Evaluation of WGS-subtyping methods for epidemiological surveillance of foodborne salmonellosis.

Authors: Manal Mohammed* and Salina Thapa
School of Life Sciences, College of Liberal Arts and Sciences, University of Westminster, UK.

***Corresponding author: Dr Manal Mohammed** (m.mohammed@westminster.ac.uk)

ABSTRACT:

Background: Salmonellosis is one of the most common foodborne diseases worldwide. Although human infection by non-typhoidal *Salmonella* (NTS) *enterica subspecies enterica* is associated primarily with a self-limiting diarrhoeal illness, invasive bacterial infections (such as septicaemia, bacteraemia and meningitis) were also reported. Human outbreaks of NTS were reported in several countries all over the world including developing as well as high-income countries. Conventional laboratory methods such as pulsed field gel electrophoresis (PFGE) do not display adequate discrimination and have their limitations in epidemiological surveillance. It is therefore very crucial to use accurate, reliable and highly discriminative subtyping methods for epidemiological characterisation and outbreak investigation.

Methods: Here, we used different whole genome sequence (WGS)-based subtyping methods for retrospective investigation of two different outbreaks of *Salmonella* Typhimurium and *Salmonella* Dublin that occurred in 2013 in UK and Ireland respectively.

Results: Single nucleotide polymorphism (SNP)-based cluster analysis of *Salmonella* Typhimurium genomes revealed well supported clades, that were concordant with epidemiologically defined outbreak and confirmed the source of outbreak is due to consumption of contaminated mayonnaise. SNP-analyses of *Salmonella* Dublin genomes confirmed the outbreak however the source of infection could not be determined. The core genome multilocus sequence typing (cgMLST) was discriminatory and separated the outbreak strains of *Salmonella* Dublin from the non-outbreak strains that were concordant with the epidemiological data however cgMLST could neither discriminate between the outbreak and non-outbreak strains of *Salmonella* Typhimurium nor confirm that contaminated mayonnaise is the source of infection, On the other hand, other WGS-based subtyping methods including multilocus sequence typing (MLST), ribosomal MLST

(rMLST), whole genome MLST (wgMLST), clustered regularly interspaced short palindromic repeats (CRISPRs), prophage sequence profiling, antibiotic resistance profile and plasmid typing methods were less discriminatory and could not confirm the source of the outbreak.

Conclusions: Foodborne salmonellosis is an important concern for public health therefore, it is crucial to use accurate, reliable and highly discriminative subtyping methods for epidemiological surveillance and outbreak investigation. In this study, we showed that SNP-based analyses do not only have the ability to confirm the occurrence of the outbreak but also to provide definitive evidence of the source of the outbreak in real-time.

Keywords: *Salmonella*, WGS, subtyping, SNP-typing, prophage profile, CRISPR typing, MLST, rMLST, wgMLST, cgMLST

Introduction:

Foodborne salmonellosis is an important concern for public health. It is caused by the enteric pathogen *Salmonella enterica*, which includes more than 2600 serovars (1). Human *Salmonella* infections are classically divided into diseases caused by typhoidal or non-typhoidal salmonella (NTS). Typhoid fever is caused by the human restricted *Salmonella enterica* serovars Typhi and Paratyphi (2). Although non-typhoidal *Salmonella* (NTS) serovars, predominantly cause a self-limiting diarrhoeal illness they have adapted to cause invasive extra-intestinal disease known as invasive NTS (iNTS) which can result in bacteraemia and focal systemic infections (3, 4) . There are two licenced vaccines for prevention of typhoid fever however, they are not effective against NTS (5) moreover, management of iNTS illness is complicated by the emergence of multidrug resistant (MDR) strains (6). *Salmonella* serovars responsible for typhoid fever kill over 250,000 humans per year (7) while non-typhoidal *Salmonella* (NTS) serovars responsible for diarrhoeal illness cause over 155,000 deaths annually (8). Interestingly, NTS have adapted to cause febrile bacteraemia and serious systemic infections; it has been estimated that over 680,000 people die every year as a result of infection by invasive NTS (iNTS) (3). *Salmonella* Typhimurium and *Salmonella* Dublin have been associated with systemic illness (4,5). Human outbreaks of *Salmonella* Typhimurium and *Salmonella* Dublin were reported in developed countries (10-12).

Conventional laboratory methods such as pulsed field gel electrophoresis (PFGE) do not usually provide adequate discrimination among outbreak and non-outbreak strains of *Salmonella enterica* and have their limitations in epidemiological surveillance, it is therefore crucial to use accurate, reliable and highly discriminative subtyping methods for epidemiological characterisation and outbreak investigation.

Here, we evaluate different whole genome sequence (WGS)-based subtyping methods (including single nucleotide polymorphism (SNP)-based cluster analysis, multilocus sequence typing (MLST), ribosomal MLST (rMLST), whole genome MLST (wgMLST), core genome MLST (cgMLST) as well as clustered regularly interspaced short palindromic repeats (CRISPRs), prophage sequence profiling, antibiotic resistance profile and plasmid typing) for retrospective investigation of two outbreaks of *Salmonella* Typhimurium and *Salmonella* Dublin that occurred in 2013 in UK and Ireland respectively (10, 13).

METHODS:

Retrospective analyses of the two outbreaks of *Salmonella* Typhimurium and *Salmonella* Dublin:

We carried out retrospective investigation of a human outbreak of *Salmonella* Dublin that occurred in 2013 in Ireland (10) and another human outbreak of *Salmonella* Typhimurium occurred in 2013 in UK (13). We included suspected food strains isolated from mayonnaise and raw-milk cheeses that can be linked to the outbreaks of *Salmonella* Typhimurium and *Salmonella* Dublin respectively. Non-outbreak strains were also included for comparison. Details of all *Salmonella* Dublin and *Salmonella* Typhimurium isolates analysed in this study are provided in supplementary tables 1 and 2 respectively.

PFGE was of a limited value for the investigation of the outbreak of *Salmonella* Dublin (10) since all outbreak and non-outbreak isolates of *Salmonella* Dublin were indistinguishable by PFGE. Although multiple loci VNTR analysis (MLVA) was of value in discriminating the outbreak strains from an epidemiologically unrelated isolate in 2013 it was not able to provide a conclusive link between the outbreak strain and a historical isolate from 2011

(11F310) since all outbreak strains had the same MLVA pattern (3-6-1-10-2-3-12) and the historical isolate had similar MLVA pattern (3-6-1-10-2-3-11/12).

Despite the technical limitation of phage typing, it was of value for investigating the outbreak of *Salmonella* Typhimurium (13) and confirming that mayonnaise is the source of infection.

Denovo assembly of WGS data of _Salmonella_ Dublin and _Salmonella_ Typhimurium strains:

We carried out *denovo* assembly for the raw Fastq paired end (PE) reads for all *Salmonella* Dublin and *Salmonella* Typhimurium strains using two different assemblers including Velvet available at Centre for genomic epidemiology (CGE) (http://www.genomicepidemiology.org/) and SPAdes available at Enterobase (http://enterobase.warwick.ac.uk/). We then assessed the quality of the assembly for each strain was assessed using Quast assessment tool (http://quast.bioinf.spbau.ru/).

SNP typing analyses of _Salmonella_ Dublin and _Salmonella_ Typhimurium outbreaks:

SNP analysis was carried out using CSIPhylogeny (https://cge.cbs.dtu.dk/services/CSIPhylogeny/) where raw reads were mapped to reference sequences (strain LT2 of *Salmonella* Typhimurium; accession number: AE006468 and strain CT_02021853 of *Salmonella* Dublin; accession number: CP001144) using BWA software (http://bio-bwa.sourceforge.net). The depth at each mapped position was calculated using genomeCoverageBed, which is part of BEDTools (https://bedtools.readthedocs.io/en/latest/). High quality SNPs were called using mpileup which is part of SAMTools (http://samtools.sourceforge.net). Genome mappings were then compared and an alignment of the SNPs are then created by concatenating the SNPs. A

maximum likelihood (ML) phylogenetic tree was then created based on the concatenated alignment of the high quality SNPs.

Determination of MLST, rMLST, cgMLST and wgMLST of *Salmonella* Dublin and *Salmonella* Typhimurium strains:

The assembled sequences of each strain were analyzed to detect the MLST, rMLST, cgMLST and wgMLST available at Enetrobase (http://enterobase.warwick.ac.uk/) and CGE (http://www.genomicepidemiology.org/).

Determination of prophage sequence profiles in *Salmonella* Dublin *and Salmonella* Typhimurium genomes:

Prophages were determined with the draft genomes generated by Velevt and SPAdes for all *Salmonella* Dublin and *Salmonella* Typhimurium strains using PHASTER (http://phaster.ca/).

We then used CSI phylogeny available at CGE (http://www.genomicepidemiology.org/) to construct a phylogenetic tree based on the SNPs of detected prophages. Phylogenetic trees were constructed using assembled genomes generated by Velvet and SPAdes assemblers to check if the assembly could affect the tree.

Determination of CRISPRs within *Salmonella* Dublin and *Salmonella* Typhimurium strains:

Spacers sequence within the draft genomes of all *Salmonella* Dublin and *Salmonella* Typhimurium strains were characterized using CRISPRFinder (http://crispr.i2bc.paris-saclay.fr/Server/).

Determination of plasmids within *Salmonella* Dublin and *Salmonella* Typhimurium strains:

We determined the plasmids within the draft genomes of all *Salmonella* Dublin and *Salmonella* Typhimurium strains using the plasmid database; PLSDB (https://ccb-microbe.cs.uni-saarland.de/plsdb/).

***In silico* analyses of antibiotic resistance within *Salmonella* Dublin and *Salmonella* Typhimurium strains:**

We determined acquired antibiotic resistance genes and mutations within the draft genomes of all *Salmonella* Dublin and *Salmonella* Typhimurium strains using ResFinder (https://cge.cbs.dtu.dk/services/ResFinder/).

RESULTS:

WGS-based subtyping:

(A) SNP based cluster analyses:

SNP based tree showed conclusively that the outbreak strains of *Salmonella* Typhimurium were grouped together in two clades and they are very closely related to strains isolated from mayonnaise (figure 1) confirming the source of outbreak is due to consumption of contaminated mayonnaise.

The outbreak isolates of *Salmonella* Dublin were closely related to each other (figure 2) and distinct from the non-outbreak isolates that were not readily distinguishable by PFGE. However, the source of *Salmonella* Dublin outbreak could not be determined

and outbreak isolates showed high genetic divergence from the raw-milk cheese isolates related to other outbreaks occurred in France (11).

Figure (1): Maximum likelihood phylogenetic tree of *Salmonella* Typhiurium strains based on single nucleotide polymorphisms determined from whole genome sequences. The scale represents the number of nucleotide substitutions per site. Bootstrap support values, given as a percentage of 1000 replicates, are shown on the branches. The tree shows conclusively that myonaise (marked with arrows) is the source of outbreak.

Figure (2): Maximum likelihood phylogenetic tree of *Salmonella* Dublin strains based on single nucleotide polymorphisms determined from whole genome sequences. The scale represents the number of nucleotide substitutions per site. Bootstrap support values, given as a percentage of 1000 replicates, are shown on the branches. All *Salmonella* Dublin isolates had indistinguishable pulsed-field gel electrophoresis profiles. Confirmed outbreak cases (n = 9) in October–November 2013 are grouped together in one cluster. However, the source of the outbreak could not be determined as outbreak isolates showed high genetic divergence to bacterial strains isolated from the raw-milk cheeses (marked with arrows) including isolate 2014SAL02972 from Morbier cheese (accession number; ERS2767809) and isolate 2015LSAL00258 from St. Nectaire cheese (accession number: ERS2767808) .

(B) MLST, rMLST, cgMLST and wgMLST:

As illustrated in table (1), all *Salmonella* Dublin strains including the outbreak and non-outbreak strains showed identical MLST (type 10). Interestingly, outbreak isolates of *Salmonella* Dublin displayed identical rMLST (type 1429) however, some of the non-outbreak strains showed the same rMLST. Moreover, the wgMLST was different among the outbreak strains however, the cgMLST was unique among outbreak strains and can easily separate the outbreak strain from the non-outbreak strains including the 2011 historical isolate (11F310).

Table (1): MLST, rMLST, cgMLST and wgMLST results of *Salmonella* Dublin outbreak and non-outbreak strains

Strain ID:	MLST:	rMLST:	cgMLST:	wgMLST:
Outbreak strains:				
902637	10	1429	38665	259116
MF036933	10	1429	38665	259117
MF036980	10	1429	38665	259118
517138	10	1429	38665	259121
MF6869	10	1429	38665	259127
M26560	10	1429	38665	259123
MF7067	10	1429	38665	259122
MF7174	10	1429	38665	259128
40986	10	1429	38665	259126
Non-outbreak strains:				
MF038630	10	1429	38666	259131
M1314220	10	26829	38664	259120

M54827	10	1429	38667	259129
MB12371	10	26829	38668	259130
MF5994	10	92451	38669	259145
MB7978	10	1429	38670	259133
B289223	10	1429	38671	259134
11F310	10	1429	38655	259135
MB98550	10	3696	38657	259142
MF8409	10	1429	38658	259139
W151R0	10	1429	38659	259140
B261193	10	92450	38660	259141
MP015199F	10	1429	38661	259148
Food isolates:				
*2014LSAL02972	10	1429	230922	283421
*2015LSAL00258	10	96856	146469	283422

Same results for MLST, rMLST, cgMLST and wgMLST were obtained from CGE and Enterobase using Velvet and SPAdes assemblers respectively.

**Salmonella* Dublin strains isolated from raw milk cheeses related to other outbreaks occurred in France (11).

On the other hand, MLST, rMLST, cgMLST and wgMLST could not discriminate between the outbreak and non-outbreak strains of *Salmonella* Typhimurium as illustrated in table (2).

Table (2): MLST, rMLST, cgMLST and wgMLST results of *Salmonella* Typhimurium outbreak and non-outbreak strains

Strain ID	MLST:	rMLST:	cgMLST:	wgMLST:
Food strains:				
*H133060375	19	1392	60658	70401

*H133060376	19	1392	60660	70402
*H133060377	19	1392	36749	70514
*H133060378	19	1392	60661	70403
Outbreak strains:				
H133000654	19	1392	36749	70398
H132940743	19	1392	36749	70404
H132940744	19	1392	60662	70405
H132940745	19	1392	60663	70406
H132940746	19	1392	36749	70431
H132940748	19	1392	60683	70432
H132940749	19	1392	36749	70433
H132940750	19	1392	60684	70439
H132940751	19	1392	60685	70440
H132940753	19	1392	61002	70834
H132940754	19	1392	36754	70835
H132940756	2392	1392	61001	70833
H133000645	19	1392	36749	
H133300609	19	1392	36749	70944
H132300541	19	1391	36751	70951
Non-outbreak strains:				
H133260293	19	1392	71438	84026
H132780266	19	1391	71450	84040
H132960590	19	1391	36751	84041
H132920685	19	1392	36763	84076
H132980531	19	1391	36774	87971
H121600325	19	1391	20224	87972
H122720573	19	1391	20848	87973
H12320661	19	1391	20882	87974
H123020544	19	1391	20711	87975
H122020454	19	1391	21310	88017
H124860455	19	26127	20800	88018

H133040470	19	1392	71422	84006
H1330400611	19	1392	71438	84025

*Strains of *Salmonella* Typhimurium isolated from mayonnaise.

Same results for MLST, rMLST, cgMLST and wgMLST were obtained from CGE and Enterobase using Velvet and SPAdes assemblers respectively.

(C) CRISPR typing:

All *Salmonella* Dublin isolates including outbreak and non-outbreak strains harbour one CRISPR locus and we observed 3 to 5 unique spacers for CRISPR1 locus. Identical spacers were detected among the outbreak and non-outbreak strains as shown in table (3). Interestingly, the number of spacers in three isolates (517138, MF7067 and W151R0) changed from (4 spacers) based on Velvet to (5 spacers) based on SPAdes.

Table (3): Number of spacers within CRISPR1 locus in all *Salmonella* Dublin strains analysed in this study.

Strain ID:	**Spacers No. (Velvet)**	**Spacers No. (SPAdes)**
Outbreak strains:		
902637	**5**	**5**
MF036933	**5**	**5**
MF036980	**5**	**5**
517138	**4**	**5**
MF6869	**5**	**5**
M26560	**5**	**5**
MF7067	**4**	**5**
MF7174	**5**	**5**
40986	**5**	**5**

Non-outbreak strains:		
MF038630	**5**	**5**
M1314220	**5**	**5**
M54827	**3**	**3**
MB12371	**5**	**5**
MF5994	**5**	**5**
MB7978	**5**	**5**
B289223	**5**	**5**
11F310	**5**	**5**
MB98550	**4**	**4**
MF8409	**5**	**5**
W151R0	**4**	**5**
B261193	**3**	**3**
MP015199F	**3**	**3**

All *Salmonella* Typhimurium isolates harbour 3 CRISPR loci. Identical spacers were detected among the outbreak and non-outbreak strains as shown in table (4). There was no difference between the numbers of spacers using different assemblers.

Table (4): Number of spacers within CRISPRs loci in all *Salmonella* Typhimurium strains analysed in this study.

Strain ID	**Spacers No.**

	(Velvet & SPAdes)		
Food strains:			
***H133060375**	9	13	9
***H133060376**	9	13	9
***H133060377**	9	13	9
***H133060378**	9	13	9
Outbreak strains:			
H133300609	9	13	9
H132940743	9	13	9
H132940744	9	13	9
H132940745	9	13	9
H132940746	9	13	9
H132940748	9	13	9
H132940749	9	13	9
H132940750	9	13	9
H132940751	9	13	9
H132940753	13	9	9
H132940754	9	13	9
H132940756	9	13	9
H133000645	9	13	9
H133000654	9	13	9
Non-outbreak strains			
H121600325	9	13	9
H122020454	9	13	9
H122720573	9	13	9
H123020544	9	13	9
H123920661	9	13	9
H124860455	9	13	9
H132780266	9	13	9
H132920685	9	13	9
H132960590	9	13	9

H132980531	9	13	9
H133040470	9	13	9
H133260293	9	13	9
H133400611	9	13	9

*Strains of *Salmonella* Typhimurium isolated from mayonnaise.

(D) Prophage sequence profiling:

All *Salmonella* Dublin strains including the outbreak strains are lysogenic for three prophages (Gifsy_2, 118970_sal3 and RE_2010). However, phylogenetic analyses of *Salmonella* Dublin strains based on the SNPs of prophages showed that outbreak strains are intermixed with the non-outbreak strains based on velvet assembler (figure 3) and SPAdes assembler (figure 4).

Figure (3): Maximum likelihood phylogenetic tree of *Salmonella* Dublin strains based on prophages SNPs using Velvet

Figure (4): Maximum likelihood phylogenetic tree of *Salmonella* Dublin strains based on prophages SNPs using SPAdes

All *Salmonella* Typhimurium genomes assembled by SPAdes revealed the presence of four prophages in all outbreak and non-outbreak strains including the three Salmonella prophages (Gifsy 2, RE-2010, and 118970_sal3) and the Edwardsiella specific phage (GF-2).
On the other hand, *Salmonella* Typhimurium genomes assembled by Velvet were lysogenic for two *Salmonella* specific prophages (Gifsy 2 and RE-2010). All strains except one outbreak isolate (H132940750) harbour *Salmonella* 118970_sal3 phage.

Interestingly, all strains harbour Edwardsiella GF-2 prophage except three outbreak isolates (H132940748, H133000645 and H133060376).

Phylogenetic analyses of *Salmonella* Typhimurium strains based on the SNPs of prophages showed that outbreak strains are intermixed with the non-outbreak strains using velvet assembler (figure 5) and using SPAdes assembler (figure 6).

Figure (5): ML phylogenetic tree of *Salmonella* Typhimurium strains based on prophages SNPs using Velvet

Figure (6): ML phylogenetic tree of *Salmonella* Typhimurium strains based on prophages SNPs using SPAdes

(E) <u>Plasmid typing:</u>

All outbreak and non-outbreak strains of *Salmonella* Dublin harbour identical plasmid type (except three non-outbreak isolates; M1314220, MB12371 and B261193) as shown in table (5).

Table (5): Distribution of plasmids among outbreak and non-outbreak strains of *Salmonella* Dublin.

	pSA19992307 (NZ_CP030208)	pSE81-1705 (NZ_CP018654)	Plasmid : 4 (LN829404)	pATCC39184 (NZ_CP019180)	pSDU2-USMARC-69807 (NZ_CP032381)	Plasmid : 3 (NZ_LN868945)
Outbreak strains:						
902637	Present	Present	Present	Absent	Absent	Absent
MF036933	Present	Present	Present	Absent	Absent	Absent

MF036980	Present	Present	Present	Absent	Absent	Absent
517138	Present	Present	Present	Absent	Absent	Absent
MF6869	Present	Present	Present	Absent	Absent	Absent
M26560	Present	Present	Present	Absent	Absent	Absent
MF7067	Present	Present	Present	Absent	Absent	Absent
MF7174	Present	Present	Present	Absent	Absent	Absent
40986	Present	Present	Present	Absent	Absent	Absent
Non-outbreak strains:						
MF038630	Present	Present	Present	Absent	Absent	Absent
M1314220	**Absent**	Present	Present	**Present**	Absent	Absent
M54827	Present	Present	Present	Absent	Absent	Absent
MB12371	**Absent**	Present	Present	**Present**	Absent	Absent
MF5994	Present	Present	Present	Absent	Absent	Absent
MB7978	Present	Present	Present	Absent	Absent	Absent
B289223	Present	Present	Present	Absent	Absent	Absent
11F310	Present	Present	Present	Absent	Absent	Absent
MB98550	Present	Present	Present	Absent	Absent	Absent
MF8409	Present	Present	Present	Absent	Absent	Absent
W151R0	Present	Present	Present	Absent	Absent	Absent
B261193	**Absent**	Present	Present	Absent	**Present**	Absent
MP015199F	Present	Present	Present	Absent	Absent	Absent
Food strains:						Absent
*2014LSAL02972	Present	Present	Absent	Absent	Absent	Present
*2015LSAL00258	Present	Present	Absent	Absent	Absent	Present

*_Salmonella_ Dublin strains isolated from raw milk cheeses related to other outbreaks occurred in France (11).

Same plasmids were determined using Velvet and SPAdes assemblers.

All outbreak and non-outbreak isolates of _Salmonell_a Typhimurium harbour 3 plasmids (pATCC14028, plasmid: 4 and pSE81-1705) except the outbreak strain H133300609 which

328 did not carry plasmid pATCC14028 but it harbours a different plasmid (pSLT_VNP20009)

329 instead (table 6).

330 **Table (6): Distribution of plasmids among outbreak and non-outbreak strains of**

331 ***Salmonella*** **Typhimurium.**

Strain ID	**pATCC14028 (NZ_CP034231)**	**Plasmid: 4 (LN829404)**	**pSE81-1705 (NZ_CP018654)**	**pSLT_VNP20009 (NZ_CP008745)**
Food strains:				
***H133060375**	Present	Present	Present	Absent
***H133060376**	Present	Present	Present	Absent
***H133060377**	Present	Present	Present	Absent
***H133060378**	Present	Present	Present	Absent
Outbreak strains:				
H132300541	Present	Present	Present	Absent
H132940743	Present	Present	Present	Absent
H132940744	Present	Present	Present	Absent
H132940745	Present	Present	Present	Absent
H132940746	Present	Present	Present	Absent
H132940748	Present	Present	Present	Absent
H132940749	Present	Present	Present	Absent

H132940750	Present	Present	Present	Absent
H132940751	Present	Present	Present	Absent
H132940753	Present	Present	Present	Absent
H132940754	Present	Present	Present	Absent
H132940756	Present	Present	Present	Absent
H133000645	Present	Present	Present	Absent
H133000654	Present	Present	Present	Absent
H133300609	**Absent**	Present	Present	**Present**
Non-outbreak strains:				
H121600325	Present	Present	Present	Absent
H122020454	Present	Present	Present	Absent
H122720573	Present	Present	Present	Absent
H123020544	Present	Present	Present	Absent
H123920661	Present	Present	Present	Absent
H124860455	Present	Present	Present	Absent
H132780266	Present	Present	Present	Absent
H132920685	Present	Present	Present	Absent
H132960590	Present	Present	Present	Absent
H132980531	Present	Present	Present	Absent
H133040470	Present	Present	Present	Absent
H133260293	Present	Present	Present	Absent
H133400611	Present	Present	Present	Absent

*Strains of *Salmonella* Typhimurium isolated from mayonnaise.

(F) <u>Antibiotic resistance profile:</u>

All *Salmonella* Dublin isolates including the outbreak and non-outbreak strains are resistant to aminoglycosides due to the acquisition of the *aac(6')-Iaa* gene. No mutations were detected against *gyrA* and *parC* genes in all isolates except one isolate (MF038630) that carried a non-synonyms mutation within the gyrase protein and it is associated with bacterial resistance to nalidixic acid (Table 7).

342

343 Table (7): *In silico* analyses results of antimicrobial resistance genes and mutations within all

344 *Salmonella* Dublin strains

Strain ID:	**Acquired antibiotic resistance genes:**	**Mutations in *gyrA gene:***	**Mutations in *parC gene:***
Outbreak strains:			
902637	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MF036933	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MF036980	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
517138	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MF6869	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
M26560	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MF7067	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MF7174	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
40986	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
Non-outbreak strains:			
MF038630	Aminoglycoside (*aac(6')-Iaa*)	Absent	**Present**
M1314220	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
M54827	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MB12371	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MF5994	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MB7978	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
B289223	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
11F310	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MB98550	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MF8409	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
W151R0	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
B261193	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
MP015199F	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
Food strains:			

*2014LSAL02972	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent
*2015LSAL00258	Aminoglycoside (*aac(6')-Iaa*)	Absent	Absent

Salmonella Dublin strains isolated from raw milk cheeses related to other outbreaks occurred in France (11).

All the *Salmonella* Typhimurium isolates of both the outbreak and non-outbreak group are resistant to aminoglycosides due to the acquisition of the "*aac(6')-Iaa* gene". No known mutations were detected against *gyrA* and *parC* (Table 8).

Table (8): *In silico* analyses results of antimicrobial resistance genes and mutations within all *Salmonella* Typhimurium strains

Strain ID	**Acquired antibiotic resistance genes:**	**Mutations in *gyrA gene:***	**Mutations in *parC gene:***
Food strains:			
***H133060375**	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
***H133060376**	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
***H133060377**	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
***H133060378**	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
Outbreak strains:			
H132940743	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940744	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940745	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940746	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940748	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940749	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940750	Aminoglycoside (aac(6')-Iaa)	Absent	Absent

H132940751	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940753	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940754	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132940756	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H133000645	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H133000654	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H133300609	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
Non-outbreak strains:			
H121600325	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H122020454	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H122720573	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H123020544	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H123920661	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H124860455	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132780266	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132920685	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132960590	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H132980531	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H133040470	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H133260293	Aminoglycoside (aac(6')-Iaa)	Absent	Absent
H133400611	Aminoglycoside (aac(6')-Iaa)	Absent	Absent

*Strains of *Salmonella* Typhimurium isolated from mayonnaise

DISCUSSION:

Salmonellosis is one of the most common foodborne diseases worldwide and has been associated with high morbidity and mortality rates. It is estimated that over 680,000 humans throughout the world are killed each year by iNTS. The most predominant iNTS serovars are Typhimurium, Enteritidis and Dublin (14, 15). It is therefore very crucial to use accurate, reliable and highly discriminative subtyping methods for epidemiological surveillance and outbreak investigation.

Although PFGE is considered as current gold standard for all *Salmonella* serotypes, it has its limitations moreover, variation between laboratories has been reported when identifying the source of infection and discriminating between the outbreak and non-outbreak isolates (16).

Other phenotypic tools such as phage typing and antimicrobial resistance profiling have been crucial in the outbreak investigations (16, 17). Furthermore, MLVA has been used to distinguish between genetically closely related strains and trace back the sources of disease outbreaks related to food (16, 18).

Genotypic approaches have ameliorated the methods for carrying out outbreak investigation and epidemiological surveillance (19). The advent of whole genome sequencing (WGS) has opened the possibilities to enhance the typing approaches for outbreak investigation and epidemiological surveillance. In our study, WGS data have been analyzed to test the suitability of different approaches as subtyping tool for *Salmonella enterica* surveillance. We therefore carried out retrospective investigation of two different outbreaks of *Salmonella* Typhimurium and *Salmonella* Dublin that occurred in 2013 in UK and Ireland respectively (6, 9) using different WGS-subtyping methods.

384 In this study, single nucleotide polymorphism (SNP)-based cluster analysis of *Salmonella*
385 Typhimurium genomes revealed well supported clades, that were concordant with
386 epidemiologically defined outbreak and confirmed the source of outbreak is due to
387 consumption of contaminated mayonnaise. Although SNP-analyses of *Salmonella* Dublin
388 genomes confirmed the outbreak, however the source of infection could not be determined.
389
390 On the other the WGS-subtyping methods including MLST, rMLST, wgMLST, cgMLST
391 showed limited discrimination for the outbreak and non-outbreak isolates of *Salmonella*
392 Typhimurium strains. However, cgMLST defined the genetic relatedness among *Salmonella*
393 Dublin isolates more precisely and confirmed there is no relation among the 2013 outbreak
394 isolates and the 2011 historical isolate (11F310) of *Salmonella* Dublin.
395
396 It was reported that MLST might not be the most suitable epidemiological tool (20) but it is
397 best for analyzing the genetic diversity of the strain and analyze the core and conserved genes
398 of pathogens that are of public importance.
399
400 The cgMLST bridges the classic MLST with the novel WGS-based approach since it
401 combines the discriminatory power of MLST with large-scale data obtained from WGS
402 enabling to exploit a considerable number of gene targets throughout the bacterial genome
403 which would maximize the quality and resolution for surveillance and research works.
404 A recent study showed that cgMLST has shown the robustness of cgMLST as a tool to
405 investigate multi-country outbreak of *Salmonella* Enteritidis in Europe (21).
406 The difference between the cgMLST and wgMLST is that unlike cgMLST, wgMLST indexes
407 the variation of pre-defined set of genes from both core and accessory genes (22). Another
408 retrospective study on 8 different outbreaks associated with verotoxigenic *Escherichia coli*

(VTEC) O157:H7 in Canada showed that wgMLST provided higher discrimination than PFGE and MLVA (23).

Research studies have shown that cgMLST and wgMLST are viable typing methods for outbreak surveillance. In our study, cgMLST proved to provide higher discriminatory resolution for differentiating *Salmonella* Dublin isolates of outbreak group from the non-outbreak group. However, both cgMLST and wgMLST were unsuccessful in differentiating outbreak-related *Salmonella* Typhimurium isolates from outbreak-unrelated isolates.

Bacterial genome comprises a considerable amount (10% to 20%) of prophages integrated in their core genome (24). Prophages harbor genes for antimicrobial resistance, virulence and toxins which contribute to the genetic diversity of bacterial strains making prophages a potential marker for discriminating *Salmonella* serovars (25). However, one of the limitations of using prophage sequence profiles for *Salmonella* subtyping is the sensitivity and accuracy of the assembly as some prophage regions might be lost during assembly. We used two different *denovo* assemblers (SPAdes and Velvet) and found that prophage sequence profiling could not differentiate between the outbreak and non-outbreak isolates.

Recent studies have suggested that high throughput CRISPR typing has the potential to be used for epidemiological surveillance and investigation of *Salmonella* outbreaks (26, 27). . However, in our study, we detected identical spacers among outbreak and non-outbreak associated strains indicating that CRISPR typing is not useful for the surveillance of *Salmonella enetrica* outbreaks as we showed in our previous studies (28, 29) however, it might be useful for the discrimination among different *Salmonella* serovars.

Plasmid profiles and antimicrobial- susceptibility profiling have been used as an epidemiological tool since many decades. However, it was reported that analysis of plasmid profiles provided higher discrimination in the outbreak investigations than analysis of antimicrobial-susceptibility pattern (30, 31). In our study both plasmid typing and *in silico* analysis of antibiotic resistance were unable to discriminate between the outbreak isolates and non-outbreak isolates.

In this study, we compared several retrospective WGS-based subtyping methods and we showed that SNP-based cluster analysis is superior to other subtyping methods to define the source of outbreak in real-time.

In conclusion, foodborne salmonellosis is an important concern for public health therefore, it is crucial to use accurate, reliable and highly discriminative subtyping methods for epidemiological surveillance and outbreak investigation. The rapid development of next-generation sequencing (NGS) technology and bioinformatics tools have enabled WGS of any bacterial strain feasible. Various typing tools have been proposed by using WGS data but currently, the adoption of WGS-based methods have proved to be difficult due to lack of standardization. There are many layers on obtaining WGS data and there is need of standardization from the type of sequencers used to the bioinformatics analysis. Therefore, the emerging genetic analysis techniques should be combined with conventional phenotypic and molecular methods for routine surveillance and outbreak investigation until the WGS-based methods can be fully exploited, improved and standardized.

List of abbreviations:

CGE: Centre for Genomic Epidemiology

cgMLST: core genome multilocus sequence typing

CRISPRs: clustered regularly interspaced short palindromic repeats

iNTS: invasive NTS

ML: maximum likelihood

MLST: multilocus sequence typing

MLVA: multiple loci VNTR analysis

NGS: next generation sequencing

NTS: Non-typhoidal *Salmonella*

PE: paired end

PFGE: pulsed field gel electrophoresis

rMLST: ribosomal MLST

SNP: single nucleotide polymorphism

wgMLST: whole genome MLST

WGS: whole genome sequence

Declarations section:

- Ethical Approval and Consent to participate: Not Applicable

- Consent for publication: Authors agreed to publish

- Availability of supporting data: Available in supplementary tables 1 and 2

- Competing interests: Not Applicable

- Funding: MM is funded by a Quinton Hogg start-up award, University of Westminster.

- Authors' contributions: MM: designed the study, contributed to data analyses, wrote the initial draft of the manuscript. ST: contributed to data analyses and contributed to manuscript writing. All authors reviewed and edited the manuscript.

- Acknowledgements: Authors would like to thank Falling Walls Lab London for the invitation to present this research work.

- Authors' information: MM is a lecturer of Medical Microbiology, School of Life Sciences, University of Westminster. ST is a MSc Medical Microbiology student at School of Life Sciences, University of Westminster.

REFERENCES:

(1) Jajere SM. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. Vet World 2019;12(4):504-521.

(2) Barnett R. Typhoid fever. The Lancet 2016;388(10059):2467.

(3) Stanaway JD, Parisi A, Sarkar K, Blacker BF, Reiner RC, Hay SI, et al. The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet Infectious Diseases 2019;19(12):1312-1324.

(4) Mohammed M, Cormican M. Whole genome sequencing provides insights into the genetic determinants of invasiveness in *Salmonella* Dublin. Epidemiology Infect 2016 Aug;144(11):2430-2439.

(5) Tennant SM, MacLennan CA, Simon R, Martin LB, Khan MI. Nontyphoidal salmonella disease: Current status of vaccine research and development. Vaccine 2016 Jun 3;34(26):2907-2910

(6) Orzechowska B. The War between Bacteria and Bacteriophages. Growing and Handling of Bacterial Cultures Rijeka: IntechOpen; 2019. p. Ch. 6.

(7) Crump JA, Mintz ED. Global trends in typhoid and paratyphoid Fever. Clin Infect Dis 2010 Jan 15;50(2):241-246.

(8) Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, et al. The global burden of nontyphoidal Salmonella gastroenteritis. Clin Infect Dis 2010 Mar 15;50(6):882-889.

(9) Mohammed M, Le Hello S, Leekitcharoenphon P, Hendriksen R. The invasome of *Salmonella* Dublin as revealed by whole genome sequencing. BMC Infect Dis 2017 Aug 4;17(1):544-x.

(10) Mohammed M, Delappe N, O'Connor J, McKeown P, Garvey P, Cormican M. Whole genome sequencing provides an unambiguous link between *Salmonella* Dublin outbreak strain and a historical isolate. Epidemiol Infect 2016 Feb;144(3):576-581.

(11) Cadel-Six S, Vignaud ML, Mohammed M. Draft Genome Sequences of *Salmonella* enterica subsp. enterica Serovar Dublin Strains from St. Nectaire and Morbier Cheeses Characterized by Multilocus Variable-Number Tandem-Repeat Analysis Profiles Associated with Two Fatal Outbreaks in France. Microbiol Resour Announc 2019 Jan 3;8(1):10.1128/MRA.01361-18. eCollection 2019 Jan.

(12) Ford L, Wang Q, Stafford R, Ressler KA, Norton S, Shadbolt C, et al. Seven *Salmonella* Typhimurium Outbreaks in Australia Linked by Trace-Back and Whole Genome Sequencing. Foodborne Pathog Dis 2018 May;15(5):285-292.

(13) Ashton PM, Peters T, Ameh L, McAleer R, Petrie S, Nair S, et al. Whole Genome Sequencing for the Retrospective Investigation of an Outbreak of *Salmonella* Typhimurium DT 8. PLoS Curr 2015 Feb 10;7:10.1371/currents.outbreaks.2c05a47d292f376afc5a6fcdd8a7a3b6.

(14) Gordon MA. Invasive nontyphoidal Salmonella disease: epidemiology, pathogenesis and diagnosis. Curr Opin Infect Dis 2011 Oct;24(5):484-489.

(15) Mohammed M, Vignaud ML, Cadel-Six S. Whole-Genome Sequences of Two *Salmonella* enterica Serovar Dublin Strains That Harbor the viaA, viaB, and ompB Loci of the Vi Antigen. Microbiol Resour Announc 2019 Apr 4;8(14):10.1128/MRA.00028-19.

(16) Lienemann T, Kyyhkynen A, Halkilahti J, Haukka K, Siitonen A. Characterization of *Salmonella* Typhimurium isolates from domestically acquired infections in Finland by phage typing, antimicrobial susceptibility testing, PFGE and MLVA. BMC Microbiol 2015 Jul 2;15:131-8.

30) Mohammed M, Cormican M. Whole genome sequencing provides possible explanations for the difference in phage susceptibility among two *Salmonella* Typhimurium phage types (DT8 and DT30) associated with a single foodborne outbreak. BMC Research Notes 2015;8(1):728.

(18) Vignaud ML, Cherchame E, Marault M, Chaing E, Le Hello S, Michel V, et al. MLVA for *Salmonella* enterica subsp. enterica Serovar Dublin: Development of a Method Suitable for Inter-Laboratory Surveillance and Application in the Context of a Raw Milk Cheese Outbreak in France in 2012. Front Microbiol 2017 Feb 27;8:295.

(19) Ferrari RG, Panzenhagen PHN, Conte-Junior CA. Phenotypic and Genotypic Eligible Methods for *Salmonella* Typhimurium Source Tracking. Front Microbiol 2017 Dec 22;8:2587.

(20) Sails AD, Swaminathan B, Fields PI. Utility of multilocus sequence typing as an epidemiological tool for investigation of outbreaks of gastroenteritis caused by *Campylobacter jejuni*. J Clin Microbiol 2003 Oct;41(10):4733-4739.

(21) Pearce ME, Alikhan NF, Dallman TJ, Zhou Z, Grant K, Maiden MCJ. Comparative analysis of core genome MLST and SNP typing within a European *Salmonella* serovar Enteritidis outbreak. Int J Food Microbiol 2018 Jun 2;274:1-11.

(22) Henri C, Leekitcharoenphon P, Carleton HA, Radomski N, Kaas RS, Mariet JF, et al. An Assessment of Different Genomic Approaches for Inferring Phylogeny of Listeria monocytogenes. Front Microbiol 2017 Nov 29;8:2351.

(23) Rumore J, Tschetter L, Kearney A, Kandar R, McCormick R, Walker M, et al. Evaluation of whole-genome sequencing for outbreak detection of Verotoxigenic Escherichia coli O157:H7 from the Canadian perspective. BMC Genomics 2018 Dec 4;19(1):870-3.

(24) Manal M. Who Fights Whom? - Understanding the Complex Dynamics of Bacteria Phage Interaction Using Anderson Phage Typing System. Journal of Infectious Diseases & Therapy 1970 Jan 01(3):1-2.

(25) Mottawea W, Duceppe MO, Dupras AA, Usongo V, Jeukens J, Freschi L, et al. *Salmonella enterica* Prophage Sequence Profiles Reflect Genome Diversity and Can Be Used for High Discrimination Subtyping. Front Microbiol 2018 May 4;9:836.

(26) Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, et al. CRISPR typing and subtyping for improved laboratory surveillance of Salmonella infections. PLoS One 2012;7(5):e36995.

(27) Shariat N, Dudley EG. CRISPRs: molecular signatures used for pathogen subtyping. Appl Environ Microbiol 2014 Jan;80(2):430-439.

(28) Mohammed M. Phage typing or CRISPR typing for epidemiological surveillance of *Salmonella* Typhimurium? BMC Research Notes 2017;10(1):578.

(29) Mohammed M. Old school wins: outbreak investigation of foodborne salmonellosis. 2018; Available at: http://atlasofscience.org/old-school-wins-outbreak-investigation-of-foodborne-salmonellosis/. Accessed Jan 15, 2020.

(30) Holmberg SD, Wachsmuth IK, Hickman-Brenner FW, Cohen ML. Comparison of plasmid profile analysis, phage typing, and antimicrobial susceptibility testing in characterizing *Salmonella* Typhimurium isolates from outbreaks. J Clin Microbiol 1984 Feb;19(2):100-104.

(31) Brunner F, Margadant A, Peduzzi R, Piffaretti JC. The plasmid pattern as an epidemiologic tool for *Salmonella* Typhimurium epidemics: comparison with the lysotype. J Infect Dis 1983 Jul;148(1):7-11.