

Whole genome sequencing provides an unambiguous link between *Salmonella* Dublin outbreak strain and a historical isolate

M. MOHAMMED¹*, N. DELAPPE², J. O'CONNOR², P. McKEOWN³,
P. GARVEY³ AND M. CORMICAN¹

¹School of Medicine, National University of Ireland, Galway, Ireland

²National Salmonella, Shigella and Listeria Reference Laboratory, Galway, Ireland

³Health Protection Surveillance Centre, Dublin, Ireland

Received 28 January 2015; Final revision 12 June 2015; Accepted 26 June 2015;
first published online 13 July 2015

SUMMARY

Salmonella enterica subsp. *enterica* serovar Dublin is an uncommon cause of human salmonellosis; however, a relatively high proportion of cases are associated with invasive disease. The serotype is associated with cattle. A geographically diffuse outbreak of *S. Dublin* involving nine patients occurred in Ireland in 2013. The source of infection was not identified. Typing of outbreak associated isolates by pulsed-field gel electrophoresis (PFGE) was of limited value because PFGE has limited discriminatory power for *S. Dublin*. Whole genome sequencing (WGS) showed conclusively that the isolates were closely related to each other, to an apparently unrelated isolate from 2011 and distinct from other isolates that were not readily distinguishable by PFGE.

Key words: Molecular epidemiology, outbreaks, public health microbiology, *Salmonella* typing, surveillance.

INTRODUCTION

Salmonellosis is a common foodborne disease in humans worldwide [1]. It is caused primarily by consumption of food contaminated with non-typhoidal *Salmonella enterica* serovars [2]. Illness is often associated with self-limiting gastrointestinal disorders [3]; however, invasive disease occurs in a proportion of patients [4] causing severe systemic disease. *Salmonella enterica* serovars Enteritidis and Typhimurium account for most human infection. *S. Dublin* (1,9,12[Vi]:g,p:-) has a similar antigenic formula to *S. Enteritidis* (1,9,12:g,m:-); however, the epidemiology is distinct. *S. Dublin* is strongly associated

with cattle [5] and it can be transmitted to humans via contact with infected animals or consumption of contaminated products including raw milk and raw milk cheeses [6]. *S. Dublin* has also been detected in other host species including poultry, sheep, dog, horse and mink [7, 8]. Although human infection with *S. Dublin* is much less common than infection with *S. Enteritidis* a higher proportion of *S. Dublin* infections are invasive. The invasiveness of *S. Dublin* may be related to expression of the Vi (virulence) antigen which is also commonly found in human-adapted *S. Typhi*.

Because *S. Dublin* is a relatively infrequent serotype, specific subtyping methods such as phage typing or multiple-locus variable-number tandem repeat analysis (MLVA) are not widely available although application of the *S. Enteritidis* MLVA protocol [9] has recently been described as of some value.

* Author for correspondence: Dr M. Mohammed, School of Medicine, National University of Ireland, Galway, Ireland. (Email: drmanal20@hotmail.com)

Pulsed-field gel electrophoresis (PFGE) with *Xba*I according to the PulseNet protocol [10] is applicable; however, the indistinguishable PFGE pattern offers limited microbiological support for investigating a suspected *S. Dublin* outbreak.

Whole genome sequencing (WGS) has the potential to transform the confidence with which laboratories can demonstrate the degree of similarity between isolates of serovars such as *S. Dublin* which are difficult to type by established methods. In Ireland *S. Dublin* accounts for between two and seven clinical cases annually. However, approximately one third of human cases are invasive. In October–November 2013 nine isolates of *S. Dublin* were received at the national reference laboratory. An outbreak investigation was initiated. Here we describe the epidemiological and laboratory aspects of that investigation including the first reported application of WGS to support investigation of an outbreak of *S. Dublin*.

METHODS

Epidemiological investigations

Outbreak case definition

Outbreak cases were defined as persons with: (i) laboratory-confirmed infection with fully antimicrobial-susceptible *S. Dublin* exhibiting the MLVA pattern designated as (3-6-1-10-2-3-12); (ii) gastroenteritis symptoms including diarrhoea; (iii) onset of illness between October 2013 and November 2013 and (iv) reported in any region throughout Ireland.

Foodborne illness questionnaire

Cases were contacted and asked to provide their exposure histories using the National Salmonella Trawling Questionnaire. Supplementary information on food purchases was obtained using loyalty card information from a national supermarket chain. Consent was obtained from the loyalty card customer associated with four of the outbreak cases to look for a commonly purchased food item within a 6-week purchasing history prior to disease onset.

Microbiological investigations

The nine human faecal isolates detected during October–November 2013 were characterized as *S. enterica* serotype Dublin according to the White–Kauffmann–Le Minor scheme [11]. They were phenotypically negative for the Vi antigen. Antimicrobial

susceptibility testing against a panel of antimicrobials was determined using the broth dilution method [12]. The interpretation was conducted using the European Committee on Antimicrobial Susceptibility Testing (EUCAST) clinical breakpoints (<http://www.eucast.org>). PFGE was performed using the PulseNet protocol [10]. MLVA was also performed. The quoted MLVA pattern is based on the application of the Centers for Disease Control and Prevention (CDC) *S. enterica* serovar Enteritidis [9].

The nine outbreak-associated *S. Dublin* isolates isolated during October–November 2013 (Supplementary Table S1) in addition to the apparently epidemiologically unrelated isolate from December 2013 were selected for WGS. The other eight unrelated isolates were chosen in a blinded fashion and were also subjected to WGS for comparison. All isolates are of human origin and isolated from faeces except five isolates were invasive.

Genomic DNA was extracted using the QIAamp® DNA Mini kit (Qiagen, UK). DNA quality and quantity were checked by gel electrophoresis and the Qubit® quantification platform (Invitrogen, USA), respectively. Twenty microlitres of DNA (20–50 ng/μl) from each isolate was submitted for Illumina sequencing using 100-bp paired-end (PE) reads. The depth of coverage was estimated as 100×.

The quality of the short-read data was evaluated using the FastQC toolkit (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>). Low-quality reads were removed and adapter sequences were trimmed using the ea-utils package (<https://code.google.com/p/ea-utils/>). Sequence reads of each isolate were mapped against the reference genome of *S. Dublin* strain CT_02021853 (accession no.: NC_011205) isolated in 2004 from human faeces in Macau, Asia using BWA software [13]. Single nucleotide polymorphisms (SNPs) were identified using samtools mpileup [14]. SNPs over 60 quality threshold (i.e. a 1/1 000 000 chance of a miss-called variant) were only accepted (automated called variants for each isolate are provided in Supplementary Table S2). SNPs were compared to the reference strain and a maximum-likelihood (ML) phylogeny of the isolates was constructed using MEGA6 software [15] using the GTR+Gamma +Invariant model as the best-fit model of nucleotide substitution. *De novo* assembly of short PE reads was performed using Velvet [16]. Generated multi-contig draft genomes for each isolate were screened for acquired antimicrobial resistance genes using ResFinder [17]. Furthermore, housekeeping genes

were screened for mutations that are associated with resistance to antibiotics including *gyrA* and *parC* genes where certain mutations are linked to resistance to fluoroquinolone antibiotics including nalidixic acid [18]. Draft genomes were also screened for the Vi antigen genes [19] that are responsible for synthesis and transportation of Vi proteins (viPs).

RESULTS

Epidemiological investigations

Outbreak description

A total nine confirmed outbreak cases were reported with onset dates from October 2013 to November 2013. None was associated with invasive disease. Cases were resident throughout Ireland from North-West to South-East (Fig. 1). Cases were predominantly female (7/9) and ranged in age from 30 to 85 years (Table 1). Two of the nine cases were hospitalized. Where reported, the other seven cases had symptom duration of 3–8 days. All cases recovered.

Cases' exposure

No single food source was identified to be linked to the outbreak based on the completed National *Salmonella* Trawling Questionnaire and the loyalty card information from a national supermarket chain. Although cheese consumption was considered a possible source of infection, as it was reported by seven of eight cases, no specific cheese product emerged from the epidemiological investigation and *Salmonella* was not detected from cheese samples obtained from retail and restaurant premises that cases had reported visiting.

Microbiological investigations

The nine outbreak associated isolates were susceptible to all antimicrobial agents tested. They were indistinguishable on PFGE but they were also indistinguishable from many other *S. Dublin* isolates typed by PFGE. Isolates had the same MLVA pattern (3-6-1-10-2-3-12). A non-complete match to this MLVA pattern (3-6-1-10-2-3-11/12) was observed in a historical isolate from 2011.

WGS analysis

WGS analysis revealed that the nine *S. Dublin* isolates from October–November 2013 were very closely

related with only 1–9 SNP difference. This contrasts with 74–88 SNP difference between the outbreak isolates and an apparently epidemiologically unrelated isolate from December 2013 (O13). Other isolates showed high genetic divergence from the outbreak isolates except a historical isolate from 2011 (J11) that was very closely related to the 2013 cluster with a maximum 15 SNP difference (Fig. 2).

WGS data analysis of the outbreak isolates revealed no antibiotic resistance genes or mutations within housekeeping genes *gyrA* or *parC* consistent with the susceptible phenotype. However, the mutation detected within the *gyrA* gene in the epidemiologically unrelated isolate (O13) from December 2013 had changed one amino acid within the gyrase protein and this change is associated with resistance to nalidixic acid. No genetic determinants of antimicrobial resistance were detected within the other isolates consistent with the susceptible phenotype.

Generated multi-contig draft genomes of *S. Dublin* isolates were screened for the Vi antigen genes. Vi antigen-encoding genes including *tviA*, *tviB*, *tviC*, *tviD*, *tviE* (coding for polysaccharide capsule) and *vexA*, *vexB*, *vexC*, *vexD*, *vexE* (coding for the polysaccharide transportation proteins) were entirely absent from the outbreak isolates and other unrelated isolates including invasive isolates consistent with the phenotype since all isolated tested negative for the Vi antigen by slide agglutination using *Salmonella* Vi antisera.

DISCUSSION

A diffuse outbreak of *S. Dublin* gastroenteritis was detected in Ireland over a period of 6 weeks in October–November 2013. Unusually for *S. Dublin* all cases were non-invasive. No source of infection was identified.

The wide geographical distribution of the outbreak throughout Ireland suggests that the outbreak might be linked to a widely distributed food item (batch of dairy product), with the shape of the epidemic curve being consistent with a continuous source outbreak. Moreover, the source and/or transmission routes of the *S. Dublin* gastroenteritis outbreak might differ from the typical sources and/or transmission routes for previously identified sporadic *S. Dublin* cases within Ireland since the percentage of female cases was higher than normal and all cases suffered gastrointestinal disease while invasive infection was very common in sporadic *S. Dublin* cases in Ireland. The

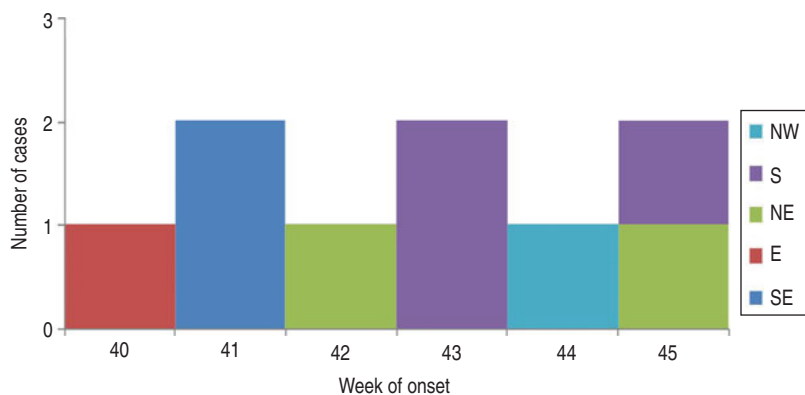


Fig. 1. *Salmonella* Dublin outbreak cases by week of onset, Ireland, October–November 2013 ($n = 9$).

Table 1. Age-sex distribution of *Salmonella* Dublin outbreak cases (October–November 2013), Ireland

Age group (years)	Male	Female	Total
25–34	–	1	1
35–44	–	1	1
45–54	1	–	1
55–64	–	4	4
≥65	1	1	2
Total	2	7	9

ability of *S. Dublin* to cause invasive infection may be related to expression of the Vi antigen which is also commonly found in human-adapted *S. Typhi*. However, in this study we found that Vi antigen is absent from all *S. Dublin* isolates including invasive isolates.

Conventional typing methods were of limited value for investigation of the *S. Dublin* outbreak. There is limited confidence in the laboratory evidence for similarity between *S. Dublin* isolates because of the limited discriminatory power of currently widely used phenotypic (antimicrobial susceptibility testing) and genotypic (PFGE) typing methods.

Since the outbreak occurred a specific MLVA scheme for *S. Dublin* has been developed [20]. MLVA is rapid, simple and suitable for processing large numbers of isolates and it can distinguish between clonal isolates that are indistinguishable by PFGE. However, it has certain limitations [21] and although MLVA was of value in discriminating an epidemiologically unrelated *S. Dublin* isolate from December 2013 from the outbreak strain it was not able to provide a conclusive link between the outbreak strain and a historical isolate.

Although PFGE and MLVA allow clustering by similarity/difference they do not provide reliable representation of phylogenetic relationships and isolates that are distantly related can have similar PFGE or MLVA as a result of chance thus limiting their effectiveness in linking historical isolates to identifying transmission routes; therefore, a powerful rapid real-time method is required.

Several retrospective WGS analyses compared to the routine typing methods during outbreaks were carried out. Studies showed that WGS is superior to conventional typing methods including PFGE and MLVA and can provide more detailed information on the genetic characteristic of the strains of foodborne and non-foodborne outbreaks [22–24]. Furthermore, WGS provides the potential to define the genetic basis for virulence and antimicrobial resistance. It has the potential to replace all currently used phenotypic and molecular typing methods.

We therefore applied WGS technology to *S. Dublin* outbreak isolates and a number of unrelated isolates to define the genetic relatedness in isolates more precisely. WGS analysis not only confirmed the very close relationship in the outbreak isolates but it also succeeded in identifying a historical isolate from 2011 that was very similar to the outbreak strain. The MLVA pattern of the outbreak isolates designated as (3-6-1-10-2-3-12) was not a complete match for the 2011 isolate (3-6-1-10-2-3-11/12), thus WGS gave greater assurance of similarity. There is no known epidemiological link between the outbreak strain and the historical isolate.

As the cost of the WGS declines and the speed and accuracy of sequencing improves, it is expected that WGS will become a standard tool and replace

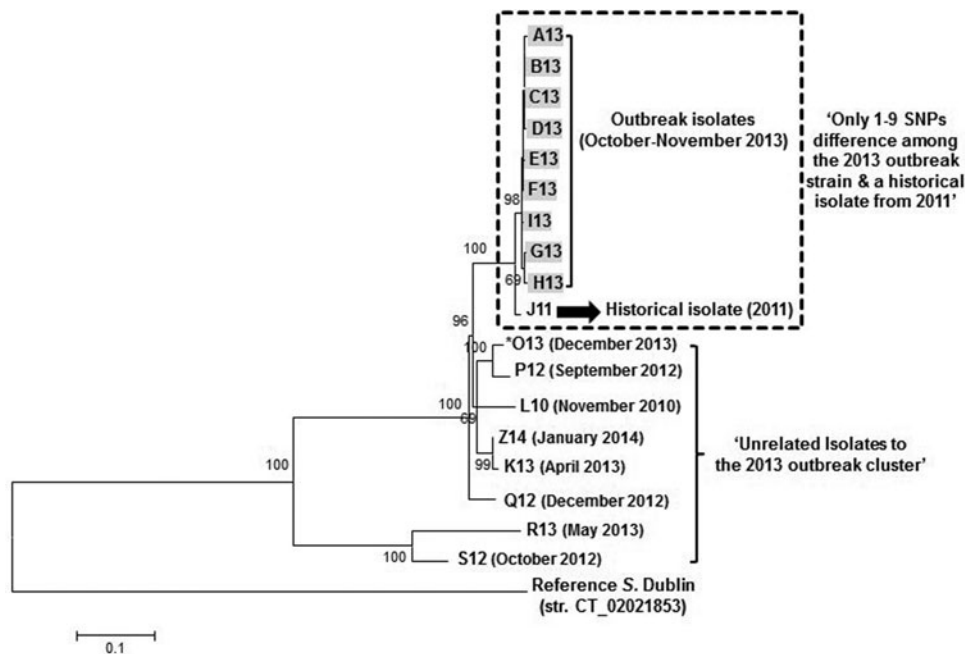


Fig. 2. Maximum-likelihood phylogenetic tree of *Salmonella* Dublin strains based on single nucleotide polymorphisms (SNP) determined from whole genome sequences. Sequence reads were mapped against the reference genome of *S. Dublin* (strain CT_02021853). The scale represents the number of nucleotide substitutions per site. Bootstrap support values, given as a percentage of 1000 replicates, are shown on the branches. All *S. Dublin* isolates had indistinguishable pulsed-field gel electrophoresis profiles. Confirmed outbreak cases ($n=9$) in October–November 2013 are highlighted in grey. Other isolates show high genetic divergence to the outbreak cluster including the epidemiologically unrelated isolate (*O13). However, a historical isolate from 2011 (J11) is very closely related to the 2013 cluster with a maximum of 15 SNP difference.

conventional typing in the detection and investigation of the outbreaks in real time. This ability to show definitive evidence of near identity between outbreak strains and historical isolates as illustrated here will grow in public health importance as databases of WGS of human and food-related isolates increase.

SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0950268815001636>.

ACKNOWLEDGEMENTS

The authors thank all members of the outbreak control team in Ireland. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

DECLARATION OF INTEREST

None.

REFERENCES

1. **de Jong B, Ekdahl K.** The comparative burden of salmonellosis in the European Union member states, associated and candidate countries. *BMC Public Health* 2006; **6**: 4.
2. **Hald T, et al.** A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Analysis* 2004; **24**: 255–269.
3. **Zhang SP, et al.** Molecular pathogenesis of *Salmonella enterica* serotype Typhimurium-induced diarrhea. *Infection and Immunity* 2003; **71**: 1–12.
4. **Gordon MA.** Invasive nontyphoidal *Salmonella* disease: epidemiology, pathogenesis and diagnosis. *Current Opinion in Infectious Diseases* 2011; **24**: 484–489.
5. **Kingsley RA, Baumler AJ.** Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. *Molecular Microbiology* 2000; **36**: 1006–1014.
6. **Vaillant V, et al.** Outbreak of *Salmonella* Dublin infection in France, November–December 1995. *Eurosurveillance* 1996; **1**: 9–10.
7. **Liebana E, et al.** Investigation of the genetic diversity among isolates of *Salmonella enterica* serovar Dublin from animals and humans from England, Wales and Ireland. *Journal of Applied Microbiology* 2002; **93**: 732–744.

8. **Dietz HH, et al.** Outbreak of *Salmonella* Dublin-associated abortion in Danish fur farms. *Canadian Veterinary Journal* 2006; **47**: 1201–1205.
9. **PulseNet.** PulseNet standard operating procedure for PulseNet MLVA of *Salmonella enterica* serotype Enteritidis – Applied Biosystems Genetic Analyzer 3500 Platform. PulseNet USA, 2013 (<http://www.cdc.gov/pulsenet/PDF/se-abi-3500-508c.pdf>).
10. **Ribot EM, et al.** Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathogens and Diseases* 2006; **3**: 59–67.
11. **Grimont PA, Weill FX.** *Antigenic Formulae of the Salmonella Serovars*, 9th edn. Paris: World Health Organization Collaborating Centre for Reference and Research on Salmonella, Pasteur Institute, 2007 (<http://www.pasteur.fr/ip/portal/action/WebdriveActionEvent/oid/01s-000036-089>).
12. **Wikler MA, et al.** Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically. Approved standard, 8th edn. Clinical and Laboratory Standards Institute (CLSI) document M07-A8. Wayne: CLSI, 2009 (<http://www.scribd.com/doc/89642226/CLSI-2009>).
13. **Li H, Durbin R.** Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
14. **Li H, et al.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
15. **Tamura K, et al.** MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 2013; **30**: 2725–2729.
16. **Zerbino DR, Birney E.** Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 2008; **18**: 821–829.
17. **Zankari E, et al.** Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy* 2012; **67**: 2640–2644.
18. **Saenz Y, et al.** Mutations in *gyrA* and *parC* genes in nalidixic acid-resistant *Escherichia coli* strains from food products, humans and animals. *Journal of Antimicrobial Chemotherapy* 2003; **51**: 1001–1005.
19. **Pulickal AS, et al.** Prevalence and genetic analysis of phenotypically Vi-negative *Salmonella typhi* isolates in children from Kathmandu, Nepal. *Journal of Tropical Pediatrics* 2013; **59**: 317–320.
20. **Kjeldsen MK, et al.** Multiple-locus variable-number tandem repeat analysis of *Salmonella enterica* subsp. *enterica* serovar Dublin. *Journal of Applied Microbiology* 2014; **116**: 1044–1054.
21. **Fabre L, et al.** CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PLoS ONE* 2012; **7**: e36995.
22. **Koser CU, et al.** Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *New England Journal of Medicine* 2012; **366**: 2267–2275.
23. **den Bakker HC, et al.** Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis. *Emerging Infectious Diseases* 2014; **20**: 1306–1314.
24. **Dallman TJ, et al.** The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing *Escherichia coli* serogroup O26:H11. *Epidemiology and Infection* 2015; **143**: 1672–1680.