

# 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

T. J. DALLMAN<sup>1</sup>†, L. BYRNE<sup>2</sup>†, N. LAUNDERS<sup>2</sup>, K. GLEN<sup>2</sup>, K. A. GRANT<sup>1</sup>  
AND C. JENKINS<sup>1</sup>\*

<sup>1</sup>Gastrointestinal and Emerging Zoonotic Infections Department, Public Health England, London, UK

<sup>2</sup>Gastrointestinal Bacteria Reference Unit, Public Health England, London, UK

Received 9 May 2014; Final revision 30 August 2014; Accepted 19 September 2014;  
first published online 15 October 2014

## SUMMARY

Many serogroups of Shiga toxin-producing *Escherichia coli* (STEC) other than serogroup O157 (non-O157 STEC), for example STEC O26:H11, are highly pathogenic and capable of causing haemolytic uraemic syndrome. A recent increase in non-O157 STEC cases identified in England, resulting from a change in the testing paradigm, prompted a review of the current methods available for detection and typing of non-O157 STEC for surveillance and outbreak investigations. Nineteen STEC O26:H11 strains, including four from a nursery outbreak were selected to assess typing methods. Serotyping and multilocus sequence typing were not able to discriminate between the *stx*-producing strains in the dataset. However, genome sequencing provided rapid and robust confirmation that isolates of STEC O26:H11 associated with a nursery outbreak were linked at the molecular level, had a common source and were distinct from the other strains analysed. Virulence gene profiling of DNA extracted from a polymerase chain reaction (PCR)-positive/culture-negative faecal specimen from a case that was epidemiologically linked to the STEC O26:H11 nursery outbreak, provided evidence at the molecular level to support that link. During this study, we describe the utility of PCR and the genome sequencing approach in facilitating surveillance and enhancing the response to outbreaks of non-O157 STEC.

**Key words:** Bacterial typing, molecular epidemiology, outbreaks, public health microbiology, Shiga-like toxin-producing *E. coli*.

## INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC), also known as Verocytotoxin-producing *E. coli* (VTEC), cause a range of symptoms from mild gastroenteritis to severe bloody diarrhoea, and about 6% of cases

develop haemolytic uraemic syndrome (HUS) [1]. HUS is the most common cause of acute kidney failure in children and can be associated with cardiac and neurological complications. Strains of *E. coli* belonging to the STEC pathotype are defined by the presence of the phage-encoded Shiga toxin genes, *stx1* and/or *stx2*. In England, national protocols for the detection of STEC are specific for serogroup O157 and focus on the isolation on selective media of non-sorbitol-fermenting colonies of *E. coli* agglutinating with antisera to the O157 antigen

\* Author for correspondence: Dr C. Jenkins, Gastrointestinal Bacteria Reference Unit, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK.  
(Email: [claire.jenkins@phe.gov.uk](mailto:claire.jenkins@phe.gov.uk))

† These authors contributed equally to this work.

(<http://www.hpa-standardmethods.org.uk/>). However, there are over 400 different serotypes of STEC and over 100 of these are known to be associated with severe disease in humans [2]. Non-O157 STEC are generally sorbitol-fermenting strains and, therefore, are not detected by the current national protocols.

STEC O26:H11 is the second most common STEC serogroup detected in England, after serogroup O157 [3] and is commonly isolated from patient with symptoms of gastrointestinal infection elsewhere [4]. Like STEC O157, STEC O26:H11 can harbour *stx1* or *stx2* or both. Prior to 1994, STEC O26:H11 in Europe almost exclusively exhibited the *stx1* genotype; however, since that time there has been a steady increase in the number of strains of STEC O26:H11 isolated harbouring *stx2*, either alone or with *stx1* [5]. This shift in genotype has been associated with strains of STEC O26:H11 that cause more severe disease [6].

In England, the guidance in the VTEC operational manual ([http://www.hpa.org.uk/webc/HPAwebFile/HPAweb\\_C/1279889252950](http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1279889252950)) recommends that faecal specimens from cases of bloody diarrhoea or HUS should be referred to Gastrointestinal Bacteria Reference Unit (GBRU) by local hospital laboratories for testing for STEC other than serogroup O157 (non-O157) [7]. About 15–20 non-O157 STEC strains were isolated each year following this testing paradigm. The selective referral of specimens from cases of the most severe disease only results in an unquantifiable under-ascertainment of non-O157 STEC in England. Recently, a number of local hospital laboratories have implemented the polymerase chain reaction (PCR) approach for the detection of gastrointestinal pathogens directly from faecal specimens using commercially available PCR assays, specifically EntericBio (Ireland) and BD Max (UK). This change in the testing paradigm has resulted in a significant increase in the number of non-O157 STEC cases identified as non-O157 STEC, previously not detected using the current national protocols described above, being detected using this approach [3]. This increase in non-O157 cases prompted a review of the current methods available at GBRU for typing isolates of non-O157 STEC to inform routine surveillance and outbreak investigations.

Multilocus variable number tandem repeat (VNTR) analysis (MLVA) is routinely used at GBRU for typing strains of STEC O157 to inform epidemiology of cases, and facilitate outbreak detection and investigation [8]. This MLVA scheme is

serotype-specific and, therefore, not applicable to strains of non-O157 STEC. Currently, multilocus sequence typing (MLST) is not discriminatory enough for outbreak investigations of non-O157 STEC and provides approximately the same resolution as serotyping. Pulsed-field gel electrophoresis is the method of choice in many national reference laboratories for the molecular typing of non-O157 STEC but it is laborious and technically demanding [9] and international laboratory comparisons of fingerprint patterns are difficult (GBRU in-house data).

Recently whole genome sequencing (WGS) has been employed at Public Health England (PHE) for typing bacterial strains associated with gastrointestinal outbreaks [10, 11]. Other groups have investigated the use of metagenomics, the direct sequencing of DNA extracted from microbiologically complex samples such as faecal specimens, to identify and characterize bacterial strains without laboratory culture [12].

In this study, we describe a complex nursery outbreak of STEC O26:H11 which highlights both the pathogenic potential of this strain, and the public health issues arising from currently employed testing strategies. Strains from the outbreak and other STEC O26:H11 from the GBRU archive were used to retrospectively evaluate the use of genome sequencing, including metagenomic analysis of a PCR-positive but culture-negative faecal specimen, as a suitable molecular typing approach for non-O157 STEC outbreak investigations.

## MATERIAL AND METHODS

### Direct detection of STEC from faecal specimens

DNA was extracted from the faecal specimens using the QiaSymphony Automated DNA extraction platform [13]. DNA from faecal extracts was tested using real-time PCR primers and probes detecting *stx1*, *stx2*, *eae* (intimin) and O157*rfbE* [7]. For all faecal specimens positive for *stx* and/or *eae* (intimin), 10 colonies were picked from either the MacConkey or SMAC plate and retested by the same PCR. Those colonies harbouring *stx* genes were identified biochemically and serotyped using antisera raised in rabbits.

### Epidemiological investigations

Local laboratories report presumptive isolates of STEC directly to PHE centres (PHEC) who undertake public health follow-up and risk assessment.

Table 1. *Molecular and epidemiological data associated with strains of E. coli O26:H11 isolated at GBRU between 2009 and 2013*

Reference no.	Stx profile	MLST	Date culture isolated	Sex/age	Travel	Additional information*
181/09	1&2	21	Feb. 2009	F/2	No travel	Outbreak 2009
461/09	1&2	21	Feb. 2009	M/14	No travel	Outbreak 2009
460/09	1&2	21	Feb. 2009	F/0	No travel	Outbreak 2009
259/10	1&2	21	Sept. 2010	M/3	No travel	HUS
605/10	1&2	21	Sept. 2010	F/3	France	HUS
467/10	2	21	Nov. 2010	M/13	Turkey	HUS
519/11	1&2	21	July 2011	M/4	No travel	
624/12	2	21	Apr. 2012	F/13	Egypt	HUS
165/12	1&2	21	May 2012	F/42	No travel	Fatal case
483/12	2	21	July 2012	M/35	Ireland and Switzerland	Outbreak 2012
482/12	2	21	July 2012	M/4	No travel	Outbreak 2012
626/12	2	21	July 2012	M/4	No travel	Outbreak 2012
627/12	2	21	July 2012	M/2	No travel	Outbreak 2012
2270–502/12	2	21	July 2012	F/1	Ireland	
2290–502/12	—	29	July 2012	M/40	No data	
670/13	—	29	Aug. 2013	F/0	Egypt	HUS
680/13	—	29	Sept. 2013	F/3	Italy	Hospitalized with severe bloody diarrhoea
075/13	2	21	Sept. 2013	M/0	Albania	
637/13	2	21	Sept. 2013	F/7	No data	

MLST, Multilocus sequence typing; HUS, haemolytic uraemic syndrome.

\* Additional information includes whether or not cases were associated with an outbreak and clinical symptoms.

Household contacts of confirmed cases are sampled and those found to be positive and belonging to risks groups are excluded from school or work until they have two negative clearance specimens taken 24–48 h apart. Each PHEC arranges for the standard enhanced surveillance questionnaire to be administered to cases in order to collect demographic details; risk status; clinical condition (including progression to HUS); household or other close contact details; laboratory results; exposures including travel, food and water consumption, contact with animals and environmental factors; case classification; outbreak/cluster status. Completed questionnaires are forwarded for inclusion in the National Enhanced Surveillance System for STEC in England (NESSS), Gastrointestinal and Emerging Zoonotic Infections (GEZI).

#### WGS and analysis of strains of STEC O26:H11

DNA was extracted from 19 strains of *E. coli* O26 using the Wizard kit (Promega, UK). These strains were isolated from faecal specimens submitted to GBRU by colleagues at local hospital laboratories

from cases with severe bloody diarrhoea or HUS between 2009 and 2013 (Table 1). Genomic DNA was fragmented and tagged for multiplexing with Nextera XT DNA Sample Preparation kits (Illumina, UK) and sequenced at the Animal Health Veterinary Laboratory Agency, Surrey, using the Illumina GAII platform with paired-end 150 bp reads. Multiplexing allowed 96 samples to be sequenced per run. Short reads were mapped to the reference STEC O26:H11 strain 11368 (NC\_011361.1) using BWA-SW [14]. The Sequence Alignment Map output from BWA was sorted and indexed to produce a Binary Alignment Map (BAM) using Samtools [15]. GATK2 [16] was used to create a variant call format (VCF) file from each of the BAMs, which were further parsed to extract only single nucleotide polymorphism (SNP) positions which were of high quality in all genomes (MQ > 30, DP > 10, GQ > 30, variant ratio > 0.9) [17]. Pseudosequences of polymorphic positions were used to create maximum-likelihood trees using RaxML [18]. Spades version 2.5.1 [19] was used to produce *de novo* assemblies of the sequenced paired-end FASTQ files.

FASTQ sequences for these 19 isolates have been deposited in the NCBI Short Read Archive under BioProject PRJNA 259 827.

MLSTs were identified by mapping the reads against all *E. coli* allele variants held in the MLST database ([www.mlst.ucc.ie/mlst/dbs/Ecoli](http://www.mlst.ucc.ie/mlst/dbs/Ecoli)) using a modification of SRST software [20].

Identification of the flagella antigen and intimin subtype were performed by BLAST [21] comparisons of the assembled genomes to a database of *fliC* genes for the determination of the flagella antigen and a database of *eae* genes for the determination of the intimin subtype.

### Metagenomic analysis of a faecal specimen from a culture-negative case epidemiologically linked to the outbreak

Faecal DNA from a PCR-positive/culture-negative faecal specimen, from a case epidemiologically linked to a nursery outbreak, was extracted on the QiaSymphony Automated DNA extraction platform [13] in triplicate to achieve the required yield for DNA sequencing. Genomic DNA was fragmented and tagged for multiplexing with Nextera XT DNA Sample Preparation kits (Illumina, UK) and sequenced on an Illumina MiSeq platform. About 18.5 m high-quality paired-end 150 bp reads were generated after trimming using Trimmomatic [22]. Reads were mapped against a diarrhoeagenic *E. coli* virulence gene panel including *stx1* and *stx2* for STEC, *eae* (encoding intimin) for STEC and enteropathogenic *E. coli* (EPEC) [23], *aggR* for enteroaggregative *E. coli* (EAEC) [24], *ipaH* for Shigella, heat-stable enterotoxin (ST) and heat-labile enterotoxin (LT) genes for enterotoxigenic *E. coli* (ETEC) and the following serotype-specific targets: *rfbE* (O157), *wzx* (O26), *wzx* (O103), *wbdI* (O111) and *ihp1* (O145) (<http://www.iss.it/vtec/index.php?lang=2&anno=2014&tipo=3>), using BWA-SW. Coverage of each gene target was calculated using Samtools depth command.

FASTQ sequences of this metagenomic sample have been deposited in the NCBI Short Read Archive under BioProject PRJNA 259 829.

Metagenomic reads were also mapped to virulence factors of pathogenic bacteria database (VFDB) [25] which contains 28 966 virulence genes from 26 pathogens including common enteric genera, such as *Salmonella*, *Listeria*, *Campylobacter*, *Shigella*, *Yersinia* and *Vibrio* species.

## RESULTS

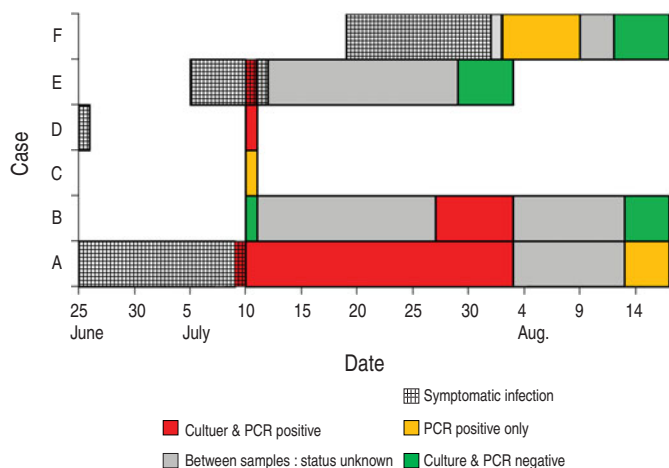
### Microbiology and epidemiology of the STEC O26:H11 outbreak

On 8 July 2012 a faecal specimen from a 1-year-old infant (case A) with symptoms of HUS was tested for bacterial gastrointestinal pathogens at a local hospital laboratory. The culture was reported as negative for *Campylobacter*, *Salmonella*, *Shigella* species and *E. coli* O157 and was submitted to GBRU for further testing for other non-O157 STEC, as recommended by the guidelines in the PHE VTEC operational manual. PCR tests at GBRU showed the faecal specimen was positive for the *stx2* and *eae* (intimin) genes and subsequently a strain of STEC O26:H11 *stx2a* was cultured from the specimen (Fig. 1 and Table 2).

Following the positive culture from case A, faecal specimens were collected from the household contacts on 10 July, including case B, a 4-year-old sibling who attended nursery and was therefore considered in a risk group, and both parents (cases C and D). All household contacts were asymptomatic at this time. The first specimen from case B was negative for STEC although subsequent specimens taken over 2 weeks later were positive by PCR and culture (Fig. 1, Table 2). Both parents (cases C and D) were positive by PCR, although STEC O26:H11 was cultured from case D only (Table 2). Cases A and B continued to excrete STEC O26:H11 until 3 August. Subsequent specimens were culture-negative although specimens from case A continued to be PCR-positive up to the point when sampling ceased on 16 August.

There was a complex exposure history within the household. In the 7 days prior to case A's onset of illness the family had visited two petting farms and consumed burgers at a family barbeque. Case D had travelled to mainland Europe on two occasions (Table 1) during the incubation period and had also experienced mild gastrointestinal illness at the same time as case A became ill on 25 June.

A 3-year-old (case E) who attended the same nursery as cases A and B experienced diarrhoea, vomiting and abdominal pain commencing on 5 July 2012. A faecal specimen collected on 10 July was positive for *stx2* and *eae* and a strain of STEC O26:H11 *stx2a* was cultured (Table 2). Those cases (A, B, E) who attended the nursery were excluded until two culture-negative results were reported. However, case F, a 6-month-old infant with symptoms of HUS, became ill on 19 July 2012. This case had direct contact with case E outside of a nursery setting during the time



**Fig. 1.** Epidemic curve for the six outbreak cases showing the timeline for symptomatic infection and STEC excretion monitored by culture and PCR. Diagnostic results for cases C and D are presented for sampling date only and not duration as only one specimen was taken from these two cases (see Table 1 for details).

case E was still excreting STEC. Faecal specimens collected from case F were PCR-positive for *stx2* and *eae* but STEC O26:H11 was never cultured from either PCR-positive specimens (Table 2).

### WGS of strains of STEC O26:H11

During the outbreak, the serotyping and *stx* profiles of the isolates along with the epidemiological data were regarded as strong evidence that these cases were linked. Retrospectively, we sequenced the whole genomes of 19 strains of *E. coli* O26:H11, including the outbreak strains, isolated between 2009 and 2013 (Table 1). The phylogenetic relationship of these strains, mapped to the reference STEC O26:H11 strain 11 368 (NC\_011 361.1), is shown in Figure 2 and comprises of 3543 polymorphic positions. Case E (482/12) was 0 SNPs different at the core genome level to case D (483/12), two SNPs different from case A (627/12) and three SNPs different from case B (626/12). The genome sequencing data indicated that the strains linked to this nursery outbreak are closely related and clearly differentiated from the sequences of the other 15 strains of *E. coli* O26:H11 from the GBRU archive. Strains from a household outbreak in 2009 were also identified as being closely related to each other with <3 SNPs between each sequence and 272 SNPs from the closest strain 637/13.

Two different MLSTs were identified, ST21 and ST29. In this collection, ST21 and ST29 were associated with the *stx*-positive and *stx*-negative strains,

respectively. Two of the three *stx*-negative strains were associated with cases with severe disease including one case of HUS and this phenomenon has been described previously [26].

All of the sequenced strains had 100% identity over the complete length to the *fliC* gene of O26:H11 strain 11 368 (NC\_011 361.1) and >99% identity to the complete length of the  $\beta$ -*intimin* gene.

We performed a metagenomic analysis on the PCR-positive/culture-negative faecal specimen from case F. Sequenced reads from the extracted faecal DNA were mapped against a panel of *E. coli* virulence genes (Table 3). Reads were successfully mapped to the STEC virulence factors *eae* and *stx* as well as *wzx*, the *E. coli* O26:H11 lipopolysaccharide gene and *fliC* encoded the flagella antigen H11. There were no reads mapped to virulence factors indicative of the presence of *Shigella* spp., ETEC or EAEC. When all metagenomic reads from the faecal specimen from case F were mapped to the assembled contigs of the isolate from case B, 75% of the genome was covered with depth greater than 10 reads comprising 52% of metagenomic reads.

Sequenced reads mapped to VFDB resulted in significant hits or matches to 1663 virulence factors. Of the 1312 that could be attributed to a species 994 (76%) were associated with *E. coli*, 24 (2%) with *Salmonella* spp., 208 (16%) with *Shigella* spp., 62 (5%) with *Yersinia* spp. and 1% with other species. Genes matching those found in other enteric pathogens could be accounted for by homology of certain genes to similar genes found in *E. coli*.

Table 2. Real-time PCR results and corresponding culture results for each faecal specimen received from each outbreak case

	Sample date (2012)	PCR	Culture
Case A	8 July	+	+
	10 July	+	+
	27 July	+	+
	1 Aug.	+	+
	3 Aug.	+	+
	13 Aug.	+	–
Case B	16 Aug.	+	–
	10 July	–	–
	27 July	+	+
	1 Aug.	+	+
	3 Aug.	+	+
Case C	13 Aug.	–	–
	16 Aug.	–	–
Case D	10 July	+	–
Case E	10 July	+	+
	27 July	–	–
Case F	31 July	–	–
	2 Aug.	+	–
	2 Aug.	+	–
	8 Aug.	+	–
	12 Aug.	–	–
	16 Aug.	–	–

PCR+/culture+, *stx* DNA detected and STEC isolated by culture; PCR+/culture–, *stx* DNA detected but STEC not isolated by culture; PCR–/culture–, *stx* DNA not detected and STEC not isolated by culture.

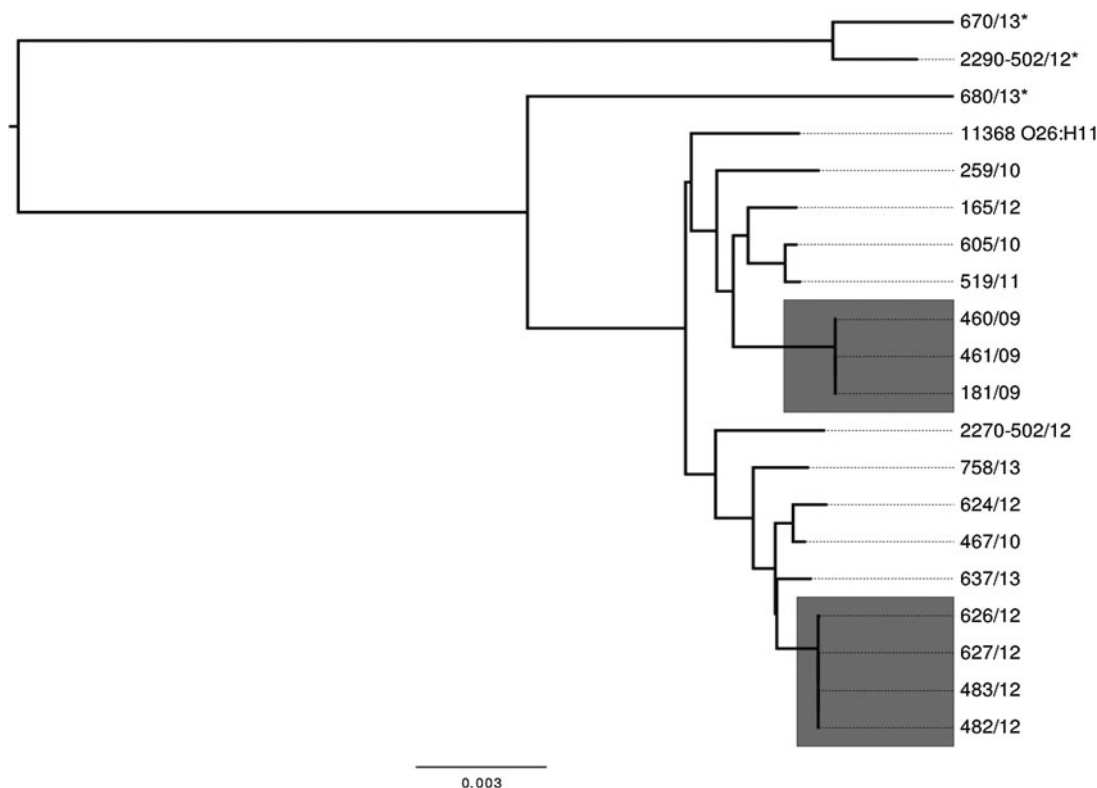
## DISCUSSION

PCR for the direct detection of STEC is rapid, sensitive and facilitates the detection of all STEC serogroups. In this outbreak, direct PCR led to the rapid identification of cases, including one PCR-positive/culture-negative case with HUS, for which STEC infection would not have been confirmed through traditional culture-dependent methods. However, this approach raises a number of public health issues. First, although the initial diagnosis is rapid, the follow-up culture and isolation of the pathogen can be labour-intensive and time-consuming and may delay public health action. Second, the increased sensitivity of the PCR can lead to PCR-positive but culture-negative results that can be difficult to interpret with respect to the risk of person-to-person transmission. Providing guidelines for public health action in such cases, especially in those belonging to risk groups including young children and food handlers, may be challenging. Following the outbreak described

here, case A returned to nursery once he was asymptomatic and had two culture-negative stools, despite the positive PCR results. No new cases were detected in the nursery cohort during this investigation following case A's return. Third, the clinical and public health significance of all serotypes of non-O157 STEC is unclear [27, 28]. STEC O26:H11, specifically those strains harbouring *stx2* only, are highly pathogenic and capable of causing HUS. This outbreak strain had *stx2* and *eae*, both pathogenicity factors associated with severe disease [29–31] and the pathogenic potential (two cases developed symptoms of HUS) and transmissibility between close contacts of this strain, was clearly demonstrated.

While two cases developed HUS, four cases had less severe or asymptomatic infection. The two cases that developed HUS were aged <5 years, the recognized risk group for progression to HUS following STEC infection [32]. While it is unclear who the index case was, case D reported mild gastrointestinal symptoms at the same time as case A was ill and it is possible that case D acquired STEC infection abroad and then transmitted it within the household. The mild/asymptomatic nature of the illness in adults in this outbreak (cases C and D) illustrate the limitations of only testing those cases with more severe disease for non-O157 STEC infection, as recommended by the current guidelines in the VTEC operational manual, and the advantages of testing all specimens by PCR at a local level. Given the high transmissibility of STEC within households, an index case in an adult with mild symptoms may go on to infect more vulnerable members of the household leading to more severe disease. The use of PCR for the direct detection of STEC at local hospital laboratories will facilitate the detection of non-O157 STEC, particularly in cases with less severe disease which may otherwise have gone undetected.

Long-term shedding of STEC is well recognized, particularly in children [33]. Given the cases of mild and asymptomatic infection described in this study and the long-term shedding of infection, it was not possible to determine the order of infection or the source of the outbreak. Transmission to humans occurs through direct or indirect contact with animals or their environment, consumption of contaminated food or water and through person-to-person contact [34, 35]. The main reservoir of STEC in England is cattle although it is carried by other animals, mainly ruminants. The family of the index case had recent contact with the farming environment on two



**Fig. 2.** Phylogenetic relationship of 19 strains of *E. coli* O26 isolated at GBRU between 2009 and 2013. Outbreak strains of STEC O26 from 2009 and 2012 are highlighted in grey. The scale represents number of nucleotide substitutions per site. \* Strains negative for the *stx* genes.

**Table 3.** Number of reads of the metagenomic sequencing data taken from the faecal specimen from case F to map to virulence genes associated with diarrhoeagenic *E. coli*

Gene	<i>Stx1</i>	<i>Stx2</i>	<i>eae</i>	<i>ipaH</i>	<i>aggR</i>	<i>bfpA</i>
No. of reads	–	25	8	–	–	–
Gene	<i>ST</i>	<i>rfbE</i> O157	<i>Wzx</i> O26	<i>Wzx</i> O103	<i>Wbdl</i> O111	<i>Ihp</i> O145
No. of reads	–	–	14	–	–	–

occasions prior to onset of symptoms and these farm visits were initially regarded as the likely source of the infection. However, the *stx* profile of the outbreak strain (*stx2a* only) is rarely seen in the UK cattle population where strains characteristically harbour *stx1* only or *stx1* and *stx2* [36]. Taking into account the travel history of case D, the possibility exists that the outbreak strain may have been acquired abroad, although there is no direct evidence for this.

As well as a cluster of cases in the same household (cases A–D), two cases (E and F) were identified at the nursery attended by cases A and B. STEC nursery outbreaks are not uncommon, and STEC O26:H11 has been previously linked to a number of such outbreaks [37–39]. In this outbreak, the epidemiological links (same household or attendance at the same nursery) between the cases were strong. However, the serotyping results did not provide conclusive evidence that the strains were microbiologically linked as STEC O26:H11 is the second most commonly isolated STEC serotype in England [3]. In contrast, the genome sequencing performed during this study, provided rapid, robust and highly discriminatory confirmation at the molecular level that the strains of STEC O26:H11 associated with the outbreak were linked (<5 SNPs) and had a common source (Fig. 2). In this example, we demonstrated that WGS was a suitable method for typing STEC O26:H11 in order to facilitate outbreak investigations.

Although the epidemiological data indicated that case F was part of the outbreak, the metagenomic analysis provided further evidence of that link in the absence of positive culture results. This analysis

confirmed the likely presence of *wzx O26*, *stx2* and *eae* (intimin) genes in the specimen. We were also able to show that there was no evidence from the metadata sequence set that other gastrointestinal pathogens were present. The number of reads mapping to each virulence gene was close to the threshold of a detectable limit suggesting that this approach is not yet sensitive enough to be used routinely as a diagnostic tool. Furthermore, PCR assays for the detection of the *E. coli* O26 target gene (*wzx* O26) from faecal specimens have been well validated and are currently used in many diagnostic laboratories [40]. However, the value of using of using metagenomic sequencing analyses as an ‘open-ended, culture-independent’ approach during outbreaks of unknown aetiology has been previously documented [12].

Non-O157 outbreaks are thought to be rare in England but as the current frontline diagnostic laboratory protocols are not designed to detect non-O157 STEC, there is no surveillance data to support this. Elsewhere, outbreaks of non-O157 STEC can involve large numbers of cases and associated symptoms can be severe and include HUS [41, 42]. The cost of WGS, including DNA extraction is currently about €50 per isolate which compares favourably with the costs of the multiple biochemical, serological and alternative molecular tests required to type non-O157 STEC to facilitate outbreak investigations currently. Furthermore, unlike the STEC O157 MLVA scheme used at GBRU, WGS is not serotype-specific. WGS is more discriminatory than MLST, as demonstrated here, and sequencing data is easily transferable. At PHE, we have established a robust and reliable pipeline for routine genome sequencing and analysis of isolates from non-O157 STEC outbreaks within 48–72 h. WGS also provided a full virulence profile of the pathogen from DNA extracted directly from a PCR-positive but culture-negative faecal specimen from a case epidemiologically linked to a nursery outbreak. This study demonstrates the potential of WGS to determine the pathogenicity profile and evolutionary origin of certain strains as well as facilitating outbreak investigations of non-O157 STEC in England.

## ACKNOWLEDGEMENTS

The authors thank all colleagues from the National Health Service and Public Health England frontline laboratories who refer specimens to GBRU. We also acknowledge Marie Chattaway, Neil Perry, Vivienne

do Nascimento, Michael Wright and Yoshini Taylor for their hard work and microbiological expertise at GBRU. Thanks are also due to Natalie Adams and Radha Patel for their contributions to the National Enhanced Surveillance System for STEC in England and to all public health practitioners who report to the scheme.

## DECLARATION OF INTEREST

None.

## REFERENCES

1. Lynn RM, *et al.* Childhood hemolytic uremic syndrome, United Kingdom and Ireland. *Emerging Infectious Diseases* 2005; **11**: 590–596.
2. Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds). In: *Bergey's Manual of Systematic Bacteriology*, 2nd edn, Vol. 2: *The Proteobacteria*. New York: Springer, 2005.
3. Byrne *et al.* The epidemiology and microbiology of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England 2009–2013. *Journal of Medical Microbiology* 2014; **639**: 1181–1188
4. Mellman *et al.* Analysis of a collection of haemolytic uremic syndrome associated enterohaemorrhagic *Escherichia coli*. *Emerging Infectious Diseases* 2008; **14**: 1287–1290
5. Zhang WL, *et al.* Molecular characteristics and epidemiological significance of Shiga toxin-producing *Escherichia coli* O26 strains. *Journal of Clinical Microbiology* 2000; **38**: 2134–2140.
6. Bielaszewska M, *et al.* Enterohemorrhagic *Escherichia coli* O26:H11/H-: a new virulent clone emerges in Europe. *Clinical Infectious Diseases* 2013; **56**: 1373–1381.
7. Jenkins C, *et al.* Assessment of a real-time PCR for the detection and characterization of verocytotoxigenic *Escherichia coli*. *Journal of Medical Microbiology* 2012; **61**: 1082–1085.
8. 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.
9. Karama M & Gyles CL. Methods for genotyping Verotoxin-producing *Escherichia coli*. *Zoonoses Public Health* 2010; **57**: 447–462.
10. 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.
11. Byrne, *et al.* A multi-country outbreak of Salmonella Newport gastroenteritis in Europe associated with watermelon from Brazil confirmed by whole genome sequencing: October 2011 to January 2012. *Euro-surveillance* 2014; **19**: 6–13.
12. Loman NJ, *et al.* A culture-independent sequence-based metagenomics approach to the investigation of an



- outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *Journal of the American Medical Association* 2013; **309**: 1502–1510.
13. Halstead FD, et al. Olympics GI Group. Universal extraction method for gastrointestinal pathogens. *Journal of Medical Microbiology* 2013; **62**: 1535–1539.
  14. Li H and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; **26**: 589–595.
  15. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
  16. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 2011; **43**: 491–498.
  17. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.
  18. Price MN, et al. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* 2010; **5**: e9490.
  19. Bankevich A, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 2012; **19**: 455–477.
  20. Inouye M, et al. Short read sequence typing (SRST): multi-locus sequence types from short reads. *BMC Genomics* 2012; **13**: 338.
  21. Altschul SF, et al. Basic local alignment search tool. *Journal of Molecular Biology* 1990; **215**: 403–410.
  22. Bolger AM, et al. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.
  23. Frankel G, et al. Intimin and the host cell – is it bound to end in Tir(s)? *Trends in Microbiology* 2001; **9**: 214–218.
  24. Nataro JP. Enteroaggregative *Escherichia coli* pathogenesis. *Current Opinion in Gastroenterology* 2005; **21**: 4–8.
  25. Chen LH, et al. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Research* 2012; **40** (database issue): D641–645.
  26. Bielaszewska M, et al. Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic *Escherichia coli* O26 infection in humans. *Applied Environmental Microbiology* 2007; **73**: 3144–3150.
  27. Feng PC & Reddy S. Prevalences of Shiga toxin subtypes and selected other virulence factors among Shiga-toxigenic *Escherichia coli* strains isolated from fresh produce. *Applied Environmental Microbiology* 2013; **79**: 6917–6923.
  28. Preußel K, et al. Shiga toxin-producing *Escherichia coli* O157 is more likely to lead to hospitalization and death than non-O157 serogroups – except O104. *PLoS ONE* 2013; **8**: 78180.
  29. Persson S, et al. Subtyping method for *Escherichia coli* Shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *Journal of Clinical Microbiology* 2007; **45**: 2020–2024.
  30. Melton-Celsa A, et al. Pathogenesis of Shiga-toxin producing *Escherichia coli*. *Current Topics in Microbiology and Immunology* 2012; **357**: 67–103.
  31. Pollock KG, et al. Highly virulent *Escherichia coli* O26, Scotland. *Emerging Infectious Diseases* 2011; **17**: 1777–1779.
  32. Karmali MA. Host and pathogen determinants of verocytotoxin-producing *Escherichia coli*-associated hemolytic uremic syndrome. *Kidney International Supplement* 2009; **112**: S4–7.
  33. Dabke G, et al. Duration of shedding of Verocytotoxin-producing *Escherichia coli* in children and risk of transmission in childcare facilities in England. *Epidemiology & Infection* 2014; **142**: 327–34.
  34. Locking ME, et al. Risk factors for sporadic cases of *Escherichia coli* O157 infection: the importance of contact with animal excreta. *Epidemiology & Infection* 2001; **127**: 215–220.
  35. Gillespie IA, et al. Foodborne general outbreaks of Shiga toxin-producing *Escherichia coli* O157 in England and Wales 1992–2002: where are the risks? *Epidemiology & Infection* 2005; **133**: 803–808.
  36. Chase-Topping ME, et al. Pathogenic potential to humans of bovine *Escherichia coli* O26, Scotland. *Emerging Infectious Diseases* 2012; **18**: 439–448.
  37. Brown JA, et al. Outbreak of shiga toxin-producing *Escherichia coli* serotype O26: H11 infection at a child care center in Colorado. *Pediatric Infectious Disease Journal* 2012; **31**: 379–383.
  38. Sonoda C, et al. An enterohemorrhagic *Escherichia coli* O26 outbreak at a nursery school in Miyazaki, Japan. *Japanese Journal of Infectious Diseases* 2008; **61**: 92–93.
  39. Werber D, et al. A multistate outbreak of Shiga toxin-producing *Escherichia coli* O26:H11 infections in Germany, detected by molecular subtyping surveillance. *Journal of Infectious Diseases* 2002; **186**: 419–422.
  40. Kagkli DM, et al. Application of the modular approach to an in-house validation study of real-time PCR methods for the detection and serogroup determination of verocytotoxigenic *Escherichia coli*. *Applied Environmental Microbiology* 2011; **77**: 6954–6963.
  41. Frank C, et al. HUS Investigation Team. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *New England Journal of Medicine* 2011; **365**: 1771–1780.
  42. Luna-Gierke RE, et al. Outbreaks of non-O157 Shiga toxin-producing *Escherichia coli* infection: USA. *Epidemiology & Infection* 2014; **142**: 2270–2280.