Whole genome sequencing improved case ascertainment in an outbreak of Shiga toxin-producing *Escherichia coli* O157 associated with raw drinking milk

H. BUTCHER[1]†, R. ELSON[2]†, M. A. CHATTAWAY[3], C. A. FEATHERSTONE[4], C. WILLIS[5], F. JORGENSEN[5], T. J. DALLMAN[3], C. JENKINS[3]*, J. McLAUCHLIN[5,6], C. R. BECK[7] AND S. HARRISON[1]

[1] *Devon, Cornwall and Somerset Public Health England Centre, Exeter, Devon, UK*
[2] *Gastrointestinal and Emerging Zoonotic Infections Department, Public Health England, UK*
[3] *Gastrointestinal Bacteria Reference Unit, Public Health England, London, UK*
[4] *Animal and Plant Health Agency, Veterinary Investigation Centre, Thirsk, Yorkshire, UK*
[5] *Food Water and Environmental Microbiology Laboratory Services, Public Health England, UK*
[6] *University of Liverpool, Institute of Infection and Global Health, Liverpool, UK*
[7] *Field Epidemiology Service, Public Health England, UK*

SUMMARY

Five cases of STEC O157 phage type (PT) 21/28 reported consumption of raw cows' drinking milk (RDM) produced at a dairy farm in the South West of England. STEC O157 PT21/28 was isolated from faecal specimens from milking cows on the implicated farm. Whole genome sequencing (WGS) showed that human and cattle isolates were the same strain. Further analysis of WGS data confirmed that sequences of isolates from an additional four cases (who did not report consumption of RDM when first questioned) fell within the same five single nucleotide polymorphism cluster as the initial five cases epidemiologically linked to the consumption of RDM. These four additional cases identified by WGS were investigated further and were, ultimately, associated with the implicated farm. The RDM outbreak strain encoded *stx2a*, which is associated with increased pathogenicity and severity of symptoms. Further epidemiological analysis showed that 70% of isolates within a wider cluster containing the outbreak strain were from cases residing in, or linked to, the same geographical region of England. During this RDM outbreak, use of WGS improved case ascertainment and provided insights into the evolution of a highly pathogenic clade of STEC O157 PT21/28 *stx2a* associated with the South West of England.

Key words: Bacterial typing, *Escherichia coli*, foodborne infections, gastrointestinal infections, public health microbiology.

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) cause a spectrum of illness ranging from mild to severe,

* Author for correspondence: Dr C. Jenkins, Gastrointestinal Bacteria Reference Unit, Public Health England, 61 Colindale Avenue, London, NW9 5EQ
(Email: claire.jenkins@phe.gov.uk)
† These authors contributed equally to this work.

bloody diarrhoea. Cardiac, neurological and renal complications, such as haemolytic uraemic syndrome (HUS) develop in 5–15% of cases, dependent on the age and sex of the case [1]. STEC are defined by the presence of the Shiga toxin-encoding genes, *stx1* and *stx2*, which can be divided in to subtypes Stx1a–1c and Stx2a–2 g [2]. The incidence of STEC infection is highest in children aged <5 years. The incubation period ranges from 6 h to 10 days, averaging 2–4

days and the infectious dose is low. The natural reservoir of STEC is the gastrointestinal tract of ruminant animals, particularly cattle. Human infection can occur via contaminated foods, beverages or water, direct contact with infected animals or their environment, or by secondary spread from cases, particularly in family groups within households [1].

In England, the most common STEC serogroup associated with human disease is O157, with around 900 cases reported each year and about 25% of cases linked to epidemiologically confirmed outbreaks [1]. Outbreaks of STEC O157 are detected through (i) routine investigation of cases by identifying common exposures between cases, (ii) detection of the same microbiological subtypes in isolates from cases that are geographically or temporally linked, and (iii) detection of an increase in the number of cases in a particular location or associated with a particular subtype [3]. Presumptive STEC O157 isolated at local or regional hospital laboratories from faecal specimens taken from cases with symptoms of gastrointestinal disease are submitted to the Gastrointestinal Bacteria Reference Unit (GBRU) at Public Health England (PHE). Prior to 2014, all confirmed STEC O157 isolates were typed by phage typing [4] and multi-locus variable number tandem repeat analysis (MLVA) [3]. Between April 2014 and March 2015, all isolates potentially linked to outbreaks of STEC O157 were also typed by whole genome sequencing (WGS) and added to the PHE STEC O157 WGS database in order to validate and evaluate the WGS approach [5].

Advances in WGS methodologies have resulted in the ability to perform high throughput sequencing of bacterial genomes at low cost, making WGS a viable alternative to traditional typing methods for public health surveillance and outbreak detection [6]. The utility of WGS for the investigation of outbreaks has already been demonstrated for several gastrointestinal pathogens [7–9]. For STEC O157, Dallman *et al.* [5] showed that WGS analysis facilitated identification of temporally distinct cases sharing common exposures and delineated those that shared epidemiological and temporal links. Furthermore, comparison with MLVA showed that while MLVA is as sensitive as WGS, WGS provides a more timely resolution to outbreak clustering [5].

The aim of this study was to describe a public health investigation into an outbreak of STEC O157 linked to the consumption of raw cows' drinking milk (RDM) and to highlight the role of WGS in prospective case ascertainment and robust resolution of the outbreak cluster. In addition, we explored the deeper phylogenetic relationship between the outbreak strain and other isolates in this dataset, and speculate on the impact this analysis might have on directing future outbreak investigations.

METHODS

Case ascertainment by enhanced epidemiological surveillance

Presumptive cases of STEC were reported directly to PHE centres by clinical microbiologists at local hospital laboratories and a standardized STEC Enhanced Surveillance Questionnaire (SESQ) (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/323423/VTEC_Questionnaire.pdf) was administered to cases either by local health protection professionals or environmental health practitioners (EHPs). Data from the questionnaires were included in the National Enhanced Surveillance System for STEC in England (NESSS) [1]. The case definition was defined as a case of STEC O157 PT21/28 with the same MLVA profile or a single-locus variant (SLV) of that profile (see below) between 1 September and 30 November 2014.

Molecular typing of STEC O157 by MLVA and WGS

At the time of the outbreak, all isolates of STEC O157 submitted to GBRU were typed using MLVA as described previously [3] and WGS (both typing methods were performed in real time concurrently for comparison purposes). Isolates with identical MLVA profiles, or with profiles that differed at one locus (SLV), were considered to be microbiologically linked. Double-locus variants (DLVs) were considered to be part of an outbreak only if an epidemiological link existed, for example if the cases had consumed the same food or had the same environmental exposure.

For WGS, DNA was extracted from cultures of STEC O157 for sequencing on the Illumina HiSeq 2500 instrument as described previously [9]. High quality Illumina reads were mapped to the STEC O157 reference genome Sakai (Genbank accession no. BA000007) using BWA-MEM [10]. Single-nucleotide polymorphisms (SNPs) were identified using GATK2 [11] in unified genotyper mode. Core genome positions that had a high-quality SNP (>90% consensus, minimum depth 10x, genotype quality $\geqslant 30$) in at

least one isolate were extracted. SNP positions that were present in at least 80% of isolates were used to derive maximum-likelihood phylogenies with RaxML [12] using the GTRCAT model with 1000 iterations.

Genomes were compared to the sequences held in the PHE STEC O157 WGS database. This database comprises genomes from more than 1500 cultures of STEC O157 submitted to GBRU between 1982 and 2015. The majority of isolates were from human cases in England reporting domestically acquired infection, although cases associated with foreign travel and isolates from domestic cattle were also included. Isolates of STEC O157 with <5 SNP differences within their core genome were considered closely related and likely to have an epidemiological link [5]. At PHE, an outbreak investigation is initiated for 5-SNP clusters comprising ⩾5 isolates identified within a 30-day time-frame [5]. Hierarchical single linkage clustering was performed on the pairwise SNP difference between all isolates at various distance thresholds (Δ250, Δ100, Δ50, Δ25, Δ10, Δ5, Δ0). The result of the clustering is a SNP address that can be used to describe the population structure based on clonal groups. Although isolates >5 SNPs apart are unlikely to be part of the same temporally linked outbreak, deeper phylogenetic relationships within the 10 or 25 SNP clusters may provide epidemiologically useful information or associations. Shiga toxin (stx) subtyping was performed as described by Ashton *et al.* [13].

Following confirmation of a temporal signal using Path-O-Gen (http://tree.bio.ed.ac.uk/software/pathogen/), timed phylogenies were constructed using BEAST-MCMC v. 1.80 [14]. Alternative clock models and population priors were computed and their suitability assessed based on Bayes factor tests. The highest supported model was a relaxed lognormal clock rate under a constant population size. All models were run with a chain length of 1 billion. A maximum clade credibility tree was constructed using TreeAnnotator v. 1.75 [14].

FASTQ reads from all sequences in this study and the PHE STEC O157 WGS data can be found at the PHE Pathogens BioProject at the National Center for Biotechnology Information (Accession PRJNA248792).

Spatial analysis methods

Postcodes were geocoded and spatially joined to the 2011 census tract data at medium super output area level. Coordinates for the centroid of each middle super output area were calculated using the British National Grid projection. All analyses were performed using ArcGIS software (ESRI, USA) or SatScan [15]. Spatial clustering was detected using a discrete Poisson model and the maximum cluster size was set at 50% of the population at risk. This analysis was based on the likely location at which the case was exposed to the pathogen. For sporadic cases, this was either the postcode of residence or postcode of UK travel destination if the cases had travelled in the 7 days prior to onset of symptoms. For cases linked to outbreaks, this was the postcode where the outbreak occurred or postcode of RDM production. Only one geographical location per event was included to best reflect each case's location of exposure to infection.

Microbiological examination of food and environmental samples

Food and environmental samples (Table 1) were collected and transported in accordance with the Food Standards Agency (FSA) Food Law Code of Practice (https://www.food.gov.uk/enforcement/codes-of-practice/food-law-code-of-practice-2015). Samples were collected from cases' homes by EHPs, and from the farm where the implicated RDM was produced by sampling officers from the FSA Dairy Hygiene Inspectorate and transported to PHE Food, Water and Environmental Microbiology Laboratories at Porton, London, York or Birmingham in cold boxes at a temperature of between 0 °C and 8 °C and tested within 24 h of collection

Tests for the detection of STEC O157, *Salmonella* spp., *Campylobacter* spp., *Listeria* spp. in 25 ml milk were performed. Enumeration of coliform bacteria, *E. coli*, coagulase-positive staphylococci, aerobic colony count and *Listeria* spp. (including *L. monocytogenes*) was performed using dilutions of milk samples. The protocols for the International Standard methods can be found at: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm.

Swabs were examined for the presence of STEC O157 by suspending in 100 ml mTSB and processed as described above.

Real-time polymerase chain reaction (PCR) was used to examine samples for the presence of STEC O157 based on CEN/ISO TS 13136 as described previously [9]. Enrichment broths that were PCR positive for *stx* were subcultured onto MacConkey agar and cefixime tellurite sorbitol MacConkey agar and up to 50 colonies retested using the same PCR assay.

Table 1. *Microbiological results obtained from milk and environmental swab samples examined during the outbreak investigation*

Date	Sampling place	Sample type*	No. of samples	Microbiological results STEC O157
1 Oct. 2014	Cases home (West Midlands)	Empty milk bottle	1	Not isolated
2 Oct. 2014	Cases home (Hampshire)	Raw milk in sealed bottles	11	Not isolated (*E. coli* O157 antigen-encoding gene (*rfb*O157) and *stx2* detected by PCR in one sample)
7 Oct. 2014	Dairy farm (South West England)	Raw bulk tank milk	7	Not detected
		Environmental swabs from tank room tables, bulk tank ladder, teat cab liners, tank room sink	9	Not isolated

* The samples tested were produced by the implicated dairy farm.

An empty milk bottle was examined by rinsing with 225 ml mTSB, followed PCR examination for the presence of STEC O157 and by culture and as described above.

Veterinary investigation and microbiological examination of animal faecal specimens

A visit to the farm was undertaken by a veterinary investigation officer from the Animal and Plant Health Agency (APHA) in order to confirm the animals on the farm were the source of the infection, and to establish if there were any farm practices which might lead to increasing the likelihood of faecal contamination of RDM. Thirty faecal specimens from milking cows were collected during this inspection and tested using immunomagnetic separation culture methodology as described by Pritchard *et al.* [16].

RESULTS

Descriptive epidemiology

On 29 and 30 September 2014, the Devon, Cornwall and Somerset (DCS) Public Health England Centre (PHEC) was alerted to two cases of STEC O157 not resident within the DCS area, both reporting consumption of RDM originating from a farm in the South West (SW) of England. The farm sold bottled RDM, cheese and other dairy products locally and nationally through online sales delivered by courier. A multidisciplinary outbreak control team (OCT) meeting was convened on 2 October to identify potential sources and implement control measures to prevent further cases. The food business operator at the farm was advised to suspend the sale of RDM and conduct a product recall. Bulk supply lines were identified and further distribution was terminated. On 3 October 2014 PHE and the FSA issued a joint press statement advising of the investigation into the farm and in November the FSA published a message on its website reiterating its advice that RDM should not be consumed by children and other vulnerable groups.

A total of four primary cases and one secondary case, all non-residents in the DCS area, were notified directly to the OCT. An additional three primary cases and one secondary case, also all non-residents in the DCS area, were identified by MLVA and WGS and subsequently linked to the outbreak (Table 2). Seven of the nine cases were male. The median age was 5 years (range 1–49 years, mean 11 years). The duration of symptoms ranged from 2 to 17 days (mean 7 days, median 6 days). Two cases of HUS were reported (Table 2). The epidemic timeline indicates that contaminated food products were available for consumption over a period of 6 weeks (Fig. 1).

Microbiological examination of animal faecal specimens, food and environmental samples

Twelve samples of RDM, including residue from an empty bottle, were examined by culture and PCR. One sample was positive for the *E. coli* O157 antigen encoding gene (*rfb*O157) and *stx2*. However, STEC O157 was not cultured from the RDM samples, possibly due to the low number of bacteria present. The environmental samples collected from the dairy farm on 6 October 2014 were negative for STEC O157 (Table 1). *Salmonella enterica* serovar Mbandaka

Table 2. *Summary of the typing data relating to the human and animal isolates linked to the RDM outbreak and the five sporadic cases within the 25-SNP cluster*

Identifier	Date presented to primary care	Region	HUS	Sex	MLVA profile*	SNP address†
Human outbreak cases within the same 5-SNP cluster						
Outbreak case 1, E	24 Sept. 2014	Hampshire, Isle of Wight and Dorset	No	M	8,7,13,5,5,3,6,9	4·4·4·611·887·929·1083
Outbreak case 2, E	27 Sept. 2014	West Midlands	Yes	F	8,2,13,5,5,3,6,9	4·4·4·611·887·929·1083
Outbreak case 3, E	3 Oct. 2014	West Midlands	No	M	8,7,14,5,5,3,6,9	4·4·4·611·887·929·1263
Outbreak case 4, W	7 Oct. 2014	Anglia and Essex	No	M	8,7,13,5,5,3,6,9	4·4·4·611·887·929·1221
Outbreak case 5, W	7 Oct. 2014	London	No	M	8,7,13,5,5,3,6,9	4·4·4·611·887·929·1221
Outbreak case 6, W‡	20 Oct. 2014	London	Yes	M	8,2,13,5,5,3,6,9	4·4·4·611·887·929·1221
Outbreak case 7, W	20 Oct. 2014	West Midlands	No	M	8,7,13,5,5,3,6,9	4·4·4·611·887·929·1221
Outbreak case 8, E	23 Oct. 2014	Hampshire, Isle of Wight and Dorset	No	F	8,7,13,5,5,3,6,9	4·4·4·611·887·929·1083
Outbreak case 9, E‡	26 Oct. 2014	Hampshire, Isle of Wight and Dorset	No	M	8,7,13,5,5,3,6,9	4·4·4·611·887·929·1083
Cattle						
Cow 1		Devon, Cornwall and Somerset			8,7,13,5,5,3,6,9	4·4·4·611·887·929·1221
Cow 2		Devon, Cornwall and Somerset			8,7,13,5,5,18,6,9	4·4·4·611·887·929·1221
Cow 3		Devon, Cornwall and Somerset			8,7,13,5,5,3,6,9	4·4·4·611·887·929·1221
Human sporadic cases within the 25-SNP cluster						
Case 10	7 Oct. 2014	Devon, Cornwall and Somerset	No	M	8,7,7,5,4,3,6,9	4·4·4·611·2458·2535·2740
Case 11	8 Oct. 2014	Devon, Cornwall and Somerset	No	M	8,7,14,5,5,3,7,9,	4·4·4·611·2459·2536·2742
Case 12	9 Oct. 2014	London	No	F	8,7,12,5,5,3,7,9	4·4·4·611·930·972·1126
Case 13‡	21 Oct. 2014	Devon, Cornwall and Somerset	No	M	8,7,12,5,5,3,7,9	4·4·4·611·930·972·2035
Case 14	29 Oct. 2014	Devon, Cornwall and Somerset	No	F	8,7,12,6,5,3,7,9	4·4·4·611·930·972·2035

RDM, Raw cows' drinking milk; SNP, single nucleotide polymorphism; HUS, haemolytic uraemic syndrome; MLVA, multi-locus variable number tandem repeat analysis.

E, Identified by analysis of data collected by NESSS; W, identified by WGS.

* Outbreak MLVA profile 8,7,13,5,5,3,6,9 or single-locus variant thereof.

† Five-SNP cluster outbreak SNP address 4·4·4·611·887·929.% (where % represents any number).
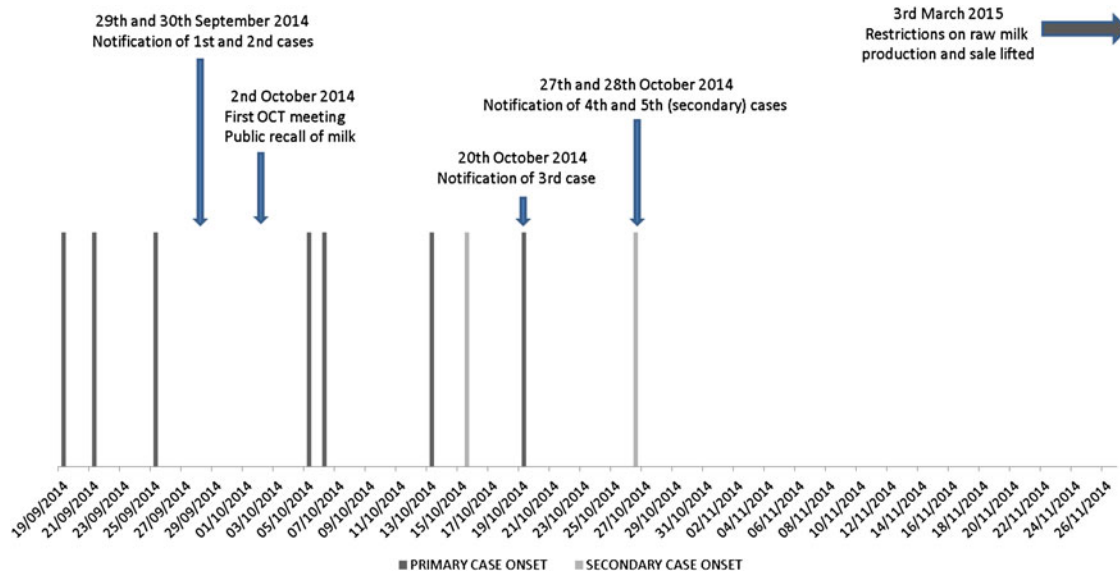
‡ Secondary case.

Fig. 1. Timeline for the raw cows' drinking milk STEC O157 PT21/28 outbreak showing dates of onset of symptoms for each case ($n = 9$), case notification and public health interventions.

was detected in two of the seven bulk tank milk samples. No other pathogens were detected in any milk or environmental swab samples. The seven bulk tank milk samples all gave results within the regulatory plate count and coliform standards as stipulated in the Food Hygiene Regulations [17].

STEC O157 *stx2* PT21/28 was identified in three of the 30 bovine faecal specimens with the same MLVA profile as the outbreak strain (or an SLV thereof). The WGS analysis showed that the isolates from the cattle fell within the same 5-SNP WGS cluster as the human cases, consistent with the farm animals being the source of the STEC O157 infection in the human cases (Fig. 2). A rigorous investigation was conducted of procedures and practices on the farm, including animal management, milking and bottling. Although cattle faeces were identified as the source, no breaches in farm practices that might have led to increasing the likelihood of faecal contamination of RDM were identified.

Case ascertainment supported by MLVA and WGS data

The MLVA outbreak profile derived from the isolates for the five cases notified directly to the OCT was 8-7-13-5-5-3-6-9 or an SLV of that profile (Table 2). Following the epidemiological identification of the initial five cases, real-time analysis of national MLVA data on all isolates submitted to GBRU identified an

additional nine cases of STEC O157 *stx2* PT21/28 with the same MLVA profile or closely related profiles (Table 2). The WGS data confirmed that the isolates from four of these nine cases fell within the same 5-SNP cluster as the initial five cases (Table 2, Fig. 2). The SNP address of the outbreak strain, representing the 5-SNP cluster, was designated 4·4·4·611·887·929.% (where % represents any number) (Table 2). There was 1 SNP difference between outbreak cases 1, 2, 8, 9 and outbreak cases 4, 5, 6, 7 (and cows 1, 2, 3). Case 3 had an additional SNP. There was no correlation between time and tree topology suggesting this diversity most likely occurred at source on the farm.

The MLVA SLVs observed within the nine confirmed outbreak cases (Table 2) were not epidemiologically informative. Of note, the MLVA profiles belonging to these four cases were identical to or SLVs of the outbreak profile, whereas the five isolates with DLVs or triple locus variants of the outbreak profile all fell outside the 5-SNP threshold (25–58 SNPs from the outbreak cluster).

The case questionnaires of all nine additional cases were reviewed by the OCT for evidence of the consumption of RDM. None of the cases could be epidemiologically linked to consumption of RDM from the implicated farm following this initial review. The nine additional cases identified by MLVA and WGS were investigated further. The cases were either re-interviewed and asked specific questions about their consumption of dairy products and UK travel in the 7
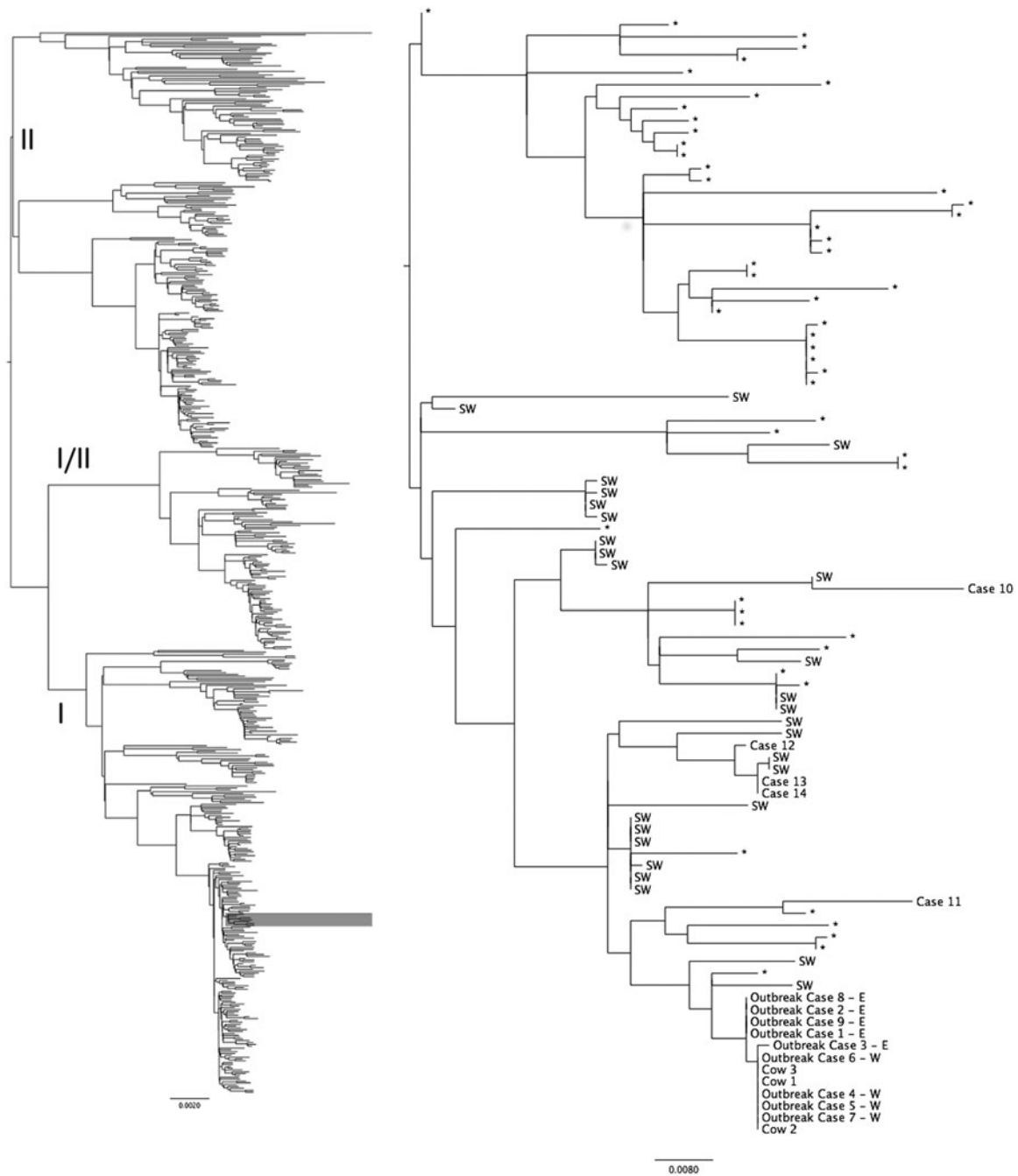
Fig. 2. The phylogeny of STEC O157 in England is illustrated on the left-hand side of the figure. Each isolate represents a 25-SNP cluster. The shaded area shows the position in the phylogeny of the 'South West' clade that includes the raw cows' drinking milk outbreak isolates and is enlarged on the right-hand side of the figure. All the isolates in the 'South West' clade encoded *stx2a* only. Further details of all the cases labelled are shown in Table 2. Those cases resident in, or reporting recent travel to, the South West of England are labelled SW. E, Identified by analysis of data collected by National Enhanced Surveillance System for STEC in England; W, identified by WGS.

days before they became ill, or their names and post-codes were found on the implicated farm distribution list. Following these in-depth interviews and subsequent follow-up investigations, three of the four cases that fell within the same 5-SNP cluster as the outbreak strain were, ultimately, linked to the consumption of RDM from the implicated farm or reported recent travel close to where the farm was situated. One case reported consumption of RDM but a direct link with the implicated farm could not be confirmed.
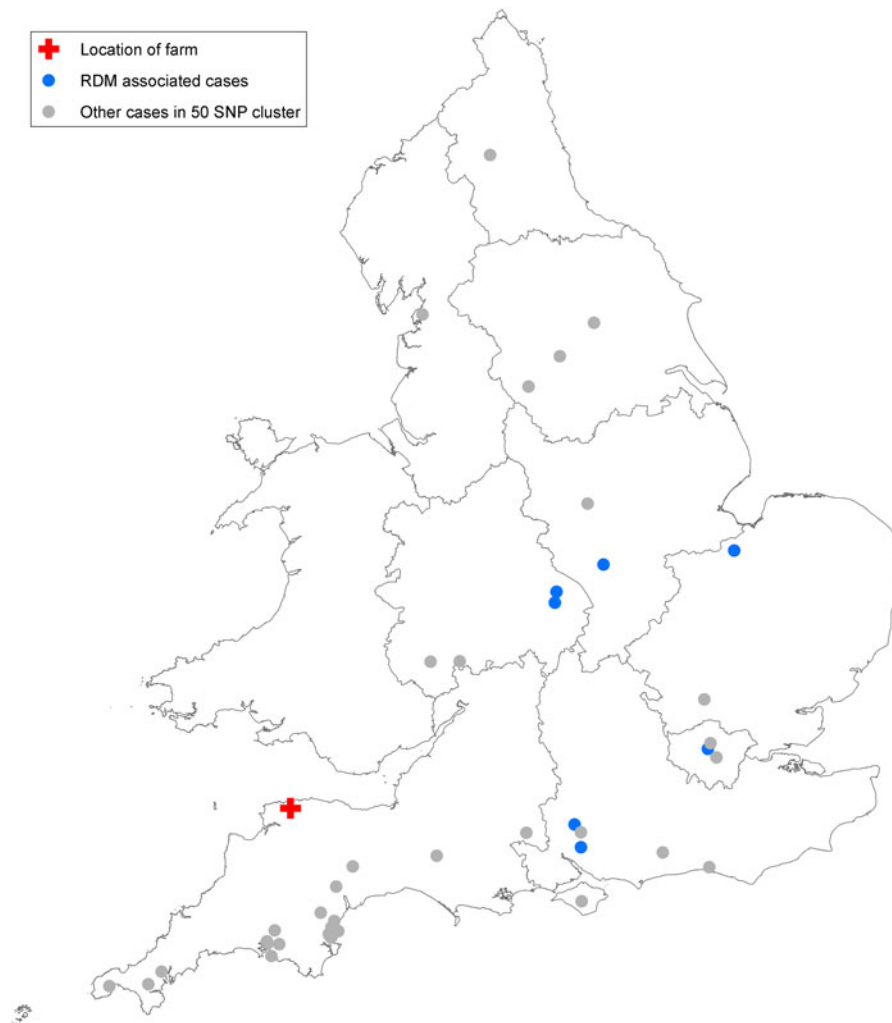
Fig. 3. Map showing the location of the cases linked to the raw cows' drinking milk (RDM) outbreak (blue) and cases within the outbreak cluster reported to National Enhanced Surveillance System for STEC in England between 2009 and 2015 (grey). Nine cases are represented by seven locations as two pairs of cases resided at the same address.

Evolutionary context of the outbreak strain

Of the five additional cases that did not fall within the 5-SNP cluster (referred to as sporadic cases in Table 2), all had isolates that were DLVs to the outbreak MLVA profile (Table 2) with no exposure to RDM. However, they did fall within a 25-SNP cluster of the outbreak cases and four of the five cases were resident in the DCS area (Fig. 2). Analysis of the wider cluster associated with the outbreak strain reported to NESSS between 2009 and 2015 showed that 64% ($n = 32$) of cases were deemed sporadic with the remainder being linked to six different outbreaks (Fig. 3). Two of these outbreaks were linked to schools, two were linked to farms (one farm was linked to two separate outbreaks in consecutive years) in SW England and one was linked to the consumption of RDM (Fig. 4).

Epidemiological analysis of the cases within the cluster, showed that of 33 primary or co-primary cases unrelated to the outbreak reported between 2009 and 2015, 70% ($n = 23$) were residents of SW England or had travelled there within 5 days before the onset of illness. Spatial analysis of the geographical location of the presumed exposure of cases revealed a highly significant cluster in the Devon and Cornwall area [observed cases 16, expected cases <1, relative risk (RR) 45, $P < 0.001$] (Fig. 4). Rates of infection with this strain were significantly lower in other parts of England (observed cases 1, expected cases 15, RR 0·04, $P < 0.001$) (Fig. 4). The strains comprising this cluster were isolated between 2000 and 2015.

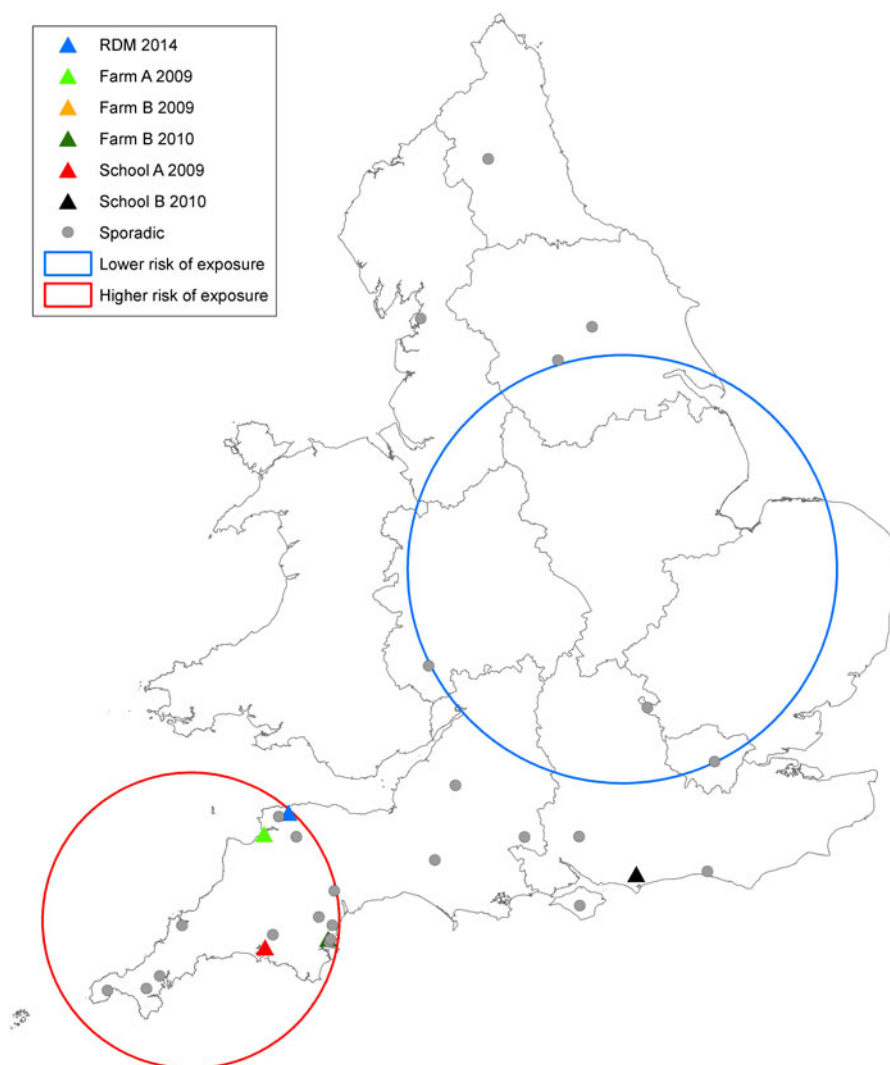The WGS *stx2* subtyping data showed that the outbreak strain, and all the isolates in the 'South West'

Fig. 4. Map showing the single point of exposure for cases falling within the outbreak cluster between 2009 and 2015. RDM, Raw cows' drinking milk.

clade, encoded *stx2a* only (Fig. 2), whereas the progenitor of this clade encoded both *stx2a* and *stx2c* [18]. Dated phylogenetic analysis indicated that the isolates in the 'South West' clade lost the bacteriophage-encoded *stx2c*-encoding gene approximately 16 years ago, shortly after the PT21/28 lineage evolved about 25 years ago [18].

DISCUSSION

RDM has a diverse microbial flora which can include pathogens transmissible to humans and, although most commonly sourced from cows, is also produced and marketed from sheep, goats, horses, donkeys and camels within the EU [19, 20]. The main microbiological hazards associated with human illness are *Campylobacter* spp., *Salmonella* spp., *Brucella melitensis, Mycobacterium bovis*, tick-borne encephalitis virus and STEC [19]. Contamination by these hazards can arise from direct excretion into the milk from animals with systemic infection as well as from localized infections, such as mastitis, and faecal contamination during milking or from the wider farm environment [19]. In England, Wales and Northern Ireland RDM may be sold directly to the consumer at the farm gate, in a farmhouse catering operation, through a milk roundsman, through the internet or through sales by farmers at farmers' markets [21]. The restrictions on the sale of RDM is governed by Food Hygiene (England) Regulations [17] which includes a microbiological standard of plate count at 30 °C of ⩽20 000 c.f.u./ml and coliforms of <100 c.f.u./ml.

Of note, all RDM samples from the implicated farm were compliant with these standards, although a pathogen (*S.* Mbandaka) was isolated from the bulk tank milk taken directly from the farm. The same *Salmonella* serovar was isolated from the faeces of one of the STEC O157 cases associated with this outbreak, and *S.* Mbandaka was later isolated from a sample of unpasteurized cheese produced on the implicated farm.

The consumption of RDM contaminated with STEC is a public health problem in countries where RDM is commercially available, with outbreaks reported in USA, Europe, Africa and Asia [22–25]. In England, outbreaks of STEC O157 linked to the consumption of RDM are rare with the last outbreak being recorded in 2002. Of the nine milkborne outbreaks of STEC O157 documented between 1992 and 2000, five were linked to the consumption of RDM and four to pasteurization failures by Gillespie and *et al.* [26]. These authors found that small farm dairies that bottled their own milk were identified as a significant problem due to the lack of regular testing of their product for pathogenic bacteria. The farm implicated in this outbreak was classed as a small farm, with 150 cattle and no other animals being farmed. There was no evidence of a lack of regular testing of the RDM product. RDM production was a small component of the business and most of the milk was sold as pasteurized. A previous study of RDM carried out by PHE (designated the Public Health Laboratory Service at that time) between 1996 and 1997, showed that 41 (3·7%) of 1097 samples were contaminated with potentially pathogenic bacteria, including *Salmonella* (five samples), *Campylobacter* (19 samples) and STEC O157 (three samples) [27].

Due to restrictions on the sale of RDM, that only allow direct sales from the farm to the final consumer (and not via an intermediate retailer), milkborne outbreaks associated with RDM are smaller than those caused by pasteurization failures and, therefore, more difficult to detect. However, trends in distribution of RDM appear to be changing, with farms more commonly using the internet for sale of their product, followed by delivery either by a farm van or a courier company. This can be seen in the outbreak described here, which included cases as widespread as Hampshire, London and the Midlands, despite the farm being located in SW England.

Four cases did not report the consumption of RDM during their initial interviews. The reasons for this may be that they did not know that the milk was unpasteurized or simply failed to recall consuming the product until prompted. Of note, were the low mean and median of the ages of the cases associated with this outbreak. Severe symptoms of gastrointestinal disease caused by STEC O157 are seen more frequently in younger children [1]. It is of concern that the families were not aware of the risk or, if they were aware, felt the risk was acceptable. This assessment may be influenced by information promoting the perceived benefits of RDM without balancing this against the risk of foodborne infection.

In this study, MLVA reliably confirmed that the initial four primary and one secondary cases with an epidemiological link to the consumption of RDM from the farm were microbiologically linked to each other and to STEC O157 isolated from cattle on the same farm. The timescale was consistent with a batch of contaminated milk that had been stored frozen over a period of 6 weeks. Real-time MLVA surveillance identified an additional nine isolates that appeared to be closely related to the outbreak and there was uncertainty as to whether or not these additional cases, not reporting RDM consumption on the SESQ, were linked to the outbreak. In contrast, the WGS provided robust, highly discriminatory typing data and confirmed that four of the nine additional cases were from the same outbreak. Subsequent epidemiological investigations, supported by the robustness of the WGS data, ultimately provided evidence that three of the four cases had consumed RDM from the implicated farm and one had consumed RDM, whereas no evidence of consumption of RDM could be uncovered for the remaining five cases. WGS analysis revealed that the isolates from these five unlinked cases had a deeper phylogenetic relationship (the same 25-SNP cluster) to the outbreak strain (Fig. 2) and subsequent epidemiological investigations revealed that four of the five cases resided in SW England. Furthermore, there was evidence that a subset of cases in the cluster were linked to SW England suggesting that isolates belonging to certain clusters are geographically restricted. Identification of the geographical origin of isolates of STEC O157 PT21/28 may assist in future outbreak investigations as it may be possible to determine the regional source of an implicated food, thus providing an evidence base to direct traceback investigations to specific locations. Further geographical analyses may also elucidate the role of the environment as a risk factor for localized transmission of STEC.

The outbreak strain belonged to a clade of STEC O157 PT21/28 characterized by the loss of *stx2c*

approximately 16 years ago leading to the evolution of a highly pathogenic strain harbouring *stx2a* only (Fig. 2). Previous studies have shown that isolates of STEC encoding *stx2a* only are more likely to be associated with more severe gastrointestinal symptoms and with the development of HUS [28, 29]. Use of WGS for routine public health surveillance of STEC O157 enables us to monitor the emergence of highly pathogenic variants and transmission routes linked to food, environmental exposures and person-to-person contact [5].

Combining WGS with enhanced epidemiological investigations improved case ascertainment and provided robust, highly discriminatory typing data during an outbreak of STEC O157 PT21/28 associated with consumption of RDM in England. Furthermore, WGS provided insights into the evolution of a highly pathogenic clade of STEC O157 PT21/28 encoding *stx2a* only associated with SW England.

ACKNOWLEDGEMENTS

DECLARATION OF INTEREST

None.

REFERENCES

1. **Byrne L, *et al.*** The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009–2012. *Epidemiology and Infection* 2015; **143**: 3475–3487.
2. **Scheutz F, *et al.*** Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *Journal of Clinical Microbiology* 2012; **50**: 2951–2963.
3. **Byrne L, *et al.*** Evaluating the use of multilocus variable number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 as a routine public health tool in England. *PLoS ONE* 2014; **9**: e85901.
4. **Khakhria R, Duck D, Lior H.** Extended phage-typing scheme for *Escherichia coli* O157:H7. *Epidemiology and Infection* 1990; **105**: 511–520.
5. **Dallman TJ, *et al.*** Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. *Clinical Infectious Disease* 2015; **61**: 305–312.
6. **Koser CU, *et al.*** Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathogens* 2012; **8**:e1002824.
7. **McDonnell J, *et al.*** Retrospective analysis of whole genome sequencing compared to prospective typing data in further informing the epidemiological investigation of an outbreak of *Shigella sonnei* in the UK. *Epidemiology & Infection* 2013; **141**:2568–2575.
8. **Quick J, *et al.*** Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biology* 2015; **16**: 114.
9. **Jenkins C, *et al.*** Public health investigation of two outbreaks of Shiga toxin-producing *Escherichia coli* O157 associated with consumption of watercress. *Applied and Environmental Microbiology* 2015; **81**: 3946–3952.
10. **Li H, Durbin R.** Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; **26**; 589–595.
11. **McKenna A, *et al.*** The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 2010; **20**, 1297–1303.
12. **Stamatakis A,** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.
13. **Ashton PM, *et al.*** Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. *PeerJ* 2015; **3**: e739.
14. **Drummond AJ, *et al.*** Bayesian phylogenetics with BEAUti and the BEAST 1·7. *Molecular Biology and Evolution* 2012; **29**: 1969–1973.
15. **Kulldorf M.** A spatial scan statistic. *Communications in Statistics: Theory & Methods* 1997; **26**: 1481–1496.
16. **Pritchard GC, *et al.*** Verocytotoxigenic *Escherichia coli* O157 in animals on public amenity premises in England and Wales, 1997 to 2007. *Veterinary Record* 2009; **164**: 545–549.
17. **The Food Hygiene (England) Regulations.** 2006 (http://www.legislation.gov.uk/uksi/2006/14/pdfs/uksi_20060014_en.pdf).
18. **Dallman TJ, *et al.*** Applying phylogenomics to understand the emergence of Shiga-toxin-producing *Escherichia coli* O157:H7 strains causing severe human disease in the UK. *Microbial Genomics* 2015; **1**: 1–13.
19. **European Food Safety Authority.** Scientific Opinion on the public health risks related to the consumption of raw drinking milk. *European Food Safety Authority Journal* 2015; **13**: 3940.

20. **Mungai EA, Behravesh CB, Gould LH.** Increased outbreaks associated with nonpasteurized milk, United States, 2007–2012. *Emerging Infectious Disease* 2015; **21**: 119–122.

21. **Food Standards Agency.** Update on Food Standards Agency review of controls for raw drinking milk (RDM) in England, Wales and Northern Ireland and proposals for future controls. 2015 (http://www.food.gov.uk/sites/default/files/multimedia/pdfs/board/board-papers2014/fsa-140704.pdf).

22. **Guh A, et al.** Outbreak of *Escherichia coli* O157 associated with raw milk, Connecticut, 2008. *Clinical Infectious Disease* 2010; **51**: 1411–1417.

23. **Giacometti F, et al.** Sale of raw milk in northern Italy: food safety implications and comparison of different analytical methodologies for detection of foodborne pathogens. *Foodborne Pathogens and Disease* 2012; **9**: 293–297.

24. **Elhadidy M, Mohammed MA.** Shiga toxin-producing *Escherichia coli* from raw milk cheese in Egypt: prevalence, molecular characterization and survival to stress conditions. *Letters in Applied Microbiology* 2013; **56**: 120–127.

25. **Mohammadi P, et al.** Isolation of Shiga toxin-producing *Escherichia coli* from raw milk in Kermanshah, Iran. *Iranian Journal of Microbiology* 2013; **5**: 233–238.

26. **Gillespie IA, et al.** Milkborne general outbreaks of infectious intestinal disease, England and Wales, 1992–2000. *Epidemiology and Infection* 2003; **130**: 461–468.

27. **de Louvois J, Rampling A.** One fifth of samples of unpasteurised milk are contaminated with bacteria. *British Medical Journal* 1998; **316**: 625.

28. **Byrne L, et al.** Epidemiology and microbiology of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England, 2009–2013. *Journal of Medical Microbiology* 2014; **63**: 1181–1188.

29. **Ethelberg S, et al.** Virulence factors for hemolytic uremic syndrome, Denmark. *Emerging Infectious Disease* 2014; **10**: 842–847.