour M. 2013. Emergence of multidrug-resistant Salmonella enterica serotype 4,[5],12:i:- involving human cases in Canada: results from the Canadian Integrated Program on Antimicrobial Resistance Surveillance (CIPARS), 2003–10. J Antimicrob Chemother 68:1982–1986.
47. Agasan A, Kornblum J, Williams G, Pratt C-C, Fleckenstein P, Wong M, Ramon A. 2002. Profile of Salmonella enterica subsp. enterica (subspecies I) serotype 4,5,12:i:- strains causing food-borne infections in New York City. J Clin Microbiol 40:1924–1929.
48. Switt AIM, Soyer Y, Warnick LD, Wiedmann M. 2009. Emergence, Distribution, and Molecular and Phenotypic Characteristics of Salmonella enterica Serotype 4,5,12:i:–. Foodborne Pathog Dis 6:407–415.
49. Amavisit P, Boonyawiwat W, Bangtrakulnont A. 2005. Characterization of Salmonella enterica Serovar Typhimurium and Monophasic Salmonella Serovar 1,4,[5],12:i:- Isolates in Thailand. J Clin Microbiol 43:2736–2740.
50. de la Torre E, Zapata D, Tello M, Mejía W, Frías N, García Peña FJ, Mateu EM, Torre E. 2003. Several Salmonella enterica subsp. enterica serotype 4,5,12:i:- phage types isolated from swine samples originate from serotype typhimurium DT U302. J Clin Microbiol 41:2395–2400.
51. Mossong J, Marques P, Ragimbeau C, Huberty-Krau P, Losch S, Meyer G, Moris G, Strottner C, Rabsch W, Schneider F. 2007. Outbreaks of monophasic Salmonella enterica serovar 4,[5],12:i:- in Luxembourg, 2006. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 12:E11-12.
52. García P, Malorny B, Hauser E, Mendoza MC, Rodicio MR. 2013. Genetic Types, Gene Repertoire, and Evolution of Isolates of the Salmonella enterica Serovar 4,5,12:i:− Spanish Clone Assigned to Different Phage Types. J Clin Microbiol 51:973–978.
53. Elnekave E, Hong S, Mather AE, Boxrud D, Taylor AJ, Lappi V, Johnson TJ, Vannucci F, Davies P, Hedberg C, Perez A, Alvarez J. Salmonella enterica serotype 4,[5],12:i:- in swine in the United States Midwest: an emerging multidrug resistant clone. Clin Infect Dis.
54. García P, Malorny B, Rodicio MR, Stephan R, Hächler H, Guerra B, Lucarelli C. 2016. Horizontal Acquisition of a Multidrug-Resistance Module (R-type ASSuT) Is Responsible for the Monophasic Phenotype in a Widespread Clone of Salmonella Serovar 4,[5],12:i:-. Front Microbiol 7.
55. Laorden L, Herrera-León S, Martínez I, Sanchez A, Kromidas L, Bikandi J, Rementeria A, Echeita A, Garaizar J. 2010. Genetic Evolution of the Spanish Multidrug-Resistant Salmonella enterica 4,5,12:i:- Monophasic Variant. J Clin Microbiol 48:4563–4566.
56. Echeita MA, Herrera S, Usera MA. 2001. Atypical, fljB-negative Salmonella enterica subsp. enterica strain of serovar 4,5,12:i:- appears to be a monophasic variant of serovar Typhimurium. J Clin Microbiol 39:2981–2983.
57. Trüpschuch S, Laverde Gomez JA, Ediberidze I, Flieger A, Rabsch W. 2010. Characterisation of multidrug-resistant Salmonella Typhimurium 4,[5],12:i:- DT193 strains carrying a novel genomic island adjacent to the thrW tRNA locus. Int J Med Microbiol 300:279–288.
58. Listeria monocytogenes (Murray et al., 1926) Pirie, 1940.

59. Piffaretti JC, Kressebuch H, Aeschbacher M, Bille J, Bannerman E, Musser JM, Selander RK, Rocourt J. 1989. Genetic characterization of clones of the bacterium Listeria monocytogenes causing epidemic disease. Proc Natl Acad Sci U S A 86:3818–3822.

60. Orsi RH, Bakker HC den, Wiedmann M. 2011. Listeria monocytogenes lineages: Genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol 301:79–96.

61. Wiedmann M, Bruce JL, Keating C, Johnson AE, McDonough PL, Batt CA. 1997. Ribotypes and virulence gene polymorphisms suggest three distinct Listeria monocytogenes lineages with differences in pathogenic potential. Infect Immun 65:2707–2716.

62. Doumith M, Buchrieser C, Glaser P, Jacquet C, Martin P. 2004. Differentiation of the Major Listeria monocytogenes Serovars by Multiplex PCR. J Clin Microbiol 42:3819–3822.

63. Chenal-Francisque V, Lopez J, Cantinelli T, Caro V, Tran C, Leclercq A, Lecuit M, Brisse S. 2011. Worldwide distribution of major clones of Listeria monocytogenes. Emerg Infect Dis 17:1110–1112.

64. Haase JK, Murphy RA, Choudhury KR, Achtman M. 2011. Revival of Seeliger's historical "Special Listeria Culture Collection." Environ Microbiol 13:3163–3171.

65. Maury MM, Tsai Y-H, Charlier C, Touchon M, Chenal-Francisque V, Leclercq A, Criscuolo A, Gaultier C, Roussel S, Brisabois A, Disson O, Rocha EPC, Brisse S, Lecuit M. 2016. Uncovering Listeria monocytogenes hypervirulence by harnessing its biodiversity. Nat Genet 48:308–313.

66. den Bakker HC, Cummings CA, Ferreira V, Vatta P, Orsi RH, Degoricija L, Barker M, Petrauskene O, Furtado MR, Wiedmann M. 2010. Comparative genomics of the bacterial genus Listeria: Genome evolution is characterized by limited gene acquisition and limited gene loss. BMC Genomics 11:688.

67. Buchrieser C, Rusniok C, The Listeria Consortium, Kunst F, Cossart P, Glaser P. 2003. Comparison of the genome sequences of Listeria monocytogenes and Listeria innocua: clues for evolution and pathogenicity. FEMS Immunol Med Microbiol 35:207–213.

68. Wiedmann M, Zhang W. 2011. Genomics of Foodborne Bacterial Pathogens. Springer Science & Business Media.

69. Swaminathan B, Gerner-Smidt P. 2007. The epidemiology of human listeriosis. Microbes Infect 9:1236–1243.

70. Drevets DA, Bronze MS. 2008. Listeria monocytogenes: epidemiology, human disease, and mechanisms of brain invasion. FEMS Immunol Med Microbiol 53:151–165.

71. Bonazzi M, Lecuit M, Cossart P. 2009. Listeria monocytogenes internalin and E-cadherin: from structure to pathogenesis. Cell Microbiol 11:693–702.

72. Nightingale KK, Ivy RA, Ho AJ, Fortes ED, Njaa BL, Peters RM, Wiedmann M. 2008. inlA Premature Stop Codons Are Common among Listeria monocytogenes Isolates from Foods and Yield Virulence-Attenuated Strains That Confer Protection against Fully Virulent Strains. Appl Environ Microbiol 74:6570–6583.

73. Stelten AV, Simpson JM, Chen Y, Scott VN, Whiting RC, Ross WH, Nightingale KK. 2011. Significant Shift in Median Guinea Pig Infectious Dose Shown by an Outbreak-Associated Listeria monocytogenes Epidemic Clone Strain and a Strain Carrying a Premature Stop Codon Mutation in inlA. Appl Environ Microbiol 77:2479–2487.

74. Cruz CD, Pitman AR, Harrow SA, Fletcher GC. 2014. Listeria monocytogenes Associated with New Zealand Seafood Production and Clinical Cases: Unique Sequence Types, Truncated InlA, and Attenuated Invasiveness. Appl Environ Microbiol 80:1489–1497.
75. Kanki M, Naruse H, Taguchi M, Kumeda Y. 2015. Characterization of specific alleles in InlA and PrfA of Listeria monocytogenes isolated from foods in Osaka, Japan and their ability to invade Caco-2 cells. Int J Food Microbiol 211:18–22.
76. Zhu Q, Gooneratne R, Hussain MA. 2017. Listeria monocytogenes in Fresh Produce: Outbreaks, Prevalence and Contamination Levels. Foods 6.
77. Elhanafi D, Dutta V, Kathariou S. 2010. Genetic characterization of plasmid-associated benzalkonium chloride resistance determinants in a Listeria monocytogenes strain from the 1998-1999 outbreak. Appl Environ Microbiol 76:8231–8238.
78. Katharios-Lanwermeyer S, Rakic-Martinez M, Elhanafi D, Ratani S, Tiedje JM, Kathariou S. 2012. Coselection of cadmium and benzalkonium chloride resistance in conjugative transfers from nonpathogenic Listeria spp. to other Listeriae. Appl Environ Microbiol 78:7549–7556.
79. Mullapudi S, Siletzky RM, Kathariou S. 2008. Heavy-metal and benzalkonium chloride resistance of Listeria monocytogenes isolates from the environment of turkey-processing plants. Appl Environ Microbiol 74:1464–1468.
80. Parsons C, Lee S, Jayeola V, Kathariou S. 2017. Novel Cadmium Resistance Determinant in Listeria monocytogenes. Appl Environ Microbiol 83:e02580-16.
81. Ratani SS, Siletzky RM, Dutta V, Yildirim S, Osborne JA, Lin W, Hitchins AD, Ward TJ, Kathariou S. 2012. Heavy Metal and Disinfectant Resistance of Listeria monocytogenes from Foods and Food Processing Plants. Appl Environ Microbiol 78:6938–6945.
82. Xu D, Nie Q, Wang W, Shi L, Yan H. 2016. Characterization of a transferable bcrABC and cadAC genes-harboring plasmid in Listeria monocytogenes strain isolated from food products of animal origin. Int J Food Microbiol 217:117–122.
83. Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel KM, Larios O, Allen V, Lee B, Nadon C. 2010. High-throughput genome sequencing of two Listeria monocytogenes clinical isolates during a large foodborne outbreak. BMC Genomics 11:120.
84. Kovacevic J, Ziegler J, Wałecka-Zacharska E, Reimer A, Kitts DD, Gilmour MW. 2016. Tolerance of Listeria monocytogenes to Quaternary Ammonium Sanitizers Is Mediated by a Novel Efflux Pump Encoded by emrE. Appl Environ Microbiol 82:939–953.
85. Ortiz S, López-Alonso V, Rodríguez P, Martínez-Suárez JV. 2015. The Connection between Persistent, Disinfectant-Resistant Listeria monocytogenes Strains from Two Geographically Separate Iberian Pork Processing Plants: Evidence from Comparative Genome Analysis. Appl Environ Microbiol 82:308–317.
86. Müller A, Rychli K, Muhterem-Uyar M, Zaiser A, Stessl B, Guinane CM, Cotter PD, Wagner M, Schmitz-Esser S. 2013. Tn6188 - a novel transposon in Listeria monocytogenes responsible for tolerance to benzalkonium chloride. PloS One 8:e76835.
87. Lebrun M, Audurier A, Cossart P. 1994. Plasmid-borne cadmium resistance genes in Listeria monocytogenes are similar to cadA and cadC of Staphylococcus aureus and are induced by cadmium. J Bacteriol 176:3040–3048.

88. Lundén JM, Miettinen MK, Autio TJ, Korkeala HJ. 2000. Persistent Listeria monocytogenes strains show enhanced adherence to food contact surface after short contact times. J Food Prot 63:1204–1207.

89. Latorre AA, Pradhan AK, Van Kessel JAS, Karns JS, Boor KJ, Rice DH, Mangione KJ, Gröhn YT, Schukken YH. 2011. Quantitative risk assessment of listeriosis due to consumption of raw milk. J Food Prot 74:1268–1281.

90. Borucki MK, Call DR. 2003. Listeria monocytogenes serotype identification by PCR. J Clin Microbiol 41:5537–5540.

91. Harter E, Wagner EM, Zaiser A, Halecker S, Wagner M, Rychli K. 2017. Stress Survival Islet 2, Predominantly Present in Listeria monocytogenes Strains of Sequence Type 121, Is Involved in the Alkaline and Oxidative Stress Responses. Appl Environ Microbiol 83.

92. Hingston P, Chen J, Dhillon BK, Laing C, Bertelli C, Gannon V, Tasara T, Allen K, Brinkman FSL, Truelstrup Hansen L, Wang S. 2017. Genotypes Associated with Listeria monocytogenes Isolates Displaying Impaired or Enhanced Tolerances to Cold, Salt, Acid, or Desiccation Stress. Front Microbiol 8:369.

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

94. Hein I, Klinger S, Dooms M, Flekna G, Stessl B, Leclercq A, Hill C, Allerberger F, Wagner M. 2011. Stress Survival Islet 1 (SSI-1) Survey in Listeria monocytogenes Reveals an Insert Common to Listeria innocua in Sequence Type 121 L. monocytogenes Strains. Appl Environ Microbiol 77:2169–2173.

95. Verghese B, Lok M, Wen J, Alessandria V, Chen Y, Kathariou S, Knabel S. 2011. comK prophage junction fragments as markers for Listeria monocytogenes genotypes unique to individual meat and poultry processing plants and a model for rapid niche-specific adaptation, biofilm formation, and persistence. Appl Environ Microbiol 77:3279–3292.

96. Jemmi T, Stephan R. 2006. Listeria monocytogenes: food-borne pathogen and hygiene indicator. Rev Sci Tech Int Off Epizoot 25:571–580.

97. Garcia JS. 2009. Microbiologically Safe Foods. John Wiley & Sons.

98. Denny J, McLauchlin J. 2008. Human Listeria monocytogenes infections in Europe - an opportunity for improved European surveillance. Eurosurveillance 13:9–10.

99. Tompkin RB. 2002. Control of Listeria monocytogenes in the food-processing environment. J Food Prot 65:709–725.

100. Gossner C. 2013. ECDC launches the second version of the EPIS-FWD platform. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 18.

101. EFSA Panel on Biological Hazards (BIOHAZ). 2014. Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 2 (surveillance and data management activities). EFSA J 12:n/a-n/a.

102. Rizzi V, Felicio TDS, Felix B, Gossner CM, Jacobs W, Johansson K, Kotila S, Michelon D, Monguidi M, Mooijman K. The ECDC-EFSA molecular typing database for European Union public health protection.

103. Félix B, Danan C, Van Walle I, Lailler R, Texier T, Lombard B, Brisabois A, Roussel S. 2014. Building a molecular Listeria monocytogenes database to centralize and share PFGE typing data from food, environmental and animal strains throughout Europe. J Microbiol Methods 104:1–8.
104. Ribot EM, Fair MA, Gautom R, Cameron DN, Hunter SB, Swaminathan B, Barrett TJ. 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of Escherichia coli O157:H7, Salmonella, and Shigella for PulseNet. Foodborne Pathog Dis 3:59–67.
105. Larsson JT, Torpdahl M, MLVA working group, Møller Nielsen E. 2013. Proof-of-concept study for successful inter-laboratory comparison of MLVA results. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 18:20566.
106. Nadon C, Van Walle I, Gerner-Smidt P, Campos J, Chinen I, Concepcion-Acevedo J, Gilpin B, Smith AM, Kam KM, Perez E, Trees E, Kubota K, Takkinen J, Nielsen EM, Carleton H. 2017. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. Eurosurveillance 22.
107. Schjørring S, Niskanen T, Torpdahl M, Björkman JT, Nielsen EM. 2016. Evaluation of molecular typing of foodborne pathogens in European reference laboratories from 2012 to 2013. Eurosurveillance 21.
108. Llarena A-K, Taboada E, Rossi M. 2017. Whole-Genome Sequencing in Epidemiology of Campylobacter jejuni Infections. J Clin Microbiol 55:1269–1275.
109. Sintchenko V, Holmes EC. 2015. The role of pathogen genomics in assessing disease transmission. BMJ 350:h1314.
110. Stasiewicz MJ, Oliver HF, Wiedmann M, den Bakker HC. 2015. Whole-Genome Sequencing Allows for Improved Identification of Persistent Listeria monocytogenes in Food-Associated Environments. Appl Environ Microbiol 81:6024–6037.
111. Holden MTG, Hsu L-Y, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H, Skov R, Westh H, Žemličková H, Coombs G, Kearns AM, Hill RLR, Edgeworth J, Gould I, Gant V, Cooke J, Edwards GF, McAdam PR, Templeton KE, McCann A, Zhou Z, Castillo-Ramírez S, Feil EJ, Hudson LO, Enright MC, Balloux F, Aanensen DM, Spratt BG, Fitzgerald JR, Parkhill J, Achtman M, Bentley SD, Nübel U. 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. Genome Res 23:653–664.
112. Wang S, Weller D, Falardeau J, Strawn LK, Mardones FO, Adell AD, Moreno Switt AI. 2016. Food safety trends: From globalization of whole genome sequencing to application of new tools to prevent foodborne diseases. Trends Food Sci Technol 57:188–198.
113. Hill AA, Crotta M, Wall B, Good L, O'Brien SJ, Guitian J. 2017. Towards an integrated food safety surveillance system: a simulation study to explore the potential of combining genomic and epidemiological metadata. R Soc Open Sci 4.
114. 2016. ECDC roadmap for integration of molecular typing and genomic typing into European-level surveillance and epidemic preparedness – Version 2.1, 2016-19. Eur Cent Dis Prev Control.
115. Suzuki P-M. 2017. Integration of Genomics Technologies in the Management of Food Safety and Outbreaks in Europe. Iafp.

116. 2016. Expert opinion on whole genome sequencing for public health surveillance. Eur Cent Dis Prev Control.

117. Salipante SJ, SenGupta DJ, Cummings LA, Land TA, Hoogestraat DR, Cookson BT. 2015. Application of Whole-Genome Sequencing for Bacterial Strain Typing in Molecular Epidemiology. J Clin Microbiol 53:1072–1079.

118. Ronholm J, Nasheri N, Petronella N, Pagotto F. 2016. Navigating Microbiological Food Safety in the Era of Whole-Genome Sequencing. Clin Microbiol Rev 29:837–857.

119. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S, Struelens M. 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect 13:1–46.

120. Trindade PA, McCulloch JA, Oliveira GA, Mamizuka EM. 2003. Molecular techniques for MRSA typing: current issues and perspectives. Braz J Infect Dis 7:32–43.

121. Struelens MJ. 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis 2:2–11.

122. Foley SL, Lynne AM, Nayak R. 2009. Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. Infect Genet Evol 9:430–440.

123. Carriço JA, Sabat AJ, Friedrich AW, Ramirez M, ESCMID Study Group for Epidemiological Markers (ESGEM). 2013. Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 18:20382.

124. Ranjbar R, Karami A, Farshad S, Giammanco GM, Mammina C. 2014. Typing methods used in the molecular epidemiology of microbial pathogens: a how-to guide. New Microbiol 37:1–15.

125. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl J m, Laurent F, Grundmann H, Friedrich AW, ESCMID Study Group of Epidemiological Markers (ESGEM). 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 18:20380.

126. Adzitey F, Huda N, Ali GRR. 2013. Molecular techniques for detecting and typing of bacteria, advantages and application to foodborne pathogens isolated from ducks. 3 Biotech 3:97–107.

127. Struelens MJ, Brisse S. 2013. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 18:20386.

128. Graves LM, Swaminathan B. 2006. PulseNet's Step-by-Step Laboratory Protocol for Molecular Subtyping of <Emphasis Type="Italic">Listeria monocytogenes</Emphasis> by Macrorestriction and Pulsed-Field Gel Electrophoresis, p. 57–72. *In* Food-Borne Pathogens. Humana Press.

129. Camargo AC, Woodward JJ, Nero LA. 2016. The Continuous Challenge of Characterizing the Foodborne Pathogen Listeria monocytogenes. Foodborne Pathog Dis 13:405–416.

130. Datta AR, Laksanalamai P, Solomotis M. 2013. Recent developments in molecular sub-typing of Listeria monocytogenes. Food Addit Contam Part A 30:1437–1445.

131. Maslanka SE, Kerr JG, Williams G, Barbaree JM, Carson LA, Miller JM, Swaminathan B. 1999. Molecular Subtyping of Clostridium perfringens by Pulsed-Field Gel Electrophoresis To Facilitate Food-Borne-Disease Outbreak Investigations. J Clin Microbiol 37:2209–2214.
132. Barrett TJ, Gerner-Smidt P, Swaminathan B. 2006. Interpretation of pulsed-field gel electrophoresis patterns in foodborne disease investigations and surveillance. Foodborne Pathog Dis 3:20–31.
133. Boxrud D, Pederson-Gulrud K, Wotton J, Medus C, Lyszkowicz E, Besser J, Bartkus JM. 2007. Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of Salmonella enterica serotype Enteritidis. J Clin Microbiol 45:536–543.
134. Hyytiä-Trees E, Smole SC, Fields PA, Swaminathan B, Ribot EM. 2006. Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing Escherichia coli O157 (STEC O157). Foodborne Pathog Dis 3:118–131.
135. Murphy M, Corcoran D, Buckley JF, O'Mahony M, Whyte P, Fanning S. 2007. Development and application of Multiple-Locus Variable Number of tandem repeat Analysis (MLVA) to subtype a collection of Listeria monocytogenes. Int J Food Microbiol 115:187–194.
136. Chen Y, Ross WH, Whiting RC, Stelten AV, Nightingale KK, Wiedmann M, Scott VN. 2011. Variation in Listeria monocytogenes Dose Responses in Relation to Subtypes Encoding a Full-Length or Truncated Internalin A. Appl Environ Microbiol 77:1171–1180.
137. Lindstedt B-A, Tham W, Danielsson-Tham M-L, Vardund T, Helmersson S, Kapperud G. 2008. Multiple-locus variable-number tandem-repeats analysis of Listeria monocytogenes using multicolour capillary electrophoresis and comparison with pulsed-field gel electrophoresis typing. J Microbiol Methods 72:141–148.
138. Kabuki DY, Kuaye AY, Wiedmann M, Boor KJ. 2004. Molecular Subtyping and Tracking of Listeria monocytogenes in Latin-Style Fresh-Cheese Processing Plants. J Dairy Sci 87:2803–2812.
139. De Cesare A, Mioni R, Manfreda G. 2007. Prevalence of Listeria monocytogenes in fresh and fermented Italian sausages and ribotyping of contaminating strains. Int J Food Microbiol 120:124–130.
140. Sant'ana AS, Landgraf M, Destro MT, Franco BDGM. 2013. Growth Potential of Salmonella and Listeria monocytogenes in Ready-to-Eat Lettuce and Collard Greens Packaged under Modified Atmosphere and in Perforated Film. J Food Prot 76:888–891.
141. Oscar TP. 1998. Identification and characterization of Salmonella isolates by automated ribotyping. J Food Prot 61:519–524.
142. De Cesare A, Krishnamani K, Parisi A, Ricci A, Luzzi I, Barco L, Lucchi A, Miccolupo A, Manfreda G. 2015. Comparison between Salmonella enterica Serotype Enteritidis Genotyping Methods and Phage Type. J Clin Microbiol 53:3021–3031.
143. Bailey JS, Fedorka-Cray PJ, Stern NJ, Craven SE, Cox NA, Cosby DE. 2002. Serotyping and ribotyping of Salmonella using restriction enzyme PvuII. J Food Prot 65:1005–1007.
144. Rodriguez A, Pangloli P, Richards HA, Mount JR, Draughon FA. 2006. Prevalence of Salmonella in diverse environmental farm samples. J Food Prot 69:2576–2580.

145. Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG. 2011. Novel Virulence Gene and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Multilocus Sequence Typing Scheme for Subtyping of the Major Serovars of Salmonella enterica subsp. enterica. Appl Environ Microbiol 77:1946–1956.
146. DiMarzio M, Shariat N, Kariyawasam S, Barrangou R, Dudley EG. 2013. Antibiotic Resistance in Salmonella enterica Serovar Typhimurium Associates with CRISPR Sequence Type. Antimicrob Agents Chemother 57:4282–4289.
147. Fabre L, Zhang J, Guigon G, Hello SL, Guibert V, Accou-Demartin M, Romans S de, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth-Jeanjean S, Achtman M, Brisse S, Sola C, Weill F-X. 2012. CRISPR Typing and Subtyping for Improved Laboratory Surveillance of Salmonella Infections. PLOS ONE 7:e36995.
148. Kuenne C, Billion A, Mraheil MA, Strittmatter A, Daniel R, Goesmann A, Barbuddhe S, Hain T, Chakraborty T. 2013. Reassessment of the Listeria monocytogenespan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. BMC Genomics 14:47.
149. Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T, Fields P. 2007. Multiplex, bead-based suspension array for molecular determination of common Salmonella serogroups. J Clin Microbiol 45:3323–3334.
150. Achtman M. 2012. Insights from genomic comparisons of genetically monomorphic bacterial pathogens. Philos Trans R Soc B Biol Sci 367:860.
151. Ragon M, Wirth T, Hollandt F, Lavenir R, Lecuit M, Monnier AL, Brisse S. 2008. A New Perspective on Listeria monocytogenes Evolution. PLOS Pathog 4:e1000146.
152. Droege M, Hill B. 2008. The Genome Sequencer FLX^{TM} System—Longer reads, more applications, straight forward bioinformatics and more complete data sets. J Biotechnol 136:3–10.
153. Bokulich NA, Lewis ZT, Boundy-Mills K, Mills DA. 2016. A new perspective on microbial landscapes within food production. Curr Opin Biotechnol 37:182–189.
154. Véghová A, Minarovičová J, Koreňová J, Drahovská H, Kaclíková E. 2017. Prevalence and tracing of persistent Listeria monocytogenes strains in meat processing facility production chain. J Food Saf 37:n/a-n/a.
155. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 17:238.
156. Kleinheinz KA, Joensen KG, Larsen MV. 2014. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4.
157. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl 30:2068–2069.
158. Bentley DR et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59.
159. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309:1728–1732.

160. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133–138.

161. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348.

162. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif 3:1–8.

163. Prospective Whole-Genome Sequencing Enhances National Surveillance of Listeria monocytogenes.

164. Moran-Gilad J. 2017. Whole genome sequencing (WGS) for food-borne pathogen surveillance and control – taking the pulse. Eurosurveillance 22.

165. Taboada EN, Graham MR, Carriço JA, Van Domselaar G. 2017. Food Safety in the Age of Next Generation Sequencing, Bioinformatics, and Open Data Access. Front Microbiol 8.

166. Mossong J, Decruyenaere F, Moris G, Ragimbeau C, Olinger CM, Johler S, Perrin M, Hau P, Weicherding P. 2015. Investigation of a staphylococcal food poisoning outbreak combining case-control, traditional typing and whole genome sequencing methods, Luxembourg, June 2014. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 20.

167. Jackson BR, Tarr C, Strain E, Jackson KA, Conrad A, Carleton H, Katz LS, Stroika S, Gould LH, Mody RK, Silk BJ, Beal J, Chen Y, Timme R, Doyle M, Fields A, Wise M, Tillman G, Defibaugh-Chavez S, Kucerova Z, Sabol A, Roache K, Trees E, Simmons M, Wasilenko J, Kubota K, Pouseele H, Klimke W, Besser J, Brown E, Allard M, Gerner-Smidt P. 2016. Implementation of Nationwide Real-time Whole-genome Sequencing to Enhance Listeriosis Outbreak Detection and Investigation. Clin Infect Dis 63:380–386.

168. Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, Björkman JT, Dallman T, Reimer A, Enouf V, Larsonneur E, Carleton H, Bracq-Dieye H, Katz LS, Jones L, Touchon M, Tourdjman M, Walker M, Stroika S, Cantinelli T, Chenal-Francisque V, Kucerova Z, Rocha EPC, Nadon C, Grant K, Nielsen EM, Pot B, Gerner-Smidt P, Lecuit M, Brisse S. 2016. Whole genome-based population biology and epidemiological surveillance of Listeria monocytogenes. Nat Microbiol 2:16185.

169. Dallman TJ, Ashton PM, Byrne L, Perry NT, Petrovska L, Ellis R, Allison L, Hanson M, Holmes A, Gunn GJ, Chase-Topping ME, Woolhouse MEJ, Grant KA, Gally DL, Wain J, Jenkins C. 2015. Applying phylogenomics to understand the emergence of Shiga-toxin-producing Escherichia coli O157:H7 strains causing severe human disease in the UK. Microb Genomics 1.

170. Köser CU, Ellington MJ, Peacock SJ. 2014. Whole-genome sequencing to control antimicrobial resistance. Trends Genet 30:401–407.

171. Franz E, Gras LM, Dallman T. 2016. Significance of whole genome sequencing for surveillance, source attribution and microbial risk assessment of foodborne pathogens. Curr Opin Food Sci 8:74–79.

172. Barco L, Barrucci F, Olsen JE, Ricci A. 2013. Salmonella source attribution based on microbial subtyping. Int J Food Microbiol 163:193–203.

173. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H. 2011. Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. PloS One 6:e22751.

174. Strachan NJC, Rotariu O, Lopes B, MacRae M, Fairley S, Laing C, Gannon V, Allison LJ, Hanson MF, Dallman T, Ashton P, Franz E, van Hoek AHAM, French NP, George T, Biggs PJ, Forbes KJ. 2015. Whole Genome Sequencing demonstrates that Geographic Variation of Escherichia coli O157 Genotypes Dominates Host Association. Sci Rep 5.

175. European Food Safety Authority, European Centre for Disease Prevention and Control. 2017. Multi-country outbreak of Salmonella Enteritidis phage type 8, MLVA profile 2-9-7-3-2 and 2-9-6-3-2 infections. EFSA Support Publ 14:n/a-n/a.

176. Clark CG, Taboada E, Grant CCR, Blakeston C, Pollari F, Marshall B, Rahn K, MacKinnon J, Daignault D, Pillai D, Ng L-K. 2012. Comparison of Molecular Typing Methods Useful for Detecting Clusters of Campylobacter jejuni and C. coli Isolates through Routine Surveillance. J Clin Microbiol 50:798–809.

177. Mughini-Gras L, Enserink R, Friesema I, Heck M, Duynhoven Y van, Pelt W van. 2014. Risk Factors for Human Salmonellosis Originating from Pigs, Cattle, Broiler Chickens and Egg Laying Hens: A Combined Case-Control and Source Attribution Analysis. PLOS ONE 9:e87933.

178. Angelo KM, Conrad AR, Saupe A, Dragoo H, West N, Sorenson A, Barnes A, Doyle M, Beal J, Jackson KA, Stroika S, Tarr C, Kucerova Z, Lance S, Gould LH, Wise M, Jackson BR. 2017. Multistate outbreak of Listeria monocytogenes infections linked to whole apples used in commercially produced, prepackaged caramel apples: United States, 2014–2015. Epidemiol Amp Infect 145:848–856.

179. Self JL. 2016. Notes from the Field: Outbreak of Listeriosis Associated with Consumption of Packaged Salad — United States and Canada, 2015–2016. MMWR Morb Mortal Wkly Rep 65.

180. Chen Y, Burall LS, Luo Y, Timme R, Melka D, Muruvanda T, Payne J, Wang C, Kastanis G, Maounounen-Laasri A, Jesus AJD, Curry PE, Stones R, K'Aluoch O, Liu E, Salter M, Hammack TS, Evans PS, Parish M, Allard MW, Datta A, Strain EA, Brown EW. 2016. Listeria monocytogenes in Stone Fruits Linked to a Multistate Outbreak: Enumeration of Cells and Whole-Genome Sequencing. Appl Environ Microbiol 82:7030–7040.

181. Burall LS, Grim CJ, Mammel MK, Datta AR. 2016. Whole Genome Sequence Analysis Using JSpecies Tool Establishes Clonal Relationships between Listeria monocytogenes Strains from Epidemiologically Unrelated Listeriosis Outbreaks. PLOS ONE 11:e0150797.

182. Bergholz TM, Bakker HC den, Katz LS, Silk BJ, Jackson KA, Kucerova Z, Joseph LA, Turnsek M, Gladney LM, Halpin JL, Xavier K, Gossack J, Ward TJ, Frace M, Tarr CL. 2016. Determination of Evolutionary Relationships of Outbreak-Associated Listeria monocytogenes Strains of Serotypes 1/2a and 1/2b by Whole-Genome Sequencing. Appl Environ Microbiol 82:928–938.
183. Chen Y, Luo Y, Carleton H, Timme R, Melka D, Muruvanda T, Wang C, Kastanis G, Katz LS, Turner L, Fritzinger A, Moore T, Stones R, Blankenship J, Salter M, Parish M, Hammack TS, Evans PS, Tarr CL, Allard MW, Strain EA, Brown EW. 2017. Whole genome and core genome multilocus sequence typing and single nucleotide polymorphism analyses of Listeria monocytogenes associated with an outbreak linked to cheese, United States, 2013. Appl Environ Microbiol AEM.00633-17.
184. Hoffmann M, Luo Y, Monday SR, Gonzalez-Escalona N, Ottesen AR, Muruvanda T, Wang C, Kastanis G, Keys C, Janies D, Senturk IF, Catalyurek UV, Wang H, Hammack TS, Wolfgang WJ, Schoonmaker-Bopp D, Chu A, Myers R, Haendiges J, Evans PS, Meng J, Strain EA, Allard MW, Brown EW. 2016. Tracing Origins of the Salmonella Bareilly Strain Causing a Food-borne Outbreak in the United States. J Infect Dis 213:502–508.
185. den Bakker HC, Allard MW, Bopp D, Brown EW, Fontana J, Iqbal Z, Kinney A, Limberger R, Musser KA, Shudt M, Strain E, Wiedmann M, Wolfgang WJ. 2014. Rapid whole-genome sequencing for surveillance of Salmonella enterica serovar enteritidis. Emerg Infect Dis 20:1306–1314.
186. Wuyts V, Denayer S, Roosens NHC, Mattheus W, Bertrand S, Marchal K, Dierick K, Keersmaecker SCJD. 2015. Whole Genome Sequence Analysis of Salmonella Enteritidis PT4 Outbreaks from a National Reference Laboratory's Viewpoint. PLOS Curr Outbreaks.
187. Deng X, Shariat N, Driebe EM, Roe CC, Tolar B, Trees E, Keim P, Zhang W, Dudley EG, Fields PI, Engelthaler DM. 2015. Comparative analysis of subtyping methods against a whole-genome-sequencing standard for Salmonella enterica serotype Enteritidis. J Clin Microbiol 53:212–218.
188. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. 2014. Evaluation of whole genome sequencing for outbreak detection of Salmonella enterica. PloS One 9:e87991.
189. Ashton PM, Peters T, Ameh L, McAleer R, Petrie S, Nair S, Muscat I, de Pinna E, Dallman T. 2015. Whole Genome Sequencing for the Retrospective Investigation of an Outbreak of Salmonella Typhimurium DT 8. PLoS Curr 7.
190. 2017. Multi-country outbreak of Salmonella Enteritidis infections linked to Polish eggs. Eur Cent Dis Prev Control.
191. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341.
192. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
193. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455–477.

194. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829.

195. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW. 2015. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 15:141–161.

196. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.

197. Rusconi B, Sanjar F, Koenig SSK, Mammel MK, Tarr PI, Eppinger M. 2016. Whole Genome Sequencing for Genomics-Guided Investigations of Escherichia coli O157:H7 Outbreaks. Front Microbiol 7.

198. Gymoese P, Sørensen G, Litrup E, Olsen JE, Nielsen EM, Torpdahl M. 2017. Investigation of Outbreaks of Salmonella enterica Serovar Typhimurium and Its Monophasic Variants Using Whole-Genome Sequencing, Denmark. Emerg Infect Dis 23:1631–1639.

199. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl 25:1754–1760.

200. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio.

201. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

202. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357.

203. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576.

204. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303.

205. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma 43:11.10.1-33.

206. Chewapreecha C, Marttinen P, Croucher NJ, Salter SJ, Harris SR, Mather AE, Hanage WP, Goldblatt D, Nosten FH, Turner C, Turner P, Bentley SD, Parkhill J. 2014. Comprehensive Identification of Single Nucleotide Polymorphisms Associated with Beta-lactam Resistance within Pneumococcal Mosaic Genes. PLoS Genet 10.

207. Revez J, Llarena A-K, Schott T, Kuusi M, Hakkinen M, Kivistö R, Hänninen M-L, Rossi M. 2014. Genome analysis of Campylobacter jejuni strains isolated from a waterborne outbreak. BMC Genomics 15:768.

208. Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, Wigley P, Barquist L, Langridge GC, Feltwell T, Harris SR, Mather AE, Fookes M, Aslett M, Msefula C, Kariuki S, Maclennan CA, Onsare RS, Weill F-X, Le Hello S, Smith AM, McClelland M, Desai P, Parry CM, Cheesbrough J, French N, Campos J, Chabalgoity JA, Betancor L, Hopkins KL, Nair S,

Humphrey TJ, Lunguya O, Cogan TA, Tapia MD, Sow SO, Tennant SM, Bornstein K, Levine MM, Lacharme-Lora L, Everett DB, Kingsley RA, Parkhill J, Heyderman RS, Dougan G, Gordon MA, Thomson NR. 2016. Distinct Salmonella Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. Nat Genet 48:1211–1217.

209. Lynch T, Petkau A, Knox N, Graham M, Domselaar GV. 2016. A Primer on Infectious Disease Bacterial Genomics. Clin Microbiol Rev 29:881–913.
210. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624.
211. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Res 39:W347–W352.
212. Stothard P, Wishart DS. 2006. Automated bacterial genome analysis and annotation. Curr Opin Microbiol 9:505–510.
213. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.
214. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.
215. Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Parkhill J, Rajandream M-A. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics 24:2672–2676.
216. Maiden MCJ, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 11:728–736.
217. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. Curr Opin Genet Dev 15:589–594.
218. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome." Proc Natl Acad Sci U S A 102:13950–13955.
219. Zhang J, Halkilahti J, Hänninen M-L, Rossi M. 2015. Refinement of Whole-Genome Multilocus Sequence Typing Analysis by Addressing Gene Paralogy. J Clin Microbiol 53:1765–1767.

220. Silva M, Machado MP, Rossi M, Moran-Gilad J, Santos S, Ramirez M, Carrico JA. 2017. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. bioRxiv 173146.

221. McNally A, Oren Y, Kelly D, Pascoe B, Dunn S, Sreecharan T, Vehkala M, Välimäki N, Prentice MB, Ashour A, Avram O, Pupko T, Dobrindt U, Literak I, Guenther S, Schaufler K, Wieler LH, Zhiyong Z, Sheppard SK, McInerney JO, Corander J. 2016. Combined Analysis of Variation in Core, Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. PLOS Genet 12:e1006280.

222. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693.

223. Thorpe HA, Bayliss SC, Sheppard SK, Feil EJ. 2017. Piggy: A Rapid, Large-Scale Pan-Genome Analysis Tool for Intergenic Regions in Bacteria. bioRxiv 179515.

224. Thorpe HA, Bayliss SC, Hurst LD, Feil EJ. 2017. Comparative Analyses of Selection Operating on Nontranslated Intergenic Regions of Diverse Bacterial Species. Genetics 206:363–376.

225. Marinier E, Zaheer R, Berry C, Weedmark KA, Domaratzki M, Mabon P, Knox NC, Reimer AR, Graham MR, Chui L, Patterson-Fortin L, Zhang J, Pagotto F, Farber J, Mahony J, Seyer K, Bekal S, Tremblay C, Isaac-Renton J, Prystajecky N, Chen J, Slade P, Van Domselaar G. 2017. Neptune: a bioinformatics tool for rapid discovery of genomic variation in bacterial populations. Nucleic Acids Res 45:e159–e159.

226. Collins C, Didelot X. 2017. A Phylogenetic Method To Perform Genome-Wide Association Studies In Microbes That Accounts For Population Structure And Recombination. bioRxiv 140798.

227. Didelot X, Lawson D, Darling A, Falush D. 2010. Inference of Homologous Recombination in Bacteria Using Whole-Genome Sequences. Genetics 186:1435–1449.

228. Henri C, Leekitcharoenphon P, Carleton HA, Radomski N, Kaas RS, Mariet J-F, Felten A, Aarestrup FM, Gerner Smidt P, Roussel S, Guillier L, Mistou M-Y, Hendriksen RS. 2017. An Assessment of Different Genomic Approaches for Inferring Phylogeny of Listeria monocytogenes. Front Microbiol 8.

229. Nei M, Stephens JC, Saitou N. 1985. Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. Mol Biol Evol 2:66–85.

230. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

231. Kristensen DM, Wolf YI, Mushegian AR, Koonin EV. 2011. Computational methods for Gene Orthology inference. Brief Bioinform 12:379–391.

232. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729.

233. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE 5:e9490.

234. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268–274.
235. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973.
236. Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol 11:e1004041.
237. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, Pascale GD, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The Comprehensive Antibiotic Resistance Database. Antimicrob Agents Chemother 57:3348–3357.
238. Humphrey B, Thomson NR, Thomas CM, Brooks K, Sanders M, Delsol AA, Roe JM, Bennett PM, Enne VI. 2012. Fitness of Escherichia coli strains carrying expressed and partially silent IncN and IncP1 plasmids. BMC Microbiol 12:53.
239. Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, Villa L, Aarestrup FM, Hasman H. 2014. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrob Agents Chemother 58:3895–3903.
240. Nascimento M, Sousa A, Ramirez M, Francisco AP, Carriço JA, Vaz C. 2017. PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. Bioinforma Oxf Engl 33:128–129.
241. Ribeiro-Gonçalves B, Francisco AP, Vaz C, Ramirez M, Carriço JA. 2016. PHYLOViZ Online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees. Nucleic Acids Res 44:W246-251.

Aims of the work

The objective of this thesis is to get insights on epidemiology and persistence of food-borne pathogens through the application of whole genome sequencing and comparative genome analysis-based approaches. In order to achieve this, two studies were developed concerning *Listeria monocytogenes* and *Salmonella* Typhimurium variant 4,[5],12:i:- (MVSTm), respectively.

Study I focused on *L. monocytogenes* ST14 and ST121 isolates repeatedly collected over one year from an Italian rabbit production plant. The aims were to (i) improve molecular subtyping of *L. monocytogenes* (e.g. cgMLST, virulome), (ii) explore SNP differences within each ST as a quantitative metric to differentiate persistent from sporadic strains and (iii) investigate the genes enrichment of observed isolates including ST14 and ST121 publicly available genomes in order to identify key molecular determinants that may contribute to the ability of *L. monocytogenes* to colonize and tolerate stresses in food processing environments.

Study II focused on monophasic *Salmonella* Typhimurium variant 4,[5],12:i:- (MVSTm). Herein, the aims were to untangle the evolutionary history of MVSTm isolates circulating in swine sources and in humans in Italy during two years of sampling (2013-2014) in an extended contest of selected publicly available *S.* Typhimurium/ MVSTm strains from several countries. Therefore, we unravel the phylogeographical structure and identified biogeographical genetic markers by applying an innovative workflow based on recently developed bioinformatics approaches including gene-by-gene analysis, Genome-Wide Association Study and core SNPs phylogeny.

Study I: Listeria monocytogenes in a rabbit meat processing plant: persistence investigations

Listeria monocytogenes is responsible pathogen for listeriosis, a severe foodborne illness caused by consumption of contaminated food. Its ability to grow at low temperatures, form bio-films and persist in food processing plants for long time makes it challenging to effectively implement control measures against the survival and adaptation of this pathogen.
Genomic characterization and ecophysiology studies can address the long-term persistence and phenotypic responses of this food-borne pathogen to environmental conditions specific of the food processing plants.
In a previous study on the prevalence of *L. monocytogenes* in four Italian rabbit meat processing plants (A, B, C, D), isolates sharing the same 7-loci MLST, Multi Locus Variable number tandem repeats Analysis (MLVA) type and *Apa*I-PFGE profiles were repeatedly collected over time from carcasses, meat cuts, meat products and the meat-processing environment (1). Two of the most represented *ApaI*-PFGE profiles in the rabbit meat plant A, belonging to subtypes ST14 and ST121 and collected over one year from the rabbit meat plant A, provided an interesting dataset to investigate persistence, virulence potential and physiological adaptation to food-processing environmental stresses. In this study, different comparative genomics approaches have been applied to investigate on 33 newly sequenced *L. monocytogenes* genomes of ST14 (27) and ST121 (6) strains isolated from different sources (environment, carcasses, rabbit meat cuts and products) within the observed rabbit plant. In a first step, the research had a specific focus on: (i) the evaluation of the discriminatory power of cgMLST in comparison to molecular typing methods and (ii) the *in silico* characterization of the virulence potential of ST14 and ST121 newly sequenced genomes in comparison to a selection of publicly available genomes, mostly from strains isolated in food-processing environments. At a later stage, the research has been focusing on: (i) assessing persistence of newly sequenced *L. monocytogenes* genomes from the previous work; (ii) selecting genes associated to physiological adaptation to food-processing environment; (iii) comparing presence/ absence/ truncation of these genes. In order to fulfil these objectives, a selection of a curated gene dataset based on a wide comprehensive literature review and additional comparative genomics approaches have been applied on the 33 newly sequenced genomes in comparison to a large set of 240 publicly available ST121 and ST14 genomes. Public genomes originated from *L. monocytogenes* strains widely distributed within time and geographical locations and isolated from food processing environments and humans. Phenotypic tests have been also performed on isolates from Italian rabbit meat processing plant for genotypic results confirmation.

Introduction

Listeria monocytogenes is a food-borne pathogen adapted to survive, grow and ultimately persist in different environments (2). In food processing plants, *L. monocytogenes* has been repeatedly isolated both from food samples and food processing environment for months or years (3–5). Based on different molecular typing methods, isolates sharing the same profile have been collected over months or years in fish, meat, dairy and vegetable processing plants (3–5). Strains repeatedly isolated over time in the same plant are considered as persistent. Unfortunately, there is not yet an agreement on specific issues related to the definition of persistence. In particular, the number of times of re-isolation, the sources as well as the period of isolation are not yet uniquely defined (6).

Persistent strains of high virulence are of major concern since they commonly colonize harbourage sites difficult to clean or to reach by sanitizing procedures. These strains typically contaminate different lots of food during several months of production and have been described as responsible of outbreaks including few to hundreds cases spread in time and geographical areas (7). The reasons why *L. monocytogenes* persists in food processing plants is still in debate. One strain can by chance colonize harbourage sites (8). In this view, the persistence of the strain was supposed as more related to characteristics of the environment than particular characteristics of the strain itself (8). In this regard, authors failed to identify associations among persistence and particular genes/ features of the strain (3, 9). Nevertheless, a differential distribution of specific subtypes in food environment and clinical samples was observed, suggesting that certain *L. monocytogenes* strains might harbour unique genotypic and phenotypic characteristics facilitating the survival and growth in food processing environments and the spread to humans.

L. monocytogenes isolates of lineage II and serotype 1/2a have been isolated from food and food processing plants more frequently than isolates belonging to lineage I (10). Clonal complex (CC) 121 belonging to serotype 1/2a has been described as the most prevalent clone with a strong association to food origin. In particular, CC121 showed an overall prevalence of 17.6 % over 6,633 tested isolates of *L. monocytogenes* collected over nine years from food and clinical sources, with a statistical significant over-representation in food sources in comparison to clinical ones (92.9% vs 7.0%) (11). Low values of prevalence of human CC121 (11/116, 9.5%) were observed also within 116 strains of the Institute Pasteur *L. monocytogenes* database (http://bigsdb.pasteur.fr/listeria/). Low frequency of *L. monocytogenes* in humans might be linked to the attenuated virulence of this CC. Premature Stop Codons (PMSC) in the virulence marker *inl*A (12–16).

Compared to CC121, CC14 has been isolated rarely in food. Maury and colleagues found 92 isolates belonging to CC14 over 6,633 tested isolates (1.4%) (11). Lower detection values were described also in 19 meat processing plants located in Northern Italy (5.7% over 69 tested isolates) in comparison to CC121 (23%) (17). Besides a low prevalence, CC14 was described to be associated to higher infection rates. The clinical frequency of CC14 reached a not negligible 29.5% among isolates of clinical sources (11). Within CC14, a ST14 strain isolated from a case of invasive listeriosis was molecular characterized as belonging to epidemic clone (EC) III (corresponding to Multi-Virulence Locus Type 1 (VT1)) previously associated to a sporadic case occurred in United States in 1988, as well as a multi-state outbreak occurred in United States in 2000 (17,18) (https://sites.google.com/site/mvlstdatabase/protocol-for-mvlst).

Physiological adaptation (ecophysiology) to environmental stresses including resistance to antimicrobials, heavy metals and quaternary ammonium compounds (QAC) as well as adaptation to cold, salt, acid, oxidative stresses and desiccation and ability of biofilm formation are often described in *L. monocytogenes* isolated from food processing plants (20–32). Studies on the frequently identified ST121 strains have been recently performed aiming at identifying genes associated to persistence and/ or adaptation to food-processing environmental stress of this subtype. In particular, deletions of *lmo02774-lmo2776* were described along with no significant association of genes. In one

study, ST121 isolates were found to harbour the *qacH* carrying transposon Tn*1688* (QAC resistance) (28, 33, 34). No similar studies are available on ST14.
Whole Genome sequencing (WGS) based analyses such as single nucleotide polymorphisms (SNPs) analysis and core genome Multi Locus Sequence Typing (cgMLST) have recently revealed a valuable potential and high discrimination power in multiple investigations. This is particularly relevant in studies of repeatedly re-isolated strains showing high genetic similarity (same Pulsed Field Gel Electrophoresis (PFGE) or 7- locus MLST) in which a superior discrimination power is required, especially to differentiate true persistent strains to sporadic strains (6). Moreover, based on WGS data, studies on novel genetic determinants as well as on genetic determinants already known as linked to specific phenotypes (i.e. virulence) can be performed at once. At present, freely available tools unable a quick screening of genomes for thousands of genes selected for their association to virulence or antimicrobial resistance (35, 36). Genes, useful to predict phenotypic traits associated to *L. monocytogenes* physiological adaptation to environmental conditions, have not been collected so far in a unique dataset.
In a previous study on prevalence of *L. monocytogenes* in four Italian rabbit meat processing plants (A, B, C, D), isolates sharing the same 7-locus MLST, MLVA and *Apa*I-PFGE profiles were repeatedly collected over time from carcasses, meat cuts, meat products and the food processing environment (1). This study provided an interesting dataset to investigate further specific ST-types over time. For this purpose, a specific focus was put on ST14 (belonging to CC14) and ST121 (belonging to CC121) isolates, repeatedly collected over one year and six months, respectively, from different rabbit meat and environmental sources of the same processing plant (A). The initial aims of the present study were: 1) to evaluate the discriminatory power of cgMLST in comparison to molecular typing methods; 2) to characterize the virulence potential of 33 newly sequenced genomes along with 23 publicly available genomes of ST14 and ST121 *L. monocytogenes* strains. Additional investigation on persistence and physiological adaptation of *L. monocytogene*s to food-processing environmental stresses were also performed. In particular, the research questions were: 1) assess persistence of *L. monocytogenes* isolates from the rabbit-plant; 2) select genes associated to physiological adaptation to food-processing environment; 3) compare presence/ absence/ truncation of these genes in newly sequenced and publicly available ST121 and ST14 genomes. A particular focus has been placed on the evaluation of putative gene enrichment in the two subtypes as well as identification of putative markers of ecophysiology associated to survival and growth of rare yet hypervirulent subtype ST14 of *L. monocytogenes*.

Materials and methods

Bacterial isolates

L. monocytogenes isolates included in the present study are a subset of isolates collected from November 2005 to November 2006 within a previous study on prevalence of *L. monocytogenes* in rabbit meat processing plants (1). Isolates were considered as potentially persistent when they belonged to the same genotype and were collected more than six times over a period of more than six months from different sources (rabbit carcasses, rabbit meat cuts, rabbit meat products and food processing environment) in the same plant. Therefore, 33 *L. monocytogenes* isolates belonging to two 7-locus Multi Locus Sequence Types (MLST), ST14 and ST121, and indistinguishable by *Apa*I-Pulsed Field Electrophoresis (PFGE), automated ribotyping and Multi Locus Variable number tandem repeat Analysis (MLVA) (1) were selected as potentially persistent. An additional sporadic isolate, belonging to lineage I, serotype 1/2b, ST224 was also included for backward comparability along with three technical replicates (LSALM63-65) as controls.

In order to better explore the phylogenetic relationship of newly sequenced *L. monocytogenes* and to investigate ST-specific genetic putative markers associated to ecophysiology a selection of a representative set of publicly available ST14 (44) and ST121 (196) draft genomes from ENA and NCBI were included in this study. In particular, public available genomes were from strains isolated from both food processing environment as well as strains of human origin widely distributed over time and geographical locations (Study1_Research2_Supplementary_Table_S2).

DNA extraction, sequencing and assembly

Genomic DNA was extracted from the 34 *L. monocytogenes* isolates using the MagAttract HMW DNA Kit (Qiagen, Hilden, Germany). The purified DNA concentration and the quality parameter ratio 260/280 were measured by BioSpectrometer fluorescence (Eppendorf). The whole genome of all isolates was sequenced using Illumina MiSeq platform (TrueSeq library, paired-end reads) at the Istituto Zooprofilattico della Lombardia ed Emilia-Romagna (IZSLER) "Bruno Ubertini" in Parma, Italy. Then, paired-end reads were quality checked and *de novo* assembled using INNUca (https://github.com/INNUENDOCON/INNUca) a QA/ QC pipeline developed within the European project INNUENDO (http://www.innuendoweb.org). Briefly, INNUca calculates if the sample raw data fulfil the expected coverage (minimum default 15X). Then, after a read quality analysis using FASTQC and trimming using Trimmomatic (37), *de novo* draft genome assembly is performed with SPAdes (38) 3.11, which is subsequently improved using PILON (39) to correct bases and fix misassembles.

Bioinformatics data analyses for comparative genomics of *L. monocytogenes*

Genomic MLST analysis

Core genome Multi Locus Sequence Typing (cgMLST) was calculated on 37 genome assemblies by the Institute Pasteur cgMLST curators (http://bigsdb.pasteur.fr/listeria/) based on a core gene schema of 1,748 loci. Briefly, after genes alignment, an allele number is assigned to each locus and a cluster type representative of all allele numbers is assigned (40). Dendrogram of cgMLST profiles was inferred based on UPGMA algorithm considering alleles values as categorical data.

Virulotyping of environmental ST14 and ST121 L. monocytogenes

The Multi Virulence Locus Sequence Type (MVLST) was inferred based on *in silico* sequence alignment of seven virulence determinant gene loci: *clpP, dal, inlB, inlC, lisR, prfA* (https://sites.google.com/site/mvlstdatabase) (41) and the analyses of virulence genes presence was performed using VirulenceFinder 1.5 (https://cge.cbs.dtu.dk/services/VirulenceFinder) (42). Therefore, a BLAST search of a database of 82 *L. monocytogenes* virulence determinant genes was performed on all ST14 and ST121 newly sequenced genomes from this study, along with a selected set of 23 publicly available genomes of *L. monocytogenes* belonging to ST14 and ST121 and isolated from humans or food processing plants. Publicly available genome of EGD-e (Genbank accession number NC_003210.1) was included as reference. The default parameters used were 90% ID threshold and 60% of minimum length. Upon detection of no gene the analysis was repeated with 85% ID threshold and 20% minimum length. Alignment of *inlA* and *prfA* genes was performed by Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo).

Statistical analysis: index of diversity

The discriminatory power of this typing method was assessed by the Simpson's Index of Diversity (ID) in comparison to the typing methods applied on the same isolates by De Cesare and colleagues (16). ID values with $P<0.05$ were considered statistically significant different (43).

Phylogenetic reconstruction based on SNPs analysis

SNPs detection was performed using the Snippy v2.6 pipeline (https://github.com/tseemann/snippy) on a total of 202 ST121 and 71 ST14 *de novo* assemblies. A newly generated draft genome for each *L. monocytogenes* ST-type was chosen as reference genome: LSALM51 for ST121 and LSALM1 for ST14. After removal of Illumina Nextera adapters and low-quality sequences (*Phred* scores of <10) assemblies were mapped with the Burrows-Wheeler Aligner (BWA) v0.7.12 (44) using default parameters against the reference genome. After mapping, average nucleotide depths were determined with SAMtools v1.3 (45) and variants were called using Freebayes v0.9.20 (46). Snippy was used to pool all identified SNP positions called in at least one isolate and investigate all isolates of the panel at that position. Alignments of SNPs detected on the whole genomes were used to infer a high-resolution phylogeny. A maximum likelihood (ML) tree was constructed using IQtree (https://github.com/Cibiv/IQ-TREE) and iTOL (https://itol.embl.de) (47) was used to visualize the tree rooted at midpoint for each ST. Draft genomes were tentatively considered as belonging to the same persistent clone if the following criteria were fulfilled: 1) the difference between draft genomes and the reference genome was equal or lower than 25 SNPs; 2) draft genomes belonged to isolates collected from different origin (food and the processing environment) for at least six times during a time frame of at least 6 months. The cut-off of 25 SNPs was used as previously proposed (48).

Dataset of putative gene markers of ecophysiology in L. monocytogenes

An extensive literature review was conducted on PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) using following keywords: “antimicrobials”, “quaternary ammonium compounds”, “heavy metals”, “bacteriocins”, “cold”, “high salt concentration”, “low pH”, “desiccation”, “blue-light”, “biofilm”, “*Listeria monocytogenes*” and “genes”. At the time of the analysis (October 2017), the system retrieved overall 664 peer-reviewed published papers. The abstract of each of these papers was read with the purpose to identify genes associated to the response of *L. monocytogenes* to each stress. Around 100 papers were selected based on the abstract. These papers were thoroughly read in order to identify and specifically select genes with a strong association to the related phenotype. For this purpose, only genes confirmed by insertional mutagenesis or mutant selection experiments were included (Study1_Research2_Supplementary_Table_S1). All GenBank accession numbers and related sequences reported in published papers were checked by nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and CLUSTAL Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) for alignment. For each gene, the following information were reported: name of the gene, locus tag, annotation, main function, localization, reference paper, GenBank accession number and direct link to the web page of the sequence in NCBI (https://www.ncbi.nlm.nih.gov/nuccore/) (Study1_Research2_Supplementary_Table_S1). Each gene sequences were screened in mass using abricate pipeline (https://github.com/tseemann/abricate) on contigs of all 273 draft genomes of *L. monocytogenes* ST121 and ST14 included in this study and statistically significant differences at 95% confidence by t-test were assessed.

Data availability

De novo assembled genomes of the 27 ST14 and 6 ST121 *L. monocytogenes* isolates included in this study were deposited at GenBank under BioProject no. PRJNA396103 with individual BioSample identification (ID) numbers SAMN07420940 to SAMN07420973.
Supplementary materials are accessible in a dedicated GitHub repository under the following link: https://github.com/fedex88/PhD_Thesis. Spreadsheets have been uploaded in PhD Thesis under file names “Study1_Research2_Supplementary_Table_S1” and “Study1_Research2_Supplementary _Table_S2”.

Phenotypic tests for confirming genotyping results

Antimicrobials, cadmium and benzalkonium chloride susceptibility tests

Newly sequenced strains carrying antimicrobial resistant genes *ampC* and *tetA* were tested for susceptibility against ampicillin (10 µg) and tetracycline (30 µg) by disk diffusion method following Clinical & Laboratory Standard Institute (CLSI) (https://clsi.org) recommendations for fastidious organisms (document M45-A2) (49). The assay was performed on Mueller Hinton Agar plates supplemented with 5% defibrinated sheep blood (Thermo Fisher Scientific, Milan, Italy).
Determination of cadmium chloride and benzalkonium chloride resistance was performed as previously described (27). Briefly, a single colony from a blood agar plate culture was suspended in 100 µl of tryptic soy broth (Thermo Fisher Scientific). Three microliters of the suspension were spotted in duplicate onto: 1) Mueller Hinton Agar Cation adjusted (MHBII) containing 2% defibrinated sheep blood (Thermo Fisher Scientific) (control); 2) MHBII (Thermo Fisher Scientific) containing 2% defibrinated sheep blood (Thermo Fisher Scientific) spiked with 70 mg/l of cadmium chloride anhydrous (Sigma, Milan, Italy); 3) MHBII (Thermo Fisher Scientific) containing 2% defibrinated sheep blood (Thermo Fisher Scientific) spiked with 35 mg/l of cadmium chloride anhydrous (Sigma, Milan, Italy); 4) Mueller Hinton Agar (MHB) (Thermo Fisher Scientific) containing 2% defibrinated sheep blood (Thermo Fisher Scientific) spiked with 10 mg/l of benzalkonium chloride (Sigma). Positive and negative control strains were included. All plates were incubated at 37 °C for 48 h.

Crystal violet staining assay

In order to test the ability of biofilm formation, a representative set of ST121 and ST14 isolates of *L. monocytogenes* were submitted to the crystal violet staining assay as previously described (50). Briefly, 20 µl of standardized optical density (OD 0.8-1.0) bacterial suspension from overnight cultures were added to 230 µl of Brain Heart Infusion broth (BHI, Thermo Fisher Scientific) into each well of a sterile 96-well not tissue treated polystyrene microplate (Sarstedt, Milan Italy). The plates were incubated at 35 °C for 24 h. The content of the plate was discarded and 300 µl of sterile distilled water were added to each well. This washing step was repeated three times. Adherent bacteria were fixed with 250 µl of methanol per well. After 15 min, methanol was discarded and plates air-dried overnight. Biofilms were stained with 250 µl per well of Crystal violet (Gram-colour staining set for microscopy; Merck) for 5 min. Excess stain was rinsed off and microplates air-dried. Finally, attached bacteria were solubilized with 250 µl of 33% (v/v) glacial acetic acid per well and OD of each well was measured at 570 nm using Infinite® F50 Absorbance Microplate Reader (Tecan Group Ltd, Männedorf, Switzerland). Based on the OD detected by bacterial films, strains were classified into: no, weak, moderate or strong biofilm producers, as previously described (50). All isolates were tested in triplicate.

Results

De novo assembly

Thirty-four isolates of *L. monocytogenes* of ST14, ST121 and ST224 collected over one year of sampling in a rabbit meat processing plant from environment, meat cuts and meat products were Illumina sequenced and *de novo* assembled using the QA/QC INNUca pipeline. The *in silico* 7-loci MLST data provided at the last step of INNUca pipeline is consistent with ST data from previous study by De Cesare and colleagues (1). Extracted DNA had a yield ranging from 14.2 to 81.3 µg/mL and a 260/280 ratio values ranging from 1.81 to 1.88, proving the purity of extracted nucleotide sequences. The sizes of *de novo* assembled draft whole genomes ranged from 2,925,197 to 3,068,560

bp. Draft genomes were characterized by a low number of contigs ranging from 13 to 28 and high N50 values (from 456298 to 580604) with final covering values ranging from 42 to 187X (Table 1). According to these parameters, this study provides high-quality draft assemblies of *L. monocytogenes* isolates belonging to different ST-types.

Table 1. Sequence type (ST) and assembly statistic of the 34 sequenced L. monocytogenes isolates.

Isolate ID	Sequence Type (ST)	No. of contigs	N50	Genome size (bp)	Fold Coverage
LM1	14	13	579785	3022050	84,03
LM2	14	14	456298	3022428	41,78
LM3	14	13	580466	3022420	114,86
LM5	14	15	579913	3022168	57,83
LM8	14	14	580478	3023685	176,68
LM9	14	13	580478	3023308	121,54
LM10	14	14	580478	3023687	167,18
LM11	224	13	511384	2925197	159
LM15	14	13	580472	3022932	146,98
LM16	14	14	580454	2986892	131,2
LM17	14	14	580454	2986900	149,23
LM18	14	14	580454	2986900	187,12
LM19	14	15	580471	3023022	172,15
LM22	14	13	580454	3022659	79,03
LM27	121	26	529438	3060499	120,24
LM29	121	28	530140	3061007	131,41
LM31	121	25	530140	3060579	127,78
LM35	14	13	580472	3022414	106,18
LM39	14	15	580604	3022475	120,54
LM41	14	13	580472	3022475	126,39
LM44	14	13	580478	3022308	144,31
LM46	14	13	580478	3022405	85,42
LM47	14	13	580478	3022625	106,27
LM50	121	24	530364	3068560	111,58
LM51	121	25	530139	3060573	137,29
LM53	121	21	530354	3062194	139,38
LM54	14	13	580472	3022424	150,5
LM55	14	13	580478	3022427	142,63
LM56	14	13	580478	3022415	121,34
LM57	14	13	580478	3022426	134,73
LM58	14	13	580478	3022498	123,61
LM59	14	13	580478	3022821	135,9
LM60	14	13	580472	3022629	132,49
LM61	14	13	580478	3022408	108,5

Core genome MLST based phylogeny

Based on cgMLST analysis, ST14 isolates were classified as different Clonal Type (CT) (Fig. 1) based on the Institute Pasteur cgMLST schema, whereas some of these isolates were indistinguishable by *Apa*I-PFGE (Fig. 2), automated ribotyping and MLVA (1). The major cluster (green coloured in Fig. 1) gathered almost all putative persistent ST14 sequenced *L. monocytogenes* isolates which were classified as CT1701. Nevertheless, two ST14 isolates, LM15 and LM10, showing the same pulso-type (Fig. 2), have been included in two close singleton clusters and classified as CT1702 and CT1703. Out of the 1,748 loci of the cgMLST scheme, these two singletons carried 12 and 11 different loci respectively in comparison to CT1701 (data not shown). This finding suggests that a persistent one and two sporadic ST14 strains might be circulating in the rabbit-meat processing plant during the time of sampling. In contrast, all ST121 isolates were indistinguishable by cgMLST (Fig. 1), confirming the high similarity of these isolates as already suggested by molecular typing methods from previous work (1).

cgMLST tree

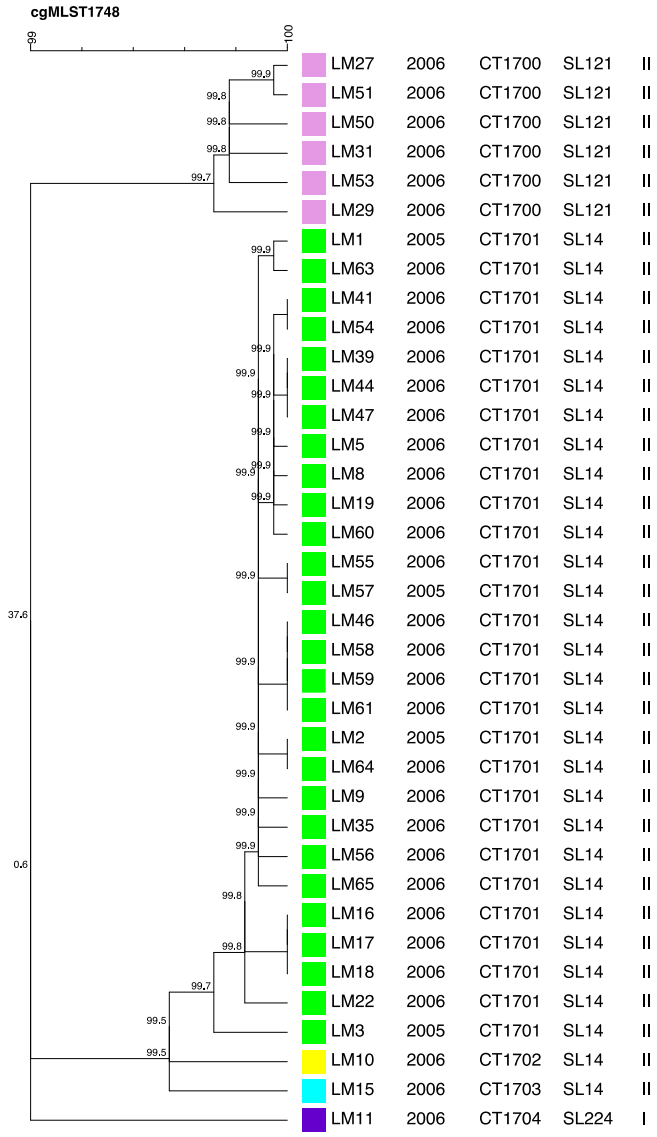

Figure 1. cgMLST based phylogenetic reconstruction of the 34 *Listeria monocytogenes* isolates sequenced in this study, including year of isolation, Clonal Type (CT), Sub-Lineage (SL) and lineage.

PFGE dendrogram

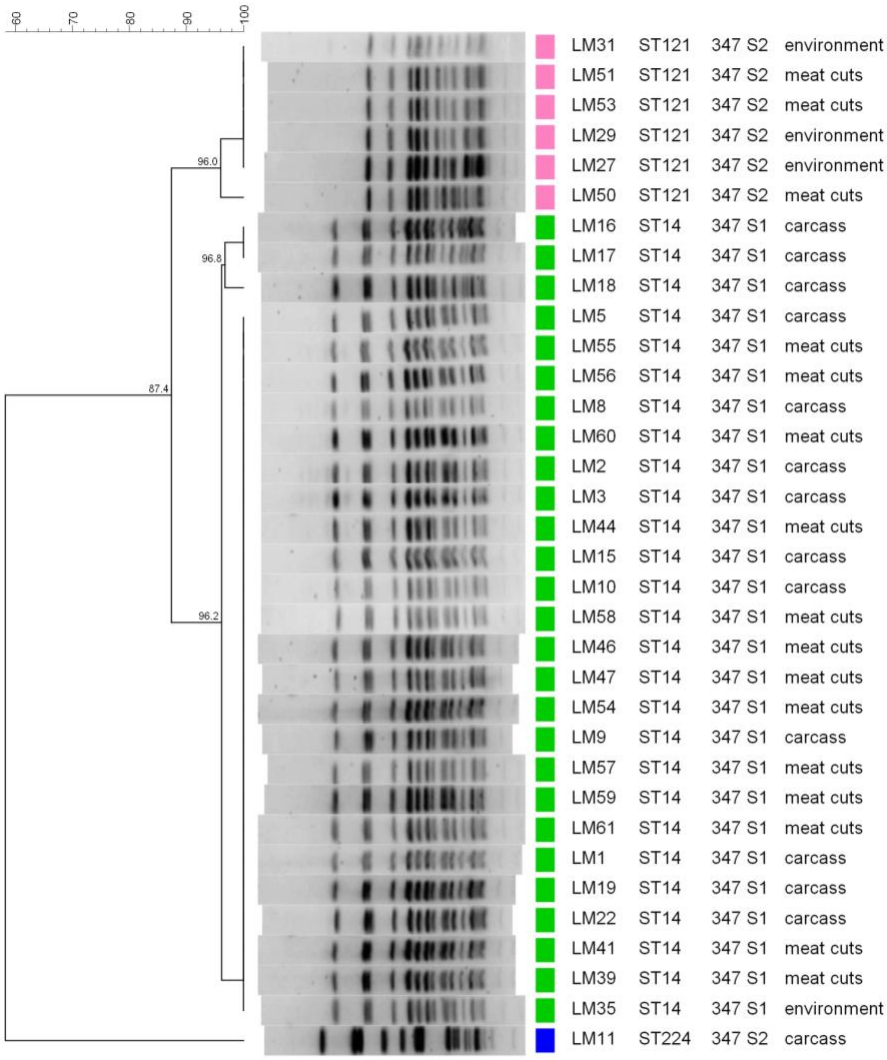

Figure 2. UPGMA based denrdogram relative to PFGE profiles of the 34 *L. monocytogenes* isolates sequenced in this study, including 7-loci MLST and *Eco*RI-ribotype from previous work, and source of isolation (1).

Statistical evaluation of typing methods discriminatory power

The cgMLST analysis showed a superior discriminatory power in comparison to PFGE, ribotyping, 7-loci MLST and MLVA based on Simpson's index of diversity (Simpson's ID) (Table 2), although statistically significant only in comparison to ribotyping (P=0.01).

Table 2. Simpson's ID of cgMLST analysis in comparison to molecular typing methods on the 34 newly sequenced *L. monocytogenes* isolates.

Typing method	**N profiles**	**Simpson's ID**	**CI (95%)**
cgMLST	5	0.439	(0.253-0.624)
Ribotyping	2	0.214	(0.045-0.383)
***Apa*I-PFGE**	3	0.348	(0.170-0.525)
MLST	3	0.348	(0.170-0.525)
MLVA	3	0.348	(0.170-0.525)

In silico characterization of virulence potential

In order to investigate the virulence potential of 33 ST14 and ST121 isolates, repeatedly isolated over one year and six months respectively, in the observed rabbit meat processing plant, MVLST was assessed. All sequenced ST14 isolates belonged to virulence type VT107, whereas all ST121 to VT94.
The *in silico* detection of 82 virulence-related genes was performed with VirulenceFinder web-service on the 33 newly sequenced genomes along with the reference *L. monocytogenes* genome EGD-e and 23 additional publicly available *L. monocytogenes* genomes of ST14 and ST121, isolated from humans or food processing environment, in order to investigate potential associations of genes presence/ absence with ST-types. The EGD-e reference genome and the ST14 and ST121 genomes carried 82, 80 and 80 virulence genes respectively, indicating that those genes are core virulence genes of *L. monocytogenes*. The genes *inlF* and *lmo2026* were not detected in the 56 ST14 and ST121 genomes analysed, whereas the *inlJ* gene was not found or found truncated. Moreover, all ST14 carried a full-length version of the a*ctA* gene (1920 nucleotides), which was detected truncated in all the ST121 genomes. The gene *ami* was detected in a truncated version in part of ST14 and ST121 genomes. A full version of the *prfA* gene was detected in all the newly sequenced genomes while *inlA* gene was found in full-length only in ST14 genomes. In contrast, all ST121 genomes showed a mutation in the the *inlA* gene leading to a premature stop codon (PMSC) predicting the translation of a truncated InlA protein of 492 aa (instead of 800 aa full-length InlA). No relevant differences in carriage of virulence genes were found in relation to food or human origin.

Phylogenetic trees based on SNPs analysis

To better understand the genetic relationship of genomes belonging to ST14 and ST121 strains isolated from environmental sources, humans and foods showing similar genetic content a SNPs analysis was performed for each ST-type. Assemblies of ST121 and ST14 sequenced isolates were mapped separately against *de novo* assemblies of LSALM51 and LSALM1 strains, respectively, using Snippy v 3.2. The maximum likelihood (ML) trees inferred based on a whole genome SNPs alignment for each ST-type are reported in Figures 1a and 1b. Form the ML tree inferred on ST14 genomes two major clade have been identified, a major one including all but two of the public ST14 isolates and a smaller one gathering together the 27 isolates from the Italian rabbit meat plant. Within this last clade, 23 out of 27 isolates collected from rabbit meat carcasses, meat cuts and products from November 2005 to November 2006, shared SNPs counts ranging from 0 to 25 SNPs in comparison to the reference genome LSALM1 and were considered as belonging to the same persistent clone. Three isolates (LSALM8, LSALM9, LSALM10), collected from rabbit carcasses from June to August 2006, showed 29 to 33 SNPs whereas one isolate (LSALM22), collected in November 2006 showed 359 SNPs (Fig. 3). Comparing LSALM1 to public ST14 draft genomes, pairwise SNPs differences ranged between 59 (environmental swab, USA) and 1183 (RTE product, USA).
The ML tree inferred on ST121 isolates differentiated genomes in two distinct clades (Fig. 3), one including only 6 isolates from fish and environmental samples showing a pairwise SNPs distances ranging from 1424 to 1584 in comparison to ST121 reference genome LSALM51. In contrast, the major clade gathers the 97% (196) of ST121 genomes including that isolated from the Italian rabbit meat plant. These last were clearly differentiated in two distinct clusters (Fig. 3) of two and four isolates respectively. The two isolates (LSALM50 and LSALM53) were collected from rabbit meat cuts in May 2006 and shared 185 and 195 SNPs. Whereas, the others four isolates counted from 0 to 24 SNPs and were collected over six months (December 2005 - May 2006) from the rabbit meat-processing environment as well as from meat cuts (Fig. 3). In comparison to LSALM51 reference genome, public ST121 draft genomes showed pairwise SNPs distances ranging from 45 (processing environment, Denmark) to 1042 (fish product, China) (Fig. 3). These results confirm that within the same STs *L. monocytogenes* genomes are characterized by high clonality even considering the wide temporal range of isolation of public genomes (1996-2017).

wgSNPs trees

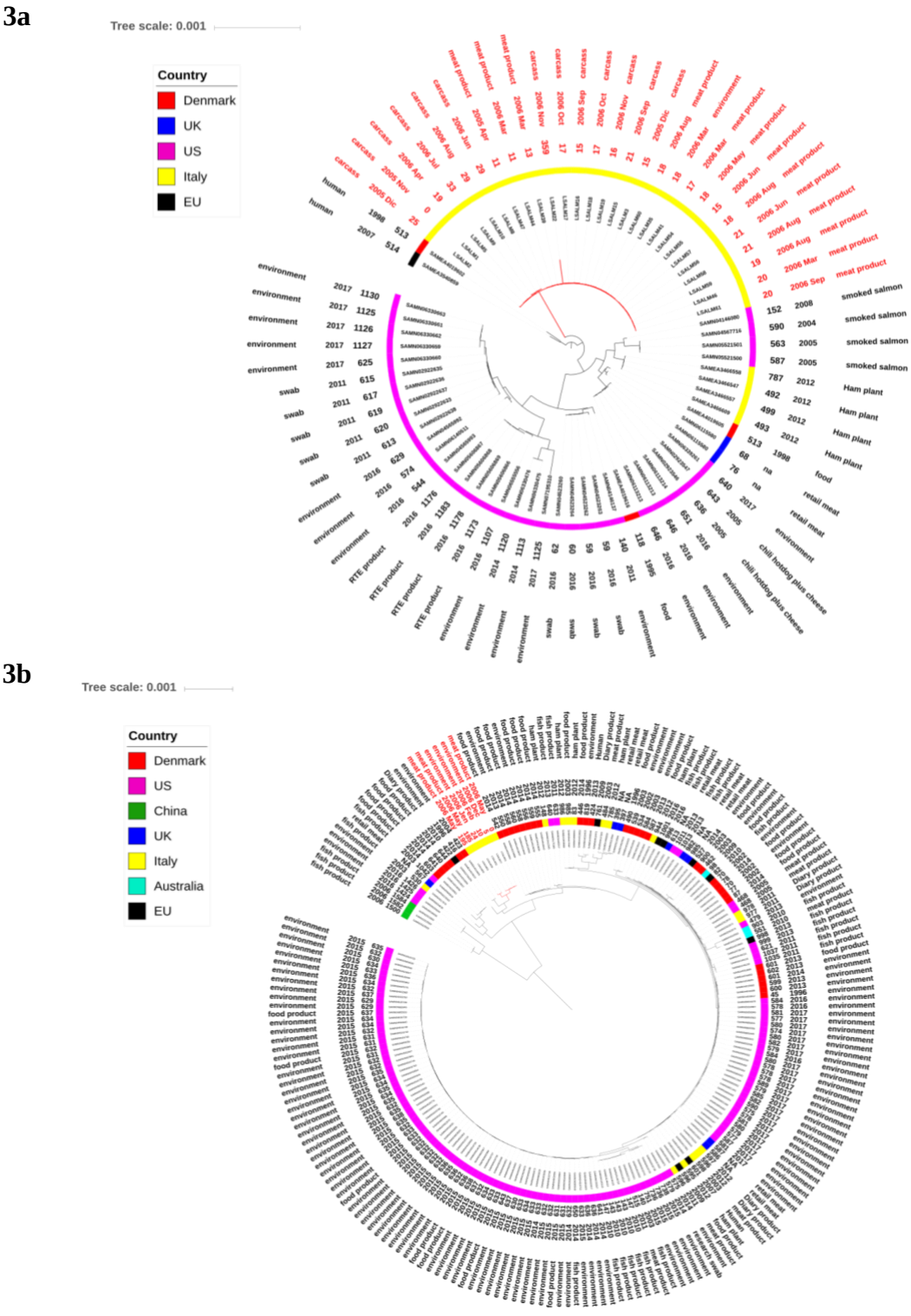

Figure 3. ML trees showing genetic distances between (**3a**) ST14 and (**3b**) ST121 *L. monocytogenes* isolates based on SNPs detected on the whole genomes. In both figures, the internal circles indicate the originating country of each isolates indicated by colours as in the legend. Externally, the pairwise SNPs distances in comparison to reference the genomes LSALM1 (**3a**) and LSALM51 (**3b**), collection date, source of isolation. The scale bar refers to the branch length representing the number of nucleotide substitutions per site.

Dataset of putative gene markers of ecophysiology in L. monocytogenes

Overall 94 genes, described in 41 published papers, were included in the dataset (Study1_Research2_Supplementary_Table_S1). All selected genes were identified as strongly associated to specific phenotypes related to physiological adaptation of *L. monocytogenes* to environmental stresses encountered in food processing plants. In particular the genes were associated to: resistance to antimicrobials, quaternary ammonium compounds, heavy metals and bacteriocins, adaptation to cold, high salt concentration, low pH, desiccation and biofilm formation (Study1_Research2_Supplementary_Table_S1). In particular, fourteen genes were included related to resistance to different antimicrobial classes such as tetracycline, ampicillin, vancomycin, streptomycin, chloramphenicol/ florfenicol, sulphonamides, erythromycin and fluoroquinolones (31, 49, 51–55). Seven genes associated to resistance to benzalkonium chloride were included (*qacH, qacA, qacC, bcrA, bcrB, bcrC* and *emrE*). These genes are associated to the active efflux pump of the QAC (22, 26, 48, 56, 57). Sixteen genes associated to resistance to cadmium and arsenic were also included. In particular, three gene cassettes, *cadA1C*, *cadA2C* and *cadA3C* were associated to resistance to 70 mg/l of cadmium chloride, whereas *cadA4C* was associated to resistance to 35 mg/l (58, 59). Regarding resistance to arsenic, *ars* genes were associated to resistance to 500 mg/l of sodium (meta) arsenite (60). Regarding bacteriocins resistance, 5 genes associated to the response of *L. monocytogenes* to cell-envelope stress were included: *virR*, *virS*, *mprF*, *lia*R, *anrB* (61–64) Resistance to bacteriocins such as nisin is associated to modification of the cell envelope composition. As far as stress adaptation is considered, different mutant selection experiments demonstrated the important role of the gene *sigB* (stressosome-regulated sigma factor B) in the modulation of expression of several genes associated to the adaptation to different environmental stresses. In particular the knockout of *sigB* was directly associated to adaptation to desiccation in *L. monocytogenes* (65).

As far as blue-light is considered, the gene *lmo079*9, coding for a blue-light receptor, was strongly associated to the adaptation of *L. monocytogenes* to this particular stress (66, 67).

L. monocytogenes can adapt to cold following different pathways. One pathway includes cold shock proteins cspB and cspD (68). Another pathway includes the glycine/ betaine transporter system which mediates the uptake of osmolytes such as glycine, betaine and carnitine, important for adaptation to both cold and high salt concentration (69). Overall 13 genes related to adaptation of *L. monocytogenes* to cold and/or high salt concentration were included in the dataset (68–71).

The Stress Survival Islet 1, corresponding to a cassette of five genes (*lmo0444-lmo0448*), was associated to the survival and growth of *L. monocytogenes* under suboptimal conditions. In particular, the knockout of the entire SSI-1 was associated to an impaired ability of this food-borne pathogen to grow at low pH and high salt concentrations (30). Within this gene cassette, *gadD1* (*lmo0447*) and *gadT1* (*lmo0448*) encode for a glutamate decarboxylase and an amino acid transporter both described as specifically involved in the adaptation to low pH (72). Along the GAD System, ADI system might be involved in response to low pH. The *arc* gene is involved in transformation of arginine into ornithine with ammonia as by product, which increases the pH. The ADI system has been described in response to mild acid pH (72). Overall, 16 genes associated to low pH adaptation were included in the dataset (30, 72–74). The Stress Survival Islet 2 (SSI-2) was firstly described as a cassette of two genes of *Listeria innocua* often present in place of SSI-1 in *L. monocytogenes* ST121 (24). SSI-2 was more recently associated to alkaline and oxidative stress in *L. monocytogenes* (23).

Regarding adaptation to desiccation, seven genes related to the motility of *L. monocytogenes*, were recently associated to this specific phenotype (75). In particular, these genes were found to be downregulated in desiccation tolerant *L. monocytogenes*.

As far as the ability of biofilm is concerned, 8 genes associated to biofilm formation were included in the dataset. Biofilm formation is essential for survival of *L. monocytogenes* and further contributes to bacterial persistence in the processing environment (76).

Screening of putative gene markers of ecophysiology and related phenotypic tests

Mass screening of 94 putative gene markers of ecophysiology was locally performed using abricate pipeline (https://github.com/tseemann/abricate) on 202 and 71 *L. monocytogenes* ST121 and ST14 draft genomes respectively (Fig. 4). All contigs were positive for AR related genes *ampC* and *tetA*, phenotypically associated to penicillin and tetracycline resistance respectively. However, phenotypic tests performed on isolates from Italian rabbit meat plant did not confirm genetic results. Although disk diffusion breakpoints are not available for *Listeria*, all isolates showed zone diameters equal or higher than 30 mm to tetracycline and ampicillin, suggesting susceptibility (data not shown). Further analysis should be performed in order to assess the reason behind these discordant results.

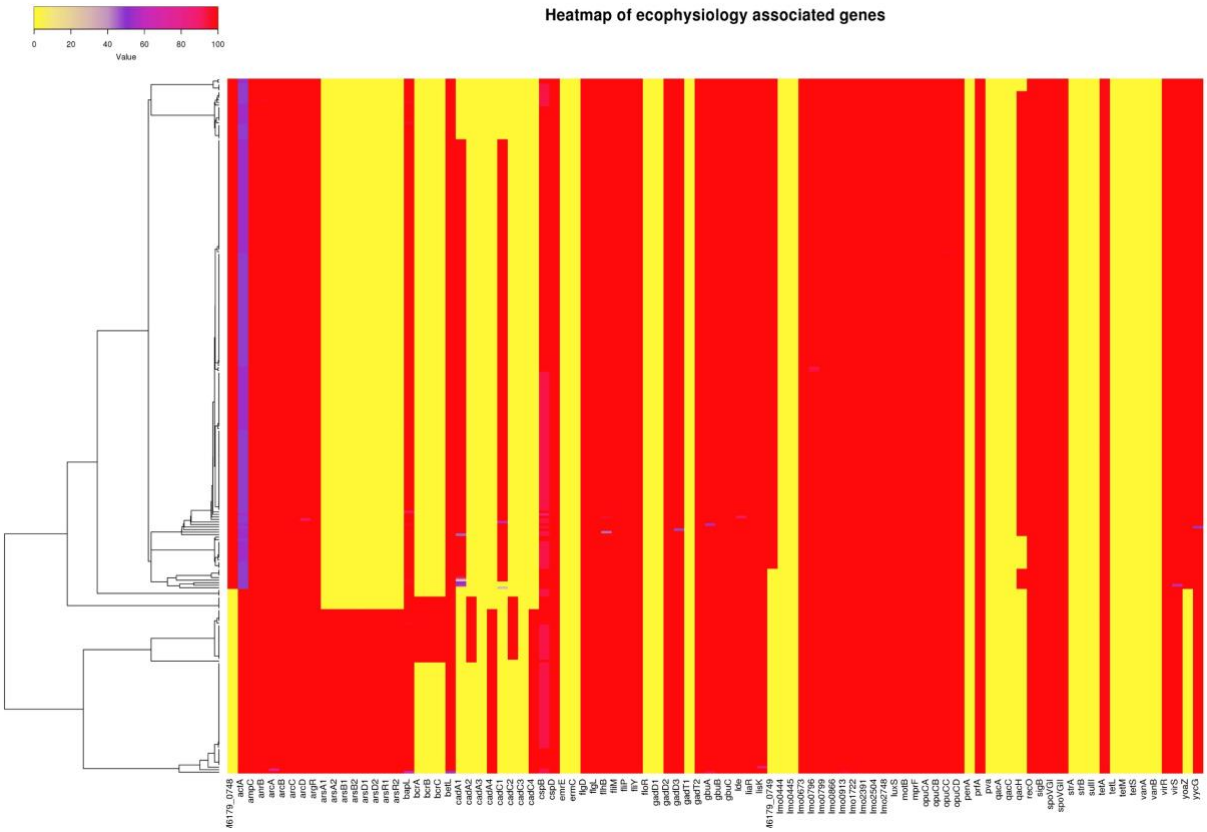

Figure 4. Heatmap of the 202 ST121 and 71 ST14 genomes function to the ecophysiology gene dataset. Genomes are represented by single linkage clustering based dendrogram on the left while at the bottom each gene of the dataset is reported under the corresponding column. The colours represent the gene coverage percentage with a minimum nucleotide identity of 80%. Yellow represents the lowest percentage (0) while red the highest one (100).

Regarding resistance to QAC, 25 of the ST14 (35.2 %) genomes and 184 (91.6 %) ST121 genomes were positive for the *bcrABC* and *qacH* gene respectively. All ST14 genomes but 8 (88.7 %), and none of the ST121 genomes were positive for the *ars* operon (arsenic resistance) and for the *cadA4C* gene cassette (cadmium chloride resistance). Further analyses on the localization of these two gene clusters within the genome, revealed that both the *ars* operon and the *cadA4C* gene cassette are located in a genomic region already identified as *Listeria* genomic Island 2 (LGI-2) (59, 60) (data not shown). In ST14 isolates included in the present study, LGI-2 was found in 24 out of 27 isolates and it was inserted within the *gtfA2* gene (data not shown). This gene codes for a sucrose phosphorylase involved in O-glycosylation of proteins. Glycosylation of flagellins is essential for bacterial flagellar

assembly, motility, virulence, and host specificity (77). The three isolates lacking LGI-2 (LSALM16, LSALM17, LSALM18) belonged to the persistent clone and were collected in September and November 2006 at the end of the sampling period. Further analyses should be performed in order to confirm whether these three isolates are natural mutants which lost the accessory genome sequence of SGI-2.
Twenty of *cadA4C* positive ST14 genomes additionally carried *cadA2C* gene cassette. The *cadA1C* gene cassettes was found only in ST121 and in particular in 177 out of the 202 tested genomes (87.6 %). Phenotypic tests conducted on the 33 *L. monocytogenes* strains from the rabbit meat-plant, confirmed that all *qacH* positive ST121 isolates were resistant to benzalkonium chloride and that all *cadA1C* positive ST121 isolates were resistant to 70 mg/l of cadmium chloride, whereas all *cadA4C* positive ST14 isolates were resistant to 35 mg/l of cadmium chloride.
As for environmental stress adaptation, no genomes were positive for the Stress Survival Islet 1 (SSI-1) and all ST121 but no ST14 were positive for the Stress Survival Islet 2 (SSI-2). All tested genomes carried a full-length version of genes associated to: cell-envelope stress response linked to bacteriocins resistance, adaptation to cold and/or high salt concentration, low pH, blue-light and desiccation.
Regarding biofilm formation, the *actA* gene was truncated in all ST121 and in none of the ST14 genomes. Phenotypic tests confirmed the differential biofilm forming ability of ST121 in comparison to ST14. In particular, tested ST121 isolates were classified as weak biofilm producers with ODC (OD Control) median value of 0.15 and OD values ranging from 0.17 to 0.24 OD, whereas all ST14 isolates but three (LSALM8, LSALM9, LSALM10) were categorized as moderate biofilm producers, with OD values ranging from 0.30 to 0.52 OD at 570 nm. OD values of ST121 isolates were statistically significant different in comparison to ST14 ones (P=0,00238).

Discussion

In the present study, the persistence of *L. monocytogenes* ST121 and ST14 repeatedly isolated within one year of sampling in a rabbit meat processing plant was investigated by different comparative genomic approaches.
Based on the WGS data, cgMLST typing, *in silico* detection of 82 virulence genes and *in silico* MVLST typing were performed in parallel. Moreover, pairwise SNPs distances based phylogenetic relationship and presence/ absence of 94 putative gene markers of *L. monocytogenes* physiological adaptation to the food processing environment stresses were investigated in 33 newly sequenced in comparison to 240 publicly available genomes of ST121 and ST14 strains. However, it must be noted that, compared to standard molecular techniques, WGS based approach requires greater bioinformatics skills, standardized protocols and validated pipelines, agreed cut off and parameter values in order to obtain robust and comparable results.
Based on PFGE-Typing, persistence of *L. monocytogenes* in dairy, meat, fish and vegetable sectors was extensively observed (4, 5, 28). However more recently, cgMLST and whole genome SNPs analysis revealed analogous superior discriminatory power in comparison to conventional typing techniques for *L. monocytogenes* investigations (78). This is particularly relevant in studies in which highly similar or clonal strains have to be differentiated in order to distinguish true persistent from sporadic strains (6). In this study, both cgMLST and SNPs based phylogenetic analysis revealed a superior discriminatory power in comparison to the molecular typing methods (*Apa*I-PFGE, 7-loci MLST, MLVA, ribotyping) applied on the same isolates in a previous study (1). Indeed, the cgMLST confirmed only in part results of molecular typing, suggesting that most but not all ST14 isolates, collected over one year sampling in the same rabbit-meat processing plant, belonged to the same persistent event. Two cgMLST clusters included only one isolate each, suggesting the detection of these two isolates as a sporadic event in the rabbit-meat processing plant. Accordingly, SNPs calling

revealed that the majority but not all of the ST14 isolates from the Italian rabbit meat plant sharing the same PFGE-typing, belonged to the same clone. This clone gathering 23 out of 27 isolates, was re-isolated more than six times within one year from rabbit carcasses, meat cuts, meat products and the processing environment in the same rabbit meat processing plant. Although an agreement has not been yet achieved on the definition of persistence, for the purposes of the present study, the ST14 clone was considered as persistent (6). This persistent clone included ST14 genomes sharing a maximum of 25 SNPs differences, a SNPs cut off already proposed for definition of genetically related strains belonging to a single *L. monocytogenes* persistent clone (48). The persistence of a *L. monocytogenes* strain for long periods endorses a higher risk of food contamination and human exposure to specific pathogen strains (79). The concern is even higher when the persistent clone belongs to subtype ST14, which was described as including hypervirulent strains (11, 16, 18, 80). Concerning ST121 isolates, SNPs analysis gathered genomes from Italian rabbit meat plant in two clusters of 2 and 4 isolates each, confirming its higher discriminatory power in comparison to molecular typing methods which identified all ST121 isolates as belonging to the same clone (1). The analysed dataset suggests that one ST121 clone gathering four isolates, survived over five months in the same rabbit meat processing plant. However, since these isolates were repeatedly collected in a relative short time frame, this clone cannot be considered as persistent.

Persistent *L. monocytogenes* isolates collected in food processing plants represent a public health concern, due to their potential transfer to humans via the food chain. In order to investigate the virulence potential of sequenced ST14 and ST121 *L. monocytogenes* isolates, MVLST and *in silico* virulo-typing were performed. All newly sequenced ST14 were classified as VT107 whose concatenated 7 virulence loci sequence differs to VT1 (Epidemic Clone III) for only 4 nucleotides, suggesting the high similarity of sequenced ST14 isolates to this epidemic clone (81). Interestingly, three years after sampling of the rabbit-meat processing plant, a case of invasive Listeriosis in the North of Italy associated to ST14 strain belonging to VT1 has been described (18). Eighty out of 82 virulence determinant genes were found in the analyses of 57 *L. monocytogenes* genomes. This finding suggests that the presence/ absence of these 82 genes can hardly explain the high diversity between ST14 and ST121 which have been shown to occur with a high and low frequency respectively in clinical samples (10,16). The *inlF* gene belonging to the internalin family, which includes genus *Listeria* exclusive genes associated to the adhesion and invasion of host cells (82) and the gene *lmo2026* (recently identified as internalin gene *inlL*), associated to biofilm formation and adhesion to mucin (76), were not detected. Besides presence/ absence of virulence genes the regulation of expression as well as the presence of truncated genes or the presence of mutations leading to truncated versions of their translated proteins might have a crucial role in the virulence potential of ST14 and ST121. Truncated virulence genes *inlJ*, *actA* and *ami* were found in ST14 and ST121 genomes included in this analysis. The *inlJ* gene belongs to the internalin family (82, 83) whereas the gene *ami* codes for an autolysin which contributes to the adhesion of *L. monocytogenes* to eukaryotic cells by anchoring its cell wall (84). The *actA* gene, found truncated only in ST121 genomes, encodes the surface protein ActA, the factor responsible for actin-based motility and cell-to-cell spread (85). Regarding point mutations, the pre-mature stop codon (PMSC) detected in all the ST121 genomes included has been already described as PMSC of type 6, a mutation specifically associated to attenuated virulence of *L. monocytogenes* (15). The attenuated virulence potential of sequenced ST121 isolates is also supported by the detection of a truncated version of the gene *actA*, which was described as indispensable for *L. monocytogenes* pathogenicity (85). An attenuated virulence of ST121 might explain the low frequency of its corresponding clonal complex CC in clinical samples (11).

In order to gain more insights on the genomic bases behind the differential frequency of ST121 and ST14 in food processing plants, a comprehensive literature review was performed in order to identify genes associated to physiological adaptation of *L. monocytogenes* to specific food-processing environmental stresses. In the literature, different dataset of genes associated to particular phenotypes, such as antimicrobial resistance and virulence of *L. monocytogenes* have been described (35, 36).

However, to the best of author's knowledges, no dataset on genes associated to ecophysiology of *L. monocytogenes* is available.
In the present study, 94 genes associated to ecophysiology were accurately selected based on a comprehensive literature review and gathered in a unique dataset available for the public (Study1_Research2_ Supplementary_Table_S2). Genes were included in the dataset only if their association to the specific phenotype was confirmed by insertional mutagenesis or deletion mutant experiments. In particular, the dataset includes genes associated to resistance to antimicrobials, quaternary ammonium compounds (QAC), heavy metals and bacteriocins as well as associated to adaptation to cold, high salt concentration, low pH, desiccation and biofilm formation. Mass screening of these genes on 202 ST121 and 71 ST14 *L. monocytogenes* genomes outlined interesting findings with particular reference to a significant enrichment of certain subtype-specific genes within each ST-type (Fig. 4). In particular genes *qacH*, *cadA1C*, Stress Survival Islet -2 and a truncated version of the *actA* gene were significantly enriched within the ST121 genomes ($P < 0{,}000001$), whereas the *ars* operon and *cadA4C* gene cassettes included in the Listeria Genomic Islands 2, the *bcrABC* locus and a full-length version of the *actA* gene were significantly enriched in ST14 genomes ($P < 0{,}000001$). The presence of a QAC associated genetic determinant (either the *qacH* gene or the *bcrABC* locus) was significantly enriched in ST121 genomes in comparison to ST14 ($P < 0{,}000001$). These observations underline that ST121 and ST14 subtypes have different patterns of genes associated to ecophysiology. Further studies should be performed to confirm whether other subtypes with high and low frequency show the same pattern of genes as identified in ST121 and ST14 subtypes from this study. Comparing the two patterns it appears that ST121 showed high adaptation to sanitizing procedures (resistance to QAC and adaptation to alkaline stress) (*qacH;* SSI-2) along with adaptation to high cadmium concentration (*cadA1C*) (23, 26, 58).
Concerning biofilm-related genes, a truncated version of *actA* gene was detected in all ST121 genomes whereas all ST14 harboured a full-length version of the same gene. This virulence associated gene is involved in the polymerization of actin, a multifunctional-protein important for the motility of *L. monocytogenes* within the host cell both in the first steps of biofilm formation and in cell-to-cell aggregation (85, 86). Since deletion mutants ΔactA have been associated to attenuated virulence and inability to form biofilms (86), phenotypic test were performed in this study providing evidence on the higher biofilm forming ability of ST14 in comparison to ST121. Thus, the presence of a full-length *actA* gene and the low adaptation potential of ST14 to sanitizing procedures might suggest the ability of ST14 to form biofilms in harbourage sites where sanitizing procedures are difficult to be performed (i.e. cutting equipment which do not allow disassembly). From this harbourage site, hypervirulent ST14 subtype might repeatedly contaminate meat cuts of different food lots over month or years and finally be responsible of outbreaks scattered in time and geographical areas (7).
A full-length version of genes associated to: cell-envelope stress response linked to bacteriocins resistance, adaptation to cold and/ or high salt concentration, low pH and desiccation was detected in all *L. monocytogenes* genomes included in this study, therefore, these genes were not considered as informative for differentiating ST121 and ST14 subtypes. Nevertheless, specific allele types can be identified for those genes which are mostly located in the *L. monocytogenes* core genome. Genome Wide Association Studies (GWAS) can be further performed to associate specific allele types to specific phenotypes. Allele types with a strong association to the phenotype could be then included in the dataset. The usefulness of GWAS was already demonstrated in a study on *Campylobacter jejuni*, where authors discovered that, based on overrepresented genetic elements, different *C. jejuni* subtypes show distinct genotypes associated with survival from farm to fork (87).

Conclusions

Although not yet fully standardized, comparative genomics approaches can be successfully applied for multiple investigations, such as enhanced subtyping and prediction of virulence. In the present study, cgMLST and wgSNPs based phylogenies revealed comparable higher discriminatory power in comparison to conventional molecular typing methods confirming that most but not all ST14 and all ST121 isolates belong to two persistent strains. Nevertheless, a wider representative geographically and temporally distributed set of ST14 and ST121 isolates fulfilling the persistent clone definition criteria should be analysed to provide stronger evidence on the persistence phenotype of specific *L. monocytogenes* ST-types in rabbit meat processing plants. Based on Multi Virulence Locus Sequence typing (MVLST) and *in silico* virulence typing a higher virulence potential of ST14 in comparison to ST121 newly sequenced isolates was predicted although additional *in vitro* investigations should be performed to confirm the virulence of ST14 and ST121 isolates.
Besides relative low frequency of clinical ST121 strains, the resistance-related gene enrichment found in ST121, supported by phenotypic confirmations, confirmed that this ST should be taken in account when performing sanitizing procedures within food processing environments. Moreover, heavy metals and disinfectants resistance ability might in part explain the high frequency of detection of this subtype in food processing plants. On the other hand, attention should be given to biofilm producer ST14 strains which might occasionally contaminate harbourage sites where sanitizing procedure are difficult to be performed and potentially spread to different food lots scattered in months or years. If confirmed, this route of contamination might be even of higher concern since the driver of this repeated event might be a hypervirulent strain.

Manuscripts submission and acknowledgements

1. Published manuscript in: *Italia journal of Food Safety*, 2017. doi: 10.4081/ijfs.2017.6879.

Title:

Whole genome sequencing for typing and characterization of Listeria monocytogenes isolated in a rabbit meat processing plant.

Authors:

Federica Palma[1], Frédérique Pasquali[1], Alex Lucchi[1], Alessandra De Cesare[1], Gerardo Manfreda[1].
[1]DISTAL, Department of Agricultural and Food Sciences, University of Bologna, Italy

Acknowledgments:

This study have been supported by COMPARE project (https://www.compare-europe.eu) co-funded by the *European Union's Horizon 2020 research and innovation programme* under grant agreement N° 643476. We acknowledge the team of curators of the Institut Pasteur cgMLST system (Paris, France) (http://bigsdb.pasteur.fr/listeria/) for assistance in CT assignments.
We gratefully acknowledge Prof. Mirko Rossi from the Department of Food Hygiene and Environmental Health, University of Helsinki (Finland) for technical and bioinformatics support, critical and helpful discussions. The authors wish to thank CSC- Tieteen tietotekniikan keskus Oy (Helsinki) and Department of Physics and Astronomy (Bologna) for providing access to cloud computing resources.

2. In Press manuscript: *Frontiers in microbiology*, 2018. doi: 10.3389/fmicb.2018.00596.

Title:

Listeria monocytogenes sequence types 121 and 14 repeatedly isolated within one year of sampling in a rabbit meat processing plant: persistence and ecophysiology.

Authors:

Frédérique Pasquali[1], Federica Palma[1], Laurent Guillier[2], Alex Lucchi[1], Alessandra De Cesare[1], Gerardo Manfreda[1].

[1]DISTAL, Department of Agricultural and Food Sciences, University of Bologna, Italy.
[2]ANSES, Maisons-Alfort Laboratory for Food Safety, University Paris-Est, Maisons-Alfort, France.

Acknowledgments:

This study has been supported by COMPARE project (https://www.compare-europe.eu) co-funded by the *European Union's Horizon 2020 research and innovation programme* under grant agreement N° 643476.
We would like to thank Prof. Mirko Rossi from the Department of Food Hygiene and Environmental Health, University of Helsinki (Finland) for technical and bioinformatics support, critical and helpful discussions. We thank Dr. Jani Halkilahti for bioinformatics support in collecting publicly available *Listeria monocytogenes* genomes. The authors wish to thank CSC- Tieteen tietotekniikan keskus Oy (Helsinki) and the Department of Physics and Astronomy (Bologna) for providing access to cloud computing resources.

References

1. De Cesare A, Parisi A, Mioni R, Comin D, Lucchi A, Manfreda G. 2017. Listeria monocytogenes Circulating in Rabbit Meat Products and Slaughterhouses in Italy: Prevalence Data and Comparison Among Typing Results. Foodborne Pathog Dis 14:167–176.
2. Gray MJ, Freitag NE, Boor KJ. 2006. How the Bacterial Pathogen Listeria monocytogenes Mediates the Switch from Environmental Dr. Jekyll to Pathogenic Mr. Hyde. Infect Immun 74:2505–2512.
3. Stasiewicz MJ, Oliver HF, Wiedmann M, den Bakker HC. 2015. Whole-Genome Sequencing Allows for Improved Identification of Persistent Listeria monocytogenes in Food-Associated Environments. Appl Environ Microbiol 81:6024–6037.
4. Véghová A, Minarovičová J, Koreňová J, Drahovská H, Kaclíková E. 2017. Prevalence and tracing of persistent Listeria monocytogenes strains in meat processing facility production chain. J Food Saf 37:n/a-n/a.
5. Leong D, Alvarez-Ordóñez A, Jordan K. 2014. Monitoring occurrence and persistence of Listeria monocytogenes in foods and food processing environments in the Republic of Ireland. Front Microbiol 5.
6. Ferreira V, Wiedmann M, Teixeira P, Stasiewicz MJ. 2014. Listeria monocytogenes persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. J Food Prot 77:150–170.

7. Tompkin RB. 2002. Control of Listeria monocytogenes in the food-processing environment. J Food Prot 65:709–725.
8. Carpentier B, Cerf O. 2011. Review — Persistence of Listeria monocytogenes in food industry equipment and premises. Int J Food Microbiol 145:1–8.
9. Ferreira V, Barbosa J, Stasiewicz M, Vongkamjan K, Moreno Switt A, Hogg T, Gibbs P, Teixeira P, Wiedmann M. 2011. Diverse geno- and phenotypes of persistent Listeria monocytogenes isolates from fermented meat sausage production facilities in Portugal. Appl Environ Microbiol 77:2701–2715.
10. Orsi RH, den Bakker HC, Wiedmann M. 2011. Listeria monocytogenes lineages: Genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol IJMM 301:79–96.
11. Maury MM, Tsai Y-H, Charlier C, Touchon M, Chenal-Francisque V, Leclercq A, Criscuolo A, Gaultier C, Roussel S, Brisabois A, Disson O, Rocha EPC, Brisse S, Lecuit M. 2016. Uncovering Listeria monocytogenes hypervirulence by harnessing its biodiversity. Nat Genet 48:308–313.
12. Beier S, Bertilsson S. 2013. Bacterial chitin degradation-mechanisms and ecophysiological strategies. Front Microbiol 4:149.
13. Olier M, Pierre F, Lemaître J-P, Divies C, Rousset A, Guzzo J. 2002. Assessment of the pathogenic potential of two Listeria monocytogenes human faecal carriage isolates. Microbiology 148:1855–1862.
14. Olier M, Pierre F, Rousseaux S, Lemaître J-P, Rousset A, Piveteau P, Guzzo J. 2003. Expression of truncated Internalin A is involved in impaired internalization of some Listeria monocytogenes isolates carried asymptomatically by humans. Infect Immun 71:1217–1224.
15. Van Stelten A, Simpson JM, Ward TJ, Nightingale KK. 2010. Revelation by Single-Nucleotide Polymorphism Genotyping That Mutations Leading to a Premature Stop Codon in inlA Are Common among Listeria monocytogenes Isolates from Ready-To-Eat Foods but Not Human Listeriosis Cases. Appl Environ Microbiol 76:2783–2790.
16. Palma F, Pasquali F, Lucchi A, Cesare AD, Manfreda G. 2017. Whole genome sequencing for typing and characterisation of Listeria monocytogenes isolated in a rabbit meat processing plant. Ital J Food Saf 6.
17. Morganti M, Scaltriti E, Cozzolino P, Bolzoni L, Casadei G, Pierantoni M, Foni E, Pongolini S. 2016. Processing-Dependent and Clonal Contamination Patterns of Listeria monocytogenes in the Cured Ham Food Chain Revealed by Genetic Analysis. Appl Environ Microbiol 82:822–831.
18. Mammina C, Parisi A, Guaita A, Aleo A, Bonura C, Nastasi A, Pontello M. 2013. Enhanced surveillance of invasive listeriosis in the Lombardy region, Italy, in the years 2006-2010 reveals major clones and an increase in serotype 1/2a. BMC Infect Dis 13:152.
19. Kathariou S. 2002. Listeria monocytogenes virulence and pathogenicity, a food safety perspective. J Food Prot 65:1811–1829.
20. Lebrun M, Audurier A, Cossart P. 1994. Plasmid-borne cadmium resistance genes in Listeria monocytogenes are similar to cadA and cadC of Staphylococcus aureus and are induced by cadmium. J Bacteriol 176:3040–3048.
21. Kathariou S. 2002. Listeria monocytogenes virulence and pathogenicity, a food safety perspective. J Food Prot 65:1811–1829.

22. Kovacevic J, Ziegler J, Wałecka-Zacharska E, Reimer A, Kitts DD, Gilmour MW. 2016. Tolerance of Listeria monocytogenes to Quaternary Ammonium Sanitizers Is Mediated by a Novel Efflux Pump Encoded by emrE. Appl Environ Microbiol 82:939–953.
23. Harter E, Wagner EM, Zaiser A, Halecker S, Wagner M, Rychli K. 2017. Stress Survival Islet 2, Predominantly Present in Listeria monocytogenes Strains of Sequence Type 121, Is Involved in the Alkaline and Oxidative Stress Responses. Appl Environ Microbiol 83.
24. Hein I, Klinger S, Dooms M, Flekna G, Stessl B, Leclercq A, Hill C, Allerberger F, Wagner M. 2011. Stress Survival Islet 1 (SSI-1) Survey in Listeria monocytogenes Reveals an Insert Common to Listeria innocua in Sequence Type 121 L. monocytogenes Strains. Appl Environ Microbiol 77:2169–2173.
25. Hingston P, Chen J, Dhillon BK, Laing C, Bertelli C, Gannon V, Tasara T, Allen K, Brinkman FSL, Truelstrup Hansen L, Wang S. 2017. Genotypes Associated with Listeria monocytogenes Isolates Displaying Impaired or Enhanced Tolerances to Cold, Salt, Acid, or Desiccation Stress. Front Microbiol 8:369.
26. Müller A, Rychli K, Muhterem-Uyar M, Zaiser A, Stessl B, Guinane CM, Cotter PD, Wagner M, Schmitz-Esser S. 2013. Tn6188 - a novel transposon in Listeria monocytogenes responsible for tolerance to benzalkonium chloride. PloS One 8:e76835.
27. Mullapudi S, Siletzky RM, Kathariou S. 2008. Heavy-metal and benzalkonium chloride resistance of Listeria monocytogenes isolates from the environment of turkey-processing plants. Appl Environ Microbiol 74:1464–1468.
28. Ortiz S, López-Alonso V, Rodríguez P, Martínez-Suárez JV. 2015. The Connection between Persistent, Disinfectant-Resistant Listeria monocytogenes Strains from Two Geographically Separate Iberian Pork Processing Plants: Evidence from Comparative Genome Analysis. Appl Environ Microbiol 82:308–317.
29. Ratani SS, Siletzky RM, Dutta V, Yildirim S, Osborne JA, Lin W, Hitchins AD, Ward TJ, Kathariou S. 2012. Heavy Metal and Disinfectant Resistance of Listeria monocytogenes from Foods and Food Processing Plants. Appl Environ Microbiol 78:6938–6945.
30. Ryan S, Begley M, Hill C, Gahan CGM. 2010. A five-gene stress survival islet (SSI-1) that contributes to the growth of Listeria monocytogenes in suboptimal conditions. J Appl Microbiol 109:984–995.
31. Srinivasan V, Nam HM, Nguyen LT, Tamilselvam B, Murinda SE, Oliver SP. 2005. Prevalence of antimicrobial resistance genes in Listeria monocytogenes isolated from dairy farms. Foodborne Pathog Dis 2:201–211.
32. Xu D, Nie Q, Wang W, Shi L, Yan H. 2016. Characterization of a transferable bcrABC and cadAC genes-harboring plasmid in Listeria monocytogenes strain isolated from food products of animal origin. Int J Food Microbiol 217:117–122.
33. Holch A, Webb K, Lukjancenko O, Ussery D, Rosenthal BM, Gram L. 2013. Genome sequencing identifies two nearly unchanged strains of persistent Listeria monocytogenes isolated at two different fish processing plants sampled 6 years apart. Appl Environ Microbiol 79:2944–2951.
34. Knudsen GM, Nielsen JB, Marvig RL, Ng Y, Worning P, Westh H, Gram L. 2017. Genome-wide-analyses of Listeria monocytogenes from food-processing plants reveal clonal diversity and date the emergence of persisting sequence types. Environ Microbiol Rep 9:428–440.

35. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, Pascale GD, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The Comprehensive Antibiotic Resistance Database. Antimicrob Agents Chemother 57:3348–3357.
36. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644.
37. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
38. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455–477.
39. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE 9:e112963.
40. Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, Björkman JT, Dallman T, Reimer A, Enouf V, Larsonneur E, Carleton H, Bracq-Dieye H, Katz LS, Jones L, Touchon M, Tourdjman M, Walker M, Stroika S, Cantinelli T, Chenal-Francisque V, Kucerova Z, Rocha EPC, Nadon C, Grant K, Nielsen EM, Pot B, Gerner-Smidt P, Lecuit M, Brisse S. 2016. Whole genome-based population biology and epidemiological surveillance of Listeria monocytogenes. Nat Microbiol 2:16185.
41. Zhang W, Jayarao BM, Knabel SJ. 2004. Multi-virulence-locus sequence typing of Listeria monocytogenes. Appl Environ Microbiol 70:913–920.
42. Kleinheinz KA, Joensen KG, Larsen MV. 2014. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4.
43. Hunter PR, Gaston MA. 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26:2465–2466.
44. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl 25:1754–1760.
45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl 25:2078–2079.
46. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio.
47. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242-245.
48. Nielsen EM, Björkman JT, Kiil K, Grant K, Dallman T, Painset A, Amar C, Roussel S, Guillier L, Félix B, Rotariu O, Perez-Reche FJ, Forbes K, Strachan N. 2017. 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.

49. Jamali H, Paydar M, Ismail S, Looi CY, Wong WF, Radmehr B, Abedini A. 2015. Prevalence, antimicrobial susceptibility and virulotyping of Listeria species and Listeria monocytogenes isolated from open-air fish markets. BMC Microbiol 15.
50. Stepanović S, Cirković I, Ranin L, Svabić-Vlahović M. 2004. Biofilm formation by Salmonella spp. and Listeria monocytogenes on plastic surface. Lett Appl Microbiol 38:428–432.
51. Charpentier E, Gerbaud G, Courvalin P. 1993. Characterization of a new class of tetracycline-resistance gene tet(S) in Listeria monocytogenes BM4210. Gene 131:27–34.
52. Godreuil S, Galimand M, Gerbaud G, Jacquet C, Courvalin P. 2003. Efflux pump Lde is associated with fluoroquinolone resistance in Listeria monocytogenes. Antimicrob Agents Chemother 47:704–708.
53. Lungu B, O'Bryan CA, Muthaiyan A, Milillo SR, Johnson MG, Crandall PG, Ricke SC. 2011. Listeria monocytogenes: antibiotic resistance in food production. Foodborne Pathog Dis 8:569–578.
54. Poyart-Salmeron C, Trieu-Cuot P, Carlier C, MacGowan A, McLauchlin J, Courvalin P. 1992. Genetic basis of tetracycline resistance in clinical isolates of Listeria monocytogenes. Antimicrob Agents Chemother 36:463–466.
55. Roberts MC, Facinelli B, Giovanetti E, Varaldo PE. 1996. Transferable erythromycin resistance in Listeria spp. isolated from food. Appl Environ Microbiol 62:269–270.
56. Elhanafi D, Dutta V, Kathariou S. 2010. Genetic characterization of plasmid-associated benzalkonium chloride resistance determinants in a Listeria monocytogenes strain from the 1998-1999 outbreak. Appl Environ Microbiol 76:8231–8238.
57. Xu D, Li Y, Zahid MSH, Yamasaki S, Shi L, Li J, Yan H. 2014. Benzalkonium chloride and heavy-metal tolerance in Listeria monocytogenes from retail foods. Int J Food Microbiol 190:24–30.
58. Mullapudi S, Siletzky RM, Kathariou S. 2010. Diverse cadmium resistance determinants in Listeria monocytogenes isolates from the turkey processing plant environment. Appl Environ Microbiol 76:627–630.
59. Parsons C, Lee S, Jayeola V, Kathariou S. 2017. Novel Cadmium Resistance Determinant in Listeria monocytogenes. Appl Environ Microbiol 83:e02580-16.
60. Lee S, Rakic-Martinez M, Graves LM, Ward TJ, Siletzky RM, Kathariou S. 2013. Genetic Determinants for Cadmium and Arsenic Resistance among Listeria monocytogenes Serotype 4b Isolates from Sporadic Human Listeriosis Patients. Appl Environ Microbiol 79:2471–2476.
61. Bergholz TM, Tang S, Wiedmann M, Boor KJ. 2013. Nisin Resistance of Listeria monocytogenes Is Increased by Exposure to Salt Stress and Is Mediated via LiaR. Appl Environ Microbiol 79:5682–5688.
62. Collins B, Curtis N, Cotter PD, Hill C, Ross RP. 2010. The ABC transporter AnrAB contributes to the innate resistance of Listeria monocytogenes to nisin, bacitracin, and various beta-lactam antibiotics. Antimicrob Agents Chemother 54:4416–4423.
63. Kang J, Wiedmann M, Boor KJ, Bergholz TM. 2015. VirR-Mediated Resistance of Listeria monocytogenes against Food Antimicrobials and Cross-Protection Induced by Exposure to Organic Acid Salts. Appl Environ Microbiol 81:4553–4562.

64. Thedieck K, Hain T, Mohamed W, Tindall BJ, Nimtz M, Chakraborty T, Wehland J, Jänsch L. 2006. The MprF protein is required for lysinylation of phospholipids in listerial membranes and confers resistance to cationic antimicrobial peptides (CAMPs) on Listeria monocytogenes. Mol Microbiol 62:1325–1339.
65. Huang Y, Ells TC, Truelstrup Hansen L. 2015. Role of sigB and osmolytes in desiccation survival of Listeria monocytogenes in simulated food soils on the surface of food grade stainless steel. Food Microbiol 46:443–451.
66. Ondrusch N, Kreft J. 2011. Blue and Red Light Modulates SigB-Dependent Gene Transcription, Swimming Motility and Invasiveness in Listeria monocytogenes. PLOS ONE 6:e16151.
67. O'Donoghue B, NicAogáin K, Bennett C, Conneely A, Tiensuu T, Johansson J, O'Byrne C. 2016. Blue-Light Inhibition of Listeria monocytogenes Growth Is Mediated by Reactive Oxygen Species and Is Influenced by σB and the Blue-Light Sensor Lmo0799. Appl Environ Microbiol 82:4017–4027.
68. Schmid B, Klumpp J, Raimann E, Loessner MJ, Stephan R, Tasara T. 2009. Role of Cold Shock Proteins in Growth of Listeria monocytogenes under Cold and Osmotic Stress Conditions. Appl Environ Microbiol 75:1621–1627.
69. Angelidis AS, Smith GM. 2003. Three transporters mediate uptake of glycine betaine and carnitine by Listeria monocytogenes in response to hyperosmotic stress. Appl Environ Microbiol 69:1013–1022.
70. Markkula A, Mattila M, Lindström M, Korkeala H. 2012. Genes encoding putative DEAD-box RNA helicases in Listeria monocytogenes EGD-e are needed for growth and motility at 3°C. Environ Microbiol 14:2223–2232.
71. Pöntinen A, Markkula A, Lindström M, Korkeala H. 2015. Two-Component-System Histidine Kinases Involved in Growth of Listeria monocytogenes EGD-e at Low Temperatures. Appl Environ Microbiol 81:3994–4004.
72. Feehily C, Finnerty A, Casey PG, Hill C, Gahan CGM, O'Byrne CP, Karatzas K-AG. 2014. Divergent Evolution of the Activity and Regulation of the Glutamate Decarboxylase Systems in Listeria monocytogenes EGD-e and 10403S: Roles in Virulence and Acid Tolerance. PLOS ONE 9:e112649.
73. Abram F, Starr E, Karatzas K a. G, Matlawska-Wasowska K, Boyd A, Wiedmann M, Boor KJ, Connally D, O'Byrne CP. 2008. Identification of components of the sigma B regulon in Listeria monocytogenes that contribute to acid and salt tolerance. Appl Environ Microbiol 74:6848–6858.
74. Ryan S, Begley M, Gahan CGM, Hill C. 2009. Molecular characterization of the arginine deiminase system in Listeria monocytogenes: regulation and role in acid tolerance. Environ Microbiol 11:432–445.
75. Hingston PA, Piercey MJ, Truelstrup Hansen L. 2015. Genes Associated with Desiccation and Osmotic Stress in Listeria monocytogenes as Revealed by Insertional Mutagenesis. Appl Environ Microbiol 81:5350–5362.
76. Popowska M, Krawczyk-Balska A, Ostrowski R, Desvaux M. 2017. InlL from Listeria monocytogenes Is Involved in Biofilm Formation and Adhesion to Mucin. Front Microbiol 8.
77. Merino S, Tomás JM. 2014. Gram-Negative Flagella Glycosylation. Int J Mol Sci 15:2840–2857.

78. Henri C, Leekitcharoenphon P, Carleton HA, Radomski N, Kaas RS, Mariet J-F, Felten A, Aarestrup FM, Gerner Smidt P, Roussel S, Guillier L, Mistou M-Y, Hendriksen RS. 2017. An Assessment of Different Genomic Approaches for Inferring Phylogeny of Listeria monocytogenes. Front Microbiol 8.
79. Lambertz ST, Ivarsson S, Lopez-Valladares G, Sidstedt M, Lindqvist R. 2013. Subtyping of Listeria monocytogenes isolates recovered from retail ready-to-eat foods, processing plants and listeriosis patients in Sweden 2010. Int J Food Microbiol 166:186–192.
80. Voronina OL, Kunda MS, Ryzhova NN, Aksenova EI, Semenov AN, Kurnaeva MA, Ananyina YV, Lunin VG, Gintsburg AL. 2015. [Regularities of the ubiquitous polyhostal microorganisms selection by the example of three taxa]. Mol Biol (Mosk) 49:430–441.
81. Murugesan L, Kucerova Z, Knabel SJ, LaBorde LF. 2015. Predominance and Distribution of a Persistent Listeria monocytogenes Clone in a Commercial Fresh Mushroom Processing Environment. J Food Prot 78:1988–1998.
82. Vázquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Domínguez-Bernal G, Goebel W, González-Zorn B, Wehland J, Kreft J. 2001. Listeria pathogenesis and molecular virulence determinants. Clin Microbiol Rev 14:584–640.
83. Sabet C, Toledo-Arana A, Personnic N, Lecuit M, Dubrac S, Poupel O, Gouin E, Nahori M-A, Cossart P, Bierne H. 2008. The Listeria monocytogenes virulence factor InlJ is specifically expressed in vivo and behaves as an adhesin. Infect Immun 76:1368–1378.
84. Milohanic E, Jonquières R, Cossart P, Berche P, Gaillard JL. 2001. The autolysin Ami contributes to the adhesion of Listeria monocytogenes to eukaryotic cells via its cell wall anchor. Mol Microbiol 39:1212–1224.
85. Smith GA, Portnoy DA. 1997. How the Listeria monocytogenes ActA protein converts actin polymerization into a motile force. Trends Microbiol 5:272–276.
86. Travier L, Guadagnini S, Gouin E, Dufour A, Chenal-Francisque V, Cossart P, Olivo-Marin J-C, Ghigo J-M, Disson O, Lecuit M. 2013. ActA promotes Listeria monocytogenes aggregation, intestinal colonization and carriage. PLoS Pathog 9:e1003131.
87. Yahara K, Méric G, Taylor AJ, de Vries SPW, Murray S, Pascoe B, Mageiros L, Torralbo A, Vidal A, Ridley A, Komukai S, Wimalarathna H, Cody AJ, Colles FM, McCarthy N, Harris D, Bray JE, Jolley KA, Maiden MCJ, Bentley SD, Parkhill J, Bayliss CD, Grant A, Maskell D, Didelot X, Kelly DJ, Sheppard SK. 2017. Genome-wide association of functional traits linked with Campylobacter jejuni survival from farm to fork. Environ Microbiol 19:361–380.

Study II: Salmonella enterica ser Typhimurium variant 4,[5],12:i:- enhanced surveillance

Salmonella enterica ser. Typhimurium monophasic variant 4,[5],12:i:- has been associated with foodborne epidemics worldwide and it appeared to be related to swine production in most of the country of isolation. However, the monomorphic nature of this serovar has, so far, hindered identification of the source due to expansion of clonal lineages in multiple hosts and food producing systems. Therefore, efforts for identifying the transmission sources of this serovar are steadily increasing public health concerns. Since geographically structured genetic signals can shape bacterial populations, identification of biogeographical markers in *S.* 1,4,[5],12:i:- genomes can contribute to improving traceability and source attribution investigations.

In this study, the phylogeographical structure and micro-evolution of 148 geographically and temporally related Italian *S.* 1,4,[5],12:i:- circulating in human and swine have been investigated in comparison to an extended contest of selected publicly available *S.* Typhimurium/ *S.* 1,4,[5],12:i:- strains collected in Italy and worldwide. An innovative workflow based on recently developed bioinformatics approaches was applied to untangle the phylogenetic structure and identify biogeographical genetic markers on a large-scale dataset of genetically related S. Typhimurium/ S. 1,4,[5],12:i:- genomes. The Italian isolates resulted as belonging to a large population of ~1,300 clonal *S.* Typhimurium/1,4,[5],12:i:- isolates collected worldwide in two decades showing up to 2.5% of allele differences. Moreover, highly supported monophyletic groups resulting from phylogenetic reconstruction suggested discrete geographical segregations which have had a strong impact on the gene content of a large SopE-containing prophage associated to isolates of Italian origin, as revealed by GWAS. These data suggest that geographical segregation of certain lineages may result in the acquisition of specific accessory genetic markers useful to improve identification of the source in ongoing epidemics.

Introduction

Salmonella enterica serovar Typhimurium with the antigenic formula 4,[5],12:i:- is considered a monophasic variant of *S.* Typhimurium (MVSTm) lacking the second phase flagellar antigen (1). MVSTm has recently emerged in food-borne epidemics of multi-drug resistance (MDR) strains responsible for several outbreaks in Europe (EU) (2) as well as in other continents (1). Since first time this serovar was detected as far back in 1997 (3), it has been repeatedly associated to humans and swine production, but also to environmental samples and other food-producing animals, such as avian and cattle (2, 4–8). The increasing spread of MVSTm in EU, the growing number of food-borne outbreak in recent years (6) and the difficulties in identifying the source due to the monomorphic nature of this serovar continue to be a public health concerns. The existence of at least two distinct clones (European and Spanish clone) emerged independently from ancestral *S.* Typhimurium strains has been described (2, 5, 9–11). Additionally, different antimicrobial resistance (AR) patterns have been associated to the two clones. The prevalence of simultaneous resistance to ampicillin, streptomycin/ spectinomycin, sulphonamides and tetracycline (R-type ASSuT) has been described in strains from EU clone (12), while in strains from the Spanish clone an additional resistance to chloramphenicol, gentamycin and trimethoprim has been reported (13).

It has been argued that traditional typing methods are not well suited to unravel the evolution dynamics of MVSTm population as well as the source attribution and epidemiology of this monomorphic bacterial pathogen (14). Moreover, the misclassification of this serovar due to the technically-demanding serotyping protocols and the evolution of multiple monophasic genotypes make tackling the phylogenetic differentiation of MVSTm from serovar Typhimurium more challenging (15, 16). On the other hand, as recently reviewed by Taboada and colleagues (17) currently developed large-scale genomic approaches based on core genome multi-locus sequence typing (cgMLST) and single nucleotide polymorphisms (SNPs) phylogenies showed a surprising potential for molecular subtyping of genetically closely related strains, at a range of different resolution levels and for multiple purposes. Recent studies have showed that combining core genome analysis with accessory genes pool analysis, such as pan-genome wide association studies (pan-GWAS), has improved understanding on evolutionary and phylogeographic patterns of several food-borne bacterial pathogens (18–21).

Geographical structure of bacterial population is well documented for several pathogens such as *Mycobacterium tuberculosis* (22) and *Helicobacter pylori* (23). However, for foodborne pathogens the local phylogeographical signals can be deteriorated by the rapid movement of lineages across the globe due to international trade of food and animals, and human travelling. Nevertheless, the adaptation of certain lineages to specific hosts or food production systems which are more relevant on a local geographical scale, can result in the expansion of successful epidemic clones harbouring unique gene clusters that constitute specific biomarkers to improve source attribution in strains circulating in different countries. Therefore, in this study the phylogeographical structure of a set of 148 geographically and temporally related MVSTm isolates collected in Italy between 2012 to 2014 from human and swine was investigated in an extended contest of selected publicly available *S.* Typhimurium/ MVSTm strains from several countries.

Combining phylogenetic analysis and genome-wide association study we identified strong evidence of the phylogeographical structure of the Italian MVSTm isolates, identifying a specific SopEϕ-like phage as geographical biomarker for most these isolates.

Materials and methods

Bacterial strains and genome sequencing

A total of 148 *Salmonella enterica* serovar Typhimurium variant 4,[5],12:i:- (MVSTm) have been collected from different Italian regions between 2012 and 2014 during a surveillance study. For the aim of this study, the database has been named STY. Pig faecal samples (11), pork carcass isolates (23) and pork meat at retail isolates (27) were obtained from the Italian National Reference Laboratory for Salmonella (NRL *Salmonella*, Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro, Italy) while isolates from human with gastroenteritis (82) were obtained from the National Institute of Health (Instituto Superiore della Sanitá, Roma). Genomic DNA of the 148 STY isolates was extracted and purified using the HWD DNA minikit (QIAGEN) according to the manufacturer's instruction. Index-tagged paired-end Illumina™ sequencing libraries were prepared using NexteraXT™ library preparation kit and whole genome sequencing was performed on Illumina Miseq™ platform generating tagged 250 bp paired-end reads.

De novo assembly and *in silico* MLST

The paired-end raw reads were assembled using the INNUca pipeline (https://github.com/INNUENDOCON/INNUca), which consists of several modules and QA/ QC steps. In brief, INNUca starts by calculating if the sample raw data fulfil the expected coverage (min 15x). After subjecting reads to quality analysis using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and cleaning with Trimmomatic (24), INNUca proceeds to *de novo* draft genome assembly with SPAdes (25) 3.11 and checking assembly depth of coverage (min 30x). Finally, Pilon (26) improves the draft genome by correcting bases, fixing misassembles, and filling gaps, prior species confirmation and seven genes MLST Sequence Type (ST) is assigned with mlst software (https://github.com/tseemann/mlst).

Salmonella reference genomic dataset

A reference dataset of 4,312 public available draft or complete genome assemblies and available metadata of *Salmonella enterica* have been downloaded from public repositories (i.e. EnteroBase - https://enterobase.warwick.ac.uk/, National Center for Biotechnology Information NCBI - https://www.ncbi.nlm.nih.gov/ and The European Bioinformatics Institute EMBL-EBI - https://www.ebi.ac.uk/; accessed April 2017). The large collection includes 1,465 *Salmonella enterica* ser. Enteritidis, 1,425 ser. Typhimurium, 985 MVSTm, and 437 *Salmonella enterica* genomes of others frequently isolated serovars in Europe (6). All available assemblies for MVSTm have been chosen based on public metadata. For each of the other serovars, genomes have been selected to maintain the same proportions of genetic diversity (based on rMLST) as existing in EnteroBase at the date of collection (April 2017) and reflecting EFSA monitoring zoonosis activities report data on incidence percentages of different *Salmonella enterica* serovars (6).

Whole genome MLST (wgMLST) schema creation, validation and allele calling

Population structure analysis of *Salmonella* genomes have been performed using wgMLST methodology (27). Schema curation, validation and allele calling have been carried out using the chewBBACA (28) suite (https://github.com/B-UMMI/chewBBACA). Briefly, the wgMLST schema V2 from EnteroBase, including 21,064 loci, have been downloaded and curated using *chewBBACA AutoAlleleCDSCuration* for removing all alleles that are not coding sequences (CDS). The quality of the remain 21,032 loci have been assessed using *chewBBACA Schema Evaluation* (28) and loci with single alleles, those with high length variability (i.e. if more than 1 allele is outside the mode +/- 0.05 size) and those present in less than 0.5% of the *Salmonella* genomes in EnteroBase at the date of the

analysis (~81,000 genomes April 2017) have been removed. A total of 9,127 loci have been used as schema for calling alleles in the 4,460 Salmonella genomes using *chewBBACA Allele Calling* engine. The wgMLST schema have been further curated, excluding all those loci detected as "Repeated Loci" and loci annotated as "non-informative paralogous hit (NIPH/ NIPHEM)" or "Allele Larger/ Smaller than length mode (ALM/ ASM)" by the chewBBACA in more than 1% of the dataset. The final wgMLST schema included a total of 8,558 loci. The core genome MLST profile (cgMLST - defined as the loci presence in at least the 99% of the samples) has been extracted using *chewBBACA ExtractCgMLST* and goeBURST algorithm (29) implemented in Phyloviz 2.0 (30). cgMLST was used to calculate the globally optimal genomes clusters. Concordance between partitions obtained at different goeBURST cut-off and serotyping have been evaluated using Adjusted Wallace Coefficient (AWC) (31).

Genome annotation, pangenome analyses and genome-wide association studies (GWAS)

Pangenome analysis was performed on a set of samples based on the goeBURST clustering. *Salmonella* genomes were annotated with Prokka (32) (https://github.com/tseemann/prokka) and the produced GFF3 files were used to generate the pan-genome matrix with Roary (33) (https://github.com/sanger-pathogens/Roary) using default parameters. Maximum likelihood tree based on the binary matrix of presence and absence of accessory genes was constructed using IQtree (34) (https://github.com/Cibiv/IQ-TREE). The resulting tree was visualized on iTOL (35) (https://itol.embl.de) along with metadata information to individuate clusters of closely related isolates sharing core and accessory genes. A GWAS was performed based on Roary (33) results using Scoary v 1.6.16 (36) (https://github.com/AdmiralenOla/Scoary). Patterns of genes were reported as significantly associated to geographical origin (e.g. Italy) if they attained Benjiamini-Hochberg-corrected P-value less than 0.05 and were present in at least the 20% of the selected isolates (e.g. Italian) and absent in at least the 70% of the rest of the dataset (e.g. non-Italian isolates). The synteny of the associated loci were visually assessed using Artemis (37) annotation tool on a selection of STY isolates and further manually annotated. If the associated genes were annotated as hypothetical protein the gene was manually curated by searching homologs sequencing in non-redundant (nr) NCBI protein sequences collection using blast+ v 2.7.1 (38) (https://blast.ncbi.nlm.nih.gov/).

In silico antimicrobial resistance, plasmid and phage typing

Antibiotic resistance and plasmid prediction was performed with abricate pipeline (https://github.com/tseemann/abricate) using ResFinder (39) and PlasmidFinder (40) as reference database, respectively. The typical AR profile of *S.* 4,[5],12:i:- "EU clone" was defined as the simultaneous presence of bla_{TEM-1}, *strA* (and its synonymous *aph(3")-Ib*), *strB* (and its synonymous *aph(6)-Id*), *sul1*/ *sul2*/ *aad12* and *tet(B)* genes (R-type ASSuT) (12). In addition, when *cmlA1*, *aac(3)-IV* and *dfrA12* genes were detected in that isolates harbouring ASSuT related genes they were predicted as distinct R-type ASSuTCGTp, a specific pattern associated to S. 4,[5],12:i:- from the "Spanish clone" (13).
The presence of pSLT-genes encoding virulence factor in a 94-kb plasmid (AE006471) from *S.* Typhimurium LT2 was investigated using blastn implemented in blast+ (38) V 2.7.1. The PHAge Search Tool (PHAST) (41) was used to identify the positions of putative phage elements. For PHAST analysis, genomes have been annotated using RAST annotation server (42). Hence, the annotated GBK file was uploaded to the public PHAST web server (http://phast.wishartlab.com/) (41).

Single-nucleotide polymorphism (SNP) analysis

To establish the phylogenetic relationship between closely related strains based on the goeBURST clustering, SNP analysis have been performed using Snippy v 2.5 (https://github.com/tseemann/snippy) pipeline using the assembled genomes as input files. As

reference, the best assembled draft genome (based on N50 values and coverage) harbouring the larger set of geographical associated genes have been selected within the same goeBURST cluster. A core alignment of all the conserved nucleotide variant sites present in all genomes was used to build a maximum-likelihood tree using IQ-tree (34) with 1,000 bootstrap replicates to support the nodes. hierBAPS (43) was used for clustering the samples based on the core SNP alignment and iTOL (https://itol.embl.de) to display phylogenetic trees.

Data availability

Genome assemblies are accessible at the European Nucleotide Archive (https://www.ebi.ac.uk/ena) under the project accession number: PRJEB23875.
List of the genome EnteroBase accession number are available in Supplementary Information. Schema and allele profile of the genomes used in this study are available in https://github.com/INNUENDOCON/chewBBACA_schemas.

Results

Quality of *de novo* assembly

All the draft genome sequences from 148 Italian STY MVSTm isolates originating from human and swine passed the QA/QC measures as defined in INNUca pipeline (https://github.com/INNUENDOCON/INNUca). *In silico* MLST classified 142 (96%) samples as Sequence Type (ST)34, 3 (2%) as ST19, 1 (0.7%) as ST11 and 1 (0.7%) as ST1995. For one isolate MLST ST was not found.
Supplementary materials are accessible in a dedicated GitHub repository under the following link: https://github.com/fedex88/PhD_Thesis. Spreadsheets have been uploaded in PhD Thesis under file names ´Study2_Supplementary_Table_S1´ and ´Study1 _Supplementary _Table_S2´.

Population structure *Salmonella* genomes

Population structure analysis has been performed using core genome gene-by-gene approach using the chewBBACA (28) suite (https://github.com/B-UMMI/chewBBACA), to untangle the geospatial evolution of the 148 Italian STY MVSTm isolates in the context of a representative set (4,312) of publicly available *Salmonella enterica* genomes including the most common serovars. A total of 3,255 out of the 8,558 loci in the wgMLST schema have been detected in > 99% of the samples and used for the cgMLST study. No concordance has been found between serotyping and any goeBURST clusters based on the 3,255 loci cgMLST schema (Adjusted Wallace Coefficient (AWC) < 0.6). However, considering serovar 1,4,[5],12:i:- as Typhimurium, concordance between cgMLST clustering and serotyping have been found at ~30% (965) of allele differences (bidirectional AWC > 0.97), including 35 groups. At 965 cut-off, 141 out of 148 STY MVSTm isolates belong to a single group along with 2,595 genomes including the majority of *S.* Typhimurium and other MVSTm public available strains. Besides, a large part of the STY MVSTm isolates (136 out of 148; ~92%) cluster in a single goeBURST group at ~2.5% (75) of allele differences (goeBURST^{75} 1) along with a mixed population of 241 *S.* Typhimurium and 912 MVSTm publicly available genomes (Study2_Supplementary_Table_S1 in GitHub repository https://github.com/fedex88/PhD_Thesis). The MVSTm strains were isolated between 2001 and 2017 from human (52%), swine (14%) and other sources (22%) (124 environment, 50 avian and 24 cattle). For 111 isolates no source of isolation was available. Almost the 94% of these strains were collected in Western Europe (433) and North America (427) while the remaining isolates were from, Northern Europe (35), Southern Europe (10), Asia (4) and Eastern Europe (1). Public available MVSTm belong to ST34 (870), ST19 (25) and

ST2379 (12), ST2956 (1), ST3168 (1) and ST3224 (1). Most *S.* Typhimurium genomes (227) were classified as ST34 of which more than half (115) were human isolates mainly from North America and Western Europe. For "goeBURST[75] 1", the Minimum Spanning Tree (MST) based on a new cgMLST schema, including a total of 3,591 loci, has been calculated using Phyloviz 2.0 (30) and the country of origin of the strains have been visualized on the tree (Fig. 1). Figure 1 showed a partial enrichment of geographical linked strains in certain parts of the MST (e.g. Italian cluster, yellow circle).

cgMLST tree

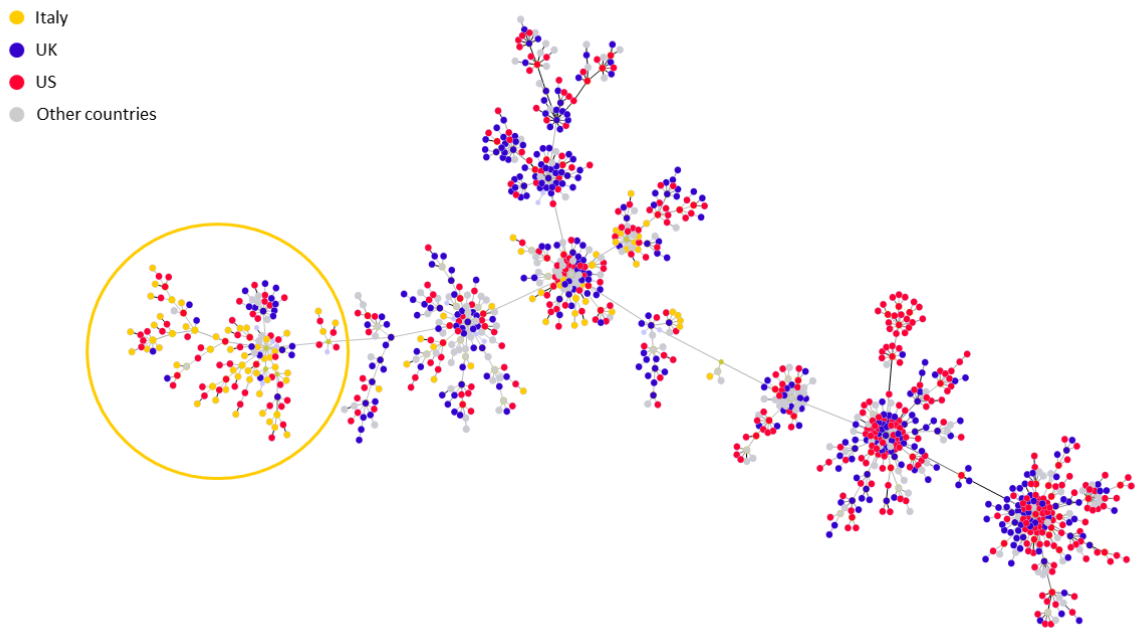

Figure 1: Minimum Spanning Tree of 1,326 genomes based on a 3,591 loci cgMLST schema calculated using Phyloviz 2.0 (30). The country of origin of the strains have been visualized on the tree by colours (Italian cluster, yellow circle).

Pangenome analysis

Pangenome analysis has been performed using Roary (33) on a total of 1,326 genomes comprising all 1,289 genomes belonging to "goeBURST[75] 1" (including the 136 STY MVSTm isolates) and an outgroup composed by a set of 38 *S.* Typhimurium/MVSTm strains, including 27 Italian MVSTm, 19 of which isolated during a large outbreak in south Italy (44). The pangenome consist in a matrix of 13,135 group of orthologues of which 9,085 are accessory (present in < 99% of genomes): 8,588 present in less than 15% of strains, 333 present in more than 15% but less than 95% of strains, and 164 genes present in more than 95% of strains but less than 99%. A maximum-likelihood tree was inferred based on binary data of presence/absence of accessory gene using IQtree (34) (Fig. 2). Two major clusters have been identified: (a) including 1,264 strains mainly of ST34, and (b) containing 62 strains mainly of ST19.

Within the large ST34 cluster, *S.* Typhimurium/MVSTm strains are aggregated in clades irrespectively of the year as well as of the source of isolation. However, even if isolates originating from different countries are gathered together across big clades, several small clusters including isolates of the same geographical area could be visually identified across the tree. In particular, the clade named herein STY-clade (Fig. 2) gathered roughly the 46% of the Italian STY MVSTm strains representing more than 71% of the isolates included in the clade. The clade is populated by 98 isolates,

70 (71%) of which were Italian, 24 Western European and 4 North American, collected between 2007 and 2017 and obtained from human, swine, and cattle. These data suggest that as result of geographical segregation, certain lineages might have acquired specific accessory genetic markers, as a result of geographical segregation.

Pangenome tree

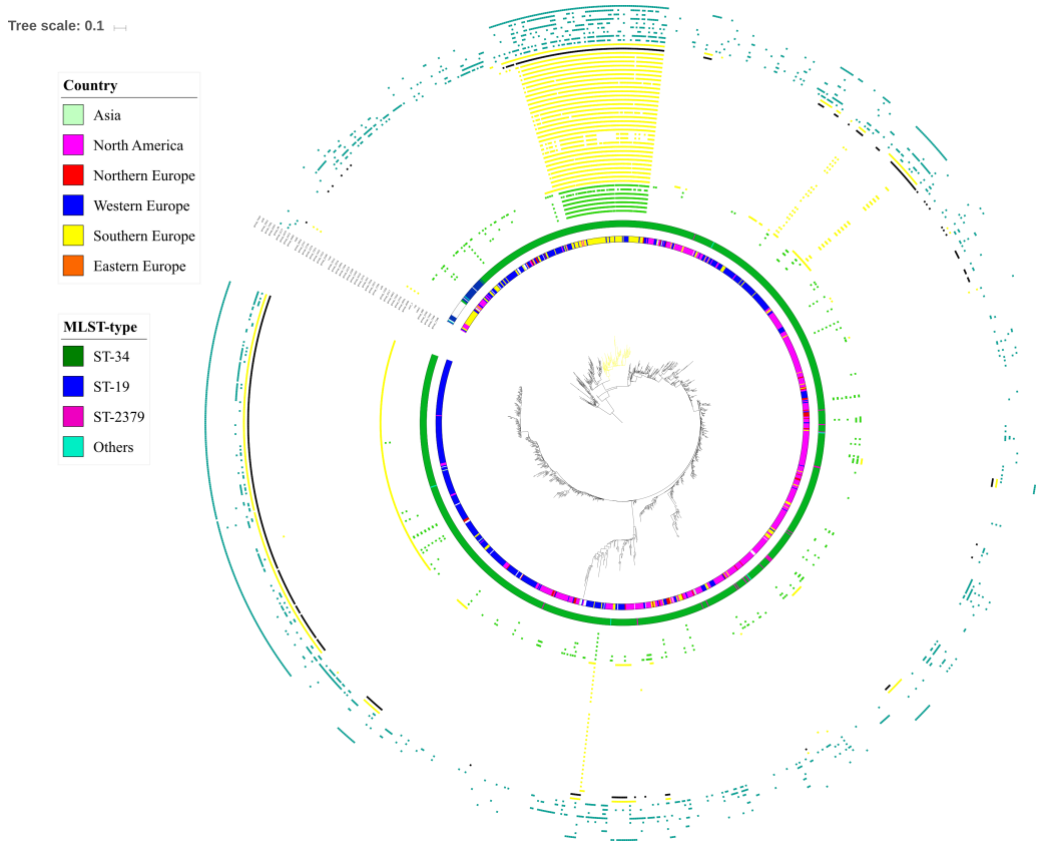

Figure 2. Maximum likelihood inferred based on binary data of presence/ absence of accessory gene of 1,326 *S.* Typhimurium/ MVSTm strains included in the study. Isolates from Asia, North America, Northern Europe Western Europe, Southern Europe and Eastern Europe are indicated by colour as reported in the legend in the first circle. The second circle indicate ST-type by colour as reported in the legend. Externally, clusters of genes statistically associated with Italian strains are labelled as from pangenome analysis and divided by colours in plasmid-related contiguous loci (green); prophage related contiguous loci (yellow); and associated loci spread across the genome (light blue). The black hits are indicating *sopE* gene presence while missing data are in white. The yellow coloured branch is indicating STY-clade.

In silico characterization of the isolates

A comprehensive list of the plasmids detected in the 1,326 genomes is available in Github repository under file name Study2_Supplementary_Table_S2 (https://github.com/fedex88/PhD_Thesis). The presence of plasmids has been detected in most of the genomes (1214; 91.5%) based on the positive match against PlasmidFinder database (40). The most frequently reported plasmid was IncQ1 (959:72.3%) followed by ColRNAI (506; 38,2%), Col156 (244; 18,4%). In total, the simultaneous presence of these three plasmids was observed in 77 ST34 strains of which ~48% (37) clusters within the above described STY-clade. Overall, less than 6% of isolates were positive for further plasmid

including several incompatibility groups Inc. Although none of the tested strains harboured the complete 94kb virulent pSLT plasmid, remains of the plasmid have been detected exclusively in the ST19 clade which includes 21 MVSTm isolates possessing from 4 to 32 of the pSLT CDSs, including the virulent markers *spvB* and *spvC* (45). All the MVSTm outbreak isolates characterized by Cito and colleagues (44) and 2 out 6 STY MVSTm isolates clustered within ST19 clade does not possess the pSLT plasmid. In contrast, the other 4 STY MVSTm within ST19 clade isolates possess from 4 to 30 pSLT CDSs.

Almost all genomes (1,279; 96.4%) are positive for at least one antimicrobial resistance associated gene (ARG) and most of them (1,011; 76.2%) possess three or more ARGs, while a limited number (3.6%) did not have any positive match in Resfinder database (39) (Study2_Supplementary_Table_S3 in the GitHub repository at https://github.com/fedex88/PhD_Thesis). The most prevalent ARGs were related to resistance to tetracycline (89,7%), sulphonamides (76,6%), ampicillin (74%), streptomycin (75,6%). More specifically, the simultaneous presence of genes for resistance to ampicillin (bla_{TEM-1B}), streptomycin (*strA*, *strB*, or *aph(3'')-Ib*, *aph(6)-Id*), sulphonamide (*sul1*, *sul2* or *aad12*) and tetracycline (*tet(A)* or *tet(B)*) predicting the ASSuT resistotype (R-type) and characterizing the so called European clone have been found in 67% of the positive isolates which were originated from human and swine sources and collected between 2004 and 2017. Particularly, a total of 105 over 167 isolates of Italian origin (~83% of the genomes within the STY-clade) exhibit R-type ASSuT. Only 18 over 1,326 genomes were classified as R-type ASSuTCGTp due to the simultaneous presence of genes for resistance to gentamycin (*aac(3)-IVa*), trimethoprim-sulfamethoxazole (*dfrA12*), and chloramphenicol (*cmlA1*). These isolates, predicted to harbour the R-type typical of MVSTm described as part of the Spanish clone, were collected from human source and interspersed among clusters including isolates from different sources and with R-type ASSuT.

Colistin resistance related genes *mcr-1*, *mcr-3*, *mcr-4* or *mcr-5* were revealed in only 10 of the ST34 genomes mainly classified as MVSTm (8/10), collected in Italy, UK and Thailand from swine (5/10) and human. No of the genomes belonging to the STY-clade are positive for colistin resistance genes.

Genome-wide association study identified geographical segregated genetic markers in Italian MVSTm

To investigate which genetic traits have been associated with the successful local expansion of specific MVSTm genotype in Italy we used Scoary (36). Each gene cluster in the accessory genome was scored according to its apparent correlation to a predefined trait defined as Italian population and Benjiamini-Hochberg (BH) P-value was calculated. Of the 9,085 accessory gene clusters, Scoary reported a total of 49 loci with a BH value under 0.05, and present in more than 20% of Italian and less than 30% of non-Italian isolates (Table 1). Loci form clusters are located in separate fragments of the genomes most of which exhibited homology to various genetic regions including phages, prophages and plasmid-associated genes originating from different bacterial species (*Salmonella enterica*, *E. coli* and *Shigella*). Thus, gene clusters have been divided in three groups: group 1 includes 7 contiguous loci belonging to a putative plasmid; group 2 includes 33 contiguous loci belonging to a large 42.9 kb prophage region; and group 3 includes the remaining 9 genes which are spread across the genome. The 49 Italian-associated loci are overrepresented in the STY-clade genomes (Fig. 2). Particularly, only a single Italian isolate harbouring both group 1 and group 2 loci, and five UK isolates possessing a significant amount of group 2 loci are located outside the STY-clade. Finally, although clearly dominant in STY-clade, group 3 loci have been found quite frequently across the tree.

Table 1: Accessory genes associated with Italian strains rated by Benjiamini Hochberg (BH) P-value.

Roary gene name	Prokka annotation	BH P-value	Italian isolates*	Non-italian isolates**	Gene details***
group_3215	Hypothetical protein	1,94E-48	70	15	Phage BRO family/N-terminal domain protein
group_5738	Hypothetical protein	2,97E-47	70	17	Phage lambda NC_001416: cell lysis protein/endopeptidase
group_7354	Hypothetical protein	3,54E-47	69	16	Phege Clostr CDMH1 NC_024144: putative signalling/NTPase protein
group_7352	Hypothetical protein	3,54E-47	69	16	Phage hypotetical protein
group_5725	Hypothetical protein	4,43E-47	70	18	Phage Erwini phiEt88_NC_015295: DNA N-6-adenine-methyltransferase
group_2349	Hypothetical protein	4,43E-47	70	18	Phage Entero SfI NC_027339: Ren protein
group_5737	Hypothetical protein	4,43E-47	70	18	Phage Entero lambda NC_001416: Bor protein precursor
group_2253	Hypothetical protein	4,43E-47	70	18	Phage Entero c_1 NC_019706: lysozyme
group_7353	Hypothetical protein	8,17E-47	69	17	Phage Gifsy_1 NC_010392: bacteriophage antiterminator protein Q
group_7359	Hypothetical protein	3,28E-46	69	18	Phage Entero 933W NC_000924: hypothetical protein
group_3054	Hypothetical protein	4,28E-46	72	23	Phage Shigel SfII NC_021857: hypothetical protein
group_4040	Hypothetical protein	1,98E-45	71	23	Phage hypotetical protein
rusA_2	Crossover junction endodeoxyribonuclease	3,80E-45	70	22	Phage Entero mEp237 NC_019704: Holliday junction resolvase RusA
group_7356	Hypothetical protein	5,10E-45	63	12	Outer membrane protein assembly factor BamE
group_4041	Hypothetical protein	9,35E-45	70	23	Phage Entero BP 4795_NC_004813: hypothetical protein
group_3216	Hypothetical protein	9,35E-45	70	23	Phage Entero SfI NC_027339: replication protein P; -; phage
group_3214	Hypothetical protein	9,35E-45	70	23	Phage hypotetical protein
group_3213	Hypothetical protein	9,35E-45	70	23	Phage hypotetical protein
group_1275	Hypothetical protein	9,35E-45	70	23	Phage Salmon SEN34 NC_028699: replication protein O
group_4491	Hypothetical protein	9,35E-45	70	23	Phage Rha protein
group_4493	Hypothetical protein	9,35E-45	70	23	Phage Stx2 II NC_004914: hypothetical protein
group_7358	Hypothetical protein	9,35E-45	70	23	Phage Entero 933W NC_000924: host-nuclease inhibitor protein Gam
group_7351	Hypothetical protein	9,35E-45	70	23	Phage Salmon ST64T NC_004348: holin protein
group_932	Hypothetical protein	9,35E-45	70	23	Phage Entero phi80 NC_021190: CII decision making protein
kilR	Killing protein KilR	9,35E-45	70	23	Phage Entero HK225 NC_019717: Kil protein
group_3075	Hypothetical protein	2,45E-44	134	233	Hypotetical protein
group_3217	Hypothetical protein	2,37E-43	69	24	Phage Entero 933W NC_000924: Bet protein
group_2348	Hypothetical protein	1,65E-42	63	16	Predicted NTPase, NACHT family domain [Signal transduction mechanisms]/Ecoli
group_686	Hypothetical protein	3,66E-42	71	30	Putative plasmid associated gene
group_2346	Hypothetical protein	4,12E-42	67	23	Phage hypotetical protein
group_7349	Hypothetical protein	5,77E-42	68	25	Putative plasmid associated gene
group_7350	Hypothetical protein	1,87E-41	68	26	Putative plasmid associated gene
group_7348	Hypothetical protein	3,53E-40	66	25	Putative plasmid associated gene
group_3117	Hypothetical protein	9,90E-39	51	6	Genomic DNA
group_1265	Hypothetical protein	4,26E-35	70	45	Putative plasmid associated gene
group_7650	Hypothetical protein	4,42E-35	52	12	Genomic DNA
group_6567	Hypothetical protein	1,48E-33	60	28	Genommic DNA

Roary gene name	Prokka annotation	BH P-value	Italian isolates*	Non-italian isolates**	Gene details***
group_3072	Hypothetical protein	1,47E-32	111	194	Genomic DNA
group_7355	Hypothetical protein	2,21E-26	42	12	Genomic DNA
prtR	Putative HTH-type transcriptional regulator	1,55E-25	41	12	O antigen synthesis gene
group_7816	Hypothetical protein	3,86E-24	40	13	Phage hypotetical protein
group_7817	Putative HTH-type transcriptional regulator	2,65E-23	39	13	Phage Salmon ST160 NC_014900: C2 phage
rop	Regulatory protein rop	6,27E-20	79	140	Putative plasmid associated gene
mbeC	Mobilization protein MbeC	2,84E-10	45	78	Putative plasmid associated gene
group_48	Hypothetical protein	6,61E-04	44	141	Genomic DNA
xerC_1	Tyrosine recombinase XerC	7,22E-04	80	334	Phage Shigel SfII_NC_021857: integrase
group_7044	Hypothetical protein	1,00E-03	79	333	Genomic DNA
group_4380	Hypothetical protein	1,48E-03	79	337	Phage protein flxA
sopE	Guanine nucleotide exchange factor SopE	4,38E-02	73	345	Phage G-nucleotide exchange factor SopE

*Number of Italian strain positive for the gene
** Number of non-Italian strain positive for the gene
***Gene details based on PHAST and on BLASTn against NCBI database

Characterization of the prophage region

A total of 10 prophages regions have been annotated by PHAST (41) in the genome of the strain STY194, selected herein as reference genome for STY-clade. Of those 10 regions, 6 were intact, 3 were incomplete and one questionable. Among the intact regions, the one from nucleotide position 1,153,230 to 1,196,161 (42.9Kb in total) includes 62 loci of which 33 have previously been classified by Scoary (36) as strongly associated with Italian MVSTm (group 2, see above). The phage has been integrated downstream the *tRNA-thrW* gene homolog of *S.* Typhimurium LT2 (Fig. 3). A similar phage has been detected also in the UK strain SAL_JA6411AA (Fig. 3), belonging to the STY-clade but missing the group 2 loci. The phage in SAL_JA6411AA shows identical 3' -end but a divergent 5'-end sequence compared with STY194. The divergent part includes the 33 Italian associated loci of group 2 in STY194, substituted in SAL_JA6411AA by 22 different genes. In STY194, the 5'-half of the phage comprises genes involved in transcription and regulation, in integration-recombination and cell division, prophage repression, cellular lysis and serum resistance. The conserved 3'-end of the pro-phage, which accounts for the 45.2% of the entire prophage genome, shows high similarity to the *Shigella flexneri* prophage SfII and mainly encodes proteins involved in capsid formation and DNA packaging (head, tail, and terminase). Most of the 24 loci of this region were found in more than 75% of all genomes. The final portion of this phage genetic region shows homology to *Salmonella* phage SP_004 tail fiber assembly protein followed by *sopE*, a G-nucleotide exchange factor protein from SopEϕ. Genes encoding for these two proteins were shared by roughly the 35% of analysed genomes and *sopE* has been found to be negatively associated with North American origin (Fisher's exact test; $P<0.0001$).

MVSTm prophage region alignment

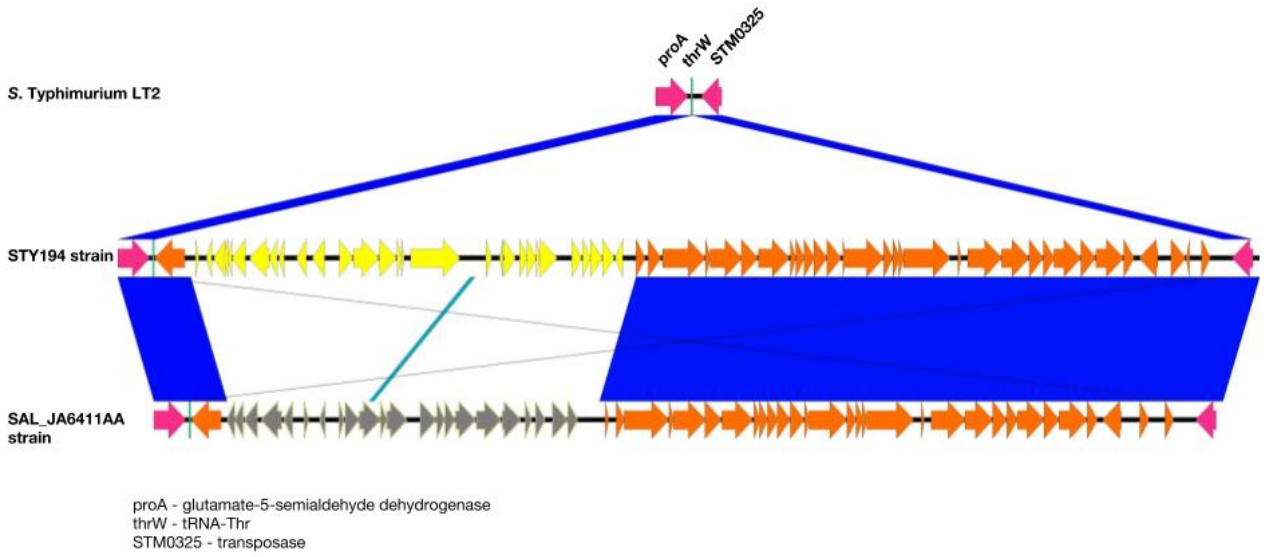

Figure 3: Alignment of 42.9Kb prophage region of STY194 strain (in the middle) including 62 loci of which 33 (yellow) are that classified by Scoary(36) as strongly associated with Italian MVSTm (group 2 in the text). At the top, S. Typhimurium LT2 showing the insertion of the prophage between the tRNA-thrW locus downstream of proA and the transposase STM0325. At the bottom, similar prophage region of UK strain SAL_JA6411AA with divergent loci coloured in grey. The full match of shared loci (orange and magenta arrows) is showed in blue.

Phylogenomic reconstruction of the MVSTm strains

To better understand the genetic relationship of isolates characterized by different accessory genes profiles but gathered into the same goeBURST group we performed a single nucleotide polymorphisms (SNPs) calling

on all isolates included in goeBURST group 1 along with a selected outgroup of four Italian isolates. Therefore, 1,293 genome assemblies were mapped against the STY-clade reference genome using Snippy v 3.2. Pairwise SNPs differences ranged between 0 to 1,793 with a median of 334 and a median percentage of bases aligned to the reference of the 98,43%. The maximum likelihood phylogeny and population structure were inferred based on 11,278 core SNPs. Two populations were identified corresponding to two major clades in the ML tree (Fig. 4). Clade I is characterized by long branches and includes 25 MVSTm and one *S.* Typhimurium isolates of ST19 mainly isolated in North American (~65%). More than half of the genomes exclusively included in clade I harboured from 10 to 32 pSLT-related genes. On the contrary, clade II is characterized by very short branches and includes 1,267 MVSTm (1,025) and *S.* Typhimurium (242) isolates mainly belonging to ST34 (97,9%) and ASSuT genotype (66,3%), and collected in Europe (60%) and North America (38%) (Fig. 5). At 0.013 (nucleotide substitutions per site) distance from the root, clade II was divided in 61 subclades (containing at least two genomes) and 56 singletons. The subclades 10, 41 and 61, composed by 91, 195 and 474 genomes, respectively, account for ~60% of the genomes in clade II and are significantly associated to the origin of the isolates.

SNPs trees

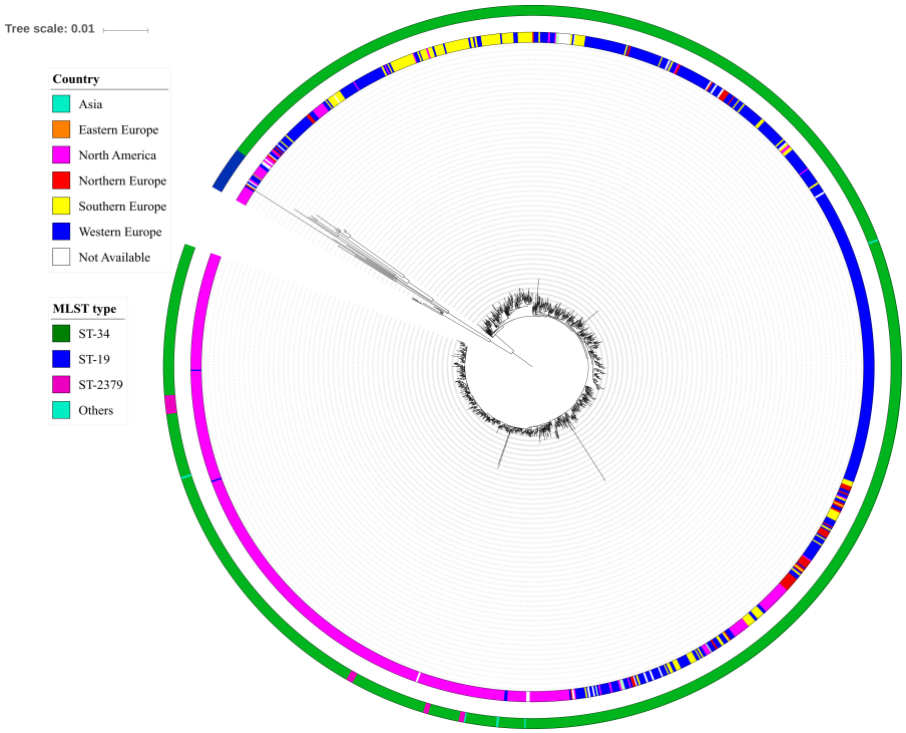

Figure 4. Maximum likelihood inferred based on 11,278 core genome SNPs. The tree shows the *S.* Typhimurium/ MVSTm population forming two major clades. The long branch indicates clade I including ST19 strains (blue) mainly from North America. On the contrary, clade II mainly includes ST34 (external circle). Countries of origin are indicated in the internal circle by colours as reported in the legend.

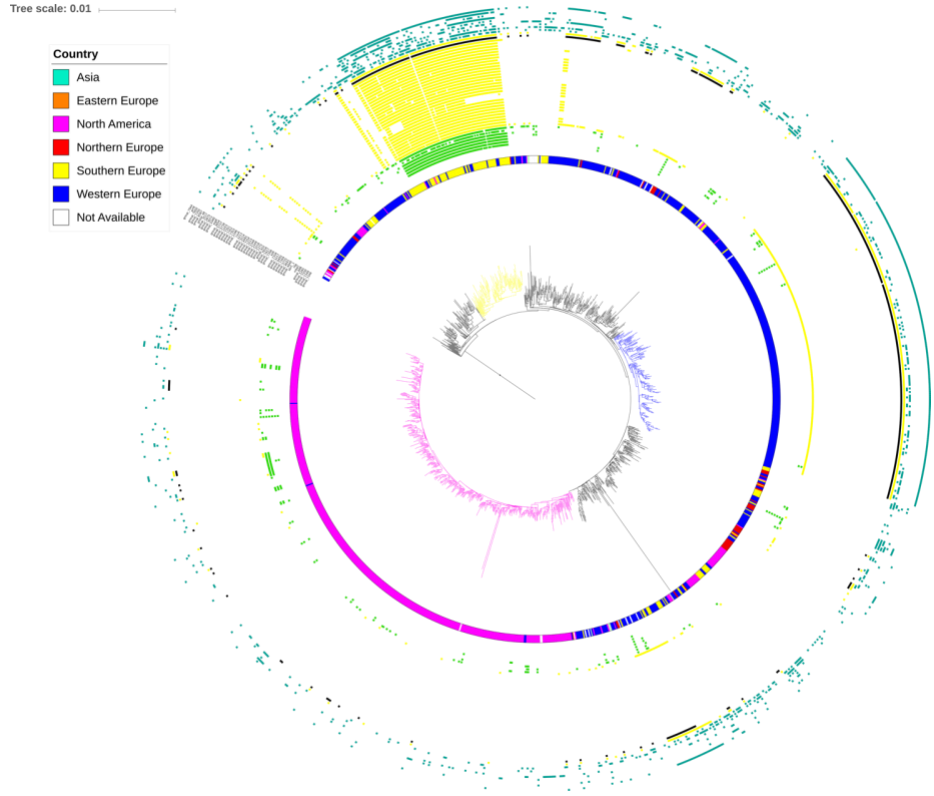

Figure 5. Maximum likelihood tree inferred based on 11,278 core SNPs on 1,293 isolates. Figure 3 shows the tree pruned on clade II with coloured branches for subclade 10 (yellow), 41 (blue) and 61 (magenta). The internal circles indicate the originating geographical area of each isolates indicated by colours as in the legend. Externally, clusters of genes statistically associated with Italian strains divided by colours in plasmid-related contiguous loci (green); prophage related contiguous loci (yellow); and associated loci spread across the genome (light blue). The black hits are indicating *sopE* gene presence.

Subclade 10 is significantly associated with Italian origin (Fisher's exact test; P<0.0001) and includes ~50% of the STY-Italian isolates available in the dataset and 5 additional Italian MVSTm isolates that were collected prior to 2012. In subclade 10, pairwise distance was from 0 to a max of 368 (Fig. 6). In comparison to the accessory genome clustering, subclade 10 contains 85/91 isolates belonging to STY-clade and all but 5 possess the SopE-containing phage. The distribution of the Italian associated loci on the core SNP tree, as shown in Figure 5, indicate a clear association with subclade 10, particularly for the group 1 and 2 loci. Subclade 41 is significant associated with Western Europe origin (UK and Ireland; Fisher's exact test; P<0.0001) while subclade 61 is significant associated with North American origin (Fisher's exact test; P<0.0001), comprising 88% of the North American isolates included in the study.

Pairwise cgSNPs distance matrix

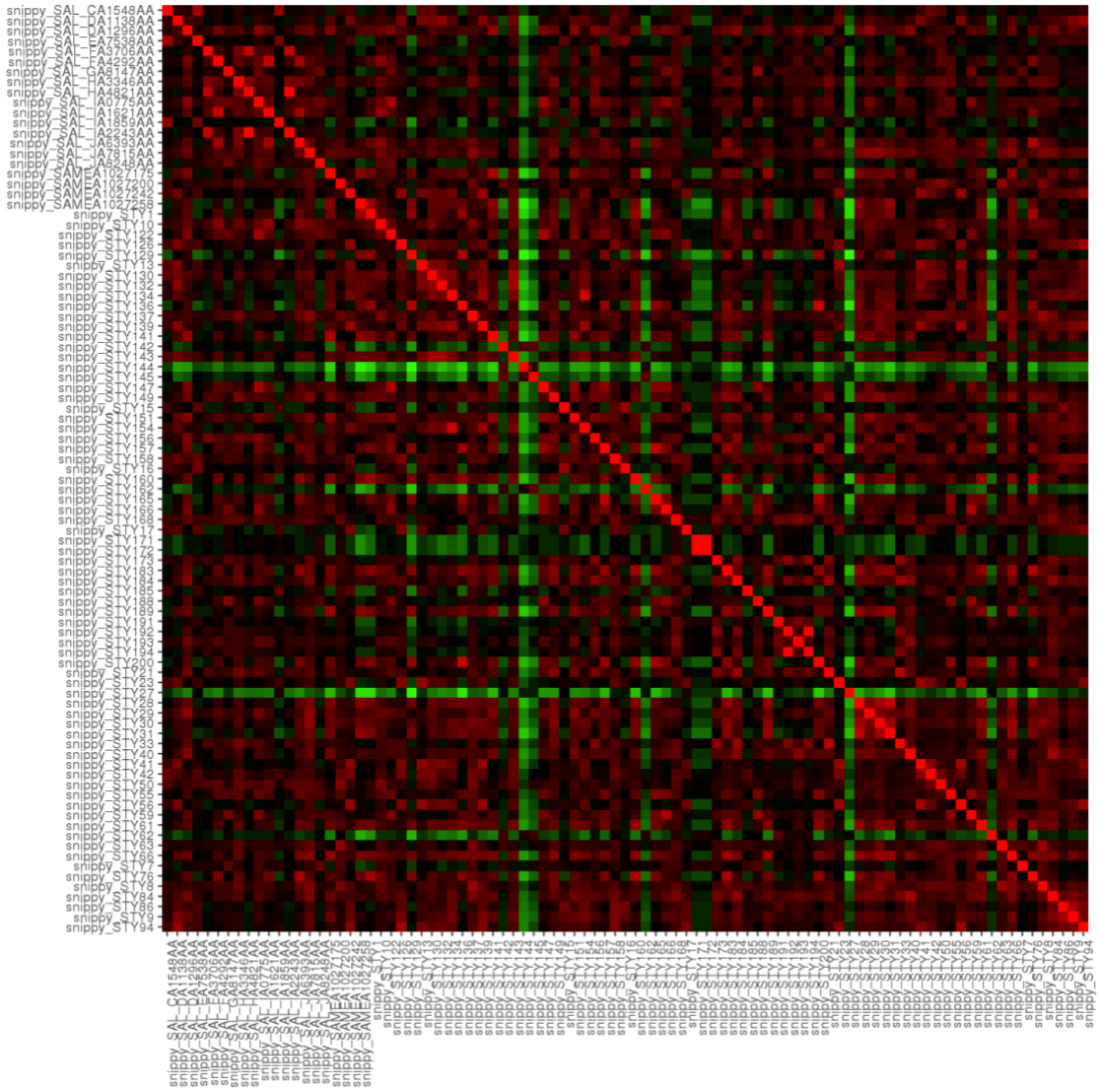

Figure 6. Pairwise SNPs distance matrix of the core genome SNPs alignment of the 91 strains from subclade 10. The pairwise distance ranges from 0 to a max of 368. SNPs distance was calculated with snps-dists (https://github.com/tseemann/snp-dists) and visualized on http://www.heatmapper.ca

Discussion

The importance of *S.* 4,[5],12:i:- (MVSTm) arose when it climbed up the charts of *Salmonella* serovar responsible of foodborne outbreaks worldwide. In particular, in Europe the joint report on zoonosis monitoring by EFSA/ ECDC described *S.* 1,4,[5],12:i:- as the third serovar among *Salmonella* already in 2013 (6). Swine appears to be the main reservoir of this peculiar *S. enterica* serovar. However, the monomorphic nature of MVSTm has been an obstacle for identifying the relative importance of other animal species as sources of human infections (46).
Several studies have been conducted in recent years to elucidate the phylogenetic relationship, transmission dynamics as well as the virulence and resistance key determinants of epidemics MVSTm (11, 12, 44, 47–52). Genomic analyses have suggested the emerging of multiple independent clones in the United States and Europe (4). Specifically, three distinct epidemics have driven the micro-evolution of MVSTm across the globe resulting in the expansion of different clones which tend to be dominant in specific geographical locations (47). Recent data on *Salmonella* serovar Cerro suggested that several genomic markers associated to geographically segregated phylogroups may contribute to the ability of *Salmonella* to rapidly diverge and adapt to a specific niche (53). Therefore, geographical segregation can play an important role in the micro-evolution of emerging clones, leaving enough genetic signal in the population which can contribute to improve source attribution of clinical MVSTm infections. In the present study, we tested this hypothesis by focusing on identifying biogeographical markers in MVSTm genomes of a geographical and temporal related set of isolates obtained in Italy between 2012 and 2014. To mine phylogenetic related isolates from different part of the world, we applied an inexperienced genomic approach by comparing the 148 Italian isolates with a large set (> 4,000) of representative public available genomes of several *S. enterica* serovars using the gene-by-gene methodology. Thus, we identified ~1,300 *S.* Typhimurium and MVSTm from a broad geographical area collected in almost 20 years from several sources showing up to 2.5% allele diversity with most of the Italian isolates. Integrating phylodynamics with genome-wide association analysis, we have shown that within this peculiar population of very similar *S.* Typhimurium/ MVSTm isolates, the expansion of genotypes in a specific geographical region is facilitated by the acquisition of unique accessory genetic markers. Phylogenetic reconstruction revealed that isolates from the same geographical origin form several highly supported monophyletic groups, providing discrete evidence of the phylogeographical structure of this population. Isolates from human and from swine related sources clustered in these groups indicating that humans are exposed to the same genotypes circulating among pigs. The presence of most of the Italian isolates within a single monophyletic clade characterized by specific repertoire of plasmid- and phage-related loci support the hypothesis that this genotype endured a substantial evolution under geographical segregation. Particularly, geographical isolation appeared to have a strong impact on the gene content of a large prophage detected in most of the isolates. Within this prophage, genome-wide comparison identified 30 genes to be strongly associated with Italian origin. Among associated prophage elements, we found *sopE* gene, a virulence factor recently described in European strains (47), enriched in genomes from subclades 10 and subclades 41, significant associated with Italian and UK origin respectively. This is consistent with findings by Petrovska and colleagues suggesting an increase of *sopE* gene frequency since 2007 in monophasic epidemic isolates from UK and Italy (47); increase confirmed by the overrepresentation of *sopE* gene in the European MVSTm collected after 2010 (350/727; ~48%) analysed in this study. Although our data show a negative association of this gene with isolates from North America, a recent study by Elnekave and colleagues (48) reported the presence of *sopE* gene in US isolates from swine samples collected during 2014-2016. The fact that MVSTm isolates harbouring *sopE* gene have being collected in Europe for several years before might raise the question of whether *sopE* positive isolates in US are most likely originating from European strains circulating in swine production chain. Further studies are needed at present to elucidate with higher resolution the genetic relationship of *sopE* positive isolates on a more representative set of US and European swine-related MVSTm strains.

Noteworthy, whereas prophage virulence gene *sopE* was mostly located in strains from Italian and Western Europe phylogroups in ST34 clade II, *spvC* and *spvB* and other virulence markers of pSLT plasmid with reported contribution to pathogenicity of *S*. Typhimurium (45, 54) were located in strains from clade I belonging to ST19. The presence of these virulence genes exclusively in ST19 strains provided evidence that these strains most likely originated from *S*. Typhimurium ancestors distinct from that of the European clone. This is consistent with results of García studies where the presence of *spvC* gene and virulence-plasmids genes is reported only in MVSTm ST-19 strains. Although *spvC* positive ST19 MVSTm strains have been described as similar to the hepta-resistant Spanish clone, antimicrobial resistance (AR) genes were found in only 4 out of 15 of the virulence-harbouring isolates of our dataset. In the current study, we showed that the most common AR profile for the majority of *Salmonella* Typhimurium/ MVSTm strains isolated from humans, animals, environment and animal foodstuffs included in ST34 clade II predict ASSuT genotype. The wide diffusion of multiple genotypes of R-type ASSuT MVSTm in European countries as well as in North America and Asia constitute a growing risk that can be associated to increased hospitalization, development of a bloodstream infection, or treatment inefficacy in patients (55). However, the local expansion of specific clones can also result in the loss of AR genes, as we observed in the UK-associated subclone 41. The loss of ASSuT or ACGSSuTTp genotypes emerged towards the terminal branches of the subclade 41 tree populated exclusively by isolates harbouring a single tetracycline resistance gene. Since these mono-resistant genotypes have been isolated in a successive timeframe than multi-resistant genotypes within subclade 41, we presume that the dynamic genome plasticity of *S*. Typhimurium/ MVSTm serovars may lead to the formation and successful expansion of clones suffering the loss of particular adaptive traits.

Petrovska and colleagues (47) have investigated the micro-evolution of MVSTm clones responsible for recent UK epidemic ways (from 2005 to 2012). The authors discovered that monophasic epidemic clones circulating in UK and Italy are characterized by the acquisition of multiple novel genes, including a *sopE*-containing prophage mTmV, and formed a single clade with remarkable genetic variation from North American and Spanish epidemics clones. They identified three distinct subclades one of which (i.e. subclade C) being preferentially associated with Italian livestock production. In our study, four of the Italian isolates collected up to 2010 included in the study of Petrovska and colleagues (47) belong to the Italian associated subclade 10 as shown in Figure 3, together with ~57% of the Italian STY isolates analysed in this study and collected from 2012 to 2014. Particularly, the subclade 10 represent a significant expansion of the *sopE* positive monophyletic group within subclade C as described by Petrovska and colleagues (47). Similarly, the UK associated subclade 41, as shown in Figure 3 herein, represent the recent expansion (>90% of the samples within subclade 41 were collected in UK or Ireland after 2014) of *sopE* positive monophyletic group within subclade A as described by Petrovska and colleagues (47). The ongoing clonal expansion of these *sopE* positive MVSTm subpopulations shows that the acquisition of this gene has conferred a clear competitive advantage in the ongoing European MVSTm epidemics. As previously suggested (47), the acquisition of *sopE* has happened in multiple independent events. This theory is confirmed by the gene contents of the *sopE*-containing prophage. Indeed, in the UK associated subclade 41, *sopE* is located at the 3'-end of a prophage mTmV as previous described (47), while in the Italian subclade 10 the prophage containing *sopE* shared only half of the mTmV genes. This novel prophage mTmV2 contains the majority of the Italian associated biogeographical markers, characterizing subclade 10, and, therefore, it can be used to enhance strains traceability in ongoing epidemics.

Conclusions

Until internationally validated WGS-derived nomenclatures and transferable protocols will be available for *Salmonella enterica* typing and phenotype prediction (e.g. AR, virulence) the immediate operationalisation of WGS-based surveillance in Europe remains a challenge. Nevertheless, the genomic analysis applied in this study on a large dataset of newly sequenced genomes in comparison to publicly available genomes have proved to be well suited to untangle the phylogeographical structure and micro-evolution of MVSTm serovar. Thanks to the growing number of publicly available genomes and metadata, further investigations on additional geographically and temporally distributed monophasic and non-monophasic variants of *S.* Typhimurium genomes would increase the power of the analyses for the search of additional geographically segregated markers, and elucidate how these markers are distributed among this populations. The gain and loss of mobile genetic elements may "unlock the secrets" for the optimization of infection-control strategies and effective containment of this emergent pathogen as was already discussed in a recent study on the transmission dynamics of *Enterococcus faecium* (56).

Manuscripts submission and acknowledgements

1. Manuscript in major revision in: *Scientific reports*

Title:

Genome-wide identification of geographical segregated genetic markers in *Salmonella enterica* serovar Typhimurium variant 4,[5],12:i:-.

Authors*:*

Federica Palma[1], Gerardo Manfreda[1], Mickael Silva[2], Antonio Parisi[3], Frédérique Pasquali[1], Mirko Rossi[4].

[1]DISTAL, Department of Agricultural and Food Sciences, University of Bologna, Italy.
[2]Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa Lisbon, Portugal.
[3]Isituto Zooprofilattico Sperimentale della Puglia e della Basilicata, Foggia, Italy.
[4]Department of Food Hygiene and Environmental Health, University of Helsinki, Helsinki, Finland

Acknowledgments:

This study has been supported by COMPARE project (https://www.compare-europe.eu) co-funded by the *European Union's Horizon 2020 research and innovation programme* under grant agreement N° 643476. This study has also been supported by INNUENDO project (https://www.innuendoweb.org) co-funded by the European Food Safety Authority (EFSA), grant agreement GP/EFSA/AFSCO/2015/01/CT2 ("New approaches in identifying and characterizing microbial and chemical hazards"). The conclusions, findings, and opinions expressed in this review papers reflect only the view of the authors and not the official position of the European Food Safety Authority (EFSA). We would like to thank Dr. Antonia Ricci from Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro (Italy) and Dr. Ida Luzzi from Istituto Superiore della Sanitá, Roma (Italy) for providing MVSTm strains. The authors wish to thank CSC- Tieteen tietotekniikan

keskus Oy (Helsinki) and Department of Physics and Astronomy (Bologna) for providing access to cloud computing resources.

References

1. Switt AIM, Soyer Y, Warnick LD, Wiedmann M. 2009. Emergence, Distribution, and Molecular and Phenotypic Characteristics of Salmonella enterica Serotype 4,5,12:i:–. Foodborne Pathog Dis 6:407–415.
2. Hopkins KL, Kirchner M, Guerra B, Granier SA, Lucarelli C, Porrero MC, Jakubczak A, Threlfall EJ, Mevius DJ. 2010. Multiresistant Salmonella enterica serovar 4,[5],12:i:- in Europe: a new pandemic strain? Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 15:19580.
3. Echeita MA, Aladueña A, Cruchaga S, Usera MA. 1999. Emergence and Spread of an Atypical Salmonella enterica subsp. enterica Serotype 4,5,12:i:− Strain in Spain. J Clin Microbiol 37:3425–3425.
4. Soyer Y, Switt AM, Davis MA, Maurer J, McDonough PL, Schoonmaker-Bopp DJ, Dumas NB, Root T, Warnick LD, Gröhn YT, Wiedmann M. 2009. Salmonella enterica Serotype 4,5,12:i:−, an Emerging Salmonella Serotype That Represents Multiple Distinct Clones. J Clin Microbiol 47:3546–3556.
5. Zamperini K, Soni V, Waltman D, Sanchez S, Theriault EC, Bray J, Maurer JJ. 2007. Molecular characterization reveals Salmonella enterica serovar 4,[5],12:i:- from poultry is a variant Typhimurium serovar. Avian Dis 51:958–964.
6. European Food Safety Authority, European Centre for Disease Prevention and Control. 2016. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2015. EFSA J 14:n/a-n/a.
7. Mossong J, Marques P, Ragimbeau C, Huberty-Krau P, Losch S, Meyer G, Moris G, Strottner C, Rabsch W, Schneider F. 2007. Outbreaks of monophasic Salmonella enterica serovar 4,[5],12:i:- in Luxembourg, 2006. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 12:E11-12.
8. Dionisi AM, Graziani C, Lucarelli C, Filetici E, Villa L, Owczarek S, Caprioli A, Luzzi I. 2009. Molecular Characterization of Multidrug-Resistant Strains of Salmonella enterica Serotype Typhimurium and Monophasic Variant (S. 4,[5],12:i:–) Isolated from Human Infections in Italy. Foodborne Pathog Dis 6:711–717.
9. Hauser E, Tietze E, Helmuth R, Junker E, Blank K, Prager R, Rabsch W, Appel B, Fruth A, Malorny B. 2010. Pork Contaminated with Salmonella enterica Serovar 4,[5],12:i:−, an Emerging Health Risk for Humans. Appl Environ Microbiol 76:4601–4610.
10. Echeita MA, Herrera S, Usera MA. 2001. Atypical, fljB-negative Salmonella enterica subsp. enterica strain of serovar 4,5,12:i:- appears to be a monophasic variant of serovar Typhimurium. J Clin Microbiol 39:2981–2983.
11. Ido N, Lee K, Iwabuchi K, Izumiya H, Uchida I, Kusumoto M, Iwata T, Ohnishi M, Akiba M. 2014. Characteristics of Salmonella enterica Serovar 4,[5],12:i:- as a Monophasic Variant of Serovar Typhimurium. PLoS ONE 9.
12. García P, Malorny B, Rodicio MR, Stephan R, Hächler H, Guerra B, Lucarelli C. 2016. Horizontal Acquisition of a Multidrug-Resistance Module (R-type ASSuT) Is Responsible for

the Monophasic Phenotype in a Widespread Clone of Salmonella Serovar 4,[5],12:i:-. Front Microbiol 7.

13. García P, Malorny B, Hauser E, Mendoza MC, Rodicio MR. 2013. Genetic Types, Gene Repertoire, and Evolution of Isolates of the Salmonella enterica Serovar 4,5,12:i:− Spanish Clone Assigned to Different Phage Types. J Clin Microbiol 51:973–978.
14. Achtman M. 2012. Insights from genomic comparisons of genetically monomorphic bacterial pathogens. Philos Trans R Soc B Biol Sci 367:860.
15. Hopkins KL, de Pinna E, Wain J. 2012. Prevalence of Salmonella enterica serovar 4,[5],12:i:- in England and Wales, 2010. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 17.
16. Wain J, O'Grady J. 2017. Genomic Diversity in Salmonella enterica, p. 91–107. *In* Applied Genomics of Foodborne Pathogens. Springer, Cham.
17. Taboada EN, Graham MR, Carriço JA, Van Domselaar G. 2017. Food Safety in the Age of Next Generation Sequencing, Bioinformatics, and Open Data Access. Front Microbiol 8.
18. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, Brisse S, Cao H, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Nguyen KV, Nguyen TV, Dao TT, Mensink M, Minh VL, Nhu NTK, Schultsz C, Kuntaman K, Newton PN, Moore CE, Strugnell RA, Thomson NR. 2015. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proc Natl Acad Sci U S A 112:E3574–E3581.
19. Skallerup P, Espinosa-Gongora C, Jørgensen CB, Guardabassi L, Fredholm M. 2015. Genome-wide association study reveals a locus for nasal carriage of Staphylococcus aureus in Danish crossbred pigs. BMC Vet Res 11.
20. Reuter S, Corander J, de Been M, Harris S, Cheng L, Hall M, Thomson NR, McNally A. 2015. Directional gene flow and ecological separation in Yersinia enterocolitica. Microb Genomics 1.
21. Bazinet AL. 2017. Pan-genome and phylogeny of Bacillus cereus sensu lato. BMC Evol Biol 17.
22. Achtman M. 2008. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu Rev Microbiol 62:53–70.
23. Moodley Y, Linz B. 2009. Helicobacter pylori Sequences Reflect Past Human Migrations. Genome Dyn 6:62–74.
24. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
25. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455–477.
26. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE 9:e112963.
27. Maiden MCJ, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 11:728–736.

28. Silva M, Machado MP, Rossi M, Moran-Gilad J, Santos S, Ramirez M, Carrico JA. 2017. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. bioRxiv 173146.
29. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrio JA. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13:87.
30. Nascimento M, Sousa A, Ramirez M, Francisco AP, Carriço JA, Vaz C. 2017. PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. Bioinforma Oxf Engl 33:128–129.
31. Severiano A, Pinto FR, Ramirez M, Carriço JA. 2011. Adjusted Wallace Coefficient as a Measure of Congruence between Typing Methods ▿. J Clin Microbiol 49:3997–4000.
32. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl 30:2068–2069.
33. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693.
34. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268–274.
35. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242-245.
36. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 17:238.
37. Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Parkhill J, Rajandream M-A. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics 24:2672–2676.
38. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
39. Kleinheinz KA, Joensen KG, Larsen MV. 2014. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4.
40. Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, Villa L, Aarestrup FM, Hasman H. 2014. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrob Agents Chemother 58:3895–3903.
41. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Res 39:W347–W352.
42. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.
43. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. 2013. Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. Mol Biol Evol 30:1224–1228.

44. Cito F, Baldinelli F, Calistri P, Di Giannatale E, Scavia G, Orsini M, Iannetti S, Sacchini L, Mangone I, Candeloro L, Conte A, Ippoliti C, Morelli D, Migliorati G, Barile NB, Marfoglia C, Salucci S, Cammà C, Marcacci M, Ancora M, Dionisi AM, Owczartek S, Luzzi I, outbreak investigation group. 2016. Outbreak of unusual Salmonella enterica serovar Typhimurium monophasic variant 1,4 [5],12:i:-, Italy, June 2013 to September 2014. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 21.
45. Gulig PA, Danbara H, Guiney DG, Lax AJ, Norel F, Rhen M. 1993. Molecular analysis of spv virulence genes of the Salmonella virulence plasmids. Mol Microbiol 7:825–830.
46. Barco L, Barrucci F, Cortini E, Ramon E, Olsen JE, Luzzi I, Lettini AA, Ricci A. 2015. Ascertaining the relationship between Salmonella Typhimurium and Salmonella 4,[5],12:i:- by MLVA and inferring the sources of human salmonellosis due to the two serovars in Italy. Front Microbiol 6.
47. Petrovska L, Mather AE, AbuOun M, Branchu P, Harris SR, Connor T, Hopkins KL, Underwood A, Lettini AA, Page A, Bagnall M, Wain J, Parkhill J, Dougan G, Davies R, Kingsley RA. 2016. Microevolution during the emergence of a monophasic Salmonella Typhimurium epidemic in the United Kingdom. Emerg Infect Dis 22.
48. Elnekave E, Hong S, Mather AE, Boxrud D, Taylor AJ, Lappi V, Johnson TJ, Vannucci F, Davies P, Hedberg C, Perez A, Alvarez J. Salmonella enterica serotype 4,[5],12:i:- in swine in the United States Midwest: an emerging multidrug resistant clone. Clin Infect Dis.
49. Yang X, Wu Q, Zhang J, Huang J, Guo W, Cai S. 2015. Prevalence and Characterization of Monophasic Salmonella Serovar 1,4,[5],12:i:- of Food Origin in China. PLOS ONE 10:e0137967.
50. Andrés-Barranco S, Vico JP, Marín CM, Herrera-León S, Mainar-Jaime RC. 2016. Characterization of Salmonella enterica Serovar Typhimurium Isolates from Pigs and Pig Environment-Related Sources and Evidence of New Circulating Monophasic Strains in Spain. J Food Prot 79:407–412.
51. Seixas R, Santos TR, Machado J, Tavares L, Bernardo F, Semedo-Lemsaddek T, Oliveira M. 2016. Phenotypic and Molecular Characterization of Salmonella 1,4,[5],12:i:- R-Type ASSuT Isolates from Humans, Animals, and Environment in Portugal, 2006–2011. Foodborne Pathog Dis 13:633–641.
52. Gymoese P, Sørensen G, Litrup E, Olsen JE, Nielsen EM, Torpdahl M. 2017. Investigation of Outbreaks of Salmonella enterica Serovar Typhimurium and Its Monophasic Variants Using Whole-Genome Sequencing, Denmark. Emerg Infect Dis 23:1631–1639.
53. Kovac J, Cummings KJ, Rodriguez-Rivera LD, Carroll LM, Thachil A, Wiedmann M. 2017. Temporal Genomic Phylogeny Reconstruction Indicates a Geospatial Transmission Path of Salmonella Cerro in the United States and a Clade-Specific Loss of Hydrogen Sulfide Production. Front Microbiol 8.
54. Fàbrega A, Vila J. 2013. Salmonella enterica Serovar Typhimurium Skills To Succeed in the Host: Virulence and Regulation. Clin Microbiol Rev 26:308–341.
55. Multistate Outbreak of Multidrug-Resistant Salmonella I 4,[5],12:i:- and Salmonella Infantis Infections Linked to Pork (Final Update) | Multistate Outbreak of Multidrug-Resistant Salmonella I 4,[5],12:i:- or Salmonella Infantis Infections Linked to Pork (Final Update) | August 2015 | Salmonella | CDC.

56. van Hal SJ, Ip CLC, Ansari MA, Wilson DJ, Espedido BA, Jensen SO, Bowden R. 2016. Evolutionary dynamics of Enterococcus faecium reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance. Microb Genomics 2.

Useful links

https://github.com/fedex88/PhD_Thesis
http://www.compare-europe.eu
https://www.innuendoweb.org
http://www.ncbi.nlm.nih.gov/genome
http://www.ebi.ac.uk/ena.
https://enterobase.warwick.ac.uk
http://WGSA.net
https://ec.europa.eu/food/safety/rasff_en
https://github.com/INNUENDOCON/INNUca
https://github.com/B-UMMI/chewBBACA
https://github.com/tseemann/snippy
https://www.ncbi.nlm.nih.gov/genome/annotation_prok
http://bigsdb.pasteur.fr/listeria
https://enterobase.warwick.ac.uk/species/index/senterica
https://github.com/harry-thorpe/piggy.
https://github.com/AdmiralenOla/Scoary
https://github.com/tseemann/abricate
https://cge.cbs.dtu.dk/services/VirulenceFinder/, http://www.mgc.ac.cn/VFs
www.cbs.dtu.dk/services/MLST
https://cge.cbs.dtu.dk/services/PlasmidFinder/
https://github.com/tseemann/snippy
http://phast.wishartlab.com
https://github.com/tseemann/mlst
https://github.com/Cibiv/IQ-TREE
https://www.ncbi.nlm.nih.gov/nuccore
https://blast.ncbi.nlm.nih.gov/Blast.cgi
https://github.com/tseemann/abricate
https://github.com/tseemann/prokka
https://github.com/sanger-pathogens/Roary
https://www.ebi.ac.uk/Tools/msa/clustalo

Discussion and conclusions

Listeria monocytogenes and *Salmonella enterica* serovar Typhimurium variant 4,[5],12:i:- are foodborne pathogens with relevant rate of associated foodborne illness cases reported in Europe. Species-specific genetic bacterial determinants and matrix-specific ecological factors (e.g. physical factors, microbial competition, predation, etc.) can contribute to the persistence of these foodborne pathogens (FBPs) in the food chain. (7). To extend our understanding on microbial evolution, ecology, population dynamic, and persistence of FBPs, multiple investigative approaches based on a combination of genomic analyses may be performed bearing up an enhanced surveillance.
Currently, there is not yet a fully standardize genomic methodology. However, albeit bioinformatics expertise and software, standardised and validated protocols and common nomenclatures are required, whole genome Sequencing (WGS) data based analyses can be successfully used for multiple investigations at once and revealed an unparalleled potential as a one-serve-all approach. In the study presented in this thesis, comparative genomics have been essential in investigating (i) the persistence, virulence potential and physiological adaptation of *Listeria monocytogenes* in food processing plants and (ii) the micro-evolution and phylogeography of *Salmonella enterica* serovar 4,[5],12:i:-.
A major aim of the present thesis was to get insights on persistence of *L. monocytogenes* isolates belonging to ST14 and ST121 collected over one year from an Italian rabbit production plant. For this purpose, 34 *L. monocytogenes* isolates belonging to three genotypes ST14, ST121, ST224 were whole genome sequenced. cgMLST analysis on all sequenced genomes showed a significantly superior discriminatory power in comparison to molecular methods (e.g. ribotyping), allowing the detection of two singletons belonging to ST14 that were not observed by other molecular methods. The Multi Virulence-Locus Sequence Typing (MVLST) and the investigation of a representative database of virulence determinant genes was performed on ST121 and ST14 sequenced genome in comparison to a set of publicly available genomes. All ST14 isolates were assigned to a Virulence-type VT107 very close to VT1 strains, identified as part of an epidemic clone responsible of outbreaks in US. Virulence genes screening showed the presence of a full-length *inlA* version in all ST14 isolates and of a mutated version including a premature stop codon (PMSC) associated to attenuated virulence in all ST121 isolates. *In silico* analyses were useful to predict the potential virulence of *L. monocytogenes* isolates. MVLST results, as well as analysis of virulence genes, suggest a higher virulence potential of ST14 sequenced isolates in comparison to ST121, besides additional *in vitro* confirmations should be performed to confirm the virulence of ST14 and ST121 isolates.
So far, the persistence and ecophysiology of specific subtypes of *L. monocytogenes* (e.g. hypervirulent ST14) able to adapt and colonise harbourage sites in food-processing environment is still under investigation. In this study, even though the number of isolates was limited, wgSNPs based phylogeny inferred on the 33 ST121 and ST14 sequenced isolates collected in the Italian rabbit-meat plant was able to predict that a persistent ST14 clone was circulating in the plant along with not persistent strains. However, it would be favourable to extend this research, in order to include a wider representative diversity of persistent and not persistent ST14. The novel dataset of accurately selected genes designed to investigate physiological adaptation to food-processing environment of *L. monocytogenes* showed a significant enrichment in ST121 genomes concerning genetic features related to adaptation to sanitizing procedures. Whereas within ST14, a genetic enrichment was observed concerning genes that confer an enhanced ability to form biofilm (*actA* gene full-length version) along with heavy metals resistance genetic features (e.g. *ars* operon and *cadA4C* gene), previously related to persistent *L. monocytogenes* strains. These results suggest that ST14 strains-specific genetic repertoire may represent a selective advantage for this genotype to colonize harbourage sites where sanitizing procedures are difficult to occur. In conclusion, this study highlights that the incidence and persistence of ST14 *L. monocytogenes* strains within food-processing plant should be taken into greater consideration, given its hypervirulent potential.

At present, the emerging incidence of *Salmonella* Typhimurium variant 4,[5],12:i:- foodborne epidemics in last years and the efforts for identifying the transmission sources due to monomorphic nature of this serovar are steadily increasing public health concerns. Therefore, another major aim of this thesis was to untangle the evolutionary history of monophasic variant *Salmonella* Typhimurium (MVSTm) strains circulating in human and swine in comparison to an extended contest of selected publicly available *S.* Typhimurium/ MVSTm strains collected worldwide, using an innovative genome-wide investigative approach. Population structure analysis revealed that Italian isolates belong to a large population of ~1,300 clonal *S.* Typhimurium/ MVSTm isolates (maximum of 2.5% of allele differences), collected from a wide-range of countries in last two decades, and showing genetic similarity to antimicrobial resistance profiles previously associated to strains from European clone. Moreover, cgSNPs based phylogenetic reconstruction revealed that isolates from the same geographical origin form several highly supported monophyletic groups, suggesting discrete geographical segregation while GWAS provided statistical associations with Italian origin for a number of genetic markers (e.g. plasmid- and prophage-related genes). These results highlight that geographical isolation has had a strong impact on the accessory gene content mainly concerning a large SopE-containing prophage detected in most of the isolates. In conclusion, these data suggest that the expansion of successful epidemic clones harbouring unique gene clusters, that constitute specific biomarkers, is driven by the adaptation of certain lineages to specific hosts or food production systems, with regard to a local geographical scale. Furthermore, in order to improve identification of the source in course of large epidemics, studies focusing on optimal dataset of genomes, representative of all the source of isolation of MVSTm in the food chain, should be performed even taking account of biogeographical genetic markers.

The recent trends for molecular typing of several foodborne pathogens including L. monocytogenes and S. enterica are now focused on the shift from traditional molecular typing to multiple WGS-data based comparative genomics typing approaches. In the light of the proposal to phase out the use of several official molecular methodologies (e.g. PFGE, serotyping, ribotyping), it was a great and challenging chance to interface with bioinformatics pipelines and other Linux-based tools, cloud-based and publicly available computing for foodborne pathogens sequence data analysis. A lot of work is in progress worldwide between laboratories engaged in standardization, harmonization and validation of protocols and curated nomenclatures facilitating the routine application and full comparability of WGS not only for surveillance and outbreak investigation purposes, but also for answering further research questions (e.g. source attribution, virulence and resistance characterization, transmission dynamics, population structure, etc.). Despite future perspectives point to give up traditional typing methodologies, phenotypic tests still have to be carried out in order to confirm the genome based feature prediction and to assess concordance between in silico typing and phenotype.

Further work remains to be done in terms of metadata curation and real-time sharing in publicly available database in order to reduce the frequent bias for collection date and/ or geographical origin of public genomes with the perspective of defining epidemiological break points for enhanced for surveillance and outbreak investigations at national and international level. A cost-effectiveness approach should be adopted in WGS applications for minimizing sequencing costs while maintaining high quality data to be interpreted in context with laboratory results.

Acknowledgments

The research performed in this thesis was supported by COMPARE project (http://www.compare-europe.eu) co-funded by the *European Union's Horizon 2020 research and innovation programme* under grant agreement N° 643476. The study presented in this thesis has also been supported by INNUENDO project (https://www.innuendoweb.org) co-funded by the European Food Safety Authority (EFSA), grant agreement GP/EFSA/AFSCO/2015/01/CT2 ("New approaches in identifying and characterizing microbial and chemical hazards"). The work was performed at the Food Safety unit of the Department of Agricultural and Food Sciences (DISTAL), Alma Mater Studiorum - University of Bologna (Italy). Nevertheless, a huge part of the work was also performed at the Department of Food Hygiene and Environmental Health, University of Helsinki, Helsinki, (Finland).

I gratefully acknowledge Dr. Maria Fredriksson-Ahomaa and Dr. João Carriço for the contribution and time they offered as referees and for the constructive comments and suggestions.

During these of doctoral studies, a lot of people have guided and trained me with multiple tasks and I would like to express my appreciation to all of you. I would like to thank all people at DISTAL and all people at the Food Hygiene and Environmental Health for their great contribution in different ways in performing this work.

I would like to thank my supervisor Prof. Gerardo Manfreda for giving me the opportunity to perform these interesting research activities and being involved in COMPARE project together with DISTAL working team. In particular, I would like to thank my friends/ colleagues Alex (Dr Lucchi), Frédérique (Dr. Pasquali), Alessandra (Dr. De Cesare), Pietro (Dr. Olivi) my colleagues from upstairs and those that have been in my office, you all helped me and taught me so much and also for all good laughs, coffee breaks and mirtillino time.

A special thanks to Prof. Mirko Rossi from Department of Food Hygiene and Environmental Health, University of Helsinki (Finland), for his essential role during last year of my studies, for the patient and time expended to teach me that much, for the funny moments, fruitful discussions during my stay in Helsinki and beyond; I have learned from you a lot more than I expected during my PhD. I would like to thank Joana (Dr. Revez), an exceptional lecturer always ready to support me with stimulating brainstorming sessions and fruitful discussions. During my stay in Helsinki, I would like in particular to thank Gil, I learned a lot on *Salmonella* symptomatology and enjoy so much with you. To Mickael Silva, Miguel Machado and Bruno Gonçalves from the Instituto de Microbiologia, Universidade de Lisboa (Portugal) and Dr. Ed Taboada and Dillon Barker from Public Health Agency of Canada for your indispensable bioinformatics and technical support. To Professor Gastone Castellani and Daniel Remondini and all the people from the Department of Physics and Astronomy, University of Bologna (Italy) for the constructive collaboration and going ahead together in the COMPARE project. I also wish to thank CSC- Tieteen tietotekniikan keskus Oy (Helsinki), the curators Enrico Giampieri, Diego Zuccato from Department of Physics and Astronomy of Bologna for providing access to cloud computing resources and Italo (Dr. Faria do Valle) for your help and bioinformatics support. I would to special thank Ola (Dr Brynildsrud) for statistical support, interesting discussions and generous hospitality. Thanks to Thomas H Conner for the great help you have been giving to me during the years of my PhD and for the funny time spent together.

I would like to thank all the people involved in COMPARE and INNUENDO project for the great and challenging work they are pursuing and for organising and coordinating interesting workshops and meetings.

I would to special thank my family and my big family of friends for supporting me whatever crazy things I do and without whom none of this would be possible.

I learned and enjoy so much with all of you and it has been an endless pleasure to work and share special moments with you. Thank you!