Shiga toxin-producing *Escherichia coli* clonal complex 32, including serotype O145:H28, in the UK and Ireland

Ella V. Rodwell[1,2,3], Yung-Wai Chan[1], Clare Sawyer[4], Anne Carroll[5], Eleanor McNamara[5], Lesley Allison[6], Lynda Browning[7], Anne Holmes[6], Gauri Godbole[1], Noel McCarthy[3,8] and Claire Jenkins[1,3,*]

Abstract

Introduction. Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 has been the most clinically significant STEC serotype in the UK for over four decades. Over the last 10 years we have observed a decrease in STEC O157:H7 and an increase in non-O157 STEC serotypes, such as O145:H28.

Gap Statement. Little is known about the microbiology and epidemiology of STEC belonging to CC32 (including O145:H28) in the UK. The aim of this study was to integrate genomic data with patient information to gain a better understanding of the virulence, disease severity, epidemic risk assessment and population structure of this clinically significant clonal complex.

Methodology. Isolates of *E. coli* belonging to CC32 (*n*=309) in the archives of public health agencies in the UK and Ireland were whole-genome-sequenced, virulence-profiled and integrated with enhanced surveillance questionnaire (ESQ) data, including exposures and disease severity.

Results. Overall, diagnoses of STEC belonging to CC32 (290/309, 94%) in the UK have increased every year since 2014. Most cases were female (61%), and the highest proportion of cases belonged to the 0–4 age group (53/211,25%). The frequency of symptoms of diarrhoea (92%), abdominal pain (84%), blood in stool (71%) and nausea (51%) was similar to that reported in cases of STEC O157:H7, although cases of STEC CC32 were more frequently admitted to hospital (STEC CC32 48% vs O157:H7 34%) and/or developed haemolytic uraemic syndrome (HUS) (STEC CC32 9% vs O157:H7 4%).

The majority of STEC isolates (268/290, 92%) had the *stx2a*/*eae* virulence gene combination, most commonly associated with progression to STEC HUS. There was evidence of person-to-person transmission and small, temporally related, geographically dispersed outbreaks, characteristic of foodborne outbreaks linked to nationally distributed products.

Conclusion. We recommend more widespread use of polymerase chain reaction (PCR) for the detection of all STEC serogroups, the development of consistent strategies for the follow-up testing of PCR-positive faecal specimens, the implementation of more comprehensive and standardized collection of epidemiological data, and routine sharing of sequencing data between public health agencies worldwide.

DATA SUMMARY

FASTQ files are submitted to the National Center for Biotechnology Information (NCBI). All data can be found under Bioproject no. PRJNA315192 and strain-specific information is found in Table S1(available with the online version of this article).

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) is a group of zoonotic, gastrointestinal pathogens transmitted to humans through the food chain, contact with animals or contact with environmental sources contaminated with animal faeces [1]. The STEC group is defined by the presence of the Shiga toxin gene (*stx*), which is located on a bacteriophage incorporated into the bacterial genome [2]. Although s*tx* is the defining virulence factor of the STEC group, the presence of additional factors, such as the pathogenicity island called the locus of enterocyte effacement (LEE) [3] and/or plasmid-encoding genes involved in attachment mechanisms [4] or survival in the host outside the gut [5] in certain STEC serotypes, are known to increase their pathogenic potential. Shiga toxin subtypes are associated with symptoms varying in severity, with *stx2a* and *stx2d* typically being more associated with severe clinical outcomes [6, 7].

Symptoms of STEC infection range from mild to severe diarrhoea, often blood-stained, vomiting, abdominal pain and fever. A subset of infected patients develop haemolytic uraemic syndrome (HUS), characterized by renal failure and neurological complications, as a result of Shiga toxins targeting epithelial, glomerular, mesangial and endothelial cells [8–10]. STEC HUS is the leading cause of renal failure in children under the age of 5.

Historically in the UK, public health surveillance systems focused on STEC serotype O157:H7, as this serotype was the causative agent of the outbreak of HUS when STEC first emerged as a threat to public health in the 1980s [11]. However, over the last decade we have seen a decrease in the incidence of STEC O157, belonging to clonal complex (CC) 11, [12] and an increase in STEC belonging to other serotypes or CCs [13].

STEC CCs have different genomic characteristics, including different virulence gene profiles, diverse animal reservoirs and/or transmission routes and variable geographical distribution; for example, STEC O117:H7 has been associated with travellers' diarrhoea and sexual transmission [14–17]. To date, the majority of reports on STEC data from the UK Health Security Agency (UKHSA) have focused on analysing non-O157 STEC as one group [13]. However, the increasing incidence and variety observed in their microbiological and epidemiological features has highlighted the need to consider each serotype or CC independently. In England, the CCs frequently associated with HUS are CC11 (O157:H7), CC21 (O26:H11), CC165 (O80:H2, O55:H9, O45:H2) and CC32 (O145:H28) [12, 13, 18].

Although previously not recognized as a major cause of gastrointestinal (GI) illness in England, globally STEC belonging to serogroup O145 are among the six most commonly detected serogroups. The 'big six' serogroups, O26, O45, O103, O111, O121 and O145, are described by the US Food and Drug Administration (FDA) as the six key non-O157 serogroups of STEC causing severe illness in humans [19]. STEC O145:H28 has been detected in the animal reservoir [20–22] and food [23, 24], and has been identified as the causative agent of a number of outbreaks [24–26].

Despite the importance of this serotype, there is limited information on the epidemiological context and the genomic composition, including the traits associated with virulence and colonization, of this serotype in the UK. The aim of this study was to integrate genomic data with epidemiology information in order to gain insight regarding the virulence profile, disease severity, epidemic risk assessment and genetic origin of this clinically significant clonal complex in the UK and Ireland.

METHODS

Microbiology

In the UK and Ireland, where STEC has been confirmed through polymerase chain reaction (PCR) at local diagnostic laboratories and/or there is clinical suspicion of STEC infection but *E. coli* O157:H7 cannot be isolated from patient stool samples with community- or hospital-acquired infection, the stool samples can be referred to the appropriate national reference laboratory in either England, Scotland or Ireland for confirmatory PCR testing and culture. The Gastrointestinal Bacteria Reference Unit (GBRU) and Scottish *E. coli* O157/STEC Reference Laboratory (SERL) employ PCR- and culture-based methods as described previously [27, 28].

The time frames for detection of the isolates included in this study from each country were as follows: English and Welsh cases were identified from 2015 to 2021, those resident in Scotland were detected between 2004 and 2021, cases from Northern Ireland were detected from 2018 to 2021 and STEC isolates from Ireland were referred to the UKHSA for whole genome sequencing (WGS) from 2016 to 2018 (Table S1).

STEC surveillance

The devolved administrations of the UK and Ireland operate independant national enhanced STEC surveillance systems. These surveillance systems comprise data from enhanced surveillance questionnaires (ESQs), which include clinical symptoms, food history, travel history, and environmental and animal exposures, and these are paired with the microbiological and genomic data for STEC infections. Aggregate data on demographics (age and sex), clinical symptoms and animal exposures were available from O145:H28 cases in Scotland. We described the demographics, clinical presentations and exposures of cases of STEC CC32 for which ESQs were available (England: 106/118, Wales: 4/18 and aggregate data from Scotland: 46/75). No ESQs were available for this study from Northern Ireland or Ireland.

For comparisons to O157:H7 STEC, clinical presentation count data were taken from previously published clinical symptom data in England [12]. Statistical analysis (chi^2 test) was performed on the count data from this study and the O157:H7 STEC study [12]. The extraction of O157:H7 STEC data from the UKHSA data warehouse was performed for the seasonality comparisons. Human isolates of O157:H7 that harboured *stx*, submitted to UKHSA between 2014 and 2021, were taken forward for this analysis.

DNA extraction and genomic processing

Genomic DNA was extracted and sequenced on Illumina HiSeq 2500 platform as described by Dallman *et al.* [29] and held in the UKHSA in-house data warehouse. At the SERL, extracted DNA was sequenced on the Illumina MiSeq [30]. Post whole-genome sequencing (WGS), isolates were processed through an in-house pipeline that includes stx and serotype profiling using GeneFinder (https://github.com/phe-bioinformatics/gene_finder), with an stx variant database and the Serotypefinder database [31], as previously described [32]. Multilocus sequence typing (MLST) was performed using Metric Orientated Sequence Typer (https://github.com/phe-bioinformatics/MOST), as described by Tewolde *et al.* [33] (Table S1).

Virulence profiling was performed using GeneFinder (https://github.com/phe-bioinformatics/gene_finder) supplemented with genes associated with the LEE and non-LEE effectors from the Centre for Genomic Epidemiology (CGE) database v2020-05-29 (https://bitbucket.org/genomicepidemiology/virulencefinder_db/src/master/virulence_ecoli.fsa) (Table S2). Antimicrobial resistance profiling was performed through UKHSA GeneFinder (https://github.com/phe-bioinformatics/gene_finder) and an in-house database (Table S3).

SnapperDB v0.2.6 [34] is the UKHSA in-house database that holds called variant data, achieved from genomic DNA sequencing, relative to an appropriate reference for *Escherichia coli* CCs. SnapperDB v0.2.6 [34] was employed to generate whole genome alignment of isolates belonging to CC32, and Gubbins v2.0.0 [35] was used to identify the recombinant regions. SnapperDB v0.2.6 was used again to generate an alignment of given variant positions that belong to a minimum of 80% of strains in the alignment, once recombinations regions had been masked, using Gubbins v2.0.0 [35]. This alignment was examined by IQTree v2.0.4 [36] which produced a maximum-likelihood phylogeny using the best fit model that was visualized in ITOL v5.7 [37].

Cluster identification

Single-nucleotide polymorphism (SNP) addresses (described by Dallman *et al.* [29]) were generated as part of the processing pipeline in UKHSA for all isolates submitted to GBRU and were used to identify closely related strains and subsequently clusters. The SNP address comprises a pairwise clustering approach to assign distance threshold levels that descend: Δ250, Δ100, Δ50, Δ25, Δ10, Δ5, Δ0. Clusters were defined as three or more isolates that fall within the same 5 SNP single linkage cluster. Commonly, isolates that belong to a cluster are likely to be epidemiologically linked and/or share a common exposure.

The program snp-dists (https://github.com/tseemann/snp-dists) (a pairwise SNP distance matrix) was used to identify any additional SNP clusters or isolates that could be added to the pre-existing SNP clusters that were not initially picked up through the SNP address investigations.

The program fastbaps [38] (https://github.com/gtonkinhill/fastbaps) was used to identify genetic clusters from the SNP alignment that was provided to generate the phylogeny. The command line script that is installed with the software was used and clustering was set to level 5 (-l 5). Level 4 was where the clusters were genetically different.

Comparison of CC32 UK data with O157:H7 UK data and global publicly available CC32 data

The epidemiological data for STEC CC32 cases was compared to the published English data for O157:H7 (12). The seasonality of STEC CC32 was compared to that of STEC O157:H7, based on the dates of sample collection for O157:H7 routinely referred to GBRU.

Assemblies submitted to Enterobase [39] from global laboratories belonging to ST32, ST6130 and ST137 were identified using Enterobase workspaces (Table S4).

Global sample assemblies (*n*=926) were downloaded from Enterobase [39] (Table S4) and where collection year was available, the dates ranged between 1983–2021. The 309 samples referred to UKHSA (collection years 2004–2021) were assembled using Spades
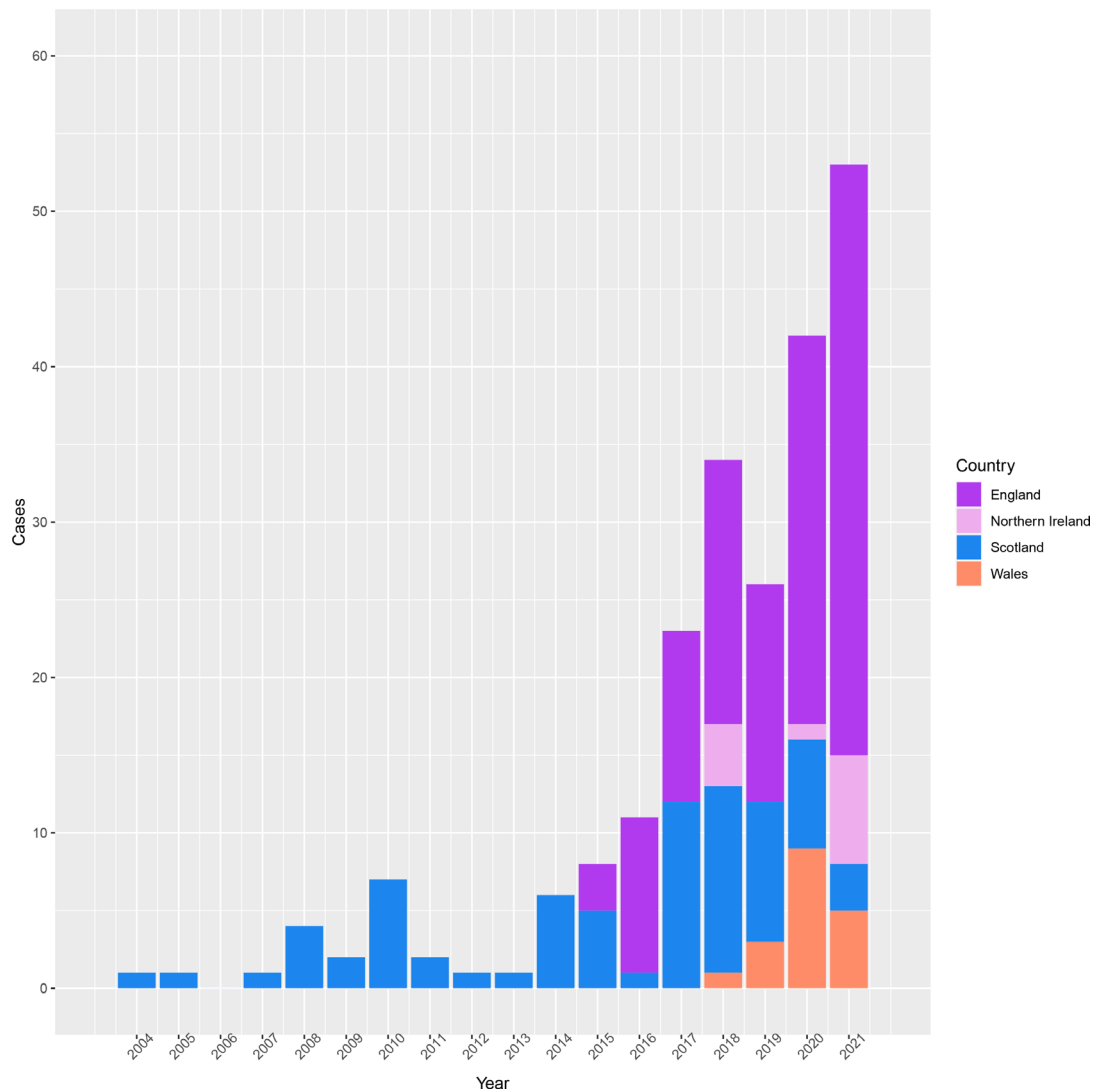
Fig. 1. Annual number of STEC cases belonging to CC32 reported to UKHSA, originating from England, Wales, Scotland and Northern Ireland (*n*=223) from 2004 to 2021. Note: isolates submitted to this study from Scotland were up until September 2021.

V 3.8.0 [40]. Quast v.5.0.2 [41] was used to investigate the quality of the assemblies from Enterobase [39] and Spades [40], and quality metrics are found in Table S5. Assemblies with a total length of ~4.8–5.7 Mbp with a number of contigs up to ~500 and a GC% of ~50% (min–max=48–52%) were included in the analysis. This resulted in 920 samples from countries outside the UK and 303 samples from the UK from this study.

Parsnp v2.3.4 [42] was used to investigate the phylogenetic relationship of UK samples in a global context built against reference CP006027.1 (ESC_GA4801AA_AS). Harvest v1.2 [42] was used to obtain the core gene alignment from Parsnp output, and fed into IQ Tree with 1000 bootstraps using the best-fit model. The tree was visualized in ITOL [37].

RESULTS

Epidemiology

Of the 309 cases of CC32 in the UK and Ireland in this study, there were 290 cases of STEC, of which 67/290 (22%) were resident in Ireland and were diagnosed between January 2016 and December 2018, and 223/290 (78%) were notified to UK public health agencies between 2004 and 2021 (England *n*=118/290, 41%; Scotland *n*=75/290, 26%; Wales *n*=18/290, 6%; Northern Ireland *n*=12/290, 4%). Of these, 286/290 (99%) belonged to STEC serotype O145:H28. Overall, diagnoses of STEC O145:H28 in the UK have increased every year since 2014 (except in 2019) (Fig. 1). Cases were detected in all regions of England, with the South East having the highest frequency (*n*=35/118, 30%), followed by London (*n*=22/118, 19%).
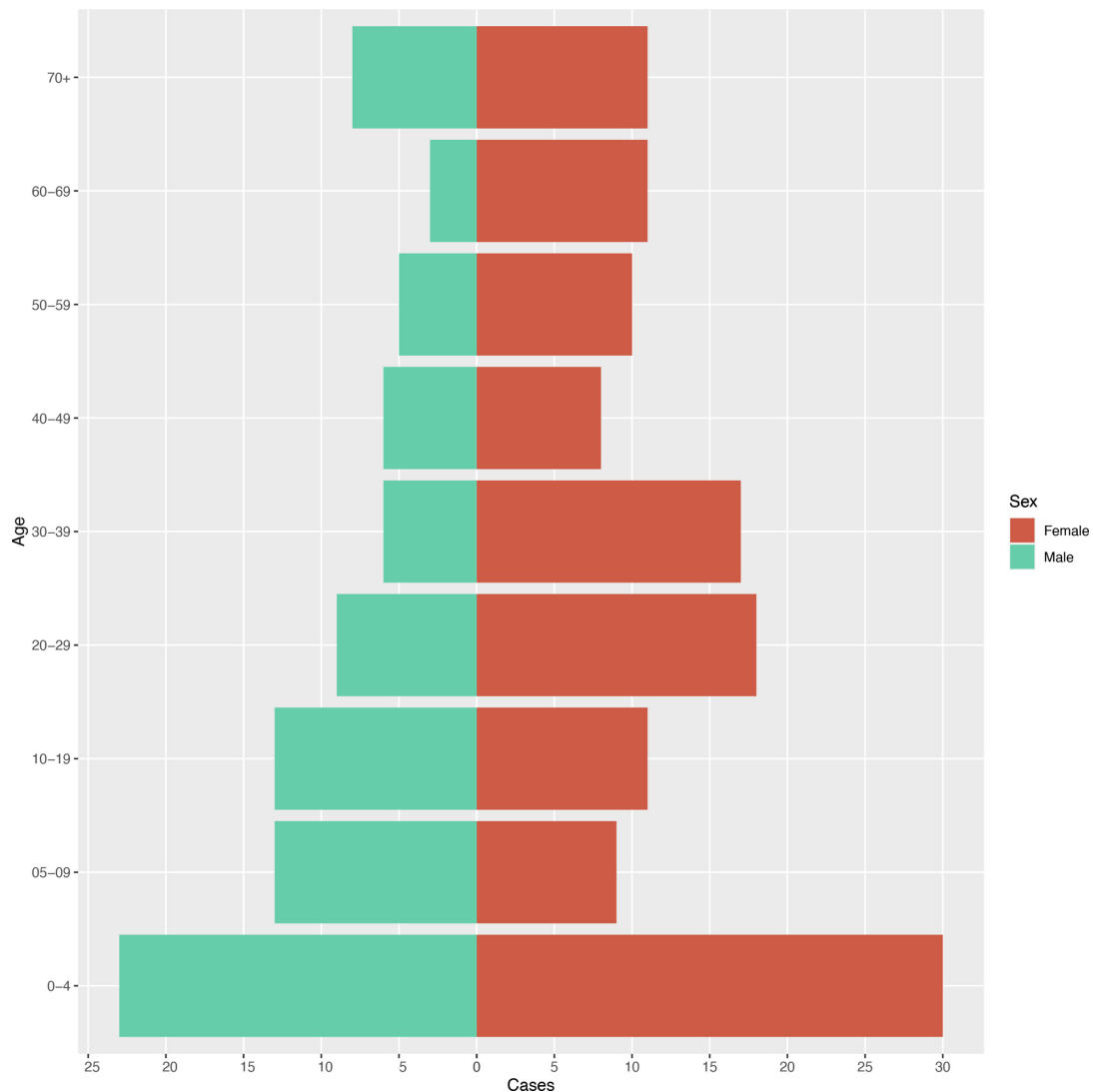
Fig. 2. Age–sex distribution of STEC cases reported to UKHSA [isolates originating from England, Wales, Northern Ireland and Ireland (*n*=211)], where date for birth and sample or receipt date were available.

For those cases of STEC CC32 where sex data were available (*n*=260/290), 61% (*n*=158/260) of the cases were female, compared to 56% for STEC O157:H7 (12). Where age and sex data were available (*n*=211/290), the median age for infections of STEC O145:H28 in females was 25 years [interquartile range (IQR): 5–50], and for males the median age was 14 years (IQR: 4–41), which is comparable with STEC O157:H7, where the female and male median ages are 26 and 17 years, respectively [12]. As with STEC O157:H7 infection, the highest proportion of cases belonged to the 0–4 age group for both males and females (53/211, 25%) (Fig. 2) [12]. Seasonal variation in the number of cases of STEC O145:H28 was reported, with numbers increasing from June and peaking in October, before decreasing to baseline levels during the winter months (Fig. 3). In contrast, the increase in case reports of STEC O157:H7 (*n*=4516) between 2014 and 2021 started earlier in April and peaked in August (Fig. 3).

Clinical outcome data were available for 156 cases of STEC O145:H28 from England, Scotland and Wales (Table 1). Diarrhoea (92%), abdominal pain (84%), blood in stool (71%) and nausea (51%) were reported in over half of the cases and the frequencies were similar to those reported for cases of STEC O157:H7 (chi^2 test: $P>0.05$) (Table 1). When compared to STEC O157:H7, cases of STEC O145:H28 were more frequently admitted to hospitals (STEC O145:H28 48% vs O157:H7 34%; chi^2 test: $P<0.05$), and HUS was more frequently reported (STEC O145:H28 9% vs O157:H7 4%; (chi^2 test: $P<0.05$) (Table 1). Among the 14 patients infected with STEC O145:H28 who developed HUS, further details were available for the 13 English cases. The sex distribution of cases revealed that eight females and five males were diagnosed with HUS. Six cases (46%) were aged between 0 and 4 (including one fatal case), two (15%) were between 5 and 9, three (23%) were between 10 and 19, and two (15%) were aged 45 years or above. Twelve of the isolates infecting the English HUS cases had *stx2a*, and one (the fatal case) had *stx1a, stx2a*.
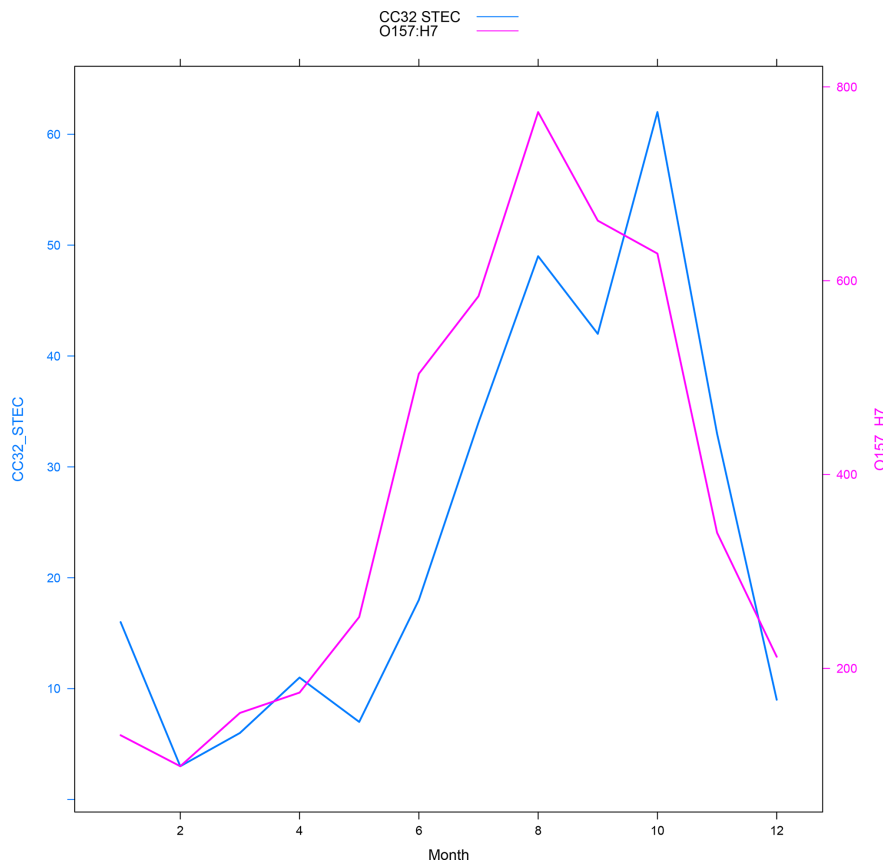
Fig. 3. Number of CC32 STEC (*n*=290) and O157:H7 (*n*=4516) by months in the UK and Ireland. The left *y*-axis and blue line indicate cases of CC32 STEC and the right *y*-axis and pink line indicate cases of O157:H7 STEC.

ESQ data, including information on animal exposures and travel histories, were available for 106 cases of STEC O145:H28. Over half of the cases reported contact with animals (Wales *n*=3/4, 75%; Scotland *n*=30/46, 65%; England *n*=55/106, 55%) (Table 2), and 12/106 (11%) reported travelling outside the UK within 7 days of onset of associated symptoms. Travel destinations were diverse, with two cases each reporting travel to Germany, Turkey and Tunisia and one case each to Armenia, France, Greece, Ireland, Italy and Portugal. Of the four cases from Wales with epidemiological data, one case confirmed travel to Ireland prior to illness onset. Three cases in Scotland had reported travel outside the UK prior to illness onset, but travel destinations were not available.

Population structure of CC32 in the UK and the global context

Overall, 309 isolates belonging to CC32 were analysed in this study, including 242/309 from the UK archives identified between 2004 and 2021, and 67/309 from Ireland identified between January 2016 and December 2018. Of these, 290/309 (94%) isolates harboured one or more *stx* and were defined as STEC. The remaining 19 isolates had *eae* but not *stx* and were defined as enteropathogenic *E. coli* (EPEC). Genome-derived serotyping revealed that O145:H28 was the dominant serotype (305/309, 99%), with serotypes O unidentifiable:H28 (2/309, 0.6%), O145:H16 (1/309, 0.3%) and O5:H9 (1/309, 0.3%) also being detected (Fig. 4). All EPEC isolates were serotype O145:H28.

Genome-derived MLST revealed four sequence types (STs): ST32 (269/309), ST137 (24/309), ST6130 (15/309) and ST8625 (1/309) (Fig. 4, Table S1). The population structure of CC32 from the UK was generated against reference CP006027 and fastbaps was used to identify genetic clusters based on the alignment. Seven genetic clusters (GCs) were identified using the algorithm (Fig. 4). The GC names are as follows; ST32-1, ST32-2 (ST8625 was part of GC2), ST1373, ST137-4, ST137-5, ST6130-6 and ST137-7. Visual inspection of ST32-2 revealed that this GC appeared to further split into three clades, where the branches have been coloured purple (ST32-2A), red (ST32-2B) and green ST32-2C).

The global population structure of CC32 includes ST32, ST137, ST6130 and ST8625 (Fig. S1). ST32 makes up the majority of the population, and this ST was pruned in ITOL to further detail the structure of ST32 globally (Fig. S1). The UK isolates of ST32 appear in two main clusters, indicating that there are two domestic clades within the UK. In the UK tree and global tree the two

Table 1. Clinical presentation of CC32 STEC cases from England (*n*=106/118), Scotland (*n*=46/75) and Wales (*n*=4/18), with comparison to O157:H7 STEC cases in England [12]

| Clinical Presentation | STEC CC32 | | | | | | | | STEC O157:H7 | | |
| | England (*n*=106) | | Scotland (*n*=46) | | Wales (*n*=4) | | Overall (*n*=156) | | 2009–2014 (*n*=4812) | 2015–2019 (*n*=2786) |
	Count	%	Count	%	Count	%	Count	%	%	%
Diarrhoea	98	92	41	89	4	100	143	92	93	95
Abdominal pain	89	84	38	83	4	100	131	84	80	83
Blood in stool*	67	63	39	85	4	100	110	71	62	61
Nausea	57	54	21	46	1	25	79	51	46	48
Vomiting	46	43	21	46	1	25	68	44	37	35
Fever	33	31	15	33	1	25	49	31	33	34
HUS†	13	13	1	2	0	0	14	9	5	3
Died	1	1	0	0	0	0	1	1	NA	NA
Admitted to hospital†	54	51	18	39	3	75	75	48	34	34

*Chi² produced a *P* value of >0.05 when the association between blood in stool and type of STEC infection was assessed.
†Chi2 produced a *P* value of <0.05 when the association between HUS and hospitalisation clinical outcome and type of STEC infection was assessed.

Table 2. Animal exposures of CC32 STEC cases from England (*n*=106), Scotland (*n*=46) and Wales (*n*=4) with comparison to O157:H7 STEC cases in England [12]

| | STEC CC32 | | | | | | | | STEC O157:H7 | |
| Animal | England (*n*=106) | | Scotland (*n*=46) | | Wales (*n*=4) | | Overall (*n*=156) | | 2009–2014 (*n*=4812) | 2015–2019 (*n*=2786) |
Animal	Count	%	Count	%	Count	%	Count	%	%	%
Contact with animals	55	52	30	65	3	75	88	56	56	59
Dogs	39	37	20	43	3	75	62	40	32	36
Cats	22	21	9	20	0	0	31	20	23	24
Visited farm	13	13	6	13	1	25	30	19	14	11

domestic clades in the UK were shaded in green and blue (Figs 4 and S1). The remaining UK isolates were dispersed across the tree and were situated with isolates originating from multiple countries, including the USA, Japan and Europe (Fig. S1). These isolate branches have also been coloured to match the branch colours in Fig. 4 to show the relationship between the UK and global populations. Associated travel data for isolates in the England were limited, but the patient in one case (SRR8087115) had travelled to France and that isolate clustered with isolates from France (Fig. S1).

Virulence genes associated with STEC CC32

Analysis of the *stx* subtyping revealed that *stx2a* was the most common subtype detected (*n*=268/290, 92%), followed by *stx1a* (*n*=11/290, 4%), *stx2c* (*n*=5/290, 2%), *stx1a,stx2a* (*n*=3/290, 1%) and *stx2d* (*n*=3/290, 1%) (Fig. 4, Table S1). Over half (15/24, 63%) of the ST137 isolates were EPEC and the remaining seven isolates harboured *stx1a* (*n*=4), *stx2c* (*n*=1) and *stx2a* (*n*=4) (Fig. 4). Twelve of the 15 ST6130 isolates harboured *stx2a*, with the remaining 3 not harbouring any Shiga toxin-encoding phage. ST32 was the largest ST and only 1/269 (0.4%) was EPEC. The remaining 268 isolates harboured *stx2a* (*n*=252), *stx1a* (*n*=6), stx2c (*n*=4), *stx1a,stx2a* (*n*=3) or *stx2d* (*n*=3).

Isolates were examined for genes that confirm the presence of the LEE, such as *espA*, *espB*, *espF* and *tir*. These 4 genes were detected in all 309 isolates, along with the presence of *eae*, confirming the LEE in CC32. We also screened the isolates for non-LEE effectors such as, *espJ*, *nleA*, *nleB* and *nleC*, where presence was varied. The genes *espJ*, *nleA*, *nleB* and *nleC* were detected in the majority of isolates (Fig. S2), with the exception of ST137-3 for *espJ* and *nleC*. The gene *iss* was detected in ST137-3, ST137-4 and all of ST32 (with two exceptions) (Fig. S2, Table S2).

Genes associated with the pO157 plasmid were investigated. Detection of these genes varied between the STs and sub-clusters of ST32. The gene *ehxA* was detected in all but six isolates and these six isolates did not harbour *espP*, *toxB* or *katP* either. The gene *espP* was not present in ST137-4, ST137-5, ST137-7 and ST6130-6. The genes *katP* and *toxB* were not present in ST137-3, ST137-5, ST137-7 and ST6130-6. The gene *katP* was also not detected in ST32-2C (one of the domestic clades) (Fig. S2). When we compare STEC CC32 virulence gene detection to that for another emerging HUS-causing clonal complex, CC165, the data show that STEC CC32 (O145:H28) has isolates that harbour *katP* and *toxB*, whereas CC165 does not [18].

Genome-derived AMR profiles

Three hundred and eight of 309 CC32 isolates were AMR-profiled *in silico*, with 48/308(16%) being multidrug-resistant (where 3 or more resistance determinants against different antimicrobial classes were detected in the genome) (Fig. S3, Table S3). Fifty-one (17%) isolates harboured resistance determinant to aminoglycosides, with the most common profile (*strA*/*strB*) being detected in 46 samples (Table S3). This was followed by 46/308 (15%) isolates with resistance determinants to sulfonamides; the most common profile was 36/46 isolates harbouring *sul-1* or *sul-2* (Table S3). Resistance to tetracyclines was held by 40/308 isolates with 25/40 harbouring *tet(A)-1* (Table S3).

Beta-lactam resistance was predicted in 33/308 (11%) isolates, with *TEM-1* detected in 27/33 isolates (Table S3). Trimethoprim resistance was predicted in 12/308 isolates, with 7/12 isolates harbouring *dfrA-1* (Table S3). Eight of 308 (2.5%) isolates held resistance to chloramphenicol, where all eight isolates had *floR,* and four isolates had an additional gene *catA-1* (*n*=3) and *cml-1* (*n*=1) (Table S3).

Outbreak detection and investigation

There were 17 SNP clusters of cases with isolate sequences that fell within a 5 SNP single linkage cluster (Figs 1 and 5) based on SNP address and SNP distance investigations. All isolates in clusters were STEC serotype O145:H28 and *stx* subtype *stx2a*. The
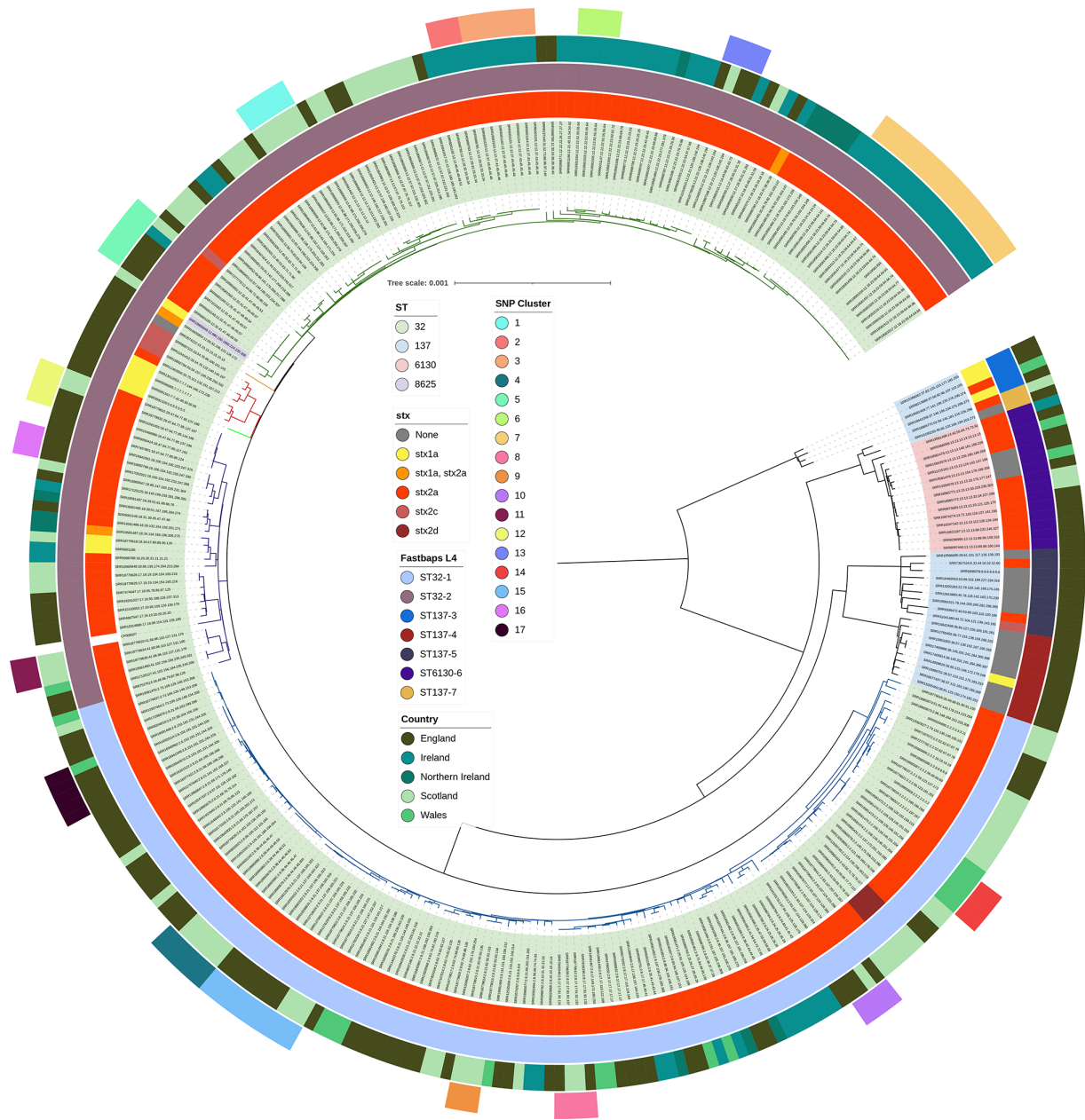
Fig. 4. Population structure of CC32 in the UK (midpoint rooted) curated in IQTree2 through an alignment of variant positions with a length of 9168 bases. The sequence type (ST) of isolates is shown through the colouring of isolate labels. The *stx* profile of isolates is indicated as the first colour strip after the ST labels. fastbaps genetic clusters are indicated by the first colour strip external of the ST labels. Where ST32-2 subdivides to ST32-2A, ST32-2B and ST32-2C, this has been indicated as coloured branches (purple, red and green, respectively). The country the isolates were referred to is indicated by the shaded green colour bar. SNP clusters within the population are also indicated through the outer ring that represents the 17 different clusters identified through the SNP address and SNP distance analysis.

median number of persons identified in each cluster was 4 (range: 3–16) (Table 3). Of the 17 clusters, 6 (38%) involved household transmission based on associated metadata taken during sample collection and referral.

Seven clusters were geographically dispersed, and likely to be associated with foodborne transmission (clusters 4, 5, 8, 12, 13, 15, 17). Eleven of the 17 clusters were temporally related, with the duration of time between cases ranging from 3 to 35 days (Table 3).

The largest cluster (cluster 7) comprised 17 cases, of which at least 10 were aged 5 and under, and the evidence indicated that this cluster was associated with a nursery outbreak and subsequent familial spread. Additionally, 16 of the 17 cases were reported in September and October 2017, and one final case was reported in January 2018.
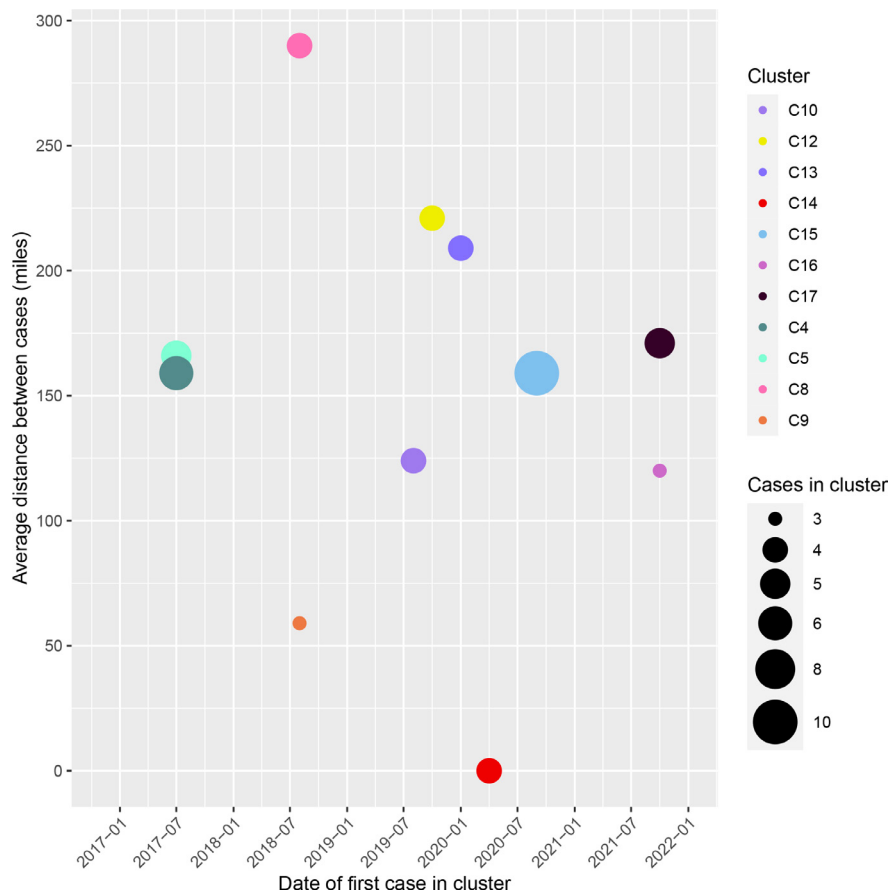
Fig. 5. SNP clusters of CC32 STEC serotype O145:H28 identified in the UK and Ireland. Each point is a 0 or 5 SNP cluster (*n*=11) (identified through 0 and 5 SNP single linkage clusters from SNP address and pairwise SNP analysis using snp-dist), where postcode data were available. Clusters are named in order of appearance: i.e. cluster 1 was the earliest date and cluster 16 was the most recently detected. Each cluster is represented by different colours, with the size of each dot proportional to the cluster size. The *x*-axis is the year–month of the detection for the first case of each cluster, with the *y*-axis showing the average distance in miles (beeline approach) between isolates within the cluster. Postcode information is missing for one isolate in cluster 5 and for all isolates in clusters 1, 2, 3, 6, 7 and 11.

There were two clusters widely dispersed over (clusters 10 and 15), both comprising cases with symptom onset dates over 12 months apart. Cluster 10 had four isolates, three from England in October/November 2021 (where 2/3 isolates had familial linkage) and one isolate from Scotland in August 2019. Of the 10 cases in cluster 15, 6 isolates were notified in September 2020 and 4 were detected in October 2021. This cluster was investigated and imported salad vegetables were the suspected vehicle, but this was not confirmed. Cluster 5 revealed that five cases that were reported in England and Scotland were referred within a 2 month range, whereas the case from Ireland was reported 4 months later.

DISCUSSION

During this study, analysis of trend data to assess incidence and prevalence of STEC CC32 was challenging due to the inconsistencies in the way surveillance is conducted across the UK and the ongoing changes to the algorithms for the diagnosis of non-O157 STEC. For Northern Ireland, Scotland and Wales, where surveillance strategies have been relatively consistent since 2018, the number of cases each year has fluctuated. Despite social distancing lockdown restrictions during the coronavirus disease 2019 (COVID-19) pandemic, the number of cases of CC32 in England has increased during the last 2 years; however, it is difficult to unpick whether this is a true increase or due to improvements in the diagnostics algorithms for STEC.

Although in England cases were reported across all the regions, the highest rates were seen in the South East and London, where a higher proportion of local diagnostic laboratories have implemented PCR. This pattern has been seen with other clonal complexes of non-O157 STEC in England [18]. In addition to the inconsistent approach to the detection of non-O157 STEC across England, strategies for referral of PCR-positive faecal specimens to the national reference laboratories for confirmation, culture and typing also vary, further confounding the surveillance of non-O157 STEC.

Table 3. SNP clusters of CC32 STEC cases isolated in the UK and Ireland identified through the generation of single linkage SNP addresses. Columns indicate country/countries of origin, number of cases in each cluster and the duration in days between the first and last case in each cluster, based on receipt date, as this was consistently reported across all cases. Asterisks (*) identify the clusters where the epidemiological data suggest a level of familial transmission. Data for familial transmission in Scottish cases were not available

Cluster	Country/countries	Cases (*n*)	Duration between first and last case (days)
1	Scotland	5	35
2*	Ireland	3	14
3*	Ireland	7	10
4	England and Scotland	6	29
5	England, Scotland and Ireland	6	165
6*	Ireland	4	114
7*	Ireland	17	109
8	England and Scotland	4	3
9	Scotland	3	3
10*	England and Scotland	4	810
11	Scotland	3	8
12	England and Scotland	4	95
13	England and Scotland	4	19
14*	Wales	4	4
15	England and Scotland	10	399
16	England	3	23
17	England and Wales	5	14

As with STEC O157:H7, there was higher proportion of female cases, most notably in the young adult age groups. Historically, this has been attributed to the increased risk of exposure to STEC during food preparation and childcare, although these are no longer regarded as traditionally female roles. Overall, as with STEC O157:H7, the majority of cases were aged 0–4 years old, which may be due to the increased likelihood of children in that age group presenting to primary healthcare when symptomatic, or may be a true indication that children under the age of 5 years old are more vulnerable and more susceptible to STEC infection.

Clinical presentation of STEC O145:H28 STEC mirrors that of O157:H7 STEC, with high rates of diarrhoea, abdominal pain and blood in the stool [12]. The risk of developing HUS for O145:H28 STEC in England was lower than the risk following infection with CC165 [18], but higher than that associated with O157:H7 STEC [12]. The higher proportion of HUS cases in patients infected with STEC O145:H28 when compared to STEC O157:H7 is likely to reflect the higher proportion of isolates harbouring *eae* and *stx2a*, a combination of virulence genes associated with more severe clinical outcomes [6]. STEC belonging to CC165 carry additional plasmid-encoded extra-intestinal virulence genes, in addition to *eae* and *stx*, which are thought to enhance their pathogenic potential [18].

While the source and transmission routes of O157:H7 STEC are well described, those for most of the non-O157 STEC clonal complexes, in the UK at least, are less clear. However, as with STEC O157:H7, there is evidence that cattle are an important zoonotic reservoir of STEC O145:H28 in the UK and elsewhere. A recent cattle survey in the UK provided evidence that STEC O145:H28 are endemic in UK cattle [21], and STEC O145:H28 has also been detected in cattle/cattle farms in Belgium and the USA [20, 22]. The most common animal contact reported by the English cases was with dogs. Recently, outbreaks of gastrointestinal disease have been associated with handling raw pet food [43]. Dog owners are also more at risk from environmental exposures while walking in rural areas inhabited by cattle and other ruminants that may be colonized with STEC or wildlife acting as transient vectors.

Although only approximately 25% of local diagnostic laboratories in England are able to detect non-O157 STEC using PCR, a number of outbreak clusters of STEC O145:H28 were detected in England. These clusters were small and geographically dispersed, and for the most part, temporally related, indicative of foodborne transmission linked to a nationally distributed product. STEC O145:H28 have been linked to foodborne outbreaks caused by contaminated flour, ice cream and leafy greens [23, 24, 26]. We also observed evidence of person-to person transmission of STEC O145:H28, as a number of household clusters were detected, together with an outbreak potentially linked to a nursery school, as also described previously [25]. These data provide evidence

that person-to-person transmission of STEC O145:H28 can occur in certain settings. As with STEC O157:H7 and STEC O26:H11, transmission of O145:H28 is likely to be facilitated by a low infectious dose and long-term asymptomatic shedding, and patients in high-risk groups, such as children aged 5 and under and food handlers, should be excluded from returning to childcare settings or work until their faecal specimen is microbiologically clear of the pathogen [20, 44].

In common with most STEC clonal complexes detected in UK residents – CC165 and CC504 being notable exceptions – the majority of isolates belonging to CC32 were predicted to be fully susceptible to the 10 classes of antibiotic included in the testing algorithm, based on the genome-derived AMR profiles. The most commonly detected genes conferred resistance to ampicillin, aminoglycosides, sulphonamides and tetracyclines, and these antimicrobials account for the majority of antimicrobials sold for veterinary use [45]. As STEC is zoonotic, it is plausible that, historically, antimicrobial use in animal husbandry was a driver for the acquisition of the resistance determinants observed in a subset of STEC in the UK.

The identification of UK domestic clades within the global context, the low number of travel-acquired infections and the recent cattle survey provided evidence that STEC O145:H28 is endemic in the UK. As with STEC O157:H7, ruminants may be the primary zoonotic reservoir, although further animal sampling surveys are required to determine the range of animal hosts and the role of transient wildlife vectors and household pets. We know that clinical outcomes of infection with STEC O145:H28 are similar to those for STEC O157:H7, and there may be a higher risk of hospitalization and progression to HUS, most likely because of the high proportion of the STEC O145:H28 population that have the highly pathogenic *stx2a* and *eae* gene combination.

We recommend the more widespread use of PCR for the detection of all STEC serogroups, as well as more consistent strategies for the follow-up confirmation/characterization of PCR-positive faecal specimens. More comprehensive and standardized collection of clinical outcomes, food and environmental exposure data, and routine sharing of sequencing data between public health agencies in the UK, will facilitate the detection and investigation of future outbreaks of STEC O145:H28.

References

1. Byrne L, Jenkins C, Launders N, Elson R, Adak GK. The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009–2012. *Epidemiol Infect* 2015;143:3475–3487.

2. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, *et al*. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin Microbiol Rev* 2013;26:822–880.

3. Stevens MP, Frankel GM. The locus of enterocyte effacement and associated virulence factors of enterohemorrhagic *Escherichia coli*. *Microbiol Spectr* 2014;2:EHEC-0007.

4. Boisen N, Melton-Celsa AR, Scheutz F, O'Brien AD, Nataro JP. Shiga toxin 2a and eteroaggregative *Escherichia coli*--a deadly combination. *Gut Microbes* 2015;6:272–278.

5. Cointe A, Birgy A, Mariani-Kurkdjian P, Liguori S, Courroux C, *et al*. Emerging multidrug-resistant hybrid pathotype Shiga toxin-producing *Escherichia coli* O80 and related strains of clonal complex 165, Europe. *Emerg Infect Dis* 2018;24:2262–2269.

6. Byrne L, Adams N, Jenkins C. Association between Shiga toxin-producing *Escherichia coli* O157:H7 stx gene subtype and disease severity, England, 2009-2019. *Emerg Infect Dis* 2020;26:2394–2400.

7. Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bover-Cid S, *et al*, EFSA BIOHAZ Panel. Pathogenicity assessment of Shiga toxin-producing *Escherichia coli* (STEC) and the public health risk posed by contamination of food with STEC. *EFS2* 2020;18.

8. Gasser C, Gautier E, Steck A, Siebenmann RE, Oechslin R. Hemolytic-uremic syndrome: bilateral necrosis of the renal cortex in acute acquired hemolytic anemia. *Schweiz Med Wochenschr* 1955;85:905–909.

9. Obrig TG. *Escherichia coli* Shiga toxin mechanisms of action in renal disease. *Toxins* 2010;2:2769–2794.

10. Lingwood CA. Verotoxin/globotriaosyl ceramide recognition: angiopathy, angiogenesis and antineoplasia. *Biosci Rep* 1999;19:345–354.

11. Karmali MA, Petric M, Lim C, Fleming PC, Arbus GS, *et al*. The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*. *J Infect Dis* 1985;151:775–782.

12. Butt S, Jenkins C, Godbole G, Byrne L. The epidemiology of shiga toxin-producing *Escherichia coli* serogroup O157 in england, 2009-2019. *Epidemiol Infect* 2022;1–30.

13. Vishram B, Jenkins C, Greig DR, Godbole G, Carroll K, *et al*. The emerging importance of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England. *J Med Microbiol* 2021;70.

14. Simms I, Gilbart VL, Byrne L, Jenkins C, Adak GK, *et al*. Identification of verocytotoxin-producing *Escherichia coli* O117:H7 in men who have sex with men, England, November 2013 to August 2014. *Euro Surveill* 2014;19.

15. Baker KS, Dallman TJ, Thomson NR, Jenkins C. An outbreak of a rare Shiga-toxin-producing *Escherichia coli* serotype (O117:H7) among men who have sex with men. *Microb Genom* 2018;4.

16. Dallman T, Cross L, Bishop C, Perry N, Olesen B, et al. Whole genome sequencing of an unusual serotype of Shiga toxin-producing *Escherichia coli*. *Emerg Infect Dis* 2013;19:1302–1304.

17. Olesen B, Jensen C, Olsen KEP, Fussing V, Gerner-Smidt P, et al. VTEC O117:K1:H7 a new clonal group of *E. coli* associated with persistent diarrhoea in Danish travellers. *Scand J Infect Dis* 2009;37:288–294.

18. Rodwell EV, Vishram B, Smith R, Browning L, Smith-Palmer A, et al. Epidemiology and genomic analysis of Shiga toxin-producing *Escherichia coli* clonal complex 165 in the UK. *J Med Microbiol* 2021;70.

19. Bertoldi B, Richardson S, Goodrich-Schneider R, Kurdmongkoltham P, Schneider K. Preventing Foodborne Illness: *E. coli* "The Big Six [Internet]. University of Florida. https://nifa.usda.gov/sites/default/files/resource/Preventing-Foodborne-Illness-E-coli-the-big-six.pdf

20. Shridhar PB, Worley JN, Gao X, Yang X, Noll LW, et al. Analysis of virulence potential of *Escherichia coli* O145 isolated from cattle feces and hide samples based on whole genome sequencing. *PLOS ONE* 2019;14:e0225057.

21. Hoyle DV, Keith M, Williamson H, Macleod K, Mathie H, et al. Prevalence and epidemiology of Non-O157 *Escherichia coli* Serogroups O26, O103, O111, and O145 and Shiga toxin gene carriage in Scottish cattle, 2014-2015. *Appl Environ Microbiol* 2021;87:e03142-20.

22. Engelen F, Thiry D, Devleesschauwer B, Mainil J, De Zutter L, et al. Occurrence of "gang of five" Shiga toxin-producing *Escherichia coli* serogroups on Belgian dairy cattle farms by overshoe sampling. *Lett Appl Microbiol* 2021;72:415–419.

23. Projahn M, Lamparter MC, Ganas P, Goehler A, Lorenz-Wright SC, et al. Genetic diversity and pathogenic potential of Shiga toxin-producing *Escherichia coli* (STEC) derived from German flour. *Int J Food Microbiol* 2021;347:109197.

24. Buvens G, Possé B, De Schrijver K, De Zutter L, Lauwers S, et al. Virulence profiling and quantification of verocytotoxin-producing *Escherichia coli* O145:H28 and O26:H11 isolated during an ice cream-related hemolytic uremic syndrome outbreak. *Foodborne Pathog Dis* 2011;8:421–426.

25. Wahl E, Vold L, Lindstedt BA, Bruheim T, Afset JE. Investigation of an *Escherichia coli* O145 outbreak in a child day-care centre-extensive sampling and characterization of eae- and stx1-positive *E. coli* yields epidemiological and socioeconomic insight. *BMC Infect Dis* 2011;11:238.

26. Taylor EV, Nguyen TA, Machesky KD, Koch E, Sotir MJ, et al. Multistate outbreak of *Escherichia coli* O145 infections associated with romaine lettuce consumption, 2010. *J Food Prot* 2013;76:939–944.

27. Jenkins C, Ling CL, Ciesielczuk HL, Lockwood J, Hopkins S, et al. Detection and identification of bacteria in clinical samples by 16S rRNA gene sequencing: comparison of two different approaches in clinical practice. *J Med Microbiol* 2012;61:483–488.

28. Jenkins C, Perry NT, Godbole G, Gharbia S. Evaluation of chromogenic selective agar (CHROMagar STEC) for the direct detection of Shiga toxin-producing *Escherichia coli* from faecal specimens. *J Med Microbiol* 2020;69:487–491.

29. Dallman TJ, Greig DR, Gharbia SE, Jenkins C. Phylogenetic context of Shiga toxin-producing *Escherichia coli* serotype O26:H11 in England. *Microb Genom* 2021;7.

30. Scottish *E. coli* O157/STEC Reference Laboratory. Whole genome sequence typing and analysis of non-O157 STEC [Internet]. Food Standards Scotland; 2020. https://www.foodstandards.gov.scot/downloads/WGS_Typing_and_Analysis_of_Non-O157_STEC_-_Jan_2020_v3.pdf

31. Joensen KG, Engsbro ALØ, Lukjancenko O, Kaas RS, Lund O, et al. Evaluating next-generation sequencing for direct clinical diagnostics in diarrhoeal disease. *Eur J Clin Microbiol Infect Dis* 2017;36:1325–1338.

32. Chattaway MA, Dallman TJ, Gentle A, Wright MJ, Long SE, et al. Whole genome sequencing for public health surveillance of Shiga toxin-producing *Escherichia coli* other than Serogroup O157. *Front Microbiol* 2016;7:258.

33. Tewolde R, Dallman T, Schaefer U, Sheppard CL, Ashton P, et al. MOST: a modified MLST typing tool based on short read sequencing. *PeerJ* 2016;4:e2308.

34. Dallman T, Ashton P, Schafer U, Jironkin A, Painset A, et al. SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. *Bioinformatics* 2018;34:3028–3029.

35. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43:e15.

36. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020;37:2461.

37. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.

38. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res* 2019;47:5539–5549.

39. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M. The enterobase user's guide, with case studies on *Salmonella* transmissions, Yersinia pestis phylogeny, and *Escherichia* core genomic diversity. *Genome Res* 2020;30:138–152.

40. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. *Curr Protoc Bioinformatics* 2020;70:e102.

41. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 2018;34:i142–i150.

42. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 2014;15:11.

43. Kaindama L, Jenkins C, Aird H, Jorgensen F, Stoker K, et al. A cluster of Shiga Toxin-producing *Escherichia coli* O157:H7 highlights raw pet food as an emerging potential source of infection in humans. *Epidemiol Infect* 2021;149:e124.

44. Reeves PR, Liu B, Zhou Z, Li D, Guo D, et al. Rates of mutation and host transmission for an *Escherichia coli* clone over 3 years. *PLoS One* 2011;6:e26907.

45. Eckford S, Grace K, Harris C, Reeves H. UK Veterinary antibiotic Resistance and Sales Surveillance. Internet. Veterinary Medicines Directorate; 2013. https://www.gov.uk/government/publications/veterinary-antimicrobial-resistance-and-sales-surveillance-2013