Identification of domestic reservoirs and common exposures in an emerging lineage of Shiga toxin-producing Escherichia coli O157:H7 in England

Identification of domestic reservoirs and common exposures in an emerging lineage of Shiga toxin-producing *Escherichia coli* O157:H7 in England: a genomic epidemiological analysis

Timothy J Dallman, Katri Jalava*, Neville Q Verlander, David Gally, Claire Jenkins, Gauri Godbole, Saheer Gharbia*

Summary

Background The zoonotic pathogen Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 emerged during the 1980s as a causative agent of foodborne outbreaks associated with haemorrhagic colitis and haemolytic uraemic syndrome, which can be fatal. We investigated the emerging lineage IIc that was causing outbreaks of STEC O157:H7, identified and quantified the domestic and non-domestic reservoirs, and quantified patient exposures across the population of England.

Methods In this genomic epidemiological analysis study, all human STEC O157:H7 lineage IIc (n=925) isolates cultured from faecal specimens submitted to the UK Health Security Agency between June 1, 2015, and Dec 31, 2020, from patients in England in the community or in hospital, were whole-genome sequenced and the genomic population structure was described. Explanatory variables were obtained from microbiological surveillance data and STEC Enhanced Surveillance Questionnaire responses. Ancestral-state reconstruction using patient travel information was used to define domestic and non-domestic clades and transmission dynamics. Exposures for patients infected with isolates from domestic clades were assessed using mixed-effects multinomial univariable and multivariable regression.

Findings Lineage IIc emerged 50 years ago, and subsequent clonal expansions have resolved into six major extant clades. We defined two English domestic clades that emerged during the past 30 years, and four non-domestic clades comprising isolates that infected or were transmitted to patients in England via international travel or the consumption or handling of imported food. Throughout the study period, non-domestic clades contributed approximately twice the number of infections as domestic clades did. Patients infected with domestic IIc clade strains reported more frequent exposure to fresh produce (raw vegetables p=0·012; prepackaged salad p=0·0009), contact with animals (cattle p=0·021), and visits to farms (p=0·0053) than patients infected with strains from other STEC O157:H7 lineages. A multivariable mixed-effects multinomial model confirmed that within the domestic clades, the major risk factors for infection were prepackaged salad (clade 2.3.3, relative risk ratio [RRR] 1·72, 95% CI 1·09–2·72; p=0·019) and visits to farms (clade 2.5.2, RRR 1·98, 1·12–3·52; p=0·020) as fixed effects. Local authority district as a random variable had a strong but variable effect for clades 2.3.3 and 2.5.2.

Interpretation Lineage IIc has emerged as the most prevalent lineage of STEC O157:H7 in England, with a sizeable domestic reservoir. Human infection is associated with the consumption of contaminated fresh produce and contact with domestic livestock. The collection of routine, detailed exposure data on patients who are infected, integrated with high-resolution microbiological typing, enables powerful reframing of our understanding of foodborne disease risk within a One Health context.

Funding National Institute for Health and Care Research Health and UK Health Security Agency.

Lancet Microbe 2022

Published Online
June 24, 2022
https://doi.org/10.1016/S2666-5247(22)00089-1

*Contributed equally

UK Health Security Agency, London, UK (T J Dallman PhD, K Jalava, PhD, N Q Verlander PhD, C Jenkins PhD, G Godbole MD, Prof S Gharbia PhD); **Division of Infection and Immunity, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, UK** (T J Dallman, Prof D Gally PhD); **Faculty of Veterinary Medicine, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, Netherlands** (T J Dallman)

Correspondence to:
Dr Timothy J Dallman, Faculty of Veterinary Medicine, Institute for Risk Assessment Sciences, Utrecht University, 3584 CL Utrecht, Netherlands
t.j.dallman@uu.nl

Introduction

Shiga toxin-producing *Escherichia coli* (STEC) serotype O157:H7 emerged in the 1980s as a zoonotic pathogen that causes human infections via the consumption of contaminated food or contact with infected animals or their environment.[1–4] Symptoms range from uncomplicated gastroenteritis to increasingly severe clinical outcomes that include haemorrhagic colitis, with some patients at risk of developing haemolytic uraemic syndrome (a condition mainly associated with renal dysfunction that can be fatal).[5] The defining genetic feature of STEC is the integration of one or more lamdoid bacteriophage encoding Shiga toxin genes, of which there are two types (*stx1* and *stx2*) and multiple subtypes (*stx1a–c*, *stx2a–k*).[6] There is a clear association between Shiga toxin profile and clinical severity, with strains harbouring *stx2a* substantially more likely to cause severe disease (including haemolytic uraemic syndrome) than other strains.[7,8] Additionally, strains encoding *stx1a* are associated with patients presenting with bloody diarrhoea.[8]

STEC O157:H7 can be delineated genetically into seven lineages: Ia, Ib, Ic, I/II, IIa, IIb, and IIc, with an

Research in context

Evidence before this study

We searched PubMed on Sept 1, 2021 for papers published in English between Jan 1, 2000, and Sept 1, 2021 using the search terms "O157" AND "phylogenetics" OR "whole genome sequencing" and identified 510 results. Several studies reported on STEC O157:H7 genetic diversity and identified many cocirculating lineages causing human infection of differing severity in a range of geographical settings. Similarly, there were multiple reports on the genetic characterisation of outbreaks of STEC O157:H7 that have been attributed to a range of transmission routes, of both animal and fresh produce origin. However, the clinical and epidemiological importance of these lineages and subpopulations within lineages is largely unknown, and the association of STEC O157:H7 outbreaks with the genomic population structure remains largely unexplored.

Added value of this study

To our knowledge, we provide the first detailed description of the epidemiological, genetic, and clinical characteristics of STEC O157:H7 lineage IIc, using 5 years of longitudinal enhanced surveillance in England. Phylodynamic analysis identified domestic and non-domestic clades with high precision and provided evidence of historical and recent international strain transmission. The virulence profile, as derived from the Shiga toxin profile, was equivalent across domestic and non-domestic clades; however, many non-domestic clades showed multidrug resistance. Consumption of prepackaged salad and visits to farms were identified as risk factors for the two domestic clades, corroborating the association between outbreaks of this lineage with fresh produce and the farming environment.

Implications of all the available evidence

Lineage IIc has increased in prevalence in England, from 30% of STEC O157:H7 cases in 2015 to 44% in 2020, and is now the dominant lineage in England. Predominantly, lineage IIc presents as a severe pathogen, with 80% of patients presenting with bloody diarrhoea, but does not commonly lead to haemolytic uraemic syndrome (<1% patients). Lineage IIc has a sizeable domestic reservoir in cattle in England, with infection often associated with contaminated fresh produce. Rapid risk assessment of the likely reservoir and mode of infection facilitates timely public health and food safety action, thus, reducing the threat of ongoing transmission and further infection.

association between lineage and *stx* profile.[8,9] STEC O157:H7 has a prominent phylogeographic signal that shows a pattern of historical strain dissemination, followed by regional expansion that has seen most lineages disseminated globally but with discernible intercountry diversity.[8,10,11] In England in the 1980s, human infections were mainly caused by strains belonging to lineage I/II,[8] whereas during the 1990s, lineage Ic replaced lineage I/II, and remained the dominant lineage for the next two decades.[8] Concurrently, from the 1990s until now, there has been a gradual increase in lineage IIc strains, which contribute to approximately a third of human infections annually.[12]

At the time of the emergence of STEC O157:H7, outbreaks were often associated with the consumption of contaminated beef,[13,14] and the consumption of contaminated beef, lamb, and dairy products and direct contact with ruminants are still common sources of outbreaks predominated by strains of lineages I/II or Ic.[15] Outbreaks associated with fresh produce (ie, fruit and vegetables) have been described as early as 1995,[16] predominantly reported in countries with a high incidence of STEC O157:H7. The lineages associated with these outbreaks vary;[17,18] however, in England, a number of outbreaks linked to fresh produce have been caused by strains that belong to lineage IIc.[19–21] Outbreaks associated with fresh produce pose challenges to food safety and public health management, especially with ready-to-eat products that have complex food-distribution chains and short shelf-lives.

Cattle surveys done in Great Britain to ascertain the prevalence of STEC O157:H7 and associated circulating strain types have revealed the presence of IIc strains in the British cattle population.[22] The proportion of lineage IIc strains isolated in a 2014 survey accounted for 17% of the total STEC O157:H7 isolated. The contribution of the domestic ruminant reservoir to the burden of clinical infection caused by lineage IIc strains is unclear.[9]

This study had two aims: first, to describe the population structure of lineage IIc and to estimate the contribution of a domestic reservoir to clinical disease, and second, to ascertain whether patients with lineage IIc were more likely to have had exposure to fresh produce compared with patients with other lineages of STEC O157:H7.

Methods

Study design and participants

In this genomic epidemiological study, all presumptive isolates of STEC O157:H7 cultured from faecal specimens submitted to the UK Health Security Agency between June 1, 2015, and Dec 31, 2020, from patients in the community or hospital in England, were characterised using whole-genome sequencing. In total, 2851 STEC O157:H7 isolates (representing a single isolate per patient) were included, representing all seven major lineages and with lineage IIc comprising 925 isolates. A further 300 isolates of lineage IIc from the UK Health Security Agency (UKHSA) archive (from June 1, 1985, to Dec 31, 2020) were included in the phylodynamic analyses, as were 21 lineage IIc isolates from British cattle surveys.

All patients with confirmed STEC O157:H7 in England were asked to complete a STEC Enhanced Surveillance Questionnaire to ascertain clinical presentation, food and animal exposures, and recent (ie, within 7 days before illness) travel history. Ethical approval for the identification, characterisation, and typing of cultures of gastrointestinal pathogens is not required as this is covered by UKHSA's surveillance mandate.

Sequence analyses

DNA extraction, whole-genome sequencing, and variant detection were done as described in a previous publication[9] and details of the methods are in the appendix (p 1).

Maximum likelihood phylogenies were computed using IQ-TREE (version 2.0.4[23]) with the best-fit model automatically selected and near-zero branches collapsed into polytomies. Clonal representatives (one per five single nucleotide polymorphisms [SNPs] single-linkage cluster) of the extended dataset (including isolates from the historical archive and cattle survey) were selected for phylodynamic analyses. Timed phylogenies were implemented in BEAST (version 1.10.4[24]), using a constant clock with the mutation rate as ascertained previously[8] using the Bayesian skyline population size prior. Ancestral state reconstruction was done using patient travel data as a discrete state. Markov chain Monte Carlo simulations were run for 50 million chains. Shiga toxin subtyping and identification of antimicrobial resistance genes were done as described in previous publications,[25,26] and the databases are available online.

Epidemiological and statistical analysis

STEC O157:H7 lineage IIc infections were classified into domestic clades[2] and non-domestic clades.[4] The demographic, clinical, and exposure characteristics of patients infected with STEC O157:H7 were described using data from the enhanced questionnaire.

We estimated the infection risk factors using a mixed-model multinomial regression with fixed and random effects, with the number of patients with a domestic clade (ie, clades 2.3.3 and 2.5.2) as the outcome variables and the total number of patients with other lineage IIc and non-lineage IIc strains as a reference group. Formal comparison tests for demographic, clinical, and exposures variables across the STEC O157:H7 lineage IIc clades 2.3.3 and 2.5.2 were obtained by initial univariable and multivariable multinomial regression.

The explanatory variables for the models were obtained from microbiological surveillance data and the STEC Enhanced Surveillance Questionnaire responses. After inclusion of random variable and adjustment factors, food and environmental variables were added one-by-one to the multivariable model. Possible multicollinearity was tested using the variance inflation factor with a cutoff of four.[27] All analyses were done with R (version 4.1.1) and Stata (version 17). Detailed methods are in the appendix (p 1).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The proportion of lineage IIc among STEC O157:H7 lineages in England is growing each year (104 [30·2%] of 344 in 2015 *vs* 147 [43·9%] of 335 in 2020; p<0·0001).

In women, a similar proportion were infected with lineage IIc strains as were infected with non-lineage IIc strains (532 [57·5%] of 925 with lineage IIc *vs* 1080 [56·1%] of 1926 with non-lineage IIc; p=0·47). There was a higher proportion of individuals older than 5 years among patients with lineage IIc than for those with non-lineage IIc (820 [88·6%] of 925 *vs* 1514 [78·6%] of 1926; p<0·0001). The regions of England with the highest incidence of lineage IIc infections were the same as those with the highest incidence of non-lineage IIc infections (ie, the South West, Yorkshire and the Humber, and the North East). Of the 925 patients infected with lineage IIc strains, 336 (38·5%) of 873 with non-missing data reported travel outside of the UK within 1 week before illness. Patients infected with STEC O157:H7 lineage IIc frequently presented with bloody diarrhoea (641 [77·0%] of 832) but rarely with haemolytic uraemic syndrome (four [0·4%] of 925; for lineage-by-lineage comparisons of travel and clinical outcomes, see table 1).

The population structure of lineage IIc can be delineated into two major sublineages (1 and 2), each with a subsequent deep-rooted substructure within them. Resolving the maximum likelihood phylogeny by the third bifurcation from the root splits the population into 15 clades, with six of these clades representing 95% of the isolates included in the study period (appendix p 3).

Lineage IIc emerged approximately 50 years ago (95% highest posterior density interval [HPDI] 1965–1976), with the most recent common ancestor emerging approximately 45 years ago (95% HPDI 1974–1978) for sublineage 1, and approximately 35 years (95% HPDI 1983–1989) for sublineage 2 (figure 1A). The most recent common ancestor for the subsequent clades was approximately 32 years previously (95% HPDI 1987–1991) for 1.3.2, 17 years (95% HPDI 2001–2003) for 2.3.3, 25 years (95% HPDI 1993–1998) for 2.4.1, 17 years (95% HPDI 2001–2004) for 2.4.2, 17 years (95% HPDI 2001–2004) for 2.5.1, and 32 years (95% HPDI 1987–1990) for 2.5.2.

Estimates of effective population size over time revealed lineage IIc to be an expanding population in terms of genetic diversity (figure 1B). An initial population expansion in the early 1990s was observed, followed by a second steady increase in population size from 2007 onwards before numbers plateaued from 2015 onwards. The first increase in effective population size was congruent with the formation of the older

For the **STEC Enhanced Surveillance Questionnaire** see https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/828287/Shiga_toxin-producing_Escherichia_coli_Enhanced_Surveillance_Questionnaire__Paper_Version_.pdf

For more on the **UKHSA's surveillance mandate** see https://www.gov.uk/government/publications/ukhsa-priorities-in-2021-to-2022/letter-from-lord-bethell-to-dr-jenny-harries-ukhsa-chief-executive

See **Online** for appendix

For the **databases** see https://github.com/phe-bioinformatics/gene_finder

	I/II (n=91)	Ia (n=68)	Ib (n=101)	Ic (n=666)	IIa (n=592)	IIb (n=408)	IIc (n=925)
Sex							
Female	46/91 (50·5%, 40·5–60·6)	38/68 (55·9%, 44·1–67·1)	61/101 (60·4%, 50·6–69·4)	346/666 (52·0%, 48·2–55·7)	366/592 (61·8%, 57·8–65·7)	223/408 (54·7%, 49·8–59·4)	532/925 (57·5%, 54·3–60·7)
Male	45/91 (49·5%, 39·4–59·5)	30/68 (44·1%, 32·9–55·9)	40/101 (39·6%, 30·6–49·4)	320/666 (48·0%, 44·3–51·8)	226/592 (38·2%, 34·3–42·2)	185/408 (45·3%, 40·6–50·2)	393/925 (42·5%, 39·3–45·7)
Risk factor							
Younger than 5 years	11/91 (12·1%, 6·7–20·5)	3/68 (4·4%, 1·0–12·7)	22/101 (21·8%, 14·8–30·8)	148/666 (22·2%, 19·2–25·5)	124/592 (20·9%, 17·9–24·4)	104/408 (25·5%, 21·5–29·9)	105/925 (11·4%, 9·5–13·6)
Travel outside the UK*	15/87 (17·2%, 10·6–26·6)	42/63 (66·7%, 54·3–77·1)	66/98 (67·3%, 57·5–75·8)	35/617 (5·7%, 4·1–7·8)	203/535 (37·9%, 33·9–42·1)	36/388 (9·3%, 6·8–12·6)	336/873 (38·5%, 35·3–41·8)
Characteristics							
Blood in stool*	76/84 (90·5%, 82·1–95·3)	39/58 (67·2%, 54·4–78·0)	20/78 (25·6%, 17·2–36·4)	484/595 (81·3%, 78·0–84·3)	248/493 (50·3%, 45·9–54·7)	191/346 (55·2%, 49·9–60·4)	641/832 (77·0%, 74·1–79·8)
Haemolytic uraemic syndrome	5/91 (5·5%, 2·1–12·5)	0/68 (0·0%, 0·0–6·4)	0/101 (0·0%, 0·0–4·4)	42/666 (6·3%, 4·7–8·4)	10/592 (1·7%, 0·88–3·1)	4/408 (1·0%, 0·29–2·6)	4/925 (0·4%, 0·13–1·1)
Shiga toxin 1a	1/91 (1·1%, 0–6·6)	26/68 (38·2%, 27·6–50·1)	0/101 (0% 0–4·4)	15/666 (2·3%, 1·3–3·7)	9/592 (1·5%, 0·76–2·9)	57/408 (14·0%, 10·9–17·7)	916/925 (99·0%, 98·1–99·5)
Shiga toxin 2c	47/91 (51·6%, 41·5–61·6)	35/68 (51·5%, 39·8–62·9)	101/101 (100%, 95·6–100)	372/666 (55·9%, 52·1–59·6)	579/592 (97·8%, 96·2–98·8)	115/408 (28·2%, 24·0–32·7)	878/925 (94·9%, 93·3–96·2)
Shiga toxin 2a	90/91 (98·9%, 93·4–100·0)	40/68 (58·8%, 47·0–69·8)	2/101 (2·0%, 0·11–7·4)	646/666 (97·0%, 95·4–98·1)	256/592 (43·2%, 39·3–47·3)	297/408 (72·8%, 68·3–76·9)	72/925 (7·8%, 6·2–9·7)

Data are n/N (%, 95% CI). *Individuals with missing data were excluded.

Table 1: **Main characteristics, virulence profiles, and risk factors of all STEC O157:H7 lineages**

subclades (1.3.2, 2.4.1, and 2.5.2) whereas the subsequent increase was congruent with the formation of more recently emerged subclades (ie, 2.3.3, 2.4.2, and 2.5.1).

Travel outside of the UK was infrequently reported in clades 2.3.3 (five [3·8%] of 131 patients) and 2.5.2 (13 [8·4%] of 154) compared with clades 1.3.2 (74 [46·8%] of 158), 2.4.1 (35 [81·4%] of 43), 2.4.2 (86 [59·3%] of 145), and 2.5.1 (100 [46·9%] of 213). The most common travel destinations were concentrated around the Mediterranean region, specifically Morocco (22 [29·7%] of 74 patients) for clade 1.3.2, Egypt (20 [57·1%] of 35) for clade 2.4.1, Turkey (62 [72·1%] of 86) for clade 2.4.2, and Spain (52 [52·0%] of 100) for clade 2.5.1 (appendix p 5).

Using travel outside of the UK status as a discrete state, ancestral reconstruction predicted the ancestral node of clades 2.3.3 and 2.5.2 to be domestic, whereas the other predominant clades (1.3.2, 2.4.1, 2.4.2, and 2.5.1) were predicted to have a non-domestic origin. Transition from non-domestic to domestic status was significantly more frequent than transition from domestic to non-domestic status (105 [12·9%] of 817 *vs* 35 [4·8%] of 723; p<0·0001; figure 1C). Of the 21 British cattle isolates included in the phylodynamic analyses, 19 (90·5%) of 21 clustered within domestic clades (16 in 2.5.2 and three in 2.3.3) and two clustered in non-domestic clades (one in 2.4.2 and one in 2.5.1). The cattle isolate in clade 2.5.1 shared a most recent common ancestor 8 years previously, in 2012 (95% HPDI 2011–2014), with a patient who reported no recent travel outside of the UK and was embedded in a clade strongly predicted to be of non-domestic origin. The cattle isolate in clade 2.4.2 shared a most recent common ancestor 14 years previously (95% HPDI 2005–2007) with a patient who reported recent travel to Turkey.

We defined all lineage IIc isolates as domestic or non-domestic on the basis of clade membership; there were 306 clinical isolates within domestic clades and 619 clinical isolates associated with non-domestic clades, with the proportions remaining reasonably stable during the study period (figure 2A). Infections from strains of a non-domestic origin declined in 2020, presumably in part owing to the restriction on international travel owing to the COVID-19 pandemic. The ratio of annual lineage IIc infections caused by non-domestic strains to domestic strains was estimated to be 2·02 (95% CI 1·55–2·70). This ratio was maintained when infections were adjusted for outbreaks (one representative per five single nucleotide polymorphism [SNP] single linkage cluster; figure 2B). The majority of domestic and non-domestic clade lineage IIc infections occurred in the summer months (July–September). The proportion of patients within clade 2.3.3 was reasonably high in 2018, whereas the proportion in clade 2.5.2 was high during 2020, largely attributed to outbreaks (appendix p 4, figure 2B).

Clustering isolates at a five-SNP level using the single linkage method has been used to capture the diversity commonly seen in outbreaks associated with a single source.[9] Within this study set, 18 five-SNP clusters had five or more patients and accounted for 183 (19·8%) of 925 of the total isolates, of which five (27·8%) clusters were of a domestic origin and 13 (72·2%) clusters were of a non-domestic origin. Of the non-domestic clusters, recent travel before symptom onset was reported by

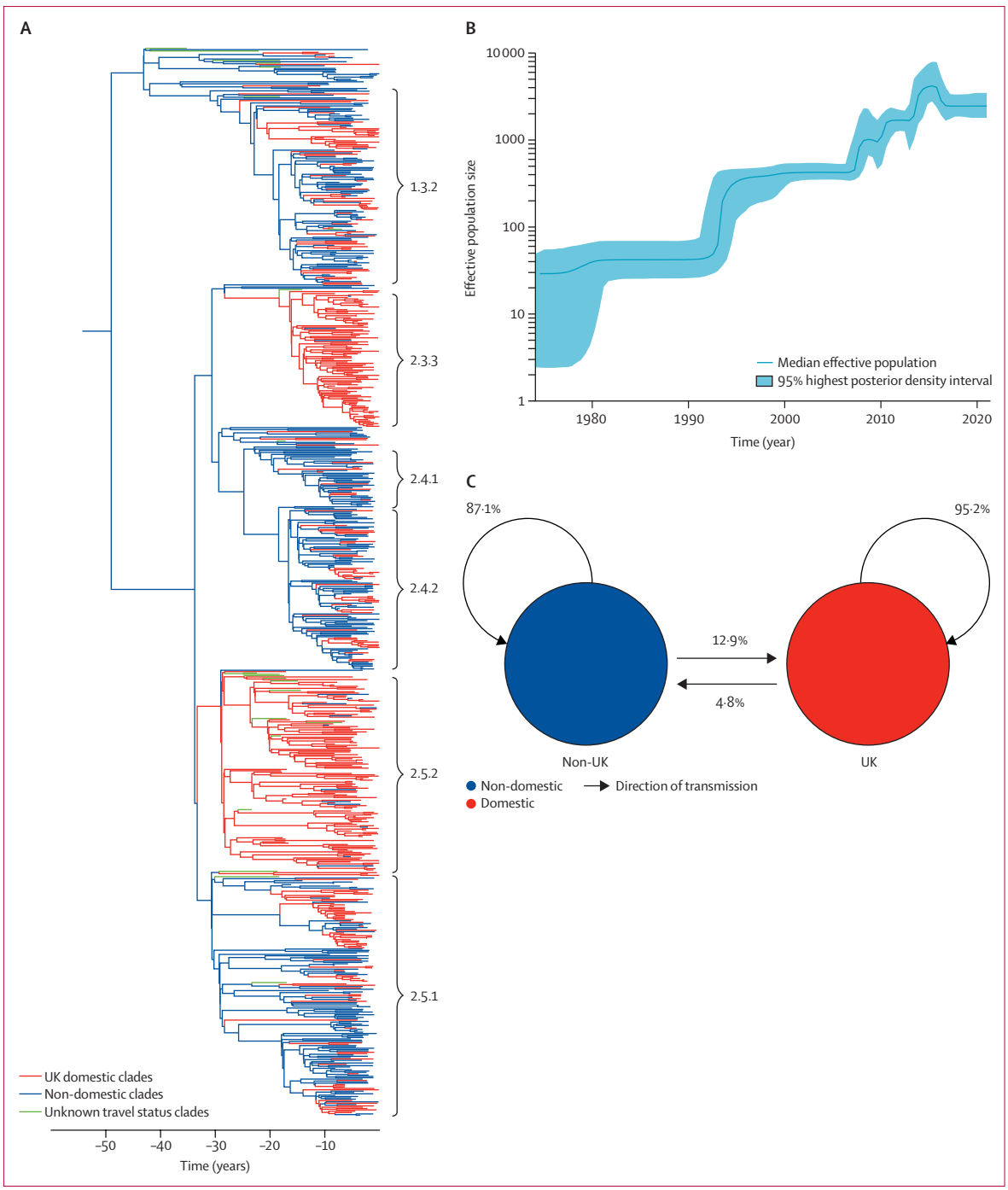

Figure 1: **Phylodynamic properties of STEC O157:H7 lineage IIc strains with respect to non-UK travel status**
(A) Maximum clade credibility tree of 797 clonal representatives taken from 925 isolates of lineage IIc from routine surveillance (June 1, 2015, to Dec 31, 2020), 300 historical clinical isolates from England, and 21 isolates from British cattle surveys, with branches coloured on the basis of predicted likely source of infection. (B) Bayesian skyline plot showed lineage IIc effective population size over time. (C) Transmission rate between domestic and non-domestic states as inferred from the ancestral state reconstruction.

most patients in only two (15·4%) of 13 clusters. As travel data were well documented in this dataset, it is plausible that the patients in the remaining clusters (11 [61·1%] of 18) acquired their infection from contaminated imported food. Isolates from domestic clades were more likely to be clustered with another case compared with isolates from non-domestic clades (54·2% *vs* 45·1%; p=0·0086).

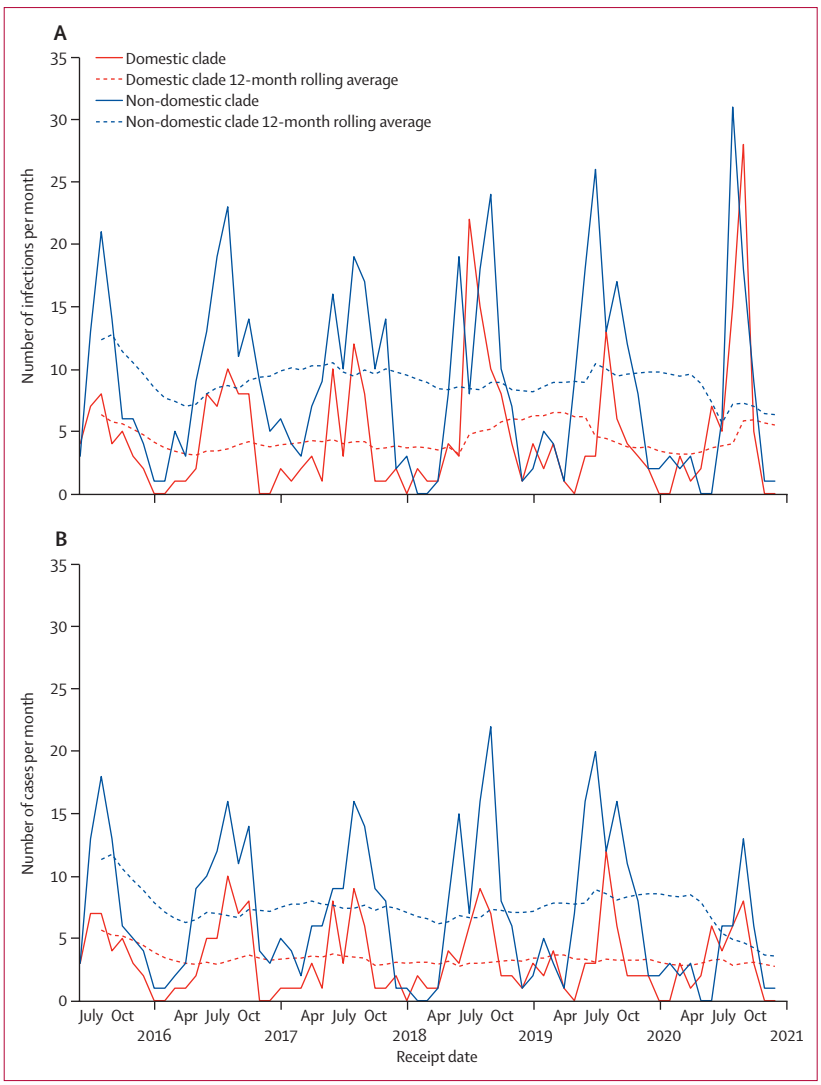

***Figure 2*: Patients per month with lineage IIc, delineated by UK and non-UK clades**
Time series showing the number of lineage IIc infections per month across the study period, delineated by domestic and non-domestic clades (A) and by domestic and non-domestic clades normalised to a single representative for each five single nucleotide polymorphism cluster (B).

The Shiga toxin profile *stx1a* or *stx2c* predominated among lineage IIc isolates (828 [89·5%] of 925). There were 72 isolates that had *stx2a*, in conjunction with *stx1a*, or *stx2c*, or both. These isolates were located across all the six major clades, of both domestic and non-domestic origin.

Among all lineage IIc isolates, 668 (72·2%) of 925 had no identifiable resistance genes and were predicted to be fully susceptible to the antimicrobial classes assayed. Altogether, 257 (27·8%) of 925 isolates had at least one genotypic antimicrobial resistance determinant, and 229 (24·8%) of 925 of isolates were multidrug resistant (ie, resistant to three or more antimicrobial classes).

The most common antibiotic resistance profile corresponded to predicted resistance against aminoglycosides, tetracycline, sulfonamides, and trimethoprim. In addition, 205 lineage IIc isolates encoded genes that were predicted to confer resistance to the β-lactams, with 191 (93·2%) of 205 harbouring bla_{TEM-1}, and nine (4·4%) of 205 harbouring $bla_{CTX-M-15}$. Resistance to chloramphenicol was predicted in 101 isolates, and 81 isolates were predicted to have reduced susceptibility or were resistant to fluroquinolones because of mutations in *gyrA* and *parC* or acquisition of plasmid-mediated quinolone resistance (*qnrS-1*). Finally, 79 isolates had predicted resistance to macrolides, most commonly with the acquisition of *mphB*.

Domestic clades had lower degrees of multidrug resistance (2 [1·2%] of 165 in clade 2.5.2 and 0 [0%] of 141 in clade 2.3.3) than some non-domestic clades (103 [61·3%] of 168 in clade 1.3.2 and 106 (47·3%) of 224 in clade 2.5.1). All of the isolates that displayed some degree of predicted resistance to fluroquinolones, and eight (88·9%) of nine isolates that were positive for $bla_{CTX-M-15}$, were restricted to clade 2.5.1, as were 64 (90·1%) of 71 of the isolates that had *mphB*. Genotypic antimicrobial profiles for all 925 isolates are provided in the appendix (pp 55–88).

For domestic clades 2.3.3 (141 patients) and 2.5.2 (165 patients; table 2, appendix pp 5–6), patients reported less travel outside of the UK (p<0·0001) and more bloody diarrhoea (p=0·050) than for non-domestic clades but there was no haemolytic uraemic syndrome within these two domestic clades (appendix pp 5–6). Clade 2.3.3 patients had more frequent exposure to fresh produce (raw vegetables, p=0·012 and prepackaged salad, p=0·0009), whereas clade 2.5.2 patients had more frequent exposure to animal contact (cattle p=0·021) or farm visits (p=0·0053), compared with other STEC O157:H7 patients.

Exposure to prepackaged salad, contact with cattle, and visiting a farm were included in a multivariable mixed-effects model as fixed variables (including all food or environmental variables with p<0·20 from the univariable analyses and adjustment variables). Cooked poultry was omitted because poultry are not a known animal reservoir for STEC O157:H7. All adjustment variables were significant and improved the fit of the model. Local authority district (n=317) was included in the model as a random variable. The final multivariable model revealed that the major risk factors were prepackaged salad (clade 2.3.3, relative risk ratio [RRR] 1·72, 95% CI 1·09–2·72; p=0·019) and visiting farms (clade 2.5.2, RRR 1·98, 1·12–3·52; p=0·020; table 3). We investigated interactions between: year and prepackaged salad; cattle contact and visiting a farm; travelling outside of the UK and prepackaged salad; and younger than 5 years and prepackaged salad. Of these interactions, for age younger than 5 years and prepackaged salad p=0·023 in the likelihood ratio test, but the numbers in each category for the clades were small and therefore omitted (table 3). The overall model did well at predicting the probability of infection by five equally spaced indices of multiple deprivation classes; for clade 2.3.3, the predictions of 8·00% (95% prediction interval 0·00%–19·67%) in

	Lineage IIc domestic clades		Comparison group*
	Proportion of people in clade 2.3.3 (n=141)	Proportion of people in clade 2.5.2 (n=165)	Proportion of people in other STEC O157:H7 (non-2.3.3 and non-2.5.2 lineage IIc and non-lineage IIc strains)
Demographic and clinical details			
Sex			
Female	92/165 (55·8%, 48·1–63·1)	85/141 (60·3%, 52·0–68·0)	1435/2545 (56·4%, 54·5–58·3)
Male	56/141 (39·7%, 32·0–48·0)	73/165 (44·2%, 36·9–51·9)	1110/2545 (43·6%, 41·7–45·5)
Younger than 5 years	19/141 (13·5%, 8·7–20·2)	26/165 (15·8%, 10·9–22·1)	472/2545 (18·5%, 17·1–20·1)
Travelled outside of the UK	5/131 (3·8%, 1·4–8·9)	13/154 (8·4%, 4·9–14·0)	715/2376 (30·1%, 28·3–32·0)
Blood in stool	100/124 (80·6%, 72·8–86·7)	111/144 (77·1%, 69·5–83·2)	1488/2218 (67·1%, 65·1–69·0)
Haemolytic uraemic syndrome	0/141 (0%, 0–3·2)	0/165 (0%, 0–2·7)	65/2545 (2·6%, 2·0–3·2)
Antibiotics	17/109 (15·6%, 9·9–23·7)	21/128 (16·4%, 10·9–23·9)	309/1956 (15·8%, 14·2–17·5)
Food handling or consumption			
Raw lamb (handling)	8/137 (5·8%, 2·8–11·3)	9/159 (5·7%, 2·9–10·5)	109/2486 (4·4%, 3·6–5·3)
Raw pork (handling)	14/137 (10·2%, 6·1–16·5)	21/159 (13·2%, 8·7–19·4)	218/2486 (8·8%, 7·7–9·9)
Cooked lamb	14/137 (10·2%, 6·1–16·5)	31/159 (19·5%, 14·0–26·4)	351/2486 (14·1%, 12·8–15·5)
Raw beef (handling)	20/137 (14·6%, 9·6–21·6)	31/159 (19·5%, 14·0–26·4)	348/2486 (14·0%, 12·7–15·4)
Raw poultry (handling)	32/137 (23·4%, 17·0–31·1)	45/159 (28·3%, 21·9–35·8)	485/2486 (19·5%, 18·0–21·1)
Cooked pork	36/137 (26·3%, 19·6–34·2)	43/159 (27·0%, 20·7–34·5)	594/2486 (23·9%, 22·3–25·6)
Prepackaged salad	45/137 (32·8%, 25·5–41·1)	33/159 (20·8%, 15·1–27·7)	549/2486 (22·1%, 20·5–23·8)
Fruit juices	52/137 (38·0%, 30·3–46·3)	63/159 (39·6%, 32·3–47·4)	839/2486 (33·7%, 31·9–35·6)
Soft fruits or berries	59/137 (43·1%, 35·1–51·4)	52/159 (32·7%, 25·9–40·3)	870/2486 (35·0%, 33·1–36·9)
Cooked beef	65/137 (47·4%, 39·3–55·8)	75/159 (47·2%, 39·6–54·9)	989/2486 (39·8%, 37·9–41·7)
Raw fruit	67/137 (48·9%, 40·7–57·2)	70/159 (44·0%, 36·5–51·8)	1056/2486 (42·5%, 40·5–44·4)
Raw vegetables (handling)	71/137 (51·8%, 43·5–60·0)	72/159 (45·3%, 37·7–53·0)	999/2486 (40·2%, 38·3–42·1)
Cooked poultry	92/137 (67·2%, 58·9–74·5)	99/159 (62·3%, 54·5–69·4)	1293/2486 (52·0%, 50·0–54·0)
Eating out	98/129 (76·0%, 67·9–82·6)	114/148 (77·0%, 69·6–83·1)	1537/2064 (74·5%, 72·5–76·3)
Animal contact			
Contact with cattle	11/136 (8·1%, 4·4–14·0)	13/153 (8·5%, 4·9–14·1)	91/2384 (3·8%, 3·1–4·7)
Visited a farm	20/114 (17·5%, 11·6–25·6)	30/136 (22·1%, 15·9–29·8)	262/2030 (12·9%, 11·5–14·4)
Contact with a cat	35/136 (25·7%, 19·1–33·7)	39/153 (25·5%, 19·2–33·0)	544/2390 (22·8%, 21·1–24·5)
Contact with a dog	56/136 (41·2%, 33·3–49·6)	69/154 (44·8%, 37·2–52·7)	879/2402 (36·6%, 34·7–38·5)
Spatiotemporal characteristics			
Year			
2015	11/141 (7·8%, 4·3–13·6)	22/165 (13·3%, 8·9–19·4)	311/2545 (12·2%, 11·0–13·6)
2016	15/141 (10·6%, 6·5–16·9)	30/165 (18·2%, 13·0–24·8)	594/2545 (23·3%, 21·7–25·0)
2017	23/141 (16·3%, 11·1–23·4)	23/165 (13·9%, 9·4–20·1)	464/2545 (18·2%, 16·8–19·8)
2018	53/141 (37·6%, 30·0–45·8)	18/165 (10·9%, 6·9–16·7)	472/2545 (18·5%, 17·1–20·1)
2019	21/141 (14·9%, 9·9–21·8)	24/165 (14·5%, 9·9–20·8)	435/2545 (17·1%, 15·7–18·6)
2020	18/141 (12·8%, 8·1–19·4)	48/165 (29·1%, 22·7–36·4)	269/2545 (10·6%, 9·4–11·8)
Region of England			
East Midlands	18/141 (12·8%, 8·1–19·4)	12/165 (7·3%, 4·1–12·4)	217/2545 (8·5%, 7·5–9·7)
East of England	13/141 (9·2%, 5·4–15·3)	14/165 (8·5%, 5·0–13·8)	214/2545 (8·4%, 7·4–9·6)
London	11/141 (7·8%, 4·3–13·6)	13/165 (7·9%, 4·6–13·1)	213/2545 (8·4%, 7·4–9·5)
North East	7/141 (5·0%, 2·2–10·1)	10/165 (6·1%, 3·2–10·9)	179/2545 (7·0%, 6·1–8·1)
North West	18/141 (12·8%, 8·1–19·4)	20/165 (12·1%, 7·9–18·1)	374/2545 (14·7%, 13·4–16·1)
South East	23/141 (16·3%, 11·1–23·4)	18/165 (10·9%, 6·9–16·7)	479/2545 (18·8%, 17·3–20·4)
South West	21/141 (14·9%, 9·9–21·8)	37/165 (22·4%, 16·7–29·4)	348/2545 (13·7%, 12·4–15·1)
West Midlands	19/141 (13·5%, 8·7–20·2)	17/165 (10·3%, 6·4–16·0)	216/2545 (8·5%, 7·5–9·6)
Yorkshire and the Humber	11/141 (7·8%, 4·3–13·6)	24/165 (14·5%, 9·9–20·8)	305/2545 (12·0%, 10·8–13·3)

(Table 2 continues on next page)

	Domestic clades		Non-domestic clades
	Proportion of people in clade 2.3.3 (n=141)	Proportion of people in clade 2.5.2 (n=165)	Proportion of people in other STEC O157:H7 (non-2.3.3 and non-2.5.2 lineage IIc and non-lineage IIc strains)
(Continued from previous page)			
Index of multiple deprivation, divided by 10 000			
Median	1·93 (1·21–2·59)	1·78 (1·10–2·42)	1·83 (0·99–2·57)
Range	0·13–3·28	<0·01–3·27	<0·01–3·28

Data are n/N (%, 95% CI), median (IQR), or range unless otherwise stated. Individuals with missing data were excluded. STEC=Shiga toxin-producing *Escherichia coli*. *All other domestic and non-domestic STEC O157:H7 except for domestic lineage IIc clades.

***Table 2*: Proportion of patients with clade 2.3.3, clade 2.5.2, and other clades, by exposure variable**

category 2 (low–mid socioeconomic status area), whereas the observed cases were 29 (6·2%) of 467; table 3). No significant multicollinearity was observed (all variance inflation factor values <4).

Discussion

Understanding reservoirs and transmission routes is imperative to inform targeted public health interventions during outbreak investigations of foodborne pathogens. Within England, STEC O157:H7 outbreaks have been caused by a range of transmission modes, originating both from a domestic reservoir[15] or via imported products that were contaminated by a reservoir outside of England.[21] In England, all people with confirmed STEC O157:H7 are asked to complete an Enhanced Surveillance Questionnaire capturing data on clinical presentation, food and environmental exposure, and recent travel and, since 2014, this questionnaire has been supplemented with routine whole-genome sequencing of all confirmed STEC O157:H7 human isolates. These initiatives provided a unique opportunity to synergise rich microbiological and epidemiological longitudinal surveillance data to explore the link between phylogenetically related isolates and likely exposures. Until now, epidemiological exposure data has mostly been used to determine common sources and transmission routes during outbreak investigations. Here, we used a fit-for-purpose statistical model with epidemiological data to explore common exposures within deeper phylogenetic relationships, within clades belonging to the same recently emerged lineage.

There is increased awareness globally that outbreaks of STEC O157:H7 associated with fresh produce represent a substantial public health risk. Salads and raw vegetables are classified as ready-to-eat foods because they are often consumed without a heat kill-step. Outbreaks associated with fresh produce present substantial challenges, because they are often caused by a single, transient contamination event and the produce often has a short shelf life, so the contaminated batch is no longer available for microbiological testing, which thereby confounds trace-back investigations. Furthermore, supply chains of fresh produce are often dynamic, with the sources of products changing from domestic farms to producers who operate overseas, and from one country to another, owing to financial margins and climate effects. The microbiological testing of complex food matrices for STEC is also challenging, as the infectious dose is low, and the causative agent is often present in low numbers and therefore can be difficult to detect among other microbiological flora.

Whole-genome sequencing provides an opportunity to delineate strain populations on the basis of phylogenetics, and to explore the hypothesis that STEC O157:H7 populations in cattle and other animal reservoirs in different countries evolve with little intercountry transmission.[9] Furthermore, structured populations allow us to test the hypothesis that different strain backgrounds might have evolved different propensities to contaminate different vehicles of transmission.

In this study, we analysed 5 years of surveillance data of STEC O157:H7, with a focus on lineage IIc, a strain previously implicated in outbreaks associated with fresh produce of both domestic and non-domestic origin. Patients infected with lineage IIc predominantly presented with severe clinical outcomes, including bloody diarrhoea in about 80% of patients, although progression to haemolytic uraemic syndrome was rare (<1%). This clinical presentation is explained by the almost ubiquitous (>95%) presence of *stx1a* and the rare occurrence of *stx2a* in the population (<8%).[8,9] Currently, although, the acquisition of *stx2a* by lineage IIc strains is a rare event, the integration of Stx2a-encoding phage has occurred at various locations across the phylogeny, highlighting the potential for this lineage to increase in pathogenicity. Evidence suggests that, following acquisition of Stx2a-encoding phage, it tends to be maintained in the population and might be associated with higher excretion concentrations in cattle.[28] As such, the emergence of STEC O157:H7 lineage IIc harbouring *stx2a* is a public health concern.

Phylodynamic analyses revealed that the lineage IIc population has undergone successive clonal expansions since emerging in the 1990s, resolving into six major extant clades. To delineate the population into potential domestic and non-domestic reservoirs, ancestral state reconstruction using a patient's recent travel history was used. This method predicted two of six clades (2.3.3 and 2.5.2) to be of a domestic origin seeded from a

non-domestic ancestor. These domestic clades were confirmed with the integration of British cattle isolates data, with the majority clustering within these clades. Two cattle isolates clustered in non-domestic clades, which suggested a reasonably recent colonisation of the domestic cattle population by imported strains. Individuals within England can be exposed to non-domestic strains, both while travelling abroad and through the consumption of contaminated food imported from a non-domestic source. This exposure is reflected in the phylodynamic modelling, in which the transient movement from a non-domestic to a domestic state occurs nearly three times as frequently as the inverse (ie, movement from a domestic to non-domestic state). Travel destinations were mostly in the Mediterranean region, and it is plausible that some of the domestically established strains originated from these areas through travellers, movement of animals, or migration of transiently colonised wildlife, such as birds.

Antimicrobial susceptibility varied across clades within lineage IIc, with domestic clades largely populated with pansusceptible isolates. By contrast, several non-domestic clades contained a reasonably high proportion of multidrug resistance isolates, which might reflect differences in agricultural antimicrobial use. Although antimicrobial therapy is contraindicated for the treatment of STEC O157:H7, the threat of antimicrobial resistant determinants entering the food chain or becoming endemic in native cattle stocks is of concern.[26]

During the study period, non-domestic clades of STEC O157:H7 lineage IIc contributed approximately twice as many infections annually than infections attributed to a domestic origin. Outbreaks of genetically monomorphic isolates were identified in both domestic and non-domestic clades, with most outbreaks associated with non-domestic clades having an absence of patient travel, suggesting an imported mode of transmission, including a recent outbreak caused by STEC O157:H7 lineage IIc in which imported fresh produce was the implicated mode.[21] Although the amounts of sporadic infection were high (>40%) for both domestic and non-domestic clades, domestic infections were significantly more likely to be clustered with another patient, presumably because of sampling from a reservoir of restricted diversity.

Plausible risk factors for domestic clades included significantly greater exposure to fresh produce, animal contact, or farm visits than for individuals with other STEC O157:H7. Prepackaged salad (clade 2.3.3), contact to cattle (2.5.2) and visiting farms (2.5.2) were all confirmed to be of importance as fixed variables, whereas the local authority district was a significant random variable, indicating the importance of geographical area to likely exposure. We found the mixed-effects multinomial model to be useful in modelling this pathogen with many potential routes of transmission. The proximity of cattle farms to irrigation water and fields in which produce is grown are known risk factors for STEC contamination,[20,21] and direct contact with cattle and farms are well established sources of STEC O157:H7 infections.[16]

	Clade 2.3.3	Clade 2.5.2
Fixed variables		
Prepackaged salad	1·72 (1·09–2·72); p=0·019	1·06 (0·65–1·72); p=0·82
Contact with cattle	1·41 (0·61–3·26); p=0·42	1·50 (0·68–3·32); p=0·31
Visited a farm	1·44 (0·76–2·74); p=0·26	1·98 (1·12–3·52); p=0·020
Adjustment variables		
Travelled outside of the UK	0·078 (0·03–0·22); p<0·0001	0·26 (0·14–0·51); p<0·0001
Younger than 5 years	0·77 (0·43–1·38); p=0·38	0·56 (0·32–0·99); p=0·046
Year		
2016	Ref	Ref
2017	3·02 (1·42–6·39); p=0·0039	1·08 (0·58–2·02); p=0·80
2018	6·71 (3·41–13·24); p<0·0001	0·80 (0·42–1·55); p=0·52
2019	3·23 (1·50–6·98); p=0·0028	1·19 (0·64–2·20); p=0·58
2020	3·03 (1·36–6·73); p=0·0065	3·65 (2·07–6·44); p<0·0001
Index of multiple deprivation		
1	Ref	Ref
2	4·47 (1·84–10·87); p=0·0010	1·10 (0·54–2·23); p=0·80
3	2·99 (1·21–7·39); p=0·018	1·32 (0·68–2·56); p=0·41
4	3·23 (1·32–7·92); p=0·010	0·92 (0·46–1·83); p=0·81
5	2·70 (1·07–6·84); p=0·036	0·83 (0·40–1·70); p=0·60
Random variable		
Local authority district	5·30 (1·03)	5·17 (1·02)
Predictions by five categories of index of multiple deprivation variable		
1	1·99% (0·00–15·81); 2·13% (8/375)	4·61% (0·00–18·58); 5·07% (19/375)
2	8·00% (0·00–19·67); 6·21% (29/467)	6·05% (0·00–18·49); 5·14% (24/467)
3	5·46% (0·00–16·67); 5·66% (30/530)	6·64% (0·00–18·30); 7·74% (41/530)
4	5·75% (0·00–16·54); 6·12% (35/572)	5·05% (0·00–16·33); 5·94% (34/572)
5	4·88% (0·00–15·83); 4·98% (28/562)	4·63% (0·00–16·02); 4·45% (25/562)

Data are relative risk ratio (95% CI), Wald p value; variance estimate (SE); or % predicted (95% CI); % observed (n/N).

Table 3: **Multivariable model**

Further studies are needed to better understand the association of domestic lineage IIc clades with fresh produce. It is plausible that such an association could be explained by an increased survival of lineage IIc strains in the environment or more efficient colonisation of plant material compared with other STEC O157:H7 lineages. Unravelling this association could provide opportunities for novel interventions that target either specific strain backgrounds or genetic elements.

In conclusion, we have identified and quantified the role of a domestic reservoir in infections of STEC O157:H7 lineage IIc in England. Exploiting patient exposure data that were collected during routine enhanced surveillance, and using a fit-for-purpose statistical model, we identified prepackaged salad and direct contact with cattle as significant exposure risks, causing predominantly domestic human infections in relation to other STEC O157:H7 lineages. This evidence base will greatly enhance the public health response to outbreak investigations, and will direct initial hypotheses generation, targeting both

likely transmission modes and reservoirs. Collecting detailed exposure data on sporadic infections and outbreak patients, and marrying that to high resolution microbiological whole-genome sequencing typing provides powerful analyses for reframing our understanding of foodborne disease risk within a One Health context. This approach represents a blueprint for the 21st century public health response in the genomic and big data era.

Contributors

TJD conceptualised the study, wrote the initial draft of the manuscript, and did the genomic analyses. GG and SG contributed to the conceptualisation of the study. KJ designed and did the statistical and epidemiological analyses and contributed to the initial draft of the manuscript. NQV contributed to the statistical analyses. DG and CJ contributed isolates for analyses. All authors edited and reviewed the final manuscript. TJD and KJ accessed and verified the raw data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

All genomes are shared via the US National Center for Biotechnology Information Sequence Read Archive under the BioProject accession number PRJNA315192.

For the **Sequence Read Archive** see https://www.ncbi.nlm.nih.gov/sra

Acknowledgments

We gratefully acknowledge the *E coli* O157 Super-shedding in Cattle and Mitigation of Human Risk (FS101055) project team for the provision of the cattle isolates and the staff of the UK Gastrointestinal Bacteria Reference Unit for the processing of isolates as part of routine surveillance and the UK Gastrointestinal Infections and Food Safety (One Health) Division for the epidemiological and statistical surveillance and for doing enhanced questionnaires. This research was partly funded by the UK National Institute for Health and Care Research (NIHR) Health Protection Research Unit in Gastrointestinal Infections at the University of Liverpool, and the NIHR Health Protection Research Unit in Genomics and Enabling Data at the University of Warwick, in partnership with UK Health Security Agency (UKHSA). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, the UK Department of Health, or UKHSA.

References

1 Mughini-Gras L, van Pelt W, van der Voort M, Heck M, Friesema I, Franz E. Attribution of human infections with Shiga toxin-producing *Escherichia coli* (STEC) to livestock sources and identification of source-specific risk factors, the Netherlands (2010–2014). *Zoonoses Public Health* 2018; **65:** e8–22.

2 Strachan NJC, Rotariu O, Lopes B, et al. Whole genome sequencing demonstrates that geographic variation of *Escherichia coli* O157 genotypes dominates host association. *Sci Rep* 2015; **5:** 14145.

3 Karmali MA. Emerging public health challenges of shiga toxin-producing *Escherichia coli* related to changes in the pathogen, the population, and the environment. *Clin Infect Dis* 2017; **64:** 371–76.

4 Riley LW, Remis RS, Helgerson SD, et al. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med* 1983; **308:** 681–85.

5 Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 2005; **365:** 1073–86.

6 Gill A, Dussault F, McMahon T, et al. Characterisation of atypical Shiga toxin gene sequences and description of Stx2j, a new subtype. *J Clin Microbiol* 2022; **60:** e0222921.

7 Koutsoumanis K, Allende A, Alvarez-Ordóñez A, et al. Pathogenicity assessment of Shiga toxin-producing *Escherichia coli* (STEC) and the public health risk posed by contamination of food with STEC. *EFSA J* 2020; **18:** e05967.

8 Dallman TJ, Ashton PM, Byrne L, et al. Applying phylogenomics to understand the emergence of Shiga-toxin-producing *Escherichia coli* O157:H7 strains causing severe human disease in the UK. *Microb Genom* 2015; **1:** e000029.

9 Dallman TJ, Greig DR, Gharbia SE, Jenkins C. Phylogenetic structure of Shiga toxin-producing *Escherichia coli* O157:H7 from sub-lineage to SNPs. *Microb Genomics* 2021; **7:** mgen000544.

10 Franz E, Rotariu O, Lopes BS, et al. Phylogeographic analysis reveals multiple international transmission events have driven the global emergence of *Escherichia coli* O157:H7. *Clin Infect Dis* 2019; **69:** 428–37.

11 Ingle DJ, Gonçalves da Silva A, Valcanis M, et al. Emergence and divergence of major lineages of Shiga-toxin-producing *Escherichia coli* in Australia. *Microb Genomics* 2019; **5:** e000268.

12 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:** 3475–87.

13 Le Saux N, Spika JS, Friesen B, et al. Ground beef consumption in noncommercial settings is a risk factor for sporadic *Escherichia coli* O157:H7 infection in Canada. *J Infect Dis* 1993; **167:** 500–02.

14 Dorn CR, Angrick EJ. Serotype O157:H7 *Escherichia coli* from bovine and meat sources. *J Clin Microbiol* 1991; **29:** 1225–31.

15 Jenkins C, Dallman TJ, Grant KA. Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing *Escherichia coli* serogroup O157:H7, England, 2013 to 2017. *Euro Surveill* 2019; **24**.

16 Kintz E, Byrne L, Jenkins C, McCarthy N, Vivancos R, Hunter P. Outbreaks of Shiga toxin-producing *Escherichia coli* linked to sprouted seeds, salad, and leafy greens: a systematic review. *J Food Prot* 2019; **82:** 1950–58.

17 Marshall KE, Hexemer A, Seelman SL, et al. Lessons learned from a decade of investigations of Shiga toxin-producing *Escherichia coli* outbreaks linked to leafy greens, United States and Canada. *Emerg Infect Dis* 2020; **26:** 2319–28.

18 Mikhail AFW, Jenkins C, Dallman TJ, et al. An outbreak of Shiga toxin-producing *Escherichia coli* O157:H7 associated with contaminated salad leaves: epidemiological, genomic and food trace back investigations. *Epidemiol Infect* 2018; **146:** 187–96.

19 Sinclair C, Jenkins C, Warburton F, Adak GK, Harris JP. Investigation of a national outbreak of STEC *Escherichia coli* O157 using online consumer panel control methods: Great Britain, October 2014. *Epidemiol Infect* 2017; **145:** 864–71.

20 Cowley LA, Dallman TJ, Fitzgerald S, et al. Short-term evolution of Shiga toxin-producing *Escherichia coli* O157:H7 between two food-borne outbreaks. *Microb Genom* 2016; **2:** e000084.

21 Mulchandani R, Brehmer C, Butt S, et al. Outbreak of Shiga toxin-producing *Escherichia coli* O157 linked with consumption of a fast-food product containing imported cucumbers, United Kingdom, August 2020. *Int J Infect Dis* 2021; **110** (suppl)**:** S62–68.

22 Pearce MC, Chase-Topping ME, McKendrick IJ, et al. Temporal and spatial patterns of bovine *Escherichia coli* O157 prevalence and comparison of temporal changes in the patterns of phage types associated with bovine shedding and human *E. coli* O157 cases in Scotland between 1998–2000 and 2002–2004. *BMC Microbiol* 2009; **9:** 276.

23 Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020; **37:** 1530–34.

24 Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018; **4:** vey016.

25 Ashton PM, Perry N, Ellis R, et al. Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. *PeerJ* 2015; **3:** e739.

26 Day M, Doumith M, Jenkins C, et al. Antimicrobial resistance in Shiga toxin-producing *Escherichia coli* serogroups O157 and O26 isolated from human cases of diarrhoeal disease in England, 2015. *J Antimicrob Chemother* 2017; **72:** 145–52.

27 Akinwande MO, Dikko HG, Samson A. Variance inflation factor: as a condition for the inclusion of suppressor variable(s) in regression analysis. *Open J Stat* 2015; **5:** 754–67.

28 Matthews L, Reeve R, Gally DL, et al. Predicting the public health benefit of vaccinating cattle against *Escherichia coli* O157. *Proc Natl Acad Sci USA* 2013; **110:** 16265–70.