Food Standards Scotland

For safe food and healthy eating

WHOLE GENOME SEQUENCE TYPING AND ANALYSIS OF NON-O157 STEC

January 2020

foodstandards.gov.scot

Scottish *E. coli* O157/STEC Reference Laboratory
Department of Laboratory Medicine
Royal Infirmary of Edinburgh
Little France
Edinburgh EH16 4SA

Tel: +44 (0)131 2426013

Main Authors:
Lesley Allison, Principal Scientist, Scottish *E. coli* O157/STEC Reference Laboratory
Anne Holmes, Clinical Scientist, Scottish *E. coli* O157/STEC Reference Laboratory
Mary Hanson, Director, Scottish *E. coli* O157/STEC Reference Laboratory

Contributing Author:
Melissa Ward, Sir Henry Wellcome Postdoctoral Research Fellowship (WT103953MA), University of Edinburgh (*E. coli* O26:H11 phylogenetic analysis)

CONTENTS

FIGURES

TABLES

1 Scientific and lay summaries

1.1 Scientific summary

Escherichia coli O157 is an established enteric pathogen with a relatively high incidence in Scotland and a propensity to cause outbreaks of infection. However, the number of Shiga toxin-producing *E. coli* (STEC) other than serogroup O157 (non-O157 STEC), detected in Scotland has increased since 2000. This is likely due to increased ascertainment, and non-O157 STEC now accounts for approximately 30% of all STEC isolated in Scotland. Although no outbreaks of non-O157 STEC with greater than five cases have been identified in Scotland, large outbreaks have been reported elsewhere, often associated with the consumption of contaminated food. In Scotland, the roll-out of PCR testing of foodstuffs at the Scottish Public Analyst laboratories since 2014 demonstrated that non-O157 STEC are present in food and water, but we still have a scant understanding of the different strains circulating in Scotland and the sources of clinical infection.

The Scottish *E. coli* O157/STEC Reference Laboratory (SERL) is responsible for the identification and typing of *E. coli* O157 and other STEC. The majority of *E. coli* O157 isolates are received from diagnostic laboratories across Scotland but the SERL also identify and type *E. coli* O157 isolated from animal, food or environmental samples thought to be associated with human infection (for example those submitted by Public Analyst laboratories). The SERL also performs an important diagnostic role in the detection and isolation of non-O157 STEC from faeces submitted, in line with current Scottish guidance. In 2017, the SERL introduced Whole Genome Sequencing (WGS) as the main typing method for all *E. coli* O157 and non-O157 STEC, and this provides unprecedented resolution to distinguish degrees of relatedness among bacterial isolates, and has proven a powerful tool for epidemiological investigations. Furthermore, the sequence data generated can be used to identify genetic markers associated with pathogenesis and antimicrobial resistance (AMR), and help understand the genomic evolution of STEC.

This research project was funded by Food Standards Scotland (FSS) to carry out WGS on the Scottish National Culture Collection of non-O157 STEC, held at the SERL, to facilitate the following:

- To provide a highly detailed description of the non-O157 STEC causing clinical infection in Scotland over a 16 year period.
- To provide a phylogenomic analysis of the strains and, where possible, to put these in an international context, with particular reference to predominant, emerging and virulent strains.

- To assess the potential of each strain to cause clinical disease based on the presence of virulence genes and with reference to current research on what might constitute a “pathogenic” STEC.
- To provide a unique and comprehensive genomic database of Scottish non-O157 STEC for future interrogation, which will be of particular value during future outbreaks of non-O157 STEC, in Scotland and elsewhere.

A total of 525 strains were analysed by WGS during this study (517 non-O157 STEC, 5 *E. coli* strains subsequently discovered to not carry *stx* genes and 3 strains subsequently identified as *Escherichia albertii*). A total of 88 different serotypes were identified, and virulence content, antimicrobial resistance profiles and phylogentic relationships elucidated. For the purpose of this report, we have focussed particularly on the predominant, emerging and hybrid strains that were discussed in detail at the 10th International Symposium on Shiga Toxin (verotoxin) Producing *Escherichia coli* Infections (VTEC2018) meeting held in May 2018[1].

This comprehensive collection of non-O157 STEC sequences from a single country is a valuable and unique resource and will be further enhanced by linking the genomic data to the clinical presentation and epidemiological information for each patient, which will now form the second phase of the study. Health Protection Scotland (HPS) undertakes enhanced surveillance of all cases of STEC infection and collates information about potential exposures, whether each case is part of an outbreak or sporadic, whether it is a primary or secondary case, or associated with foreign travel. Information is also collected on clinical presentation (e.g. bloody diarrhoea, HUS, hospitalisation), co-infections and underlying gut pathology.

Although it’s difficult to put this diversity into an international context, as most other countries have yet to undertake such a comprehensive analysis of their non-O157 STEC collections, two of the top Scottish serotypes (*E. coli* O26:H11 and *E. coli* O103:H2) are also two of the most common serotypes circulating worldwide. Due to the number of *E. coli* O26:H11 serotypes identified, we were able to carry out a full phylogenomic analysis to study their genetic relatedness and molecular characteristics, and compare with international O26:H11 strains. This group was genetically heterogeneous but, reassuringly, only two strains were identified as the new highly pathogenic “European clone”. We did not observe the newly identified French or US clones. The *E. coli* O55:H7 strains with a propensity to cause HUS, and observed annually in England, have not yet been detected in Scotland. *E. coli* O145:H28, which is rising to prominence in the Republic of Ireland, was the third most common serotype detected in Scotland and has been responsible for causing

[1] https://www.escmid.org/dates_events/calendar/calendar_event/cal/2018/05/06/event/tx_cal_phpicalendar/10th_vtec_international_symposium_on_shiga_toxinverocytotoxinproducing_escherichia_coli_infection/?no_cache=1&tx_cal_controller%5Blastview%5D=view-search_event%7Cpage_id-130%7Cquery-&cHash=386a21e3e17ab5f9c00d22db6a284ca4

small Scottish outbreaks. Its role in clinical infection will be determined during the next phase of the study.

Almost 8% of the Scottish non-O157 STEC collection carried the *stx*2f variant. This is a variant that is often overlooked when screening for STEC, as it specifically requires an additional primer set for PCR and, as far as we are aware, the majority of multiplex PCR platforms for gastrointestinal pathogens (such as EntericBio and BDMAX) do not include this as a target. Consequently, the true incidence of this in routine surveillance and research studies is likely to be underestimated or completely missed.

Internationally, there has been considerable uncertainty and a lack of consensus over what constitutes a pathogenic STEC: is it the serotype, is it the carriage of particular genes, is it based on whether an organism has previously shown to cause disease or a combination of variables, some perhaps still unidentified? However, a report was published during the course of this study which proposed a molecular risk assessment approach based on the prevalence of certain virulence genes and adherence factors and ranked risk potential for severe disease (JEMRA classifications) [1]. We were able to utilise this approach in the characterisation of Scottish non-O157 STEC in this study and compare this with Scottish *E. coli* O157:H7.

All Scottish *E. coli* O157:H7 strains (typed separately to this study) were assigned to JEMRA Levels 1-4 meaning all had the potential to cause at least bloody diarrhoea. The majority (59.5%) of Scottish non-O157 STEC strains were also assigned to Levels 1-4, with 40.5 % of non-O157 STEC strains assigned to Level 5 i.e. conferring the potential to cause diarrhoea, but not bloody diarrhoea. In order to assess how accurate a predictor for disease the JEMRA assignment might be, we must compare the predicted disease potential (the JEMRA level assigned, based on virulence gene combination) with each patient's actual symptoms. This assessment will be carried out in the second phase of this study to determine whether a molecular risk assessment approach, such as the one described in the JEMRA report, might be a viable way forward for inferring the degree of severity of illness resulting from non-O157 STEC infection.

1.2 Lay summary

Escherichia coli (known as *E. coli*) O157 is a well-known foodborne pathogen with a relatively high incidence in Scotland and a propensity to cause outbreaks of infection. However, there are a group of organisms that are very similar to *E. coli* O157 (in that they carry Shiga toxin genes and can cause severe disease) termed non-O157 Shiga toxin-producing *E. coli* (non-O157 STEC), but much less is known about them as they are more difficult to detect and isolate. However, developments in molecular methodologies are permitting the detection and isolation of these organisms and a new method called whole genome sequencing (WGS) is enabling them to be characterised at a much more detailed level which is improving our understanding of the role these organisms may play in clinical disease. In Scotland, non-O157 STEC now account for 30% of all STEC detected.

This research project was funded to generate and analyse whole-genome sequencing (WGS) data on all Scottish non-O157 STEC in the Scottish Culture Collection to provide a better understanding of the strains causing clinical infection in the hope that this may enable the identification of virulence markers and the development of a molecular risk assessment approach to help define a 'pathogenic (disease causing) STEC'.

A total of 525 strains, isolated over a 16 year period were investigated and 522 non-O157 STEC identified. A diverse number of serotypes were identified with the two most common Scottish serotypes *E. coli* O26:H11 and *E. coli* O103:H2 also being the most common non-O157 STEC serotypes in Europe and the USA. *E. coli* O26 is also the most common non-O157 STEC serogroup detected in food and animal samples tested in Europe. However, a wide diversity of serotypes were observed, some on one occasion only, and we were able to place a number of these into an international context.

As a clearer picture emerges of the non-O157 STEC strains responsible for disease and their detailed genetic make-up, we are progressing our understanding of what might consititute a pathogenic STEC and moving away from the concept of this being serotype dependent. Data is now being collected and analysed that may facilitate a prospective assessment on the level of clinical disease that a particular strain might cause and this would be a significant advance.

The data produced in this study is a valuable and unique resource which will be further enhanced by linking it to the clinical and epidemiological information collected from each patient by Health Protection Scotland (HPS), which will form the second phase of this study.

1.3 Glossary & Abbreviations

API	Analytical Profile Index – this is a way of classifying bacteria based on different biochemical reactions
AMR	Antimicrobial Resistance
AST	Antibiotic Sensitivity Testing
BAPS	Bayesian Analysis of Population Structure
BD	Bloody Diarrhoea
CC	Clonal Complex - Sequence Types (ST) that are all thought to derive from the same founding genotype
D	Diarrhoea
DLV	Double locus variant
EFSA	European Food Safety Authority
ESBL	Extended-Spectrum β-Lactamase
FAO	Food and Agriculture Organization of the United Nations
FSS	Food Standards Scotland
HPS	Health Protection Scotland
HUS	Haemolytic Uraemic Syndrome
IMS	ImmunoMagnetic Separation
JEMRA	Joint FAO/WHO Expert Meetings on Microbiological Risk Assessment
MDR	MultiDrug Resistant – refers to resistance to 3 or more antibiotic classes
ML	Maximum Likelihood
MLVA	MultiLocus Variable number tandem repeat Analysis
MLST	MultiLocus Sequence Type - a typing approach for assessing the relatedness of strains that uses the sequences of internal fragments of seven house-keeping genes. For each house-keeping gene, different sequences are assigned as distinct alleles and, for each isolate, the alleles at each of the seven loci define the allelic profile or sequence type (ST). Core-genome (cg)-MLST is a high resolution version that uses a comprehensive set of loci (n=2513).
Non-O157 STEC	Shiga toxin-producing *E. coli* (STEC) serogroups other than serogroup O157, for example *E. coli* O26 and *E. coli* O103 would be non-O157 STEC
NSF	Non-Sorbitol Fermenter – an *E. coli* strain that does not ferment sorbitol. *E. coli* O157:H7, for example, are typically non-sorbitol fermenters.
Ou	O-unidentifiable
PCR	Polymerase Chain Reaction
PFGE	Pulsed-Field Gel Electrophoresis
PHE	Public Health England
pVF	Plasmid-encoded Virulence Factors
PulseNet	A national network of public health and food regulatory agency

	laboratories that perform standardised molecular subtyping ('fingerprinting') of foodborne disease-causing bacteria, including STEC
RoI	Republic of Ireland
SERL	Scottish *E. coli* O157/STEC Reference Laboratory
Serogroup	This is based on the presence of a specific “O” antigen in the *E. coli* strain under investigation. O157 is an example of a serogroup of *E. coli*; O26 is another serogroup.
Serotype	This is based on the combination of an “O” and “H” antigen. O157:H7 is an *E. coli* serotype; O103:H2 is another serotype
SHPN	Scottish Health Protection Network
SF	Sorbitol Fermenter – an *E. coli* strain that ferments sorbitol. Many *E. coli* strains, other than serotype *E. coli* O157:H7, ferment sorbitol. These include harmless *E. coli* and also some non-O157 STEC strains.This is what makes it so difficult to detect non-O157 *E. coli* strains by culture as they look the same as all other *E. coli* on a sorbitol culture plate.
SLV	Single Locus Variant
SPATE	Serine Protease Autotransporters of *Enterobacteriaceae*
ST	Sequence Type - is assigned to an organism following 7-gene MLST. No cgMLST ST is provided when using the cgMLST scheme in BioNumerics.
STEC	Shiga toxin-producing *Escherichia coli* (see footnote)
Stx	Shiga toxin
stx	Shiga toxin gene
UKAS	United Kingdom Accreditation Service
VTEC	Verotoxigenic *E. coli* (see footnote)
vtx	Verotoxin gene (see footnote)
WGS	Whole Genome Sequencing
WHO	World Health Organisation

Nomenclature

Previously, Shiga toxin-producing *E. coli* (STEC), were referred to as Verotoxigenic *Escherichia coli* (VTEC), however there has been a widespread move to standardise the nomenclature across Europe to STEC (the preferred term in Northern America). While the two terms are interchangeable, for the purposes of this report, the term STEC will be used. Where referring to historic documentation, the original term will be retained.

2 Introduction

2.1 Background

2.1.1 Shiga Toxin-producing *E. coli*

Escherichia coli (*E. coli*) O157, a Shiga toxin-producing *E. coli* (STEC), is an important foodborne pathogen, responsible for numerous outbreaks of gastrointestinal infection worldwide. It is now apparent, however, that *E. coli* isolates of serogroups other than O157 (termed non-O157 STEC) also make a significant contribution to human diarrhoeal disease and, in an increasing number of countries including the USA, Australia and many in Europe, non-O157 STEC predominate [2,3]. Although improvements in controlling hazards associated with O157 over the past two decades are likely to have contributed to a decrease in incidence of *E. coli* O157:H7, for example, in the meat industry [4], the increase in number of non-O157 STEC serotypes has been most likely due to changes in screening protocols resulting in increased detection of these strains. This is clearly demonstrated in the Republic of Ireland (RoI), where there was a shift in dominant serogroup, from O157 to O26, which first increased above O157 levels in 2013. This coincided with the introduction of PCR detection methods in many of their regional hospitals (Anne Carroll, Public Health Laboratory, Cherry Orchard Hospital, Dublin, personal communication). An increase in numbers of non-O157 STEC detected following an alteration in screening practice has been reported in many countries worldwide [5,6], however *E. coli* O157:H7 remains the dominant serotype in Scotland [7] and England & Wales [8].

2.1.2 Assessment of pathogenicity

Pathogenicity of *E. coli* relies on a complex interaction between pathogen and host, involves a number of external factors and is not yet fully elucidated. The course of gastrointestinal infection caused by STEC involves ingestion of the bacteria followed by attachment to the epithelial lining of the intestine. There is then a requirement for the bacteria to produce toxins which can move across the intestine into the underlying tissues and bloodstream.

The most common mechanism of attachment in *E. coli* O157:H7 (and some non-O157 STEC) is the formation of attaching and effacing (A/E) lesions. The genes necessary for this histopathological alteration of the intestine are located in the Locus of Enterocyte Effacement (LEE) pathogenicity island [9]. A number of genes are associated with the LEE (for example, *eae*, *espA*, *espB and tir*), but it is the *eae* (intimin) gene which is often used as a proxy for the LEE and is carried by *E. coli* O157:H7 strains and some non-O157 STEC. A number of different subtypes of the *eae* gene have been identified [10] with some variants (beta *eae*-β, gamma *eae*-γ, theta *eae*-θ and epsilon *eae*-ε) thought to be more associated with clinical infection

than others. However, LEE-negative STEC can also cause severe disease including HUS, in humans where the A/E function normally provided by the LEE is likely to have been replaced by other adhesion proteins, such as *aggR*. This was evidenced most recently in the large European *E. coli* O104:H4 outbreak [11]. Shiga toxins are one of the key virulence factors which cause diarrhoea and haemorrhagic colitis and the life threatening complication of Haemolytic Uraemic Syndrome (HUS). There are two types of toxins involved in infection, Stx1 and Stx2, and these are further categorised into several subtypes coded for by the following genes: *stx*1a, *stx*1c, *stx*1d, *stx*2a, *stx*2b, *stx*2c, *stx*2d, *stx*2e, *stx*2f, *stx*2g [12]. Certain subtypes have been shown to be associated with more severe infection.

However, the genetic basis underlying the spectrum of clinical disease caused by non-O157 STEC is still being elucidated and is likely to involve many additional known and as yet unknown virulence factors [13].

2.1.3 Assigning pathogenic potential

We know that STEC strains differ in pathogenic potential, but how pathogenic potential is determined has been the subject of much discussion.

E. coli are identified by two surface antigens: the somatic “O” antigen and the flagellar “H” antigen. As strains can carry different combinations of O and H antigens, the number of potential O:H combinations (termed the serotype) is quite large. In 2005, in excess of 400 different STEC serotypes were reported worldwide with >100 implicated in human disease [14]. In a recent literature review, 262 different serotypes were associated with clinical cases of infection worldwide [15]. In 2003, Karmali *et al*. [16] proposed the STEC seropathotype (SPT) classification stating that STEC could be divided into 5 different seropathotype groups based on the frequency of certain serotypes in human disease, their known association with outbreaks and the severity of the outcome [16]. However, strains with the same serotype do not necessarily carry the same virulence genes so, although serotype is still is useful for epidemiological surveillance, this approach has been shown to be unreliable for predicting the pathogenic potential of a strain to cause severe disease [17].

In 2007, Scheutz *et al* [18] proposed an alternative model of classification based on virulence rather than serogroup/serotype, and a more recent EFSA Scientific Opinion [17], proposed a definition based on the presence/absence of certain molecular markers, however this approach has generally evolved in response to an incident, is reactive, and is amended/updated when a new strain emerges that doesn’t fit the existing algorithm. This was clearly demonstrated when the *E. coli* O104:H4 isolate causing the large outbreak in continental Europe did NOT possess the *eae* gene but did possess the *agg*R gene. This demonstrated that adhesion by other mechanisms is possible and highlights the evolving nature of *E. coli*. To date, there has been no agreed single combination of markers that define a pathogenic STEC [17] and

consequently, in many countries, decisions about the pathogenic potential of a particular strain have been made on a case-by-case basis, following discussion with expert international colleagues.

In October 2017, the European Commission together with Member States' experts of the European Commission Working Group asked EFSA to provide an update of the 2007 and 2013 Scientific Opinions [17,19], with regard to the following:

- Review the new body of knowledge available for pathogenicity assessment of STEC, and refine, if needed, the molecular approach for the categorisation of STEC strains proposed in the EFSA 2013 Opinion.
- Review the microbiological methods for the detection and characterisation of human pathogenic STEC in animals and food.
- Analyse available data on human foodborne STEC cases in the EU and rank different food commodities based on the public health risk.
- Provide recommendations to fill in the gaps identified above.

The report will shortly be published (F. Scheutz, WHO Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella*, Statens Serum Institute, Denmark, personal communication).

A Joint Food and Agriculture Organization of the United Nations (FAO) and the World Health Organisation (WHO) Expert Meetings on Microbiological Risk Assessment (JEMRA) report “Shiga toxin-producing *Escherichia coli* (STEC) and food: attribution, characterisation and monitoring” was published in 2018 and proposed a set of criteria for categorising the potential risk of severity of illness associated with the presence of an STEC in food [1]. It suggested a molecular risk assessment approach based on the presence of certain virulence genes and adherence factors and ranked risk potential for severe disease as “highest”, “high”, “low” and “none” where “none” represents the absence of *stx* genes (Figure 1).

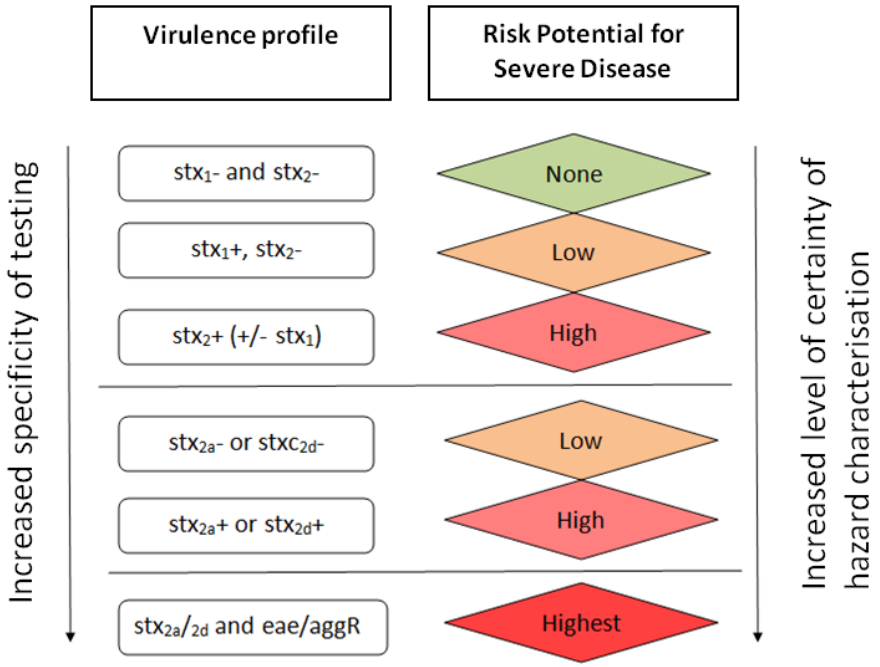

Figure 1: Testing strategy for STEC to assess health risk based on virulence genes. Adapted from JEMRA report, 2018.

Importantly, the report also stated that it is not prudent to regard any STEC strain as being non-pathogenic or not posing a health risk, as all STEC strains are likely to have the potential to cause diarrhoea and be of risk, especially to susceptible individuals.

The report acknowledges that the risk of developing HUS is complex and likely related to a number of factors including the susceptibility of the human host, virulence gene combinations, gene expression and exposure dose. For example, a child can have an STEC infection and develop HUS but a sibling can be infected with exactly the same strain but be asymptomatic. It is feasible that the child with HUS has a poor immune status, is somehow more susceptible or has ingested a larger infectious dose.

The development of new molecular technologies is revolutionising laboratory diagnosis of STEC, giving us a more comprehensive picture of the non-O157 STEC strains causing clinical disease. Coupled with the introduction of WGS technology, we can now assess the presence of thousands of different genes, plasmids carrying virulence factors and pathogenicity islands at a fine level not previously possible, improving our understanding of the role these organisms might play in clinical disease.

2.1.4 Sources of infection

An important characteristic of STEC is the ability to resist acidic pH and, together with their thermotolerant capabilities, they can survive in a variety of potentially

adverse habitats. STEC are zoonotic, with transmission occurring by direct contact with animals, their products or their environment, direct contact with an infected individual or consumption of contaminated food or water [20,21]. Although ruminants are thought to be the main reservoir of *E. coli* O157, it has also been isolated from other mammals including pigs, camels, rabbits, horses, dogs, cats, zoo animals and also birds and flies. The reservoir hosts of non-O157 STEC are not as well understood but do include ruminants. The infectious dose of *E. coli* O157 is known to be low [22] but non-O157 STEC have also been shown to have a low infectious dose [23,24] and this can be critical in the transmission of STEC and their propensity to cause outbreaks.

The WHO-Foodborne Disease Burden Epidemiology Reference Group has estimated that half of the worldwide STEC disease burden is foodborne [25]. Foodborne outbreaks caused by *E. coli* O157, are often associated with undercooked or unpasteurised products, but outbreaks are increasingly being linked to contaminated produce, including sprouted seeds, lettuce and spinach, where it may be internalised in the tissues, and therefore less susceptible to removal by washing. Indeed, the most frequently attributed sources of foodborne STEC cases, globally, were produce (13%), beef (11%) and dairy (7%), but in excess of 50% of all cases could not be attributed to any source [1].

Worringly, outbreaks have also been associated with pickled vegetables and unpasteurised cider (with a pH of 3.7 to 3.9), demonstrating an increase in acid tolerance and dispelling the belief that acidic food with a pH<4.6 is always safe. More unexpected food sources include nuts (walnuts and hazelnuts), crab meat and raw cookie dough where the organism has likely been introduced via the raw product itself (such as in the flour), or into the product during food processing [26].

A recent literature review of non-O157 STEC [15] identified a total of 674 outbreaks worldwide caused by non-O157 STEC, occurring between 1995 and 2017 (excluded the large *E. coli* O104:H4 outbreak of 2011). Sources of infection in these outbreaks included the consumption of meat products, unpasteurised drinks, unidentified foods consumed at a restaurant, and exposure to contaminated water or farm animals.

A study describing the genome sequences of >300 STEC isolated from foods in the USA has just been published [27], which will serve as a valuable dataset that can be used to compare food and clinical strains occurring and emerging worldwide.

A more detailed description of foodstuffs that have been implicated in outbreaks of non-O157 infection is presented in Table 1. This is not a comprehensive list but illustrates the variety of serogroups/serotypes involved in a range of different food products associated with outbreaks.

Table 1: Examples of non-O157 outbreaks associated with consumption of contaminated food

Year	Serogroup/ Serotype	Implicated Food/ Establishment	Country	Reference
2018	O26:H11	Raw cheese	France	[28]
2017	O103:H2	Raw cow's milk	Austria	[29]
2016	O121 & O26	Flour	USA	[30]
2015	O26	Mexican restaurant	USA	[31]
2014	O121	Raw sprouts	USA	[32]
2013	O121	Frozen food	USA	[33]
2013	O26	Dairy products	Italy/Romania	[34]
2011	O111:H8	Yukhoe (raw beef dish)	Japan	[35]
2011	O104:H4	Sprouted seeds	Germany	[36]
2010	O26	Ground beef	USA	[37]
2010	O145	Smoked game meat	USA	[38]
2010	O145	Romaine lettuce	USA	[39]
2010	O103 & O145* (**stx* negative)	Venison	USA	[40]
2008	O111	Buffet style restaurant	USA	[41]
2007	O111	Ground beef	USA	[38]
2007	O121/O26/O84	Pasteurised cheese, margarine	USA	[38]
2007	O145 & O26	Ice cream	Belgium	[42]
2006	O121	Lettuce	USA	[38]
2006	O26	Strawberries/blueberries	USA	[38]
2006	O103:H25	Fermented lamb sausage	Norway	[43]

2.1.5 The Scottish *E. coli* O157/STEC Reference Laboratory

The Scottish *E. coli* O157/STEC Reference Laboratory (SERL) is responsible for the identification and typing of *E. coli* O157 and other STEC in Scotland. The majority of *E. coli* O157 isolates are received from diagnostic laboratories across Scotland but the SERL also identify and type *E. coli* O157 isolated from animal, food or environmental samples thought to be associated with human infection (for example those submitted by Public Analyst laboratories or veterinary laboratories). The SERL also performs an important diagnostic role in the detection and isolation of non-O157 STEC from faeces submitted to SERL, in line with current Scottish guidance [44].

2.1.6 Guidance for screening for STEC in Scotland

In 1995, the UK Government's Advisory Committee on the Microbiological Safety of Food (ACMSF) recommended that all diarrhoeal stools be examined for verocytotoxin-producing *Escherichia coli* (VTEC) O157 [45], and this was echoed in the subsequent Pennington Report (1997) written in response to the Central Scotland *E. coli* O157 butcher's outbreak of 1996/7 [46]. A report by the Scottish Task Force on *E. coli* O157 (2001) [47], formed a practical action plan to improve the protection of the Scottish public from infection by *E. coli* O157 and made a number of recommendations, with some directly related to laboratory diagnosis of infection.

One recommendation proposed the referral from frontline diagnostic laboratories to the SERL of diagnostic faecal samples from all patients in specified high risk groups for more sensitive screening for both *E. coli* O157 and non-O157 STEC. There have been a number of revisions of this guidance (Table 2) and consequently this has had a significant impact on the workload and typing regimen of the SERL over the past 18 years, most significantly resulting in an increased number of faeces submitted to the SERL for more sensitive testing. In 2018, the SERL received 7,660 faeces for testing, compared to 2,714 in 2010 and 118 in 2001.

Table 2: Scottish Health Protection Network (SHPN) Guidance for Faecal Referral to the SERL

Task Force Criteria, 2001	SHPN STEC Guidance 2008 & 2013	SHPN STEC Guidance 2018
• Any case of HUS or TTP in which conventional culture has failed to yield a pathogen; • Any case of a bloody diarrhoea stool in which conventional culture has failed to yield a pathogen and more than 4 days has elapsed between the onset of diarrhoea and obtaining a stool sample ; • Any patient of less than 10 years of age or over 60 years of age with diarrhoea from whom conventional culture has failed to yield a pathogen and more than 4 days has elapsed between the onset of diarrhoea and obtaining a stool sample ; • Any contacts of a case or outbreak-associated case from whom conventional culture has failed to yield a pathogen and who are in a high risk group, i.e. under 10 years, over 60 years.	• Cases of suspected HUS/TMA, or bloody diarrhoea in whom conventional laboratory testing has failed to yield a pathogen; • All symptomatic contacts of cases of VTEC infection or any VTEC-outbreak associated case in whom conventional laboratory testing has failed to yield a pathogen; • All symptomatic contacts of cases of infection with SF VTEC O157 or non-O157 VTEC; • All asymptomatic contacts of cases of infection with SF VTEC O157.	• Cases of suspected HUS or cases of bloody diarrhoea in whom conventional laboratory testing has failed to yield a pathogen; • All symptomatic contacts of non-sorbitol-fermenting *E. coli* O157, sorbitol-fermenting *E. coli* O157 and non- O157 STEC in whom conventional laboratory testing has failed to yield a pathogen; • Any outbreak- associated case in whom conventional laboratory testing has failed to identify a pathogen.

2.1.7 Laboratory diagnosis

Laboratory screening for *E. coli* O157 in Scottish diagnostic laboratories has barely changed for the past 20 years. Faeces are plated onto Sorbitol MacConkey (SMAC) plates and/or SMAC plates supplemented with cefixime and tellurite (CTSMAC) followed by O157 latex agglutination of non-sorbitol-fermenting colonies.

Laboratories then confirm identity (usually biochemically using API or VITEK microbial identification system). Locally confirmed isolates are then forwarded to the SERL for confirmation and typing.

Non-O157 STEC are difficult to detect as they lack a unique biochemical identifier that will visually distinguish them from a background flora of commensal *E. coli* on standard isolation media. Historically, Scottish diagnostic laboratories screened for some *E. coli* serogroups (O26, O55, O111, O119, O126, O86, O114, O125, O127, O128, O44, O112, O124, O14 and O18c) in certain patient groups (normally children <2 years) by carrying out slide agglutination with polyvalent (usually polyvalents 2, 3, & 4) then monovalent antisera but, following PCR testing at the SERL, the majority of these strains did not possess *stx* genes and were therefore not STEC. All Scottish diagnostic laboratories had discontinued using this test by 2006. Despite the development and availability of commercially available PCR detection platforms for gastrointestinal pathogens, such as BD MAX (BD) and EntericBio (Serosep) [48], Scottish Diagnostic laboratories have yet to implement PCR screening of faeces.

There is currently no “gold standard” for STEC detection and isolation from clinical samples, and certain countries just screen for the top 5 or 6 serotypes. Since 2000, the SERL have been screening submitted faeces for a number of genes using PCR and in 2012, introduced real-time PCR. The current protocol involves enriching faeces overnight, a DNA extraction process and identifying the presence of *stx*1, *stx*2 (including all variants) and *rfb*O157 genes by real-time PCR in order to detect sorbitol fermenting (SF) and non-sorbitol fermenting (NSF) *E. coli* O157 and non-O157 STEC.

2.1.8 Isolation of STEC

Between April 2000 and December 2017, SERL received 52,378 faecal samples for screening, with 987 testing positive for one or both *stx* genes (in the absence of the *rfb*O157 gene) indicating the possible presence of a non-O157 STEC. Isolation of *E. coli* O157 from faeces is relatively straightforward and is achieved either by direct plating or using Immunomagnetic Separation (at SERL) but the isolation of non-O157 is much more complex. This is currently achieved using a dilution and spread-plate method followed by individual or pooled colony PCR until an individual isolate is identified with a *stx* profile matching that detected in the enrichment broth.

2.1.9 Strain typing

Historically, SERL sent all non-O157 STEC isolates to Public Health England (PHE) for O:H serotyping and comparison with non-O157 STEC isolated in England & Wales. SERL also performed PFGE on non-O157 STEC. Between October 2015 and July 2017, isolates of Scottish non-O157 STEC were sent to PHE for whole genome sequencing (WGS).

WGS offers unprecedented resolution to distinguish degrees of relatedness among bacterial isolates and this has proven a powerful tool for epidemiological investigations [49–52]. A number of national public health bodies are now using WGS for real-time surveillance of enteric bacterial pathogens and have replaced or are in the process of replacing existing typing methods such as MLVA, PFGE and phage typing with this technology. Serotypes (O:H) can be determined more accurately as WGS resolves issues with phenotypically untypable strains which would have been previous reported as O-unidentifiable (where there is an unknown O type) or O-rough (the result is masked due to auto-agglutination or hyper-mucoid). In addition to improving outbreak surveillance, sequence data can be used to identify genetic markers associated with pathogenesis and antimicrobial resistance, infer geographic origin of strains and help understand the genomic evolution of STEC.

Edinburgh Reference Laboratories have an Illumina MiSeq sequencer on site and the expertise to permit in-house WGS. In 2017, the SERL, working closely with PHE, completed a validation on the standardisation of WGS to enable the exchange and comparison of sequencing data [53]. Since August 2017, the SERL routinely performs WGS on all Scottish *E. coli* O157 and non-O157 STEC and data is exchanged with PHE for outbreak detection. The WGS and data analysis procedure used at the SERL is UKAS accredited (ISO 15189; https://www.ukas.com/wp-content/uploads/schedule_uploads/00007/9546%20Medical%20Multiple.pdf)

Once detected and isolated, the relatedness of the strains is determined. It is crucial to establish whether the strain is related to any other recently identified strains in Scotland and elsewhere. At the SERL**,** *E. coli* strains undergo a hierarchical typing process, including PCR typing, phage typing (O157 only) and WGS.

2.1.10 Prevalence of STEC in Scotland

Although the human prevalence of *E. coli* O157 in Scotland has remained consistently high compared to some countries in the UK and Europe, the Scottish figures remain relatively stable, around a baseline of approximately 220 cases of *E. coli* O157 per year. The number of non-O157 STEC detected in Scotland has increased since 2000, likely due to increased ascertainment, as more laboratories have adopted Scottish guidance and sent faeces to the SERL for more sensitive screening (see Figure 2). Since July 2014, **all** Scottish clinical laboratories routinely submit high risk faeces (with bloody diarrhoea or HUS) to the SERL for screening, if no pathogen is detected locally.

In 2017, the rate of STEC infection in Scotland was 4.1 (per 100,000) (Alison Smith-Palmer, Health Protection Scotland, personal communication) compared with a UK rate of 1.5 and an EU rate of 1.7. The highest rates in the EU in 2017 were observed in Ireland (16.6), Sweden (5.0) and Denmark (4.6) [54]. This compares with incidence rates in the US (1.7), New Zealand (11.9) and Japan (3.0).

In 2017/18, 34% (n=86) of all Scottish STEC cases were infected with a non-O157 STEC. However, this is likely to be an underestimate as the SERL only receives faeces from severe clinical cases so it is likely that Scotland has many undiagnosed and therefore unreported cases of non-O157 STEC infection. The USA's Center for Disease Control (CDC) estimate that for every non-O157 STEC illness diagnosed, 106.8 illnesses may remain undiagnosed [55].

Although our understanding of *E. coli* O157 infection has greatly improved over recent years, we still have a relatively scant understanding of non-O157 STEC and the sources of clinical infection.

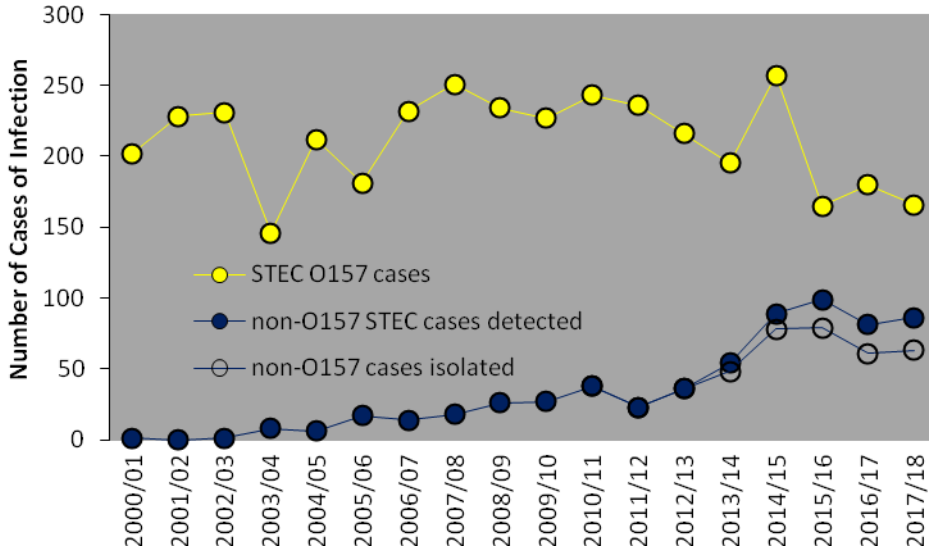

Figure 2: Scottish Cases of Infection April 2000 to March 2018

"Non-O157 STEC cases detected" - the number of cases whose faeces tested positive by PCR
"Non-O157 STEC cases isolated" - the number of cases whose faeces tested positive by PCR AND an STEC was isolated.

Although no outbreaks of non-O157 STEC with more than 6 confirmed cases have been identified in Scotland, a number of small clusters (fewer than 5 confirmed cases) have been detected, although the sources of infection have not been traced. In Scotland, the adoption of PCR testing of foodstuffs at certain Public Analyst laboratories in 2014 has demonstrated that non-O157 STECs are present in food and water. Since screening for non-O157 STEC has been implemented at these laboratories, non-O157 STEC has been isolated from private water, venison and raw cheese, and detection/isolation of these organisms from non-clinical sources is likely to continue.

Ruminants, principally cattle, are the major reservoir for human infection and contact with cattle and the environment is an important risk factor for STEC infection. Although previous studies have demonstrated that Scottish cattle shed certain non-O157 STEC serogroups [56,57] there is a lack of baseline data on the current

prevalence of non-O157 STEC and serotypes present in Scottish cattle and other ruminants.

2.1.11 The Scottish Culture Collection

The Scottish STEC culture collection comprises all *E. coli* O157 and non-O157 STEC isolates sent to the SERL for further testing or isolated at SERL from submitted faeces since 01/04/2000. Since April 2000, the SERL has received and processed in excess of 60,000 samples. The culture collection comprises >6,800 isolates, including a number of duplicate samples from individual patients and some isolates, not confirmed as *E. coli* O157 or STEC. Additionally, the collection contains archived Scottish *E. coli* O157 strains collected between 1994 and 2000.

2.2 Aims of the Study

This study was designed to generate and analyse whole genome sequencing (WGS) data on all Scottish non-O157 STEC in the Scottish Culture Collection to provide an in-depth understanding of the strain characteristics, and molecular epidemiology of Scottish non-O157 STEC related to human clinical infection. A subsequent study will be conducted following this analysis to examine the clinical and epidemiological information for each case to determine how the genetic profile of the strains relates to clinical presentation and exposure factors.

3 Methods

3.1 The study group

A total of 525 strains from clinical cases of infection, collected over a 16 year period from 26/02/02 to 05/02/18, were analysed in this study. Of these, 118 strains had already been sent to PHE for WGS as they were received between September 2015 and August 2017. We requested the sequence data files (FASTQ) from PHE so we could analyse these as part of this study. All isolates received at SERL, or isolated at SERL from submitted faeces, are stored at -80^{o}C and form part of the National STEC Culture Collection. A total of 407 isolates were recovered from storage by plating onto blood agar plates. These isolates were received from all diagnostic laboratories in the 14 different Health Board Areas (this reduced to 13 Board Areas during the course of the study due to reorganisation of Health Board boundaries). Of these strains, 317 were sequenced at the SERL and 90 were sent to PHE for sequencing (in order to complete the sequencing work within the timeframe of this project).

3.2 DNA extraction, library preparation and sequencing

Isolates were grown overnight in Difco^{TM} nutrient broth (E&O Labs., Bonnybridge, UK). Genomic DNA was extracted manually using the DNeasy Blood and Tissue Kit (Qiagen, Crawley, UK) or with the automated QIAsymphony platform using the QIA DSP DNA Mini Kit (Qiagen), both following a pre-lysis step as recommended by the manufacturer. The quantity of the DNA was measured using the Qubit Fluorimeter 3.0 (ThermoFisher Scientific) with the dsDNA Assay HS Kit. DNA libraries were prepared using the Nextera XT DNA Preparation Kit following the manufacturer's instructions (Illumina, Cambridge, UK). Paired-end sequencing was performed on the Illumina MiSeq (Illumina, CA, USA) using 500 cycle v2 reagent kits to produce 2 x 250bp reads.

3.3 Data analysis

Whole genome sequencing data was analysed using two different pipelines, the PHE bioinformatics protocol and BioNumerics v7.6 (Applied Maths) using the wgMLST and the *E. coli* genotyping plug-in tools. In combination, these methods provide a robust and complementary approach for the characterisation and typing of STEC, to identify putatively linked cases within Scotland, and to enable the rapid identification of cross-border matches. In addition, the sequences of strains with novel 7-gene MLST alleles/ST and/or unidentifiable O/H types were uploaded to Enterobase (http://enterobase.warwick.ac.uk/species/index/ecoli) for further analysis. EnteroBase is a curated database that provides new designations for novel MLST alleles and STs. It does not provide new O/H designations, however it identifies a small number of additional O groups (i.e. OX-groups, which are unique O groups

informally used by some laboratories) not present in the PHE or BioNumerics reference database.

This data analysis approach is being used routinely at the SERL for the identification, characterisation and subtyping of STEC and achieved UKAS accreditation in 2017. The data analysis protocols have been validated by our laboratory and elsewhere [8,58–61]. Furthermore, PulseNet International reported that a gene-by-gene approach (cgMLST, genes present in nearly all strains of the same species or wgMLST, all core genes plus accessory genes present in any strain used to create the allele database) was the method of choice for standardising subtyping worldwide for the real-time surveillance of foodborne pathogens [62]. Therefore our strain data is available for comparison with other laboratories in Europe and beyond without further processing.

The PHE bioinformatics protocol was developed (at PHE) using free, open source software (https://github.com/phe-bioinformatics/snapperdb) and this has been installed and validated at the SERL [53]. The protocol consists of various software to process, quality control and analyse the raw sequence data. Sequencing reads were quality trimmed with bases with a Phred score below 30 removed from the trailing edge using Trimmomatic [63]. KmerID was used to identify bacterial species and identify sample mix-ups [60]. Using the GeneFinder tool, FASTQ reads were mapped to a panel of serotype and virulence genes using Bowtie 2 [64] and the best match to each target was reported with metrics including coverage, depth, mixture and nucleotide similarity in XML format for quality assessment. Only *in silico* predictions of serotype and virulence that match a gene determinant at >80% nucleotide identity and over >80% target gene length were accepted. MLST alleles of seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA*) were determined using software called Metric-Oriented Sequence Typer (MOST [65]). Shiga toxin gene subtyping was performed using a combined mapping and BLAST approach as previously described [66].

BioNumerics v7.6 is a commercial software package that has been used in the SERL for many years to analyse microbial typing data. The software has recently been updated to accommodate WGS analysis (wgSNP and wgMLST) and is highly automated. The BioNumerics external Calculation Engine (Amazon Cloud) and the wgMLST client plug-in were used to process sequence reads (FASTQ files generated on the MiSeq). Assembly-free and assembly-based allele detection was performed to produce allelic profiles for each isolate and was stored as character data in the BioNumerics database. The assembly was performed using SPAdes integrated into the wgMLST plug-in and basic assembly metrics were calculated for quality assessment. The average read coverage was 106 ±45. Each genome assembly had an average total length of 5,334,589 ±224,816 bp, and an average

N50 (>1,000bp) of 119,208 ±53,033 bp. The average number of contigs (>1,000bp) within the assemblies was 227 ±96.

Assembled genomes were analysed using the *E. coli* genotyping plug-in, which contains reference databases for serotype (SerotypeFinder), virulence (VirulenceFinder) and antibiotic resistance (ResFinder) prediction obtained from the Center for Genomic Epidemiology (DTU, Lyngby, Denmark) https://cge.cbs.dtu.dk/services/. An overview of the genes found in the ResFinder database can be found at https://cge.cbs.dtu.dk/services/ResFinder/database.php and include those associated with resistance to aminoglycosides, sulphonamides, β-lactams, colistin, tetracyclines, macrolides, trimethoprim, phenicols, quinolones, and lincosamides. The detection parameters for gene detection were set to 90% sequence identity and 60% sequence coverage. The genotyping plug-in also has an *in silico* PCR tool for the detection of virulence genes and Shiga toxin gene subtypes using previously published primers. The *in silico* PCR settings were set to allow for 1 mismatch in the primer sequence binding sites (total length of primers between 21 - 30 base pairs). A summary of the outputs from the two bioinformatics pipelines are shown in Table 3.

Table 3. Outputs from the bioinformatics pipelines

	PHE Pipeline	**BioNumerics**
Species ID	Yes	Yes
Serotype	Yes	Yes
7-gene MLST	Yes	Yes
Sequence Type (ST)[a]	Yes	No
eae	Yes	Yes
stx subtype	Yes	Yes
Additional virulence genes	Yes[b]	Yes (Virulence Finder Database)
AMR	No	Yes (ResFinder Database)

[a] ST is based on the 7-gene MLST Achtman scheme (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA)*
[b] *bfpA, aggR, ipaH_type, aaiC, ltcA, sta1, stb*

To compare the presence of virulence genes in *eae*-positive and *eae*-negative STEC isolates, the 'N-1' chi-square test was used (https://www.medcalc.org/calc/comparison_of_proportions.php).

3.4 Phylogenomic analysis

Phylogenomic analyses were performed to study the genetic relatedness and molecular characteristics of the isolates. Genetic features superimposed on the phylogenies enabled the identification and characterisation of sublineages within different serotypes.

Core genome (cg)-MLST (see 3.4.1) analyses were used to produce phylogenies, except for the most common serotype *E. coli* O26:H11, where a more detailed

analysis using reference based assembly was used to provide a high resolution, well supported tree for evolutionary analysis (see 3.4.2). For this, 23 *E. coli* O26:H11 reference strains representative of the previously reported global O26 lineages ST29C1, ST29C2, ST21C1 and ST21C2 [67] were used (Appendix 1).

Several published genomes of other serotypes [*E. coli* O103:H2 (n=1), O104:H4 (n=1) and O145:H28 (n=2)] associated with food outbreaks in other countries were also used (Appendix 1). The raw sequence data (FASTQs) were downloaded from NCBI and analysed alongside the Scottish strains to compare their genetic relatedness and molecular characteristics.

3.4.1 Core genome (cg) MLST

cgMLST (2,513 core loci synchronised with Enterobase schema) dendograms were produced in BioNumerics v7.6 using categorical differences and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) cluster analysis. The software allows a maximum number of 200 allele differences, so as to observe the genetic distance between strains that differed more than this, a scaling factor was used. Scaling factors are noted in the figure legends. Meta-data was plotted on the dendrograms and included selected virulence genes, serotype, sequence type (ST), *stx* subtype and antibiogram length (number of antibiotic classes to which a resistant phenotype was predicted). In addition, to investigate the relationship between the serotype and sequence type of the isolates, a minimum spanning tree was created in BioNumerics v7.6.

3.4.2 Reference-based assembly of O26:H11

Further work was performed on data from the *E. coli* O26:H11 strains as this was the most common serotype detected and there were sufficient strains to support a more detailed analysis.

Consensus calling. SnapperDB was used to generate reference based consensus sequences, through the use of Picard tools (http://broadinstitute.github.io/picard) for reference preparation, BWA [68] for read mapping to the reference (Genbank Accession no. NC_013361.1), SAMtools for the formatting of SAM and BAM files [69], GATK for variant calling using UnifiedGenotyper [70], and SnapperDB [71] and PHEnix [72] for filtering of SNPs and creation of the consensus sequences. Core genome positions were defined as sites for which a base was called for each isolate.

Recombination detection. The alignment of *E. coli* O26:H11 core genome sequences was screened for recombination using the BratNextGen software [73], with 30 iterations, 100 permutations and a significance threshold of 0.05. Regions identified as recombinant were removed from the alignment prior to the phylogenetic analysis described below.

Phylogenetic analysis. Maximum-likelihood core genome phylogenies were constructed with the RaxML software (Linux version 8.2.11) [74]. For the RAxML phylogeny shown in Figure 10, the general time-reversible model of nucleotide substitution was used, with gamma-distributed rate heterogeneity across sites and 1,000 bootstrap replicates. Neighbor-joining phylogenies were also produced with MEGA version 6.06 [75]. We confirmed that, in terms of splitting into major lineages, the tree structure was robust to the use of different software frameworks and choices of evolutionary model. The tree in Figure 10 was rooted on the two O26:H11 strains we included that were identified as ST29C1 by Ogura *et al*. [67]; the suitability of these strains as an outgroup could be identified from the phylogenetic relationships inferred by Ogura *et al*. [67] and rooting on these strains produced an overall topology consistent with when a midpoint rooting was used.

Metadata were then plotted against the phylogeny. The antibiogram length was calculated for each isolate and plotted on the phylogeny. Antibiotic classes, and genes observed within the O26:H11 strains whose presence was predicted to confer resistance were β-lactams (*blaTEM-1A*, *blaTEM-1B*, *blaTEM-1C*, *blaTEM-30*), aminoglycosides (*aadA1*, *aadA12*, *aph(3')-Ia*, *strA*, *strB*), phenicols (*catA1*, *flor*), trimethoprim (*dfrA1*, *dfrA7*), sulphonamides (*sul1*, *sul2*), macrolides (*mph(B)*, and tetracyclines (*tet(A)*, *tet(B)*).

Results

3.1 Isolates confirmed as non-O157 *E. coli* by diagnostic laboratories

Between 2000 and 2006, some diagnostic laboratories used polyvalent antisera to detect certain serogroups of *E. coli* and submitted these isolates to the SERL for detection of *stx* genes. In total, 193 isolates were received over this time period but only 12 of these were confirmed as STEC.The remainder were either *stx* negative or not *E. coli*. WGS has revealed that only two (1.0%) of the strains had a serogroup that matched the original polyvalent antisera result suggesting the use of polyvalent serum was not specific for O antigen detection or highlighting issues with cross reacting antisera.

4.2. Molecular typing and characterisation of non-O157 STEC

4.2.1 Species identification and serotyping

A total of 522 (99.4%) clinical isolates were identified as *E. coli*, while three (0.6%) were *Escherichia albertii. In silico* serotyping identified 88 different serotypes (where both O and H group were identified) among the *E. coli* isolates (Table 4 and Appendix 2). There were only 15 (2.9%) isolates where the O group was not identified and these have been designated O unidentifiable (Ou). Serological O group typing was performed on 374 isolates by PHE prior to the introduction of WGS, and of these 299 (79.9%) were successfully typed, 65 (17.4%) were O unidentifiable, 9 (2.4%) were designated O rough (due to auto-agglutination or hyper-mucoid) and one (0.3%) was untyped (data not shown). A comparison with the genotypic data showed 13 of the 15 Ou by WGS were also O unidentifibable by serological typing suggesting these are novel types. In two cases (No. 483 & 484; Appendix 2) the H type was not identifed by WGS, however H antigens were not serologically typed so no comparison could be made. There were 10 discrepancies between the O group phenotypic and genotypic results (OuvO19a, OuvO8, O111vO123, O113vO8, O118vO182, O162vO174, O46vO171, O82vO174, O111vO15, O139vO87). The most likely reason for these discrepancies was the patient had a mixed infection and pure cultures were not stored so different individual organisms were typed, as these O groups are not known to cross react or have similar nucleotide sequences. DebRoy *et al*. [76] sequenced the O-antigen gene clusters of all immunologically designated O-groups (n=196; 185 O-serogroups, including the four groups that have been divided into subtypes: O18ab/ac, O28ab/ac, O112ab/ac and O125ab/ac, and the 11 OX groups) to show approximately 20 sets were highly similar (98–99.9%) in their nucleotide sequences. The phenotypic/genotypic discrepancies found in this study were not amongst those found to be highly similar.

As shown in Figure 3*, E. coli* O26:H11 was the most common serotype accounting for 27.0% (n=141/522) of all serotypes observed over the sixteen year study period. The next most common serotypes were O103:H2 (n=49/522; 9.4%) and O145:H28 (n=45/522; 8.6%). There were 47 serotypes that were observed on one occasion only (Table 4). Some isolates are named with multiple O groups (e.g. O153/O178:H7), because, as noted above, the sequences of the genes used for defining O group (e.g. *wzx,* O- antigen flippase; and *wzy,* O- antigen polymerase) are so similar that separation is not possible and many have been shown to cross-react serologically. As laboratories move away from routine serological typing there is a need for the nomenclature to be updated based on the sequence data. This would enable novel O groups [77] to be given designations and for highly similar O groups to be re-designated a single type (72).

It is important to note that the data includes all isolates, including those from cases that are epidemiologically linked.

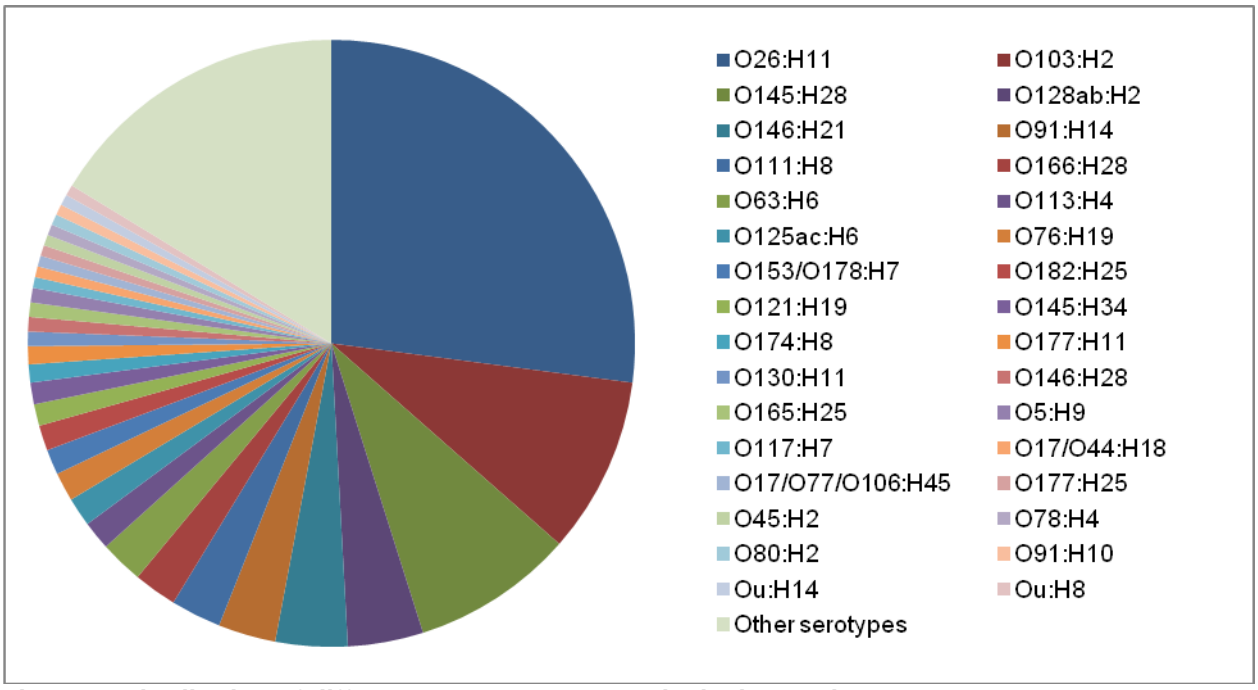

Figure 3. Distribution of different serotypes among the isolates. The top 32 serotypes are shown and the 'Other serotypes' are listed in Table 4.

Table 4: Characteristics of the non-O157 STEC serotypes

Serotype	Year(s) Isolated	No. isolates	ST(s)	*stx* subtype*	*eae* No (%) positive	AMR No (%) positive
O26:H11	2002-2017	141	21, 29, 574	*stx*1a (n=85), *stx*2a (n=9), *stx*2a *stx*1a (n=46), *stx*2c *stx*1a (n=1)	140 (99%)	33 (23%)
O103:H2	2003-2018	49	17, 20, 386, 8506	*stx*1a (n=47), *stx*2f (n=2)	48 (98%)	12 (24%)
O145:H28	2004-2018	45	32, 137, 6130, 8625	*stx*1a (n=1), *stx*2a (n=41), *stx*2d (n=3)	45 (100%)	8 (18%)
O128ab:H2	2006-2017	21	25, 811, 6265	*stx*1c (n=2), *stx*2b (n=5), *stx*2b *stx*1c (n=14)	0	3 (14%)
O146:H21	2007-2017	20	442, 8661	*stx*1c (n=6), *stx*2b (n=6), *stx*2b *stx*1a (n=1), *stx*2b *stx*1c (n=7)	0	2 (10%)
O91:H14	2003-2017	16	33	*stx*2b *stx*1a (13), *stx*1a (n=1), *stx*2b (n=2)	0	2 (13%)
O111:H8	2006-2017	14	16	*stx*1a (n=7), *stx*2a *stx*1a (n=7)	14 (100%)	9 (64%)
O166:H28	2008-2017	12	1819	*stx*2b *stx*1c (n=8), *stx*2b (n=3), *stx*2b *stx*1a *stx*1c (n=1)	0	0
O63:H6	2013-2017	12	583	*stx*2f (n=12)	12 (100%)	0
O113:H4	2006-2017	8	10	*stx*2b *stx*1c (n=7), *stx*2d (n=1)	0	0
O125ac:H6	2014-2016	8	583	*stx*2f (n=8)	8 (100%)	0
O76:H19	2007-2016	8	675	*stx*1c (n=7), *stx*2b *stx*1c (n=1)	0	0
O153/O178:H7	2009-2018	7	278, 8873	*stx*1c (n=4), *stx*2b *stx*1c (n=3)	0	0
O182:H25	2006-2018	7	300	*stx*1a (n=6), *stx*2a (n=1)	7 (100%)	0
O121:H19	2008-2014	6	655	*stx*1a (n=1), *stx*2a (n=5)	6 (100%)	0
O145:H34	2014-2017	6	722	*stx*2f (n=6)	6 (100%)	0
O174:H8	2002-2016	5	13, 8630	*stx*1a *stx*1c (n=1), *stx*2b *stx*1c (n=4)	0	0
O177:H11	2012-2017	5	29	*stx*1a (n=3), *stx*2a *stx*1a (n=2)	5 (100%)	0
O130:H11	2011-2015	4	297, 7931	*stx*2a (n=1), *stx*2a *stx*1a (n=1), *stx*2a *stx*2c *stx*1a (n=1), *stx*2a *stx*2c (n=1)	0	0
O146:H28	2016-2017	4	738, 6674	*stx*2b (n=4)	0	0
O165:H25	2014-2017	4	119	*stx*2a *stx*1a (n=3), *stx*2a *stx*2c *stx*1a (n=1)	4 (100%)	0
O5:H9	2009-2011	4	342	*stx*1a (n=4)	4 (100%)	4 (100%)
O117:H7	2011-2017	3	504, 5292	*stx*1a (n=3)	0	3 (100%)
O17/O44:H18	2012	3	69	*stx*2d (n=3)	0	0
O17/O77/O106:H45	2008-2014	3	662, 7083	*stx*1a (n=2), *stx*2d (n=1)	0	0
O177:H25	2003-2014	3	342, 659	*stx*2c (n=3)	3 (100%)	1 (33%)
O45:H2	2013-2017	3	20	*stx*2f (n=3)	3 (100%)	1 (33%)
O78:H4	2009-2017	3	3101	*stx*1c (n=3)	0	
O80:H2	2013-2014	3	301	*stx*2a (n=1), *stx*2d (n=2)	3 (100%)	3 (100%)

O91:H10	2006-2016	3	641, 8629	*stx*2a (n=1), *stx*2b *stx*2c (n=1), *stx*2d (n=1)	0	0
Ou:H14	2015-2017	3	1249	*stx*1c (n=3)	0	0
Ou:H8	2015-2016	3	26	*stx*2b (n=3)	0	0
O112ab:H2	2006-2014	2	388, 6260	*stx*1c (n=2)	0	0
O112ab:H8	2010-2017	2	75	*stx*1a (n=2)	0	0
O113:H21	2003-2008	2	223, 3695	*stx*2a (n=1), *stx*2d (n=1)	0	0
O118/O151:H12	2005-2009	2	10	*stx*2b (n=2)	0	0
O123/O186:H11	2007-2008	2	4738	*stx*2a (n=2)	2 (100%)	1 (50%)
O156:H7	2013-2015	2	504	*stx*1a (n=2)	0	2 (100%)
O172:H25	2008-2014	2	660	*stx*2a (n=2)	2 (100%)	0
O174:H21	2012-2016	2	677	*stx*2c (n=1), *stx*2b *stx*2d (n=1)	0	0
O181:H49	2011-2012	2	173	*stx*2a (n=2)	0	0
O187:H28	2007-2017	2	200, 8656	*stx*2a (n=1), *stx*2g (n=1)	0	0
O4:H2	2016	2	17, 20	*stx*1a (n=1), *stx*2f (n=1)	2 (100%)	0
O49:H-	2008	2	8264	*stx*2a (n=2)	2 (100%)	0
O50/O2:H6	2005-2006	2	141, 998	*stx*2b (n=1), *stx* neg (n=1)	0	1 (50%)
O71:H2	2010-2016	2	17	*stx*1a (n=2)	2 (100%)	2 (100%)
O8:H19	2014-2017	2	162, 201	*stx*2e (n=2)	0	0
O87:H16	2005-2016	2	2101	*stx*2b (n=2)	0	0
O98:H21	2012-2013	2	306	*stx*1a (n=2)	2 (100%)	0
Ou:H16	2014-2015	2	3188, 3236	*stx*1a (n=1), *stx*1d (n=1)	0	0
Ou:H45	2015-2016	2	656	*stx*2b (n=2)	0	1 (50%)
O102:H6	2015	1	7926	*stx*2b (n=1)	0	0
O104:H4	2011-2016	2	678	*stx*2a (n=2)	0	2 (100%)
O104:H7	2005	1	1817	*stx*1c (n=1)	0	0
O105:H20	2015	1	8638	*stx*1d (n=1)	0	0
O109:H21	2015	1	40	*stx*2f (n=1)	1 (100%)	0
O109:H32	2011	1	17	*stx*1a (n=1)	1 (100%)	0
O11:H8	2014	1	26	*stx*2b (n=1)	0	0
O113:H6	2013	1	121	*stx*2f (n=1)	1 (100%)	0
O123/O186:H10	2009	1	3695	*stx*2b *stx*1c (n=1)	1 (100%)	0
O123/O186:H2	2014	1	17	*stx*1a (n=1)	1 (100%)	0
O123/O186:H21	2015	1	4179	*stx*2f (n=1)	1 (100%)	0
O126:H34	2014	1	4101	*stx*2f (n=1)	1 (100%)	0
O137:H6	2015	1	6675	*stx*2f (n=1)	1 (100%)	0
O138:H48	2016	1	219	*stx*2e (n=1)	0	0
O149:H8	2017	1	344	*stx*2g *stx*2a (n=1)	0	0
O15:H16	2014	1	325	*stx*2g (n=1)	0	0
O15:H2	2004	1	20	*stx* neg (n=1)	1 (100%)	0
O150:H2	2014	1	306	*stx*2a *stx*1a (n=1)	1 (100%)	0
O153/O178:H19	2013	1	8369	*stx*2f (n=1)	1 (100%)	0
O171:H25	2010	1	297	*stx*2a (n=1)	0	0
O179:H8	2006	1	297	*stx*2a (n=1)	0	0
O180:H2	2013	1	301	*stx*2a (n=1)	1 (100%)	0
O21:H21	2016	1	56	*stx*2b (n=1)	0	0
O22:H8	2015	1	446	*stx*2d (n=1)	0	0
O27:H30	2012	1	753	*stx*2b (n=1)	0	0
O30:H25	2014	1	8660	*stx*2g (n=1)	0	0
O34:H4	2015	1	642	*stx*2f (n=1)	1 (100%)	0
O38:H26	2010	1	10	*stx*2b *stx*1c (n=1)	0	0
O43:H2	2008	1	937	*stx*2b (n=1)	0	0

O49:H10	2013	1	206	*stx*1a (n=1)	1 (100%)	0
O49:H21	2013	1	58	*stx*1d (n=1)	0	0
O51:H28	2013	1	8876	*stx*2b (n=1)	0	0
O55:H12	2013	1	101	*stx*1a (n=1)	0	0
O71:H8	2010	1	2836	*stx*2c *stx*1a (n=1)	1 (100%)	1 (100%)
O76:H7	2015	1	795	*stx*2a (n=1)	1 (100%)	0
O8:H8	2008	1	1656	*stx* neg (n=1)	0	0
O8:H9	2014	1	88	*stx* neg (n=1)	0	0
O81:H21	2009	1	737	*stx*1c (n=1)	0	1 (100%)
O84:H2	2010	1	306	*stx*1a (n=1)	1 (100%)	0
Ou:H2	2004	1	648	*stx* neg (n=1)	0	0
Ou:H20	2016	1	6060	*stx*1c (n=1)	0	0
Ou:H5	2015	1	1161	*stx*2f (n=1)	0	0
Ou:H6	2005	1	362	*stx*2a (n=1)	0	0
Ou:H7	2008	1	2005	*stx*1a (n=1)	0	0
OX18:H2	2014	1	847	*stx*2a *stx*2c (n=1)	0	0
OX18:H8	2014	1	8658	*stx*2a (n=1)	1 (100%)	0
OX25:H11	2016	1	8644	*stx*2a (n=1)	0	0

*The *stx* negative strains had previously tested *stx* positive. Most likely these have lost *stx* phages during subculture.

The top 5 serotypes account for 53.0% of all STEC analysed in this study. Discounting the 2018 figures, which only covered the first two months of the year, *E. coli* O26 was detected every year since 2002; *E. coli* O103:H2 was detected every year since 2005; *E. coli* O145:H28 was detected every year since 2007 (Figure 4). The increasing number of non-O157 STEC detected, particularly from 2014, likely reflects the increasing number of laboratories sending faeces to the SERL for more sensitive testing. By August 2014, all Scottish laboratories were submitting faeces to the SERL for STEC detection.

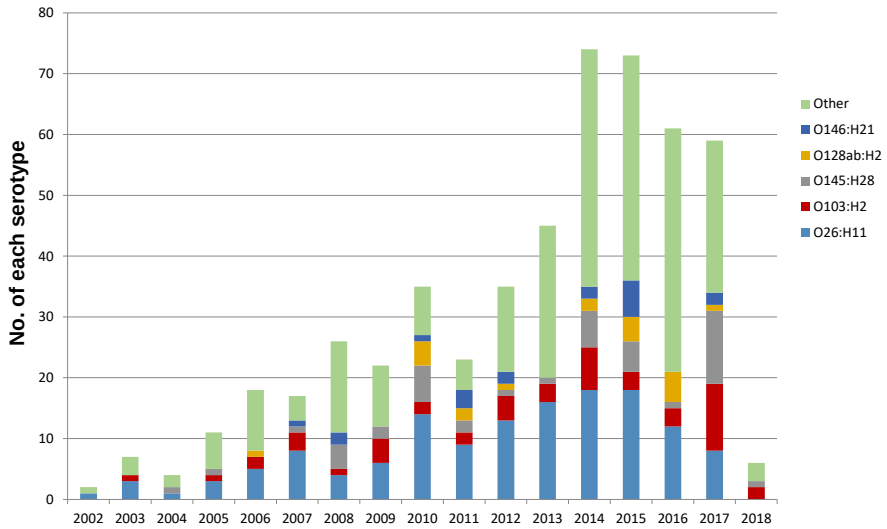

Figure 4. Distribution of the top five serotypes over the study period. The 'Other' serotypes are listed in Table 4. Please note: 2002 and 2018 were not complete study years

Eight STEC co-infections were identified during the study period, where two different non-O157 STEC strains were isolated from the same patient sample (Table 5). In addition, two different STEC were isolated from the same patient from faecal samples submitted one month apart. This could either be the same episode of infection, whereby the patient has been co-infected with two different strains and a different strain has been isolated from each faecal sample, or it could be two different episodes of infection with two different strains. Two different STEC were also isolated from one patient but two years apart, which is most likely a separate infection episode. The number of co-infections detected here most likely represents an underestimate as the sampling protocol involves the follow up of only one colony shown to be positive for Shiga toxin genes unless the screening PCR suggests otherwise.

Table 5: STEC Co-infections

Patient	Strain 1	Strain 2
1	*E. coli* O145:H28	*E. coli* O146:H21
2	*E. coli* O26:H11	*E. coli* O145:H28
3	*E. coli* O51:H28	*E. coli* O174:H8
4	*E. coli* O145:H28	*E. coli* O26:H11
5*	*E. coli* O91:H14	*E. coli* O91:H14
6	*E. coli* Ou:H8	*E. coli* O103:H2
7	*E. coli* O103:H2	*E. coli* O26:H11
8	*E. coli* O145:H28	*E. coli* O26:H11

*Strains 1 and 2 had different virulence profiles

4.2.2 7-gene MLST

A total of 104 different STs were identified by *in silico* 7-gene MLST analyses (Table 4). Figure 5 shows the distribution of serotypes, with ST21, ST17 and ST32 being the most common.

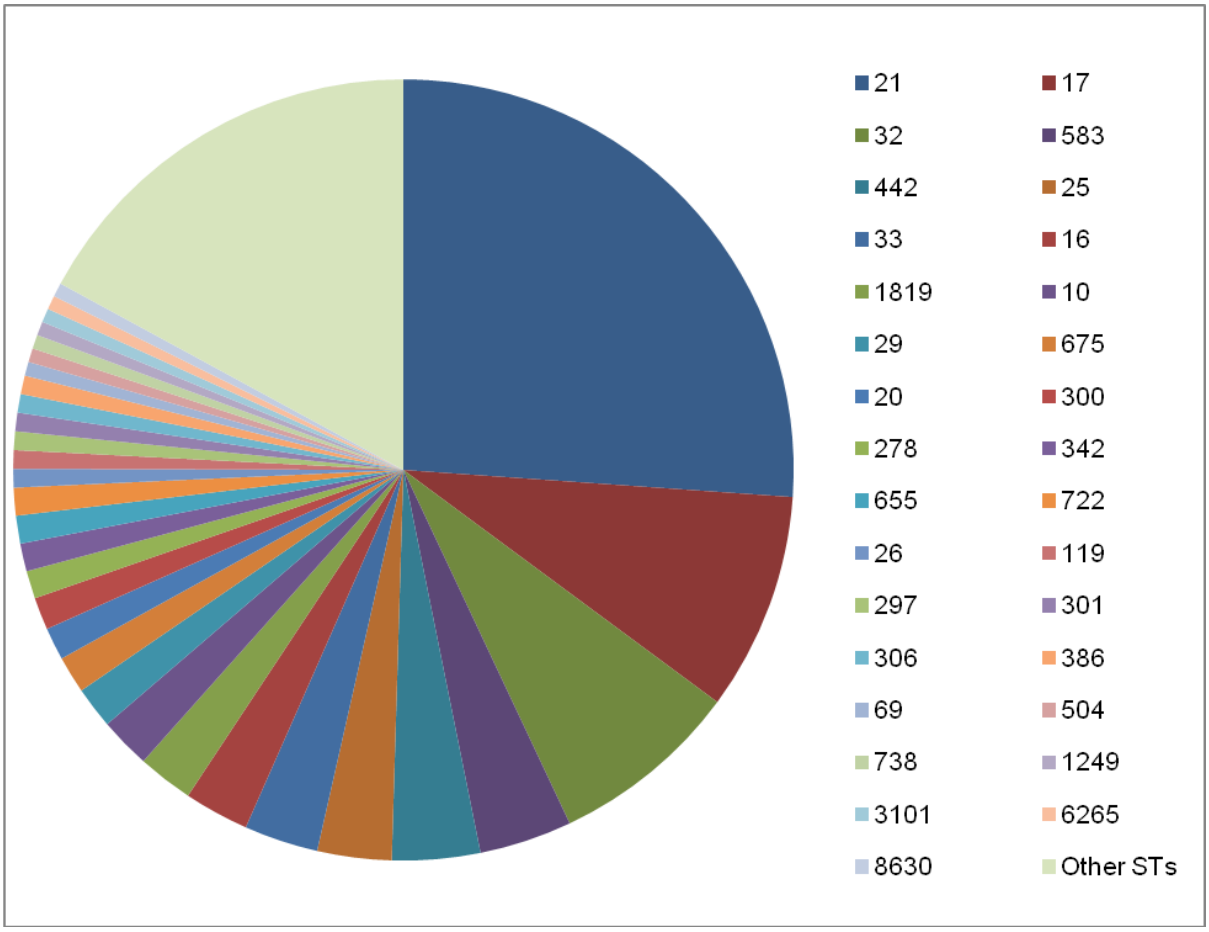

Figure 5. Distribution of different sequence types (STs) among the isolates. The top 31 sequence types (with three or more isolates in the culture collection) are shown and the 'Other STs' are listed in Table 4.

To explore the relationship between serotype and ST (determined using WGS), a minimum spanning tree (MST) based on the 7-gene MLST data was created. As shown in Figure 6, serotypes clustered with ST although, in some cases, different serotypes shared the same ST. For example, O26:H11 and O177:H11, O63:H6 and O125ac:H6, O103:H2 and O45:H2, and O177:H25 and O5:O9 (Table 4). Others have shown that closely related isolates can have different serotypes, particularly O antigens as these are subject to strong selection pressure from the mammalian host (animal/human) immune system resulting in recombination events around this locus [78].

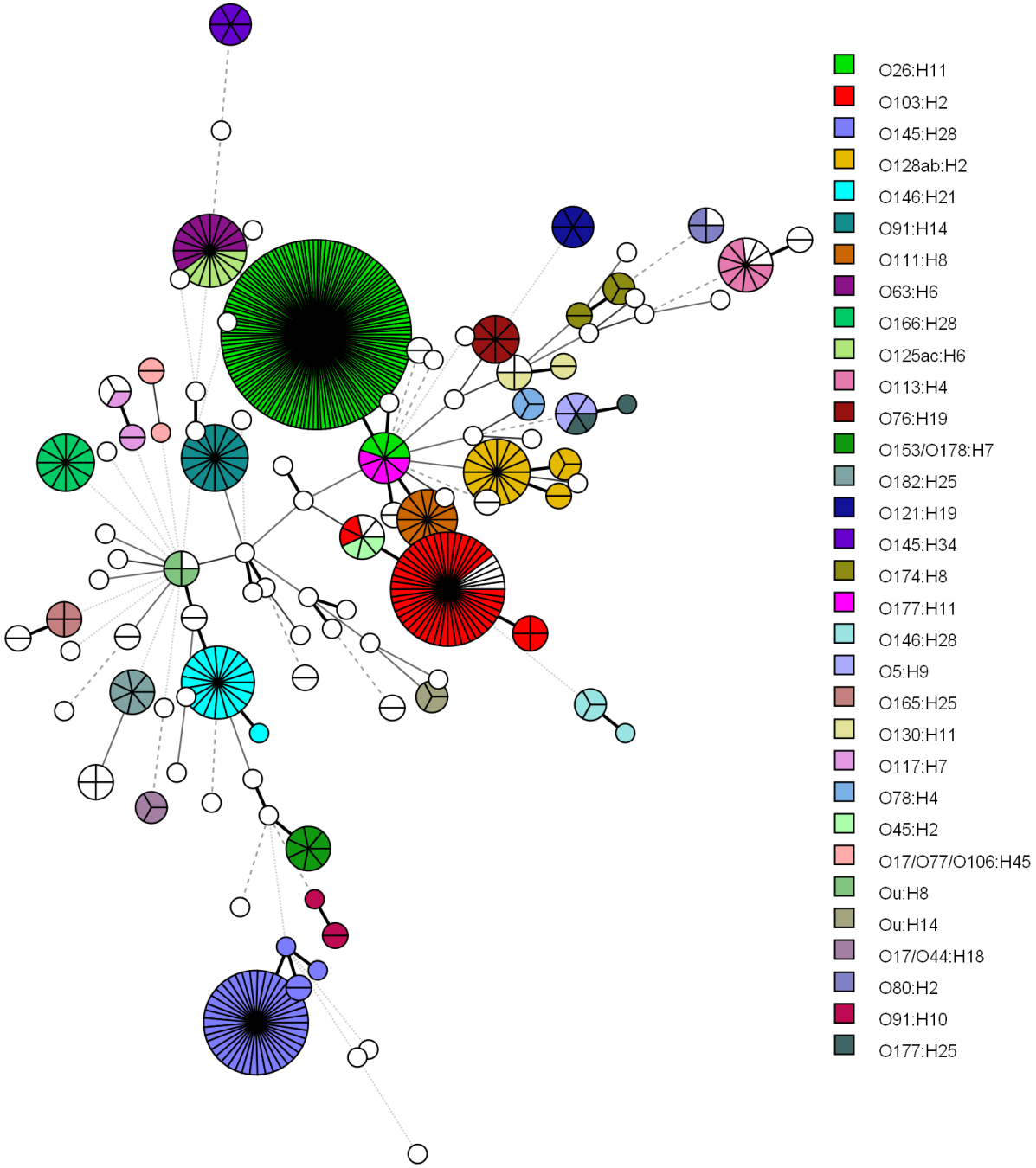

Figure 6. Minimum spanning tree (MST) based on the 7-gene MLST data of the 522 clinical isolates. The top 32 serotypes are highlighted in different colours. The MST was created using BioNumerics v7.6 using the advance cluster analysis tool and the template 'MST for categorical data'. The node sizes are related to the number of isolates and the lines infer relatedness; solid lines joining nodes denote single locus variants. Nodes and segments in white do not belong to the top 32 serotypes.

4.2.3 Shiga xoxin gene profiles

A total of 21 different *stx* subtype profiles were detected, with all known *stx* subtypes identified (Table 6). *stx*1a only was the most common subtype profile (n=176/522; 33.7%), followed by *stx*2a only (n=78/522; 14.9%) and *stx*2a *stx*1a (n=60/522; 11.5%). Of the 176 non-O157 STEC harbouring *stx1a*, 48.3% were serotype O26:H11 and 26.7% were O103:H2. Of the 78 isolates harbouring *stx2a* only, 52.5% (n=41/78) were serotype O145:H28. The most common *stx* subtype combination was *stx*2a *stx*1a harboured by six different serotypes, with O26:H11 predominating (n=46/60; 76.7%). The subtype combination of *stx*2b *stx*1a was almost exclusively detected in serotype O91:H14, while *stx*2b *stx*1c was detected in eight different serotypes including O128ab:H2, O166:H28, O113:H4 and O146:H21 (not unlike the *stx* subtype-serotype combinations reported by Chattaway *et al.* [8]). Of the 13 isolates carrying *stx*2d only, 8 different serotypes were identified including O145:H28 and O80:H2. Five isolates did not carry *stx* genes; most likely the genes were lost during subculture. All known *stx* subtypes were detected but *stx*2e was observed on just three occasions and *stx*2g detected four times.

Table 6: Shiga toxin gene profiles detected among the isolates

***stx* subtype profile**	**No. isolates**
*stx*1a	176
*stx*2a	78
*stx*2a *stx*1a	60
*stx*2b *stx*1c	46
*stx*2f	40
*stx*2b	36
*stx*1c	30
*stx*2b *stx*1a	14
*stx*2d	13
*stx*2c	4
*stx*1d	3
*stx*2e	3
*stx*2g	3
*stx*2a *stx*2c	2
*stx*2a *stx*2c *stx*1a	2
*stx*2c *stx*1a	2
*stx*1a *stx*1c	1
*stx*2a *stx*2g	1
*stx*2b *stx*1a *stx*1c	1
*stx*2b *stx*2c	1
*stx*2d *stx*2b	1
stx neg	5

As shown in section 4.2.1, a large proportion of the isolates were O26:H11. Historically, this serotype typically possessed the *stx*1 gene only [79], but over time has acquired *stx*2 [80]. In this study, the first strain carrying *stx*2a in combination with *stx*1 was isolated in 2005, while the first *stx*2a only strain was isolated in 2010 (Figure 7). However, in Scotland, *stx*1a, alone, still predominates. See section 4.5 for a more detailed analysis of the *E. coli* O26:H11 strains.

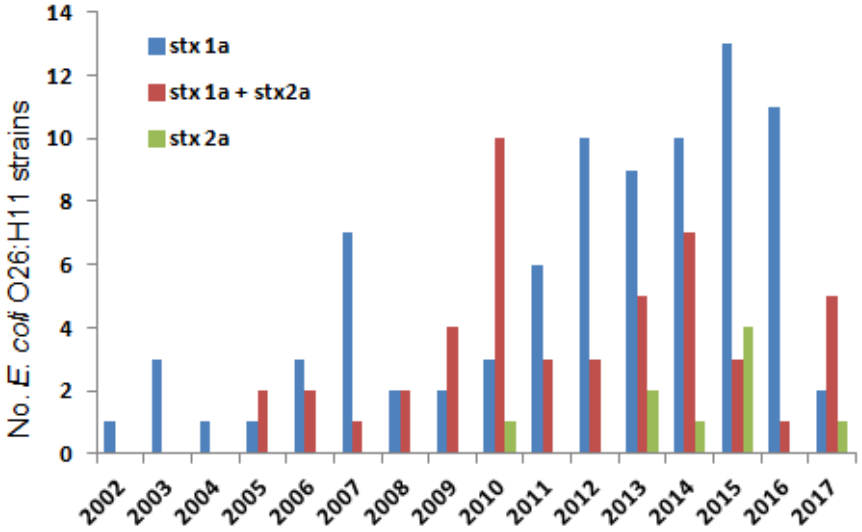

Figure 7: *stx* Subtypes in *E. coli* O26:H1

14.2.4 Virulence Gene Detection

Shiga toxin subtypes (in particular *stx*2a) and the LEE pathogencity island carrying *eae* are key virulence factors strongly associated with severe disease in patients infected with STEC. However, these factors alone do not accurately predict pathogenicity as, in their absence, some strains are able to cause severe disease [11,81]. A large number of virulence factors have been identified in STEC using various experimental (e.g. mutagenesis and *in-vitro* and *in-vivo* models) and computational approaches [82–89]. These include genes encoding proteins involved in: acid resistance, enabling bacteria to survive in the low pH conditions of the intestine; adhesion, allowing attachment to the intestinal epithelium; secretion system effectors involved in a variety of different processes including translocation and adherence; Serine Protease Autotransporters of *Enterobacteriaceae* (SPATE), secreted proteins involved in virulence functions such as adherence, aggregation, and toxicity; siderophores for transporting iron; microcins associated with bacterial competition; and other enzymes such as catalase peroxidase to reduce oxidative stress in the host intestine. To identify the combination(s) of genes that equip STEC with the ability to cause clinical disease, known virulence genes present in the VirulenceFinder reference database were interrogated. A total of 89 virulence genes (excluding the *stx* genes) were detected among the isolates (see Appendix 4 for the genes and their category). The most common genes were *Fim*H (518), *gad* (n=512),

iss (n=424), *ehx*A (n=399), *hly*D (n=398), *ipf*A (339) and *eae* (n=336). Each isolate sequenced was found to contain virulence genes (other than *stx*), and the mean number of virulence genes associated with each serotype is shown in Table 7 and ranged from three (O8:H8) to 25 (O71:H8).

Table 7. Mean number of virulence genes (excluding *stx* subtype) associated with each serotype (NB no serotypes were found to contain none).

Mean No. Virulence Genes[1]	Serotype[2]
25	O71:H8
24	O177:H11, **O26:H11**
23	O104:H4, O109:H32, O123/O186:H11, O80:H2
22	O172:H25, O165:H25, O5:H9
21	O111:H8, O177:H25, O4:H2, O81:H21
20	**O145:H28**, O71:H2, O76:H7, O50/O2:H6
19	**O103:H2**, O123/O186:H2, O84:H2
18	O121:H19, O49:H-
17	O45:H2, O182:H25
16	O146:H21*, O137:H6, O150:H2, O98:H21, O112ab:H8, O156:H7
15	O174:H8, O91:H14*, O128ab:H2*, O11:H8, O146:H28, O49:H10, O166:H28
14	O123/O186:H10, O15:H2, O171:H25, O27:H30, O51:H28, O181:H49
13	O78:H4, O130:H11, O109:H21, O113:H6, O118/O151:H12, O123/O186:H21, O17/O77/O106:H45, O179:H8, OX18:H2, OX18:H8, Ou:H14, O76:H19
12	Ou:H8, O117:H7, O63:H6
11	O105:H20, O113:H21, O113:H4, O153/O178:H7, O180:H2, O55:H12, Ou:H7, O125ac:H6, O112ab:H2
10	O104:H7, O145:H34, O187:H28, O22:H8, O34:H4, Ou:H2, Ou:H5, Ou:H6, OX25:H11
9	O126:H34, O149:H8, O17/O44:H18, O21:H21, O38:H26, O43:H2
8	O102:H6, O138:H48, O15:H16, O174:H21, Ou:H16, O91:H10,
6	O30:H25, O49:H21, O8:H19, Ou:H20, O87:H16
5	O153/O178:H19, O8:H9
3	O8:H8

[1] The number of virulence genes detected in each isolate was calculated. The mean number associated with each serotype was then determined.
[2] The top three eae-positive Scottish serotypes are in bold and the top three *eae*-negative serotypes are in red and denoted with *.

There was a large variation in the number and combination of viruence genes detected among the different serotypes highlighting the diversity of non-O157 STEC. Fourteen serotypes were associated with a mean number of virulence genes >20, which included the most common serotype O26:H11. Interestingly, some uncommon serotypes were associated with large numbers of virulence genes.

In the majority of cases, all isolates within a serotype either carried *eae* (a marker of the LEE pathogenicity island) or did not carry *eae*, however a large number of serotypes contain only small numbers of isolates. The top three serotypes that did not contain *eae* were O128ab:H2, O146:H21 and O91:H14 (Table 4). A comparison of the isolates was made based on *eae*-presence/absence to elucidate the virulence gene repertoire of these two groups. The results showed *eae*+ve strains (n=336/522; 64.4%) were more frequently isolated than *eae*-ve strains (n=186/522; 35.6%) and a greater number of different virulence genes were associated with *eae*+ve strains, 55 compared with 41 respectively. As shown in Figure 8,

- Genes significantly more often detected in *eae*+ve strains were *espA*, *tir, nleB, espF*, *espJ*, *cif*, *ehxA*, *hlyD*, *espB*, *nleA*, *astA*, *efa1*, *nleC*, *espP*, *katP*, *toxB*, *PA(MAIX)*, *fyuA, tccP, TraT, etpD, ibeA, espC*.
- Genes more predominant in *eae*-ve strains were *lpfA*, *iha*, *celb*, *cma*, *epeA*, *mchF*, *mchB*, *mchC, hra, iron, cvi-cvaC*, *hlyF*, *ireA*, *mcmA, usp*, *senB*, *iutA*, *subA, air, eilA*, *pic*, *saa*.

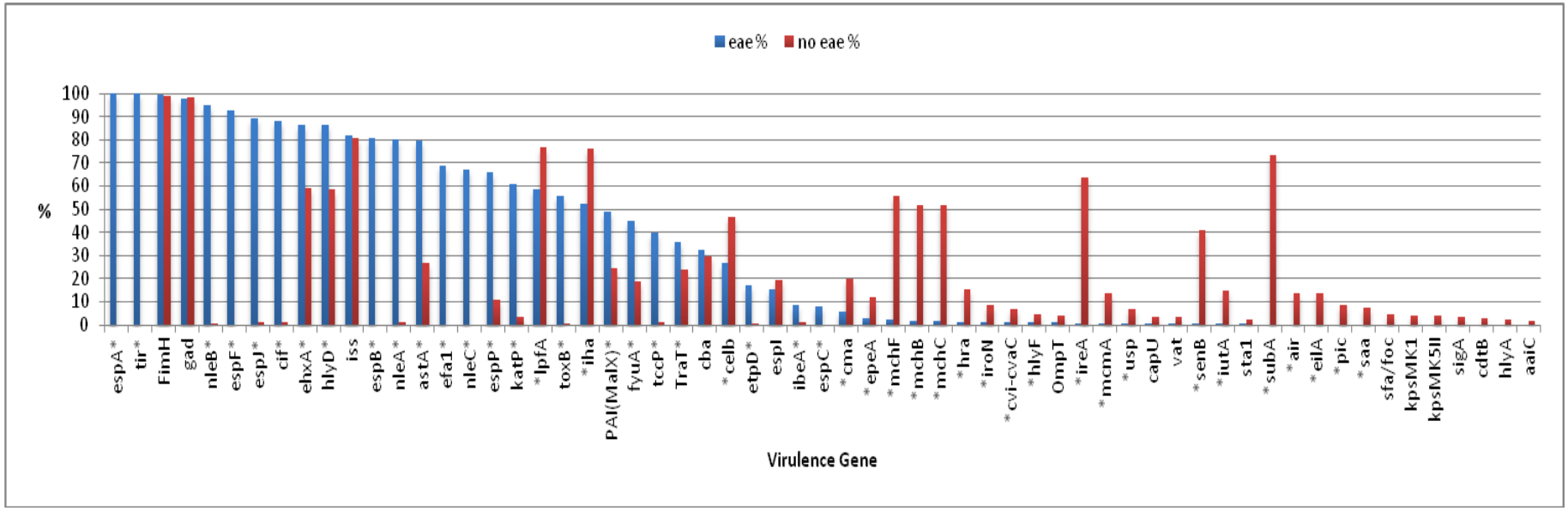

Figure 8. **Comparison of virulence genes in *eae*-positive and *eae*-negative STEC.** Blue and red bars represent the frequency (%) of virulence genes in *eae*-positive and *eae*-negative isolates, respectively. An asterix (*) before or after the gene name denotes significance ($p < 0.05$) in *eae*-negative and *eae*-positive isolates respectively. Genes carried by <2 isolates are not shown and significance tests were performed when at least 10 isolates carried a particular gene.

Thirty four genes were uniquely associated with *eae*-ve strains, however the majority of these (85%) were not widely distributed and present in only a small subset of isolates. *subA* was the most prevalent gene solely linked with *eae*-ve strains, present in 73.1% of all *eae*-ve isolates and in 98.2% of the isolates in the top 3 *eae*-ve serotypes. *subA* encodes for a cytotoxin that inhibits protein synthesis and is carried on the subtilase-encoding pathogenicity island [89].

Genes widely distributed but not exclusive to *eae*-ve strains were *lpfA* and *iha* (adhesins), *celb* (colicin), *mchF/B/C* (micocins), *ireA* (siderophore receptor) and *senB* (enterotoxin) (Figure 8). A closer look at these genes in the top *eae*+ve and *eae*-ve serotypes showed, in some cases, the genes clustered with serotype. In particular, *lpfA* was detected in the top three *eae*-ve serotypes and in O26:H11 strains but not in O145:H28 or O103:H2 strains (data not shown).

Recently others have identifed new molecular mechanisms that may compensate for the lack of the LEE that were not included in the reference database used in this study. For example, Montero *et al.* reported 60% (29/48) of LEE-negative STEC were positive for Hes (Haemagglutinin from STEC), a virulence factor present on the Locus of Adhesion and Autoaggregation (LAA) involved in colonisation phenotypes such as hemagglutination, adhesion and autoaggregation.

4.2.5 Acquired antimicrobial resistance (AMR)

A total of 92/522 (17.6%) isolates carried AMR genes present in the Resfinder database. Thirty two different genes were detected and these included genes resistant to aminoglycosides (n=10; *aadA1*, *aadA5*, *aadA12*, *aadB*, *aph(3')-Ia*, *aph(4)-Ia*, *aac(3)-Iid, aac(3)-Iva*, *strA*, *strB*), β-lactams (n=6; *blaTEM*-1A, *blaTEM*-1B, *blaTEM*-1C, *blaTEM*-30, blaCTX-M-15, *blaSHV*-48), tetracycline (n=4; *tet(*A*)*, *tet(*B*)*, *tet(*D*)*, *tet(*31*)*), trimethoprim (n=6; *dfr*A1, *dfr*A5, *dfr*A7, *dfr*A8, *dfr*A14, *dfr*A17), sulphonomide (n=2; *sul*1, *sul*2), macrolide (n=2; *mph(*A*)*, *mph(*B*)*), chloramphenicol (n=2; *catA1*, *floR*). The frequency of these acquired resistance genes among the isolates is shown in Figure 9.

Aminoglycoside resistance genes were found in 85/92 (92.3%) isolates showing resistance. The most common aminoglycoside genes detected were *strA and/or strB,* which were carried by 75/92 (81.3%) strains. Eighteen strains carried genes encoding aminoglycoside adenylyltransferase (*aad*A1, *aad*A5, *aad*A12 and *aad*B), 13 carried genes encoding aminoglycoside phosphotransferases (*aph(3')-Ia*, *aph(4)-Ia*) and three carried genes encoding aminoglycoside acetyltransferases (*aac(3)-IId, aac(3)-Iva*). Sulphonamide resistance genes were found in 75/92 (81.5%) isolates: 68 carried *sul*2 and 20 carried *sul*1. Twelve isolates carried *sul*2 in combination with *sul1*.

Genes predicted to confer resistance to β-lactam antibiotics were detected in 45/92 (48.9%) isolates. The most common genes were the penicillinase-encoding *blaTEM-1* (n=43), especially *blaTEM*-1B. blaCTX-M-15 and blaSHV$_{102}$ were only detected once.

Tetracycline resistance genes occurred in 45/92 isolates (48.9%), mostly *tet*(A) (n=34). Only two isolates carried a combination of two different genes, *tet*(A) and *tet*(B). Only one isolate carried *tet*(31) and one carried the efflux pump-encoding gene *tet*(D).

Trimethoprim resistance-conferring *dfr*A gene variants were identified in 17/92 isolates (18.4%), most commonly *dfr*A1 (n =7). The remaining isolates carried five additional variants of *dfr*A. No isolates harboured a combination of two different *dfr*A genes.

Genes linked to chloramphenicol resistance were identified in 10/92 isolates (11.0%). Efflux pump genes (*floR*) were found in 9 isolates. Chloramphenicol acetyltransferase genes (*cat*A1) were detected in 3 isolates. Two isolates (O80:H2) harboured genes encoding both an efflux pump and an acetyltransferase.

There were no genes associated with resistance to colistin.

Of the 92 isolates carrying resistance genes 63/92 (68.5%) were multi-drug resistant (MDR, resistance to three or more antibiotic classes). Four strains possessed genes conferring resistance to 6 different classes of antibiotic (serotypes O111:H8, O26:H11, O128:H2 and O103:H2) (Appendix 3). There were 18 different AMR profiles and the most common profile was resistance to aminoglycoside, sulphonamide and tetracycline (n=15).

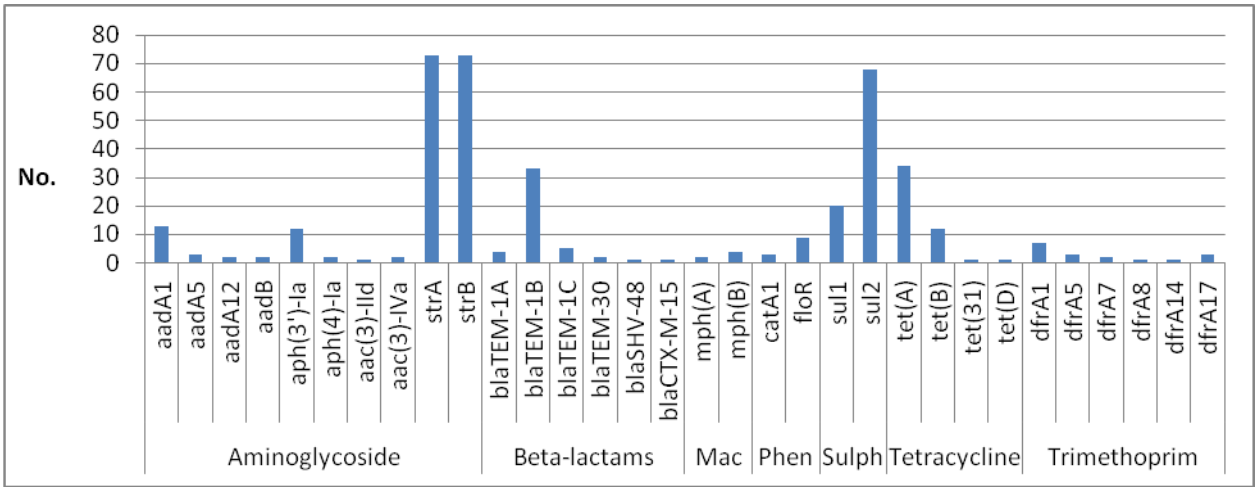

Figure 9. Number of acquired antimicrobial resistance genes detected among isolates carrying one or more resistance genes (n=92). The seven classes of antibiotics were aminoglycosides, β-lactams, macrolides (Mac), phenicols (Phen), sulphonamides (Sulph), tetracycline and trimethoprim.

4.3 Phylogenetic overview of non-O157 STEC

The phylogenetic relationship between the isolates was examined by cgMLST (Figure 10), which showed non-O157 STEC formed multiple, distinct lineages. Notably, the major *eae*+ve (O26:H11, O103:H2, O145:H28) and *eae*-ve (O128:H2, O146:H21 and O91:H14) serotypes detected in this study formed separate lineages, consistent with other reports that non-O157 STEC have evolved by parallel evolution [90,91]. Overall, the same serotypes clustered together, consistent with the comparison of 7-gene MLST with serotype shown in Figure 6.

In some cases, different serotypes clustered closely together and shared a ST, suggesting they derived from a common ancestor. Notably, O26:H11 and O177:H11, O63:H6 and O125ac:H6, and O103:H2 and O45:H2. Interestingly, in many of these cases, isolates with the same H groups clustered closely together while isolates with the same O group, e.g. O145:H28 and O145:H34, O146:H21 and O146:H28, O177:H11 and O177:H25 were often found at distinct locations on the tree and belonged to different lineages.

This has been reported previously, in some cases with different serotype combinations. For example, Ju *et al*. [92] showed O26:H11 strains clustered together closely with O111:H11 strains while different O111 serogroups (H21, H2, H8 and H11) were found at disinct locations. Iguchi *et al*., [93] suggested that STEC O103:H2, O103:H11 and O103:H25 formed three different lineages while O26:H11 strains clustered with isolates of O103:H11. More recently, Alikhan *et al.* [91] showed strains of serotype O45:H2 and O103:H2 belonged to the same lineage. These data reflect the greater immune pressure surface O antigens experience compared with H antigens, resulting in more frequent O antigen lateral gene transfer, and the H group being a better predictor of relatedness than O group.

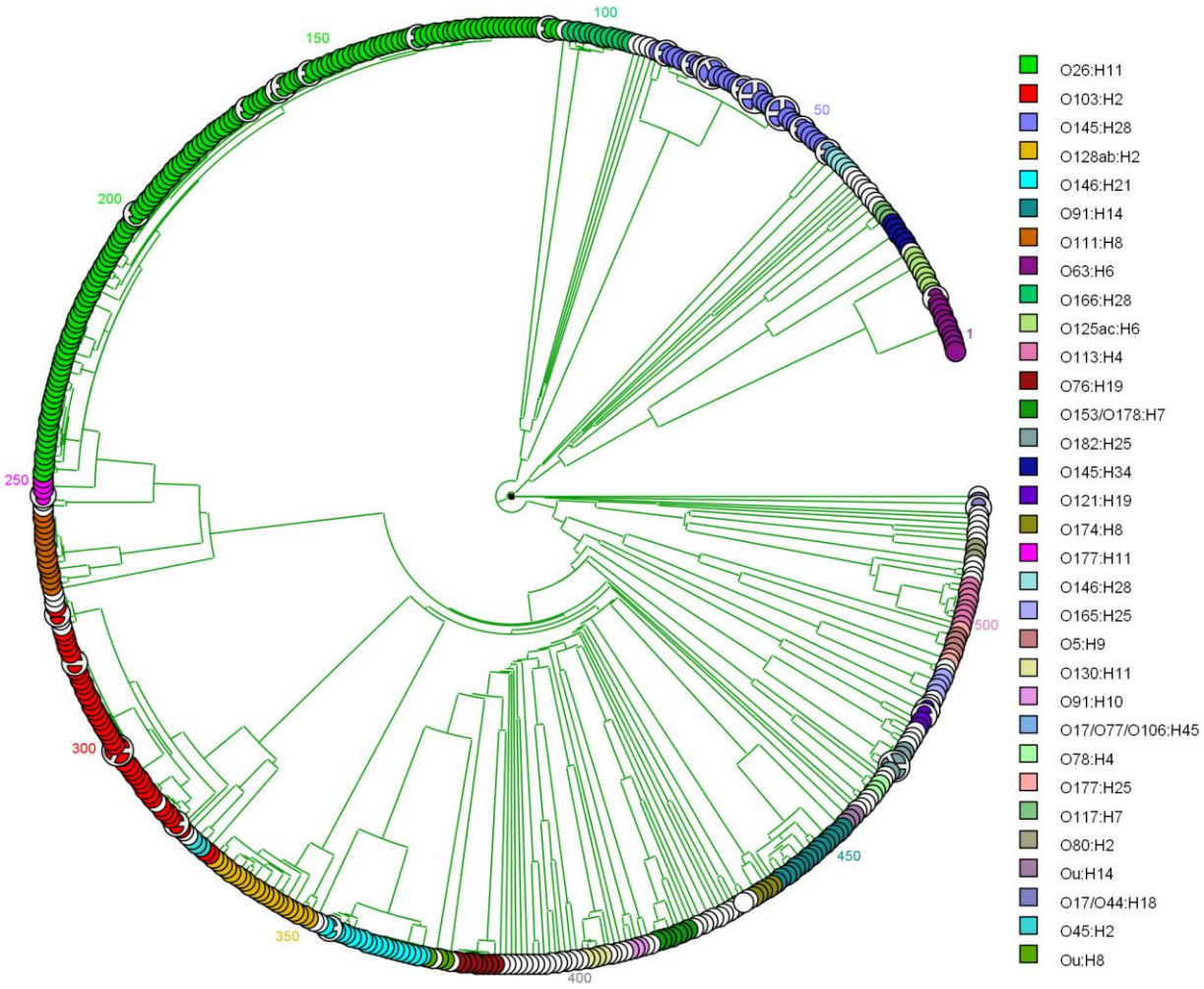

Figure 10. Dendogram based on the allelic profiles of 2513 cgMLST target genes for the 522 isolates. The tree was produced in BioNumerics v7.6 using the Advanced Cluster Analysis Tool (scaling factor of 13) and Topscore UPGMA. Serogroups (n=32) represented with greater than two isolates are highlighted in different colours and the SERL ID of every 50th sample is labelled.

4.4 Potential to cause clinical disease – JEMRA level assignment

The JEMRA report estimated the potential of different combinations of virulence genes to cause diarrhoea (D), bloody diarrhoea (BD) and haemolytic uraemic syndrome (HUS). The virulence profiles of the non-O157 STEC in this study were categorised according to virulence gene combination and assigned a JEMRA level, where Level 1 would demonstrate the potential for that strain to cause D/BD/HUS (Table 8).

Table 8: JEMRA Level Assignment of Non-O157 STEC

JEMRA Level[1]	Trait (Gene and subtype combination)[2]	No. of strains	Serotype[3]
1 (D/BD/HUS)	***stx*2a, *eae***	66	O121:H49, O123-O186:H11, **O145:H28**, O172:H25, O180:H2, O182:H25, **O26:H11**, O49:H-, O76:H7, O80:H2, Ou:H8
	stx*2a**, ***eae (with *stx*1a)	59	O111:H8, O150:H2, O165:H25, O177:H11, **O26:H11**
	stx*2a**, ***eae (with *stx*1a & *stx*2c)	1	O165:H25
	***stx*2a**, ***agg*R**	2	O104:H4
2 (D/BD/HUS[4])	***stx*2d**	13	O113:H21, O113:H4, **O145:H28**, O17-O44:H18, O17-O77-O106:H45, O22:H8, O80:H2, O91:H10
	***stx*2d** (with *stx*2b)	1	O174:H21
3 (D/BD)[5]	***stx*2c**, ***eae***	3	O177:H25
	stx*2c**, ***eae (with *stx*1a)	2	**O26:H11**, O71:H8
4 (D/BD)[5]	***stx*1a**, ***eae***	161	**O103:H2**, O109:H32, O111:H8, O121:H19, O123-O186:H2, **O145:H28**, O177:H11, O182:H25, O26:H11, O4:H2, O49:H10, O5:H9, O71:H2, O84:H2, O98:H21
5 (D)	**other *stx* subtypes**	209	O102:H6, **O103:H2**, O104:H7, O105:H20, O109:H21, O11:H8, O112ab:H2, O112ab:H8, O113:H21, O113:H4, O113:H6, O117:H7, O118/O151:H12, O123/O186:H10, O123/O186:H21, O125ac:H6, O126:H34, **O128ab:H2**, O130:H11, O137:H6, O138:H48, O145:H34, O146:H21, O146:H21, **O146:H28**, O149:H8, O15:H16, O153/O178:H19, O153/O178:H7, O156:H7, O166:H28, O17/O77/O106:H45, O171:H25, O174:H21, O174:H8, O179:H8, O181:H49, O187:H28, O21:H2, **O26:H11**, O27:H30, O30:H25, O34:H4, O38:H26, O4:H2, O43:H2, O45:H2, O49:H21, O50/O2:H6, O51:H28, O55:H12, O63:H6, O76:H19, O78:H4, O8:H19,O81:H21, O87:H16, O87:H16, O91:H10, O91:H14, Ou:H14, Ou:H16, Ou:H20, Ou:H45, Ou:H5, Ou:H6, Ou:H7, Ou:H8, OX18:H2, OX25:H11

[1]Potential to cause illness in parenthesis. This will also be dependent on host susceptibility and other factors such as antibiotic treatment.
[2] Extra genes present in parenthesis. This table does NOT include the 5 *stx* negative strains.
[3] Top 5 Scottish serotypes represented in bold
[4] Association with HUS dependent on *stx*2d variant and strain background
[5] Some subtypes have been reported to cause BD, and rarely HUS

Of those non-O157 in this study possessing *stx* genes (n=517), 142 (27.5%) were assigned Level 1 or 2, with the potential for these strains to cause HUS; 166 (32.2%) were assigned Level 3 or 4 with potential to cause diarrhoea or bloody diarrhoea and 209 (40.5%) assigned to level 5 with potential to cause diarrhoea. Interestingly, three

of the top 5 Scottish serotypes are present in more than one JEMRA level. For example, strains of *E. coli* O26:H11 are present in JEMRA Levels 1, 3 and 5; *E. coli* O145:H28 are present in JEMRA Levels 1, 2 and 4 and *E. coli* O103:H2 strains are present in JEMRA levels 4 and 5.

4.5 Predominant, emerging and hybrid strains: a comparison with published data

4.5.1 *E. coli* O26:H11

E. coli O26:H11 was the most common serotype isolated in Scotland over the study period (n=141; 27.1%). *In silico* MLST analysis identified three different STs (all belonging to CC29): ST21 (n=136; 96.4%), ST29 (n=4; 2.8%) and ST574 (n=1; 0.7%). All O26:H11s typed were positive for *eae*, except one (sample no.127, an ST21). This isolate also did not harbour the LEE genes *espA*, *espB* and *tir* suggesting the loss of this element. The majority of the ST21 were *stx*1a positive only (n=83; 61.0%), 45 (33.0%) were positive for *stx*1a and *stx*2a, 7 (5.1%) were *stx*2a only and one (0.7%) was positive for *stx*1a and *stx*2c. Of the four ST29 strains, two carried *stx*2a only, one was positive for *stx*1a and *stx*2a and one was *stx*1a only. The ST574 strain carried *stx*1a only. No O26:H11 strains harboured *stx*2d.

A total of 139/141[2] of the Scottish strains were analysed alongside sequences from the public database representative of the previously reported global O26 lineages ST29C1, ST29C2, ST21C1 and ST21C2 [67,94]. The SERL strains were labelled according to their phylogenetic position relative to the reference strains included in our analysis. Bootstrap values for the lineages and sub-lineages described below were greater than 90% in the maximum likelihood (ML) phylogeny (with the majority being 100%), as indicated in Figure 11, indicating a high level of support for the clusters observed. Strong concordance was observed between the phylogenetic clustering of strains in Figure 11 and the Bayesian analysis of population structure (BAPS) clusters inferred from the BratNextGen analysis.

The ML phylogenetic analysis revealed an overall topology consistent with that published by Ogura *et al.* [67]. The ST29 strains formed two separate clusters, each with 100% bootstrap support. One clade was denoted ST29C2 following previous studies, and the other contained two strains which had been denoted ST29C1 by Ogura *et al*. [67] but one of which was assigned to ST5172 (a single locus variant of ST29) by our 7-gene MLST analysis. The phylogeny was rooted on the two ST29C1 strains identified by Ogura *et al*. [67] since their results indicated these would be a suitable outgroup for our dataset; rooting on the ST21C1 strains had the same overall effect on the topology as midpoint rooting. The four ST29 Scottish strains clustered with the ST29C2 reference strains and all had the same plasmid-encoded

[2] The two ST21 *E. coli* O26 strains associated with dual infections were not included in this analysis

virulence factor (pVF) gene profile, *ehxA*+/*katP*-/*espP*-/*etpD*+. Two of these four strains additionally carried *stx*2a only showing they belonged to the 'new European clone' [80,94,95]. No strains with the characteristics of the 'new French clone' (ST29C3 *stx*2d *ehxA*-/*katP*-/*espP*-/*etpD*- [95]) were detected and no Scottish strains clustered with representatives of lineage ST29C1.

The ST21 strains formed a distinct lineage (100% bootstrap support for the clade) from ST29 in the ML tree, and the ST21 clade also included the single locus variant ST574 isolate. Two ST21 isolates (No.247 and No.248) were basal to all other ST21 isolates, including the publicly available ST21C1 and ST21C2 strains. One of these two strains (No.247) had the plasmid gene profile *ehxA*-/*katP*-/*espP*-/*etpD*- and was the only strain in our dataset that was found to be negative for all four of these genes. The majority (131/134; 98%) of the ST21 strains had the pVF gene profile *ehxA*+/*katP*+/*espP*+/*etpD*-. Two strains had the profile *ehxA*+/*katP*-/*espP*-/*etpD*-.

Within ST21, 16 Scottish isolates formed a monophyletic clade (100% bootstrap support) with the public ST21C2 sequences. However, the phylogenetic relationships between ST21C1 and ST21C2 strains in the ML tree would support the further subdivision of ST21C1 into two sublineages and we also note that our BAPS analysis depicted in Figure 11 split the ST21C1 isolates into different clusters. These patterns are consisent with the ML tree and BAPS clustering analysis of Ogura *et al*. and we therefore henceforth label the two sublineages of isolates which cluster with the public ST21C1 strains as ST21C1a and ST21C1b (each with 100% bootstrap support, as depicted on Figure 11). We did not assign the two ST21 isolates (No.247 and No.248) described above to a named ST21 sublineage.

The majority of the SERL isolates (107/139; 76.9%) belonged to ST21C1b, and most of these (68/107; 63.6%) carried *stx*1a only. Thirty eight of the 107 SERL ST21C1b isolates carried both *stx*1a and *stx*2a, one strain carried both *stx*1a and *stx*2c, and no strains were positive for *stx*2a only. By contrast, a monophyletic clade (100% bootstrap support) containing 5 out of the 10 (50%) ST21C1a isolates harboured *stx*2a only; 4 ST21C1a strains carried *stx*1a and *stx*2a and one carried *stx*1a only. Considering the high prevalence of *stx*2a amongst ST21C1a, it will be interesting to determine if these strains are associated with more severe disease. Sixteen strains belonged to ST21C2 and 13 harboured *stx*1a alone, while 3 were *stx*1a and *stx*2a positive. The pattern of *stx*2a presence or absence amongst SERL O26 isolates, particularly across the ST21C1b clade, is indicative of multiple gain/loss events and could warrant further quantification through genetic analysis.

A total of 32/139 (23.0%) strains carried acquired AMR genes, and 18/139 (12.9%) were MDR. Resistance was exclusively among the ST21 strains.

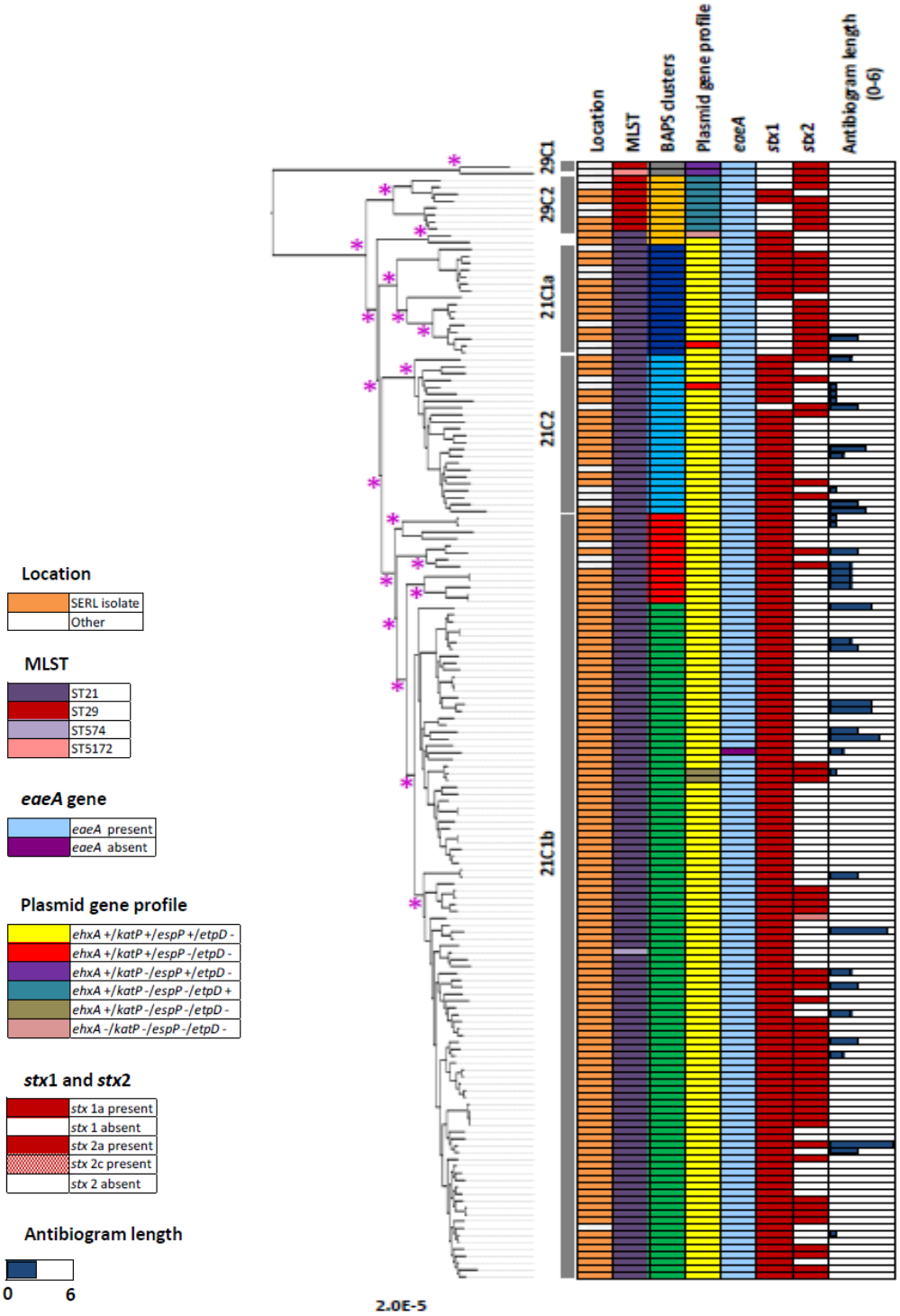

Figure 11 Maximum-likelihood core genome phylogeny for 139 *E. coli* O26:H11 isolates from the Scottish *E. coli* Reference Laboratory, plus publicly available strains from major O26 lineages. The tree was constructed with RAxML using a general time-reversible model of nucleotide substitution and gamma distributed rate heterogeneity across sites. Branch lengths are in numbers of substitutions per site. SERL isolates are indicated in orange in the 'Location' column. 7-gene multi-locus sequence types ('MLST'), clusters identified by the Bayesian Analysis of Population Structure (BAPS) software, plasmid gene profiles, *stx* subtypes and the presence or absence of the *eaeA* gene are shown. The number of antibiotic classes to which an isolate was predicted to be resistant ('Antibiogram length') based on identification of resistance genes from the sequence assembly is also displayed as a bar plot (with a longer blue bar representing resistance to a greater number of antibiotics, on a scale from 0 to 6). Sublineages, defined as described in the text, are indicated by vertical grey bars. 1000 bootstrap replicates were performed, and bootstrap values of at least 90% are indicated by asterisks at nodes corresponding to the major lineages and sublineages described in the text or identified by the BAPS clustering analysis.

4.5.2 *E. coli* O103:H2

E. coli O103:H2 was the second most frequently isolated serotype (n=49/522, 9.4%) in this study group. *In silico* 7-gene MLST identified 4 different STs (belonging to CC20): ST17 (n=42; 85.7%), ST386 (n=4; 8.2%), ST20 (n=2; 4.1%), and ST8506 (n=1; 2.0%). The ST allelic profiles are shown in Appendix 2. All but one strain was positive for *eae* (n=48; 98.0%) and the majority of strains (n=47; 96.0%) were *stx*1a positive, except for the two (4.1%) ST20 isolates that were *stx*2f positive. Twelve (24.5%) of the isolates carried AMR genes, with nine different profiles (Appendix 3).

The Scottish strains were analysed alongside a strain (17-00944) from an outbreak among German children following a school trip to Austria that was associated with the consumption of raw cow's milk [29]. Similar to the majority of the Scottish strains, the outbreak strain was ST17 and carried *stx*1a (Figure 12). The ST17 strains were associated with several different pVF gene profiles, with *ehxA*+/*katp*+/*espP*-/*etpD*+ being the most comon (n=22). Interestingly, a small cluster of ST17 strains (No.316-319) had the same pVF gene profile as the ST386 isolates (*ehxA*+/*katp*+/*espP*+/*etpD*-) and were characterised by the presence of *toxB*, a homolog of a large clostridial toxin found on the *E.coli* O157:H7 EHEC plasmid. The ST20 *stx*2f positive isolates did not carry any of the pVF genes and were genetically distinct from the other O103:H2 isolates, differing at a large number of loci (>585). The majority of the O103:H2 carried the genes encoded on the locus for enterocyte effacement (LEE) pathogenicity island, including, *eae*, *tir*, *espA* and *espB*. The outbreak strain did not carry any resistance genes, and the SERL isolates carrying AMR genes were found throughout the tree.

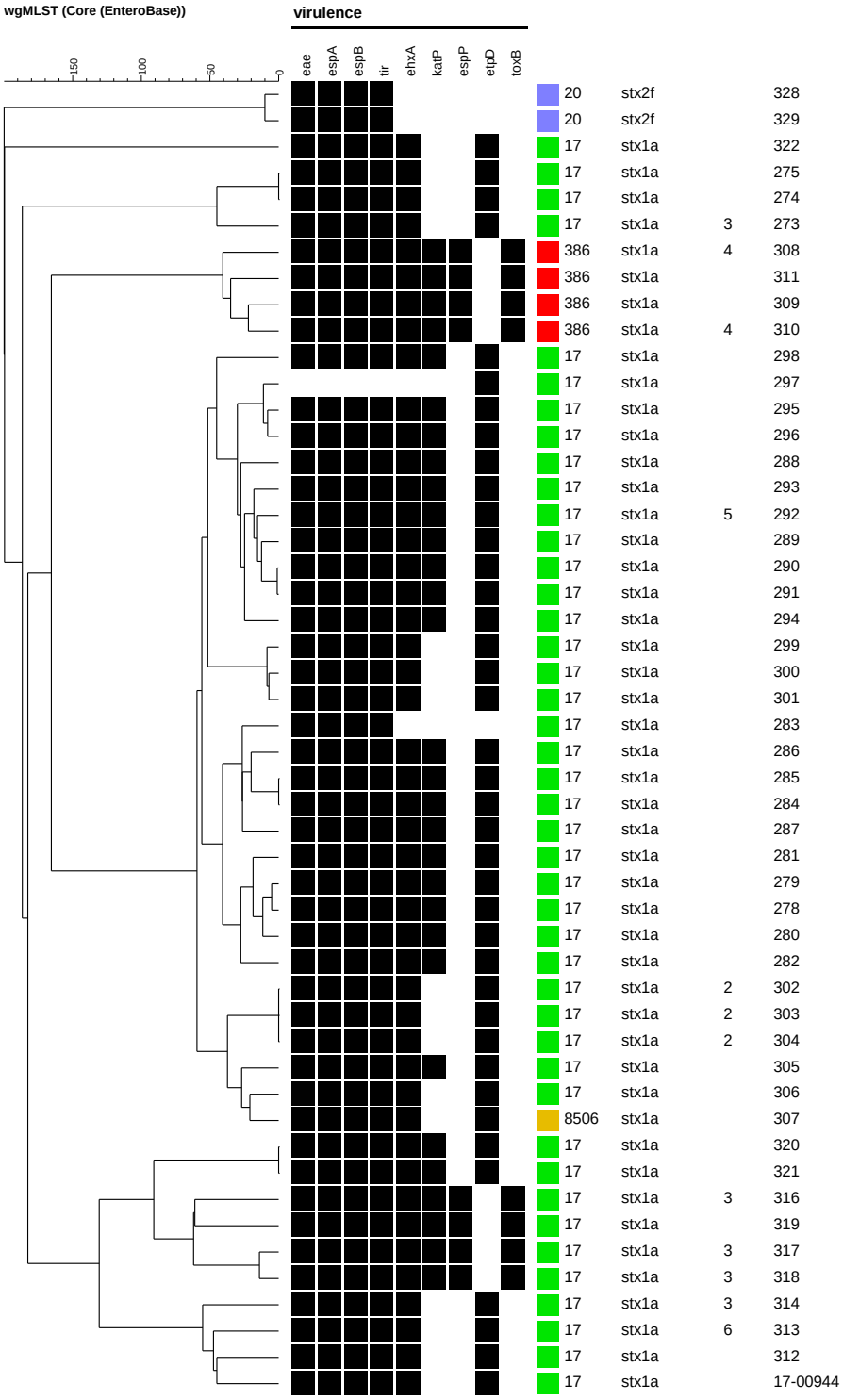

Figure 12. Dendogram based on the allelic profiles of 2513 cgMLST genes for the *E. coli* O103:H2 strains. The tree was created in BioNumerics v7.6 with the categorical (differences) coefficient of similarity and UPGMA cluster analysis method. To observe the genetic diversity among closely related strains no scaling factor was used and the maximum number of loci differences allowed was 200. Selected virulence genes are shown and the sequence types (STs) are highlighted in different colours. The columns from left to right are: *stx* subtype, antibiogram length, and ID (the raw milk outbreak strain = 17-00944).

4.5.3 *E. coli* O145:H28

E. coli O145:H28 was the third most commonly isolated serotype (n=45/522; 8.6%). *In silico* 7-gene MLST analysis identified four different STs (belonging to CC32): ST32 (n=42; 93.3%), three (6.7%) ST6130, one (2.2%) ST137 and one (2.2%) ST8625. All strains carried *eae*. The majority harboured *stx*2a (n=41, 91.1%), three (6.7%) were *stx*2d positive, and one was *stx*1a positive. A total of 8 (17.8%) isolates were shown to carry acquired AMR genes. All eight were predicted to confer resistance to aminoglycosides, β-lactams and sulphonamides (therefore MDR), and one also carried genes linked to tetracycline and trimethoprim resistance (Appendix 3).

This serotype has been associated with large foodborne outbreaks, including a large multistate outbreak in 2010 in the US associated with bagged romaine lettuce [39] and an outbreak in 2007 in Belgium linked to ice cream [96]. Cooper *et al.* reported the genome sequences of two strains [RM13514 (lettuce) and RM13516 (ice cream)] from the outbreaks and we have compared them with the SERL isolates. As shown in Figure 12, RM13516 clustered with the ST6130 isolates, while RM13514 clustered with the ST32 isolates. The SERL ST32 isolates were divided into two arbitrarily assigned clusters based on cgMLST, which we have designated ST32C1 and ST32C2. There were some notable differences between these clusters with respect to virulence and antibiotic resistance gene profiles. The majority of the ST32C1 isolates had the pVF gene profile *ehxA+/katp+/espP+/etpD-* and none harboured AMR genes. By contrast, the majority of ST32C2 were characterised by *ehxA+/katp-/espP+/etpD-* and eight of the strains carried AMR genes. We did not assign RM13514 to a named cluster, however this strain was *ehxA+*/katp-/*espP+*/*etpD-* and carried genes linked to resistance to several antibiotics including sulphonamides, streptomycin, tetracycline and chloramphenicol (due to the presence of plasmid pRM13514; Cooper *et al*. [96]), similar to the SERL ST32C2 isolates. The ST137 and ST6130 isolates shared the pVF gene profile *ehxA+/katp-/espP-/etpD+*, and did not carry *toxB*. All isolates carried the LEE associated genes *eae*, *espA*, *espB and tir*, and like the majority of the SERL O145:H28 isolates, the two outbreak strains were *stx*2a positive (Figure 13).

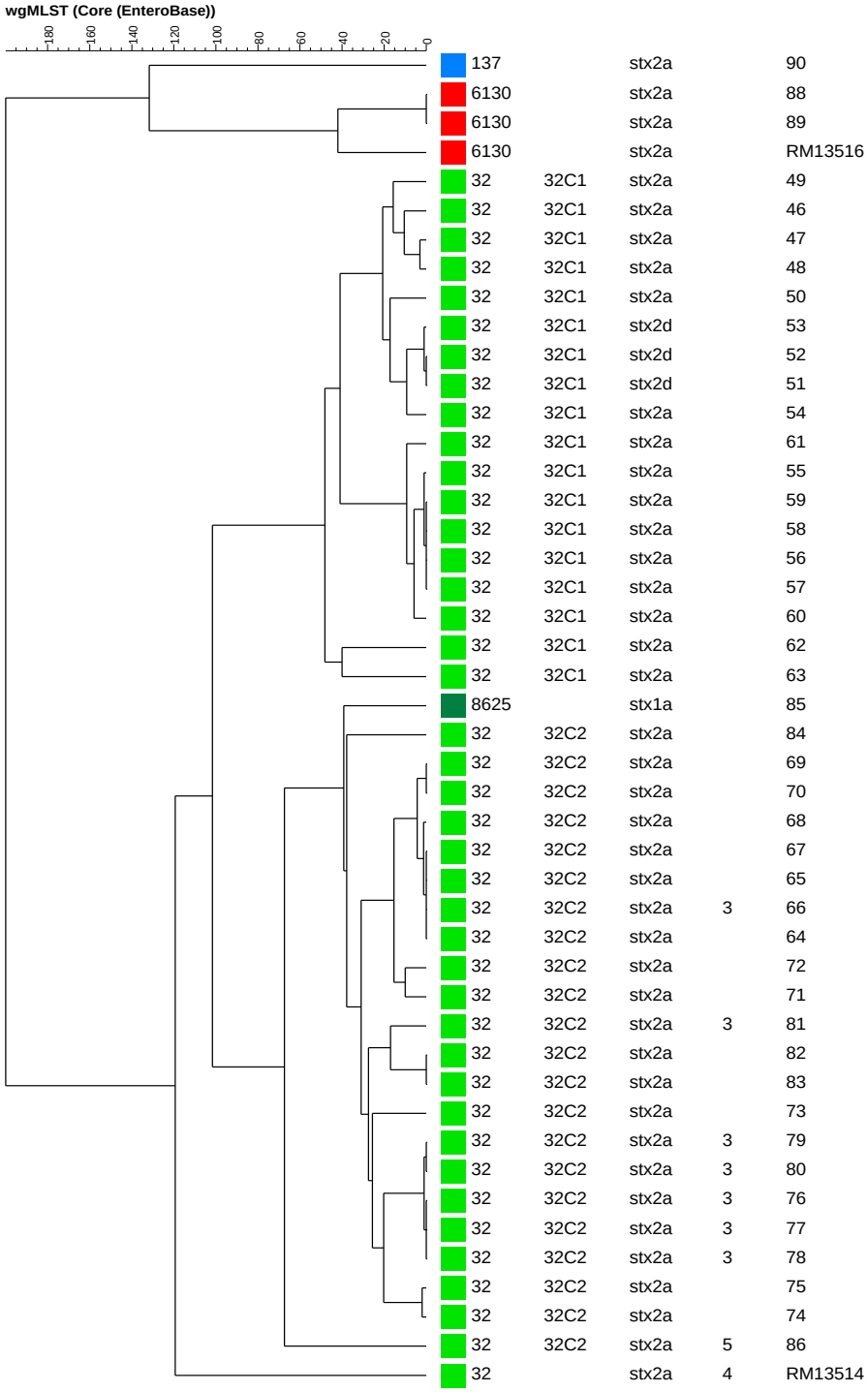

Figure 13. Dendogram based on the allelic profiles of 2513 cgMLST genes for the *E. coli* O145:H28 strains. The tree was created in BioNumerics v7.6 with the categorical (differences) coefficient of similarity (no scaling) and UPGMA cluster analysis method. Selected virulence genes are shown and the sequence types (STs) are highlighted in different colours. The columns from left to right are: cluster (ST32C1&C2), *stx* subtype, antibiogram length, and ID (representative strains from the romaine lettuce and ice cream outbreaks are RM13514 and RM13516 respectively).

4.5.4 *E. coli* O104:H4 - Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC)

Two *E. coli* O104:H4, isolated in 2011 and 2016 (No.s 431 and 432), were detected during the study period. We compared them with strain 2011C_3493 from the large German outbreak in 2011 and found both SERL strains belonged to ST678 and carried genes characteristic of both STEC (e.g. *stx2a*, *iha*, *ipf)* and EAEC (e.g. *aggR*, *pic*, *sigA*, *aap, aatA, aaiC*) [11,97–99]. Notably they did not contain the genes associated with the LEE, including *eae*. Isolate 431 had the same virulence and resistance gene profile as the outbreak strain, however, some differences were observed with isolate 432, including the absence of the SPATE protease *sepA* and the presence of the aggregative adherence fimbriae (AAF) variant III (carrying genes *agg3A*-D) rather than the AFF variant I.

There were also some differences in resistance genes; most notably 432 did not carry the extended spectrum beta-lactamase (ESBL) gene *bla*$_{\text{CTX-M-15}}$. Rasko *et al*., 2011 previously reported the plasmid carrying this gene was not identified in most of the O104:H4 genomes sequenced, suggesting that this plasmid might only be acquired by some strains or, alternatively, might be relatively unstable and consequently lost by many strains. cgMLST analysis showed 431 differed by only two alleles to the oubreak strain, whilst 432 differed at 93 loci suggesting it belonged to a different sublineage of ST678.

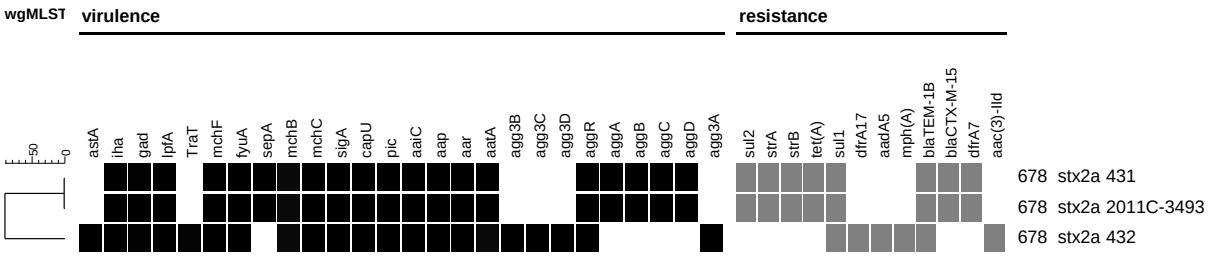

Figure 14. Dendogram based on the allelic profiles of 2513 cgMLST genes for the *E. coli* O104:H4 strains. The tree was created in BioNumerics v7.6 with the categorical (differences) coefficient of similarity (no scaling) and UPGMA cluster analysis method. Virulence and antibiotic resistance genes are shown. The columns from right to left are: ST, *stx* subtype and ID.

4.5.5 *E coli* O80:H2 - Hybrid EHEC and ExPEC Virulence genes

We isolated three *E. coli* O80:H2 over the study period (in 2013 and 2014). The strains belonged to ST301 and one carried *stx*2a, while the other two were positive for *stx*2d. Similar to the ST80 strains described by Soysal *et al.* [100], a rare variant of the intimin gene (*eae*-ξ) was detected among the strains, and genetic determinants related to the pS88 plasmid associated with extraintestinal-virulence pathogenic *E. coli* (ExPEC) were detected, including *iss*, *hlyF*, *iroN*, *ompT* (Figure 15). Soysal *et al.* [100] reported the minimal combination of plasmidic genes

common to all hybrid strains was the association of *ompT* and *hlyF*, which might represent a beneficial influence on the intestinal pathogenic virulence of *E. coli* O80:H2. The strains also possessed the genes carried on the LEE and the pVF gene profile *ehxA*+/*katp*-/*espP*+/*etpD*-. Furthermore, consistent with other O80:H2 strains described in the literature, the SERL isolates were resistant to multiple classes of antibiotics including aminoglycosides, β-lactams, sulphonimides and tetracyclines.

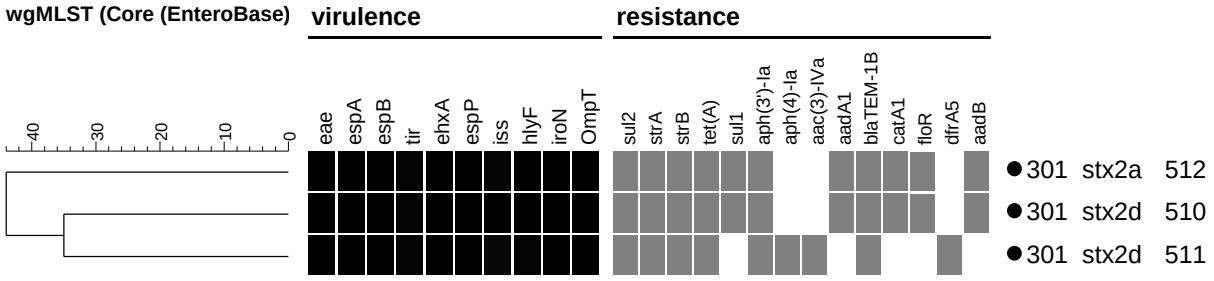

Figure 15. Dendogram based on the allelic profiles of 2513 cgMLST genes for the *E. coli* O80:H2 strains. The tree was created in BioNumerics v7.6 with the categorical (differences) coefficient of similarity (no scaling) and UPGMA cluster analysis method. Selected virulence genes and antibiotic resistance genes are shown. The columns from left to right are: ST, *stx* subtype and ID.

4.5.6 STEC O91

There have been many reports of serogroup O91 in food, including raw milk cheeses in France [101] and Switzerland [102], and retail meats in South Korea [103]. Werber *et al.* [104] reported that O91 was the second most common serogroup isolated from food in Germany. In our study, O91:H14 was the sixth most common serotype isolated from human clinical samples, while O91:H10 was less common with only three isolates detected. *E. coli* O91:H21 has also been associated with human disease [105] but this serotype was not isolated from a clinical case in Scotland. The O91:H14 isolates (n=16) belonged to ST33 and were associated with the *stx* subtype profiles: *stx*2b *stx*1a (n=13), *stx*2b only (n=2) and *stx*1a (n=1). Two of the strains carried streptomycin resistance genes, *strA* and/or *strB*.

The O91:H10 strains belonged to ST641 or ST8649 (a single locus variant of ST641). They had different *stx* subtype profiles: *stx*2b *stx*2c, *stx*2d and *stx*2a. No AMR genes were detected among the strains.

Virulence gene detection showed O91:H10 strains carried fewer genes associated with virulence compared with O91:H14 strains (Figure 16). None of the strains were positive for the LEE genes, including *eae*, however they did carry other adhesins; all strains carried the adhesin *lpfA* (long polar fimbriae) and 15 (78.9%) carried *iha* (*Irg*A homolog adhesin). Eight of the O91:H14 strains were positive for the pVF genes *ehxA* and/or *kapP*. Other genes that were variably detected, and less commonly found in *eae*+ve STEC, included *mchF*, *ireA*, *celb*, *subA*, *mchB*, *mchC*, *mcmA*, *senB*.

A previous study investigating the virulence genes and genetic diversity of STEC O91 strains from cattle, beef and poultry products also found a number of adhesins among the strains with the potential to bind to host cells, including *ipfA* and *ehxA* [106].

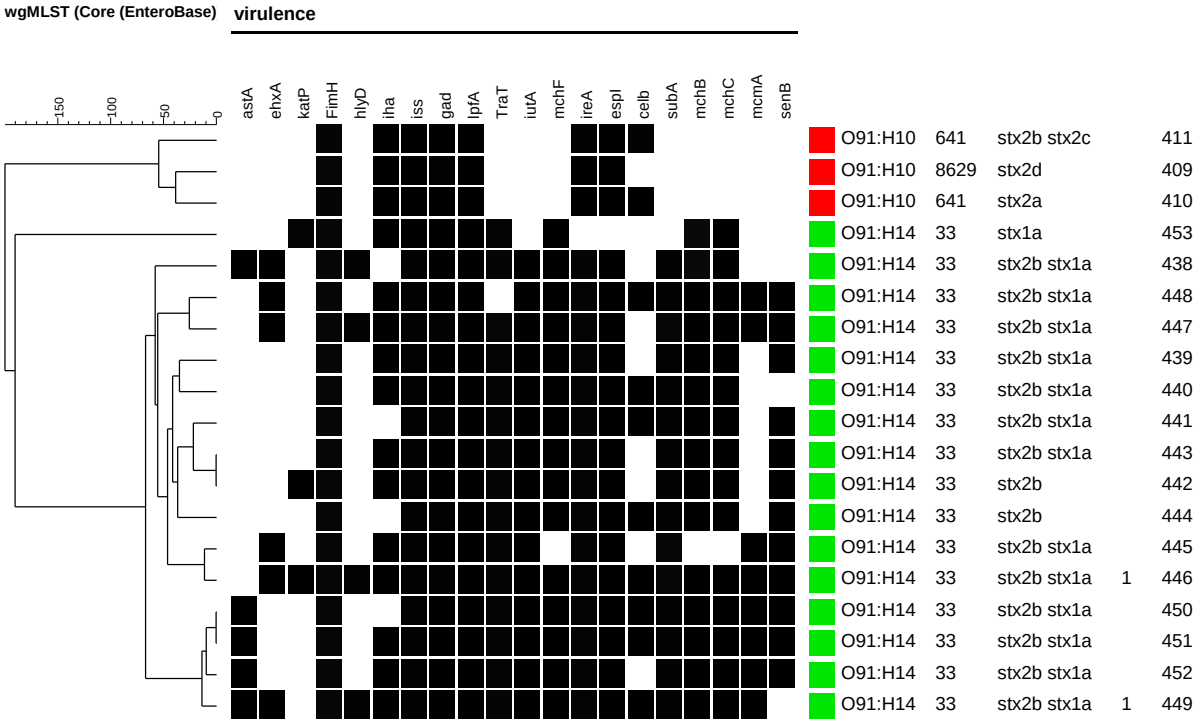

Figure 16. Dendogram based on the allelic profiles of 2513 cgMLST genes for the *E. coli* O91 strains. The tree was created in BioNumerics v7.6 with the categorical (differences) coefficient of similarity (scaling factor of 3) and UPGMA cluster analysis method. Selected virulence genes are shown and the serotypes are in different colours. The columns from left to right are sequence type, *stx* subtype, antibiogram length and ID.

4.5.7 *E. coli* encoding Shiga toxin 2f

In our collection, 40/522 (7.7%) *E. coli* isolates were positive for *stx*2f. The first Scottish isolate carrying this *stx* subtype was detected in 2013. The isolates belonged to 13 different serotypes and the most common were O63:H6 (n=12, 30.0%), O125ac:H6 (n=8, 20.0%) and O145:H34 (n=6, 15.0%), similar to a recent report from the Netherlands [107].

As shown in Figure 17, O63:H6 and O125ac:H6 shared the same ST and were closely related by cgMLST. The majority of the *stx*2f strains carried the pathogenicity island LEE genes *eae*, *espA* and *tir*, however only the ST20 strains carried *espB.* Notably the strains did not carry the pVF genes (*ehxA-/katP-/espP-/etpD-*). Only one strain carried AMR genes to aminoglycosides, sulphonamide, tetracycline and

trimethoprim (Appendix 3). In keeping with other reports, *stx*2f were not found in combination with any other *stx* genes.

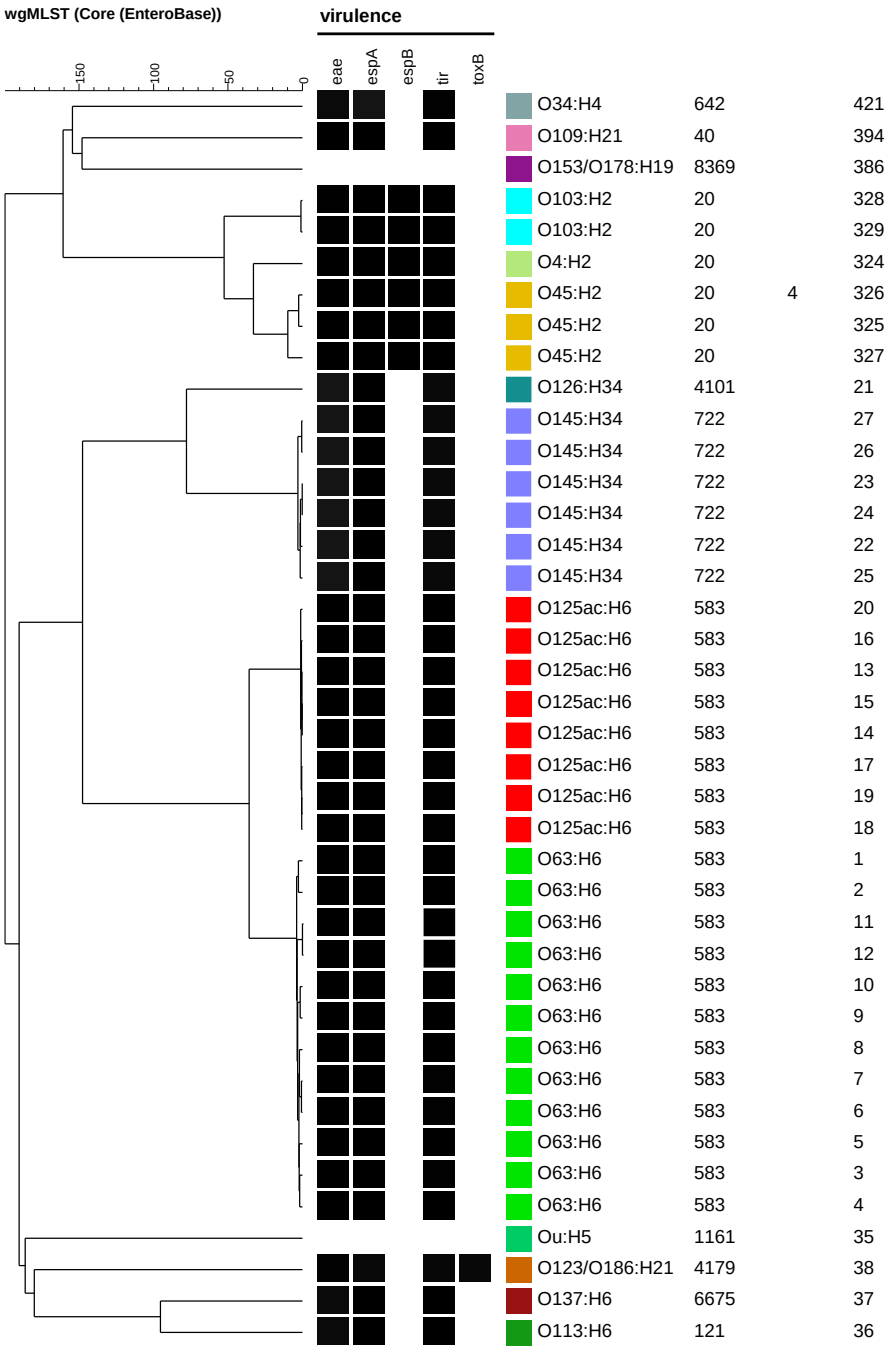

Figure 17. Dendogram based on the allelic profiles of 2513 cgMLST genes for the *stx*2f positive strains. The tree was created in BioNumerics v7.6 with the categorical (differences) coefficient of similarity (scaling factor of 10) and UPGMA cluster analysis method. Selected virulence genes are shown and serotypes are highlighted in different colours. The columns from left to right are sequence type, antibiogram length and ID.

5. Discussion

Most studies on STEC have focussed on the epidemiology and virulence of *E. coli* O157 but the importance of non-O157 STEC is now recognised, likely in part due to advances in methodology to detect these organisms and an increasing awareness of the serious infections they can cause. Consequently, in a number of countries with active surveillance systems whose frontline diagnostic laboratories are using new technology to detect all STEC, non-O157 STEC are more prevalent than *E. coli* O157. In the Republic of Ireland, for example, *E. coli* O26 is now the dominant serotype detected with numbers exceeding that of *E. coli* O157 since 2013. Although Scottish diagnostic laboratories are not yet using this new technology, the SERL has been screening faeces from high risk patient groups for STEC, focussing on the detection of *stx* genes (all variants) by PCR and since 2014, laboratories in all Health Board Areas have been sending high-risk faeces to SERL for screening.

The purpose of this study was to generate and analyse WGS data on all STEC in the Scottish Culture Collection to provide a greater understanding of the molecular epidemiology of non-O157 STEC infection in Scotland. We have also compared the Scottish data with known predominant, emerging and hybrid strains to provide an international context.

A total of 525 strains were analysed during this study (517 non-O157 STEC, 5 *E. coli* strains subsequently discovered to not carry *stx* genes and 3 strains subsequently identified as *Escherichia albertii*). The first non-O157 STEC strain (an *E. coli* O26:H11) was isolated in 2002. Although 88 different serotypes were observed (where both O and H antigens were identified), 41 were observed on one occasion only. It is difficult to put this diversity in an international context as few laboratories have published full O:H serotype results for their non-O157 collections. In 2018, colleagues published data on STEC isolated in a particular region of Brussels over a 27 year period [108]. The strains were subjected to “O” typing using traditional methods but “H” typing was not performed on all strains. They identified 72 different “O” serogroups out of 606 STEC strains, so a diverse range of serogroups were observed. For comparison, 65 different “O” serogroups were identified in this study. In 2014, colleagues in England identified 138 non-O157 STEC where a full O:H serotype was available following WGS [8]. Thirty five different serotypes were identified with *E. coli* O146 and *E. coli* O26 being the most common.

Serotyping by WGS has demonstrated the limitations of serogroup assignment by agglutination with antisera, particularly the use of polyvalent antisera by diagnostic laboratories. In addition to significantly reducing the cost per test and workload, one of the main advantages of serotype assignment by WGS is that the majority of previous O:unidentifiable and O:rough strains will have a serotype designation. We

found a small number of isolates (2.9%) that were O-unidentifiable however these were most likely novel types.

A number of STEC co-infections were identified over the course of this study, as a consequence of the screening and isolation method used to identify and isolate STEC. Multiple colonies from a culture plate are chosen for PCR in an attempt to find a colony with a virulence profile that matches that detected in the enrichment broth. However, occasionally a colony with a different virulence profile or different colony morphology is detected while screening multiple colonies. This is likely to be an under-representation of the number of STEC co-infections occurring in clinical cases but it would not be feasible or cost effective to screen multiple colonies from every case of infection. Establishing STEC co-infections would not alter the clinical management of the patient but it may be of importance during outbreaks, especially when establishing the case definition.

In 2017, the most frequent serogroups reported in confirmed cases of human STEC infection in the EU/EEA were O157 (31.9%), O26 (14.3%), O103 (6%), and O91 (4.4%). Serogroup O157 was the most frequently reported cause of HUS replacing O26, which had become the most common cause of HUS for the first time in 2016 [54]. In 2016, the top STEC serogroups associated with human STEC in the United States were O157 (42.7%), O26 (16%), O103 (15.6%), O111 (10.2%) and O121 (4.7%) [2]. Similar to other countries worldwide, by far the most common non-O157 STEC serotype detected in Scotland was *E. coli* O26:H11. *E. coli* O26 was first recognised as a cause of infantile diarrhoea in 1951 [109]. In 1977, five years before the identification of *E. coli* O157 as a pathogen, production of Shiga toxin was identified in *E. coli* O26 strains isolated from infants with diarrhoea [110]**.**

Since 2014, when all Scottish laboratories started sending faeces to SERL for screening, *E. coli* O26 has featured in the top 3 serotypes detected at SERL (Table 9).

Table 9: Top 3 serotypes detected at the SERL 2015-17

Most Common Serotype	2015	2016	2017
1	**O26:H11** n=20	**O26:H11** n=11	O145:H28 n=12
2	O146:H21 n=6	O63:H6 O128ab:H2 n=5	O103:H2 n=11
3	O145:H28 O128:H28 n=5	O125:H6 n=4	**O26:H11** n=8

E. coli O26 is also the most common non-O157 STEC detected in food and animal samples tested in Europe [54]. It has been associated with outbreaks and severe disease (10 deaths due to STEC infection were reported in the EU in 2017) and is known to survive well in bovine faeces and withstand substantial periods of stress [111].

Typing of *E. coli* O26, by various methods has revealed great genetic heterogeneity within the group [67,79]. Surveillance of *E. coli* O26 in Europe in the 1990s established a shift in *stx* subtype from *stx*1 to *stx*1 & *stx*2 to *stx*2 [79] and this has also been observed more recently with the Scottish strains (Figure 7). Phylogenetic analysis of *E. coli* O26:H11 has demonstrated the occurrence of a number of different lineages with the identification of a highly pathogenic European clone [80], which is now thought to be disseminated throughout Europe and has emerged in American and Asia [95]. More recently, a new “French” clone [95] has been described and a “US” clone [112], however, with the exception of two strains characterised as the new European clone (ST298C2), the majority of the Scottish strains belonged to EHEC O26:H11 lineage ST21C1 [80,94,95]. We did however detect a sublineage of ST21C1 (termed here ST21C1a) that was associated with a high prevalence of *stx2a*. In the next phase of the study, it will be interesting to determine whether this sublineage is linked with more severe disease.

An FSA project conducted in 2004 [113] concluded that Scottish cattle are a potential reservoir and a source for human infection of *E. coli* O26. The reasons why *E. coli* O26 may be such a successful human pathogen likely include the fact they are very dynamic, they frequently lose and acquire mobile genetic elements, for example they have acquired the *stx*2a gene, which is associated with increased pathogenic potential. *E. coli* O26 also appears to demonstrate “supershedder” behaviour when excreted from cattle [114,115].

The second predominant serotype among Scottish STEC was O103:H2. As previously mentioned, O103:H2 is also a common cause of human infection in many other countries. In 2011 it was declared one of the top six non-O157 STEC adulterants in non-intact raw beef by the US Food Safety and Inspection Service (FSIS), along with O26, O11, O121, O45 and O145 [116]. More recently, O103:H2 has been detected in flour [117] and raw milk [29]. Unlike O26:H11, we have not seen *stx*2a associated with this serotype, with the vast majority carrying *stx*1a. Other studies characterising isolates of O103:H2 in sheep and cattle have highlighted the absence of *stx*2a in this serotype [118,119]. By contrast, the rare serotype O103:H25, has been shown to carry *stx*2a and was responsible for a severe outbreak of HUS in Norway in 2006 [43].

Multiple virulence genes were identified among the *E. coli* serotypes These genes are known to be involved in various processes important for pathogenesis, including

acid resistance, bacterial adhesion, translocation of bacterial effector proteins, host invasion, iron acquisition and regulation of stress factors [82,120]. The pathogenesis of STEC is clearly complex, initially involving the attachment of STEC to intestinal cells. The most important adhesion protein among STEC is intimin, encoded by *eae*. In this study, 336/522 (64.4%) non-O157 STEC isolates possessed the *eae* gene. However STEC not possessing this gene are still capable of causing severe disease, including HUS [11,81] and therefore do still present a public health risk. We compared the virulence gene content of strains carrying *eae* with those not carrying *eae* and showed a greater number and distribution of genes were detected in *eae*+ve strains. Several adhesins were detected among the *eae*-ve strains, however most were not widely distributed (e.g. *saa* and *aggR*), except *fimH*, *lpfA* and *iha*, and these are also associated with *eae*+ve strains. Additional mechanisms for adhesion are still being elucidated.

Montero et al. [121] described the Locus of Adhesion and Autoaggregation (LAA), which is present in emerging serotypes associated with severe disease and may provide an alternative mechanism of adherence of STEC to intestinal cells. Following attachment and colonisation, many secretary proteins and toxins are released and these are important for survival and multiplication in the gut. In this study, we found the toxin *subA* was unique and widely distributed among *eae*-ve serotypes, which has been reported previously. Paton *et al.* [122] described this toxin in O113:H21 strains associated with a small outbreak of HUS in Australia. Further work has shown that this toxin is lethal for mice and induces pathological features similar to those seen in HUS [123]. Therefore possession of this gene may well increase the likelihood of life-threatening complications such as HUS in patients. The toxins responsible for the most severe damage in the intestine and to other organs, are the Shiga toxins. In this study we detected the most commonly recognised Shiga toxin subtypes (1a, 1c, 1d, 2a-g), with *stx*1a predominating. However certain subtypes have been associated with more severe disease and as a result, in combination with the adherence proteins, *eae* and *aggR*, the Shiga toxin subtypes have been used to assign STEC into different JEMRA levels based on their potential to cause disease.

Using guidance provided in the JEMRA report (2018), we estimated the potential to cause diarrhoea, bloody diarrhoea and haemolytic uraemic syndrome, based on the combinations of STEC virulence genes present in Scottish non-O157 STEC strains (Table 8). For comparison, we also assigned JEMRA categories to the strains of *E. coli* O157:H7 we have already sequenced (n=304) (Table 10). Based on current knowledge, it is generally accepted that STEC producing *stx*2a, in combination with *eae* or *agg*R, are most consistently associated with severe disease including HUS (JEMRA, 2018). In this study, 128 (24.8%) non-O157 STEC strains fell into this “Level 1” category, compared to 177 (58.2%) *E. coli* O157:H7 strains.

Table 10: JEMRA level assignment of non-O157 STEC compared with *E. coli* O157:H7, based on virulence gene combination and potential to cause diarrhoea (D), bloody diarrhoea (BD) and haemolytic uraemic syndrome (HUS).

JEMRA Level[1]	Trait - Subtype combination[2]	No. of non-O157 isolates[3]	No. of O157:H7 isolates	% non-O157 in each JEMRA level	% O157:H7 in each JEMRA level
1 (D/BD/HUS)	***stx*2a, *eae***	66	53	**24.8**	**58.2**
	stx*2a**, ***eae *(with stx1a)*	59	4		
	stx*2a**, ***eae *(with stx1a, stx2c)*	1	13		
	stx*2a**, ***eae *(with stx2c)*	0	107		
	stx2a, ***agg*R**	2	0		
2 (D/BD/HUS)	***stx*2d**	13	0	**2.7**	**0.0**
	***stx*2d** (with *stx2b*)	1	0		
3 (D/BD)	***stx*2c**, ***eae***	3	30	**1.0**	**40.1**
	stx*2c**, ***eae *(with stx1a)*	2	92		
4 (D/BD)	***stx*1a**, **eae**	161	5	**31.1**	**1.6**
5 (D)	**other *stx* subtypes**	209	0	**40.4**	**0.0**

[1]Potential clinical outcome in parenthesis
[2] Extra genes present in parenthesis
[3] This table does not include the five *stx* negative strains

Subtype *stx*2d may also be associated with severe disease, although factors other than the actual presence of the gene may affect disease outcome [1]. Consequently these were assigned “Level 2”. Although there has been one report of an *E. coli* O157:H7 strain carrying *stx*2d, this is most uncommon and has not been detected in Scottish *E. coli* O157:H7 strains. The *stx*2d subtype has, however, been detected in non-O157 STEC in this study (n=14; 5/14 with *eae*).

The majority of Scottish *E. coli* O157:H7 strains were assigned to Level 1 (n=177; 58.2%) or Level 3 (n=122; 40.1%) and no *E. coli* O157:H7 strains were assigned to Level 5 meaning all had the potential to cause at least bloody diarrhoea. The majority of Scottish non-O157 STEC strains (n=370; 71.6%) were assigned to Levels 4 and 5. Indeed, 40.4 % of non-O157 STEC strains were assigned to Level 5 with potential to cause diarrhoea, not bloody diarrhoea. This is an interesting finding as

the main criterion for sending faeces to SERL for further testing is bloody diarrhoea and a recent audit of faecal sample submission conducted at SERL suggests that this submission criterion is adhered to. Therefore, the patients infected with these strains are likely to have either had an infection sufficiently severe to warrant sending the faeces to SERL for screening or, in a smaller number of cases, will have been asymptomatic contacts of a known case. During the second phase of this study, we will be able to compare the JEMRA assigned disease potential (based on virulence gene combination) with each patient's actual symptoms to assess how accurate this assignment might be and whether there may be potential to use this or a similar risk-based approach when managing Scottish cases of non-O157 STEC infection.

Stx subtypes *stx*2e, *stx*2f and *stx*2g are normally associated with carriage of STEC in animals [124]. Subtypes *stx*2f and *stx*2g cause no ill effects in animals but STEC producing *stx*2e cause oedema disease in pigs [125]. One Belgian group of researchers have noted that there is such a high mortality rate in pigs infected with STEC producing *stx*2e that it has had a severe economic impact on the swine industry [126]. STEC carrying subtype *stx*2e have been previously isolated, albeit rarely, from patients with diarrhoea [127] and HUS [128]. We detected a total of three isolates (two *E. coli* O8:H19 and one O138:H48), which carried the *stx*2e subtype. O8:H19 is an STEC serotype known to be carried by pigs [129] and one clinical strain of *E. coli* O8:H19, carrying *stx*2e, was also identified in an English clinical case in 2014 [8].

E. coli encoding *stx*2f has been described as an emerging human pathogen [130]. Since it was first described in *E. coli* from pigeons in 2000 [131] there have been a number of reports of *stx*2f associated with human disease, including HUS [108,130,132,133]. This *stx* subtype has not been detected in *E. coli* O157.

The *stx*2f subtype was first detected in Scotland in 2013 and was carried by 7.7% (40/521) of strains in this study, in addition to three strains of *Escherichia albertii,* two of which additionally carried the *eae* gene. *E. albertii* possessing *stx*2f have only recently been reported and are thought to be associated with mild disease [134] although one clinical case of *E. albertii* harbouring the *stx*2a gene has been reported, associated with bloody diarrhoea [135].

E. albertii are frequently mischaracterised as *E. coli* because it is difficult to distinguish these strains from *E. coli* biochemically and these were misidentified as *E. coli* at SERL based on their biochemical API profile (using API 20E). They can however be detected by PCR using the following gene targets: *lys*P, *mdh* and *cdt*B (they should be positive for all these targets and negative for *uid*A). The number of *stx*2f strains detected each year can vary but has been as high as 20% (Table 11).

Table 11: % Scottish non-O157 STEC possessing *stx*2f subtype (2015-2017)

2015	2016	2017
15 % (11/73)	20% (12/59)	6.8% (4/59)

In a study conducted in the Netherlands, between 2008 and 2011, one quarter of all non-O157 STEC isolated possessed the *stx*2f gene. Approximately one half of the *stx*2f subtypes detected were carried by *E. coli* O63 strains [132]. Similarly, since 2012, 12% of all Belgian STEC strains were *stx*2f positive. Although *E. coli* O63:H6 was the most common Scottish serotype possessing the *stx*2f gene (28%) it was observed in an additional 14 different serotypes, including two strains of *E. coli* O103:H2 (which, based on submission dates and subsequent sequence analysis, are likely to be linked). This may the first report of *E. coli* O103:H2 carrying *stx*2f. The *stx*2f *E. coli* O103:H2 strains clustered quite separately from the other *stx*1a *E. coli* O103:H2 clinical strains. The majority (38/41; 92.7%) of non-O157 STEC *stx*2f producing strains also carried the *eae* gene. One strain (serotype O145:H34) also carried genes inferring resistance to four different classes of antibiotic. Although pigeons carry *E. coli* possessing *stx*2f, their involvement in zoonotic disease has not been proven and the source of *stx*2f human infection is unclear. Hoek *et a*l. [107] recently showed a lack of evidence for pigeons as the source of *stx*f2 carrying *E. coli* causing human infections.This *stx* subtype is often overlooked when screening for STEC as it specifically requires an additional primer set for PCR and, as far as we are aware, the majority of multiplex PCR platforms for gastrointestinal pathogens (such as EntericBio and BDMAX) do not include this as a target. Consequently, the true incidence of this in routine surveillance and research studies is likely to be underestimated or completely missed.

A recent publication [136] described the isolation of an EPEC/STEC hybrid *stx*2f positive *E. coli* O137:H6 from exotic companion birds. The single Scottish *E. coli* O137:H6 strain shared many similarities with these strains: *stx*2f, *eae*, *Fim*H, *ibe*A although it did not possess the *bfp*A gene and was a different sequence type. However, this further highlights the zoonotic potential of birds and specifically companion birds as a public health risk.

The variant *stx*2g was first described in the faeces of healthy cattle [137] and has subsequently been detected in humans, food and the environment [130]. Its role in human pathogenicity has not yet been determined but carriage in humans has been associated with diarrhoea, fever and abdominal pain [138] including one strain in England [8].

The prevalence of *stx*2g-harbouring STEC in this Scottish collection (0.7%) matches that of Germany (0.6% [138]) and is in keeping with the low incidence reported in

Denmark [139]. We identified four STEC strains carrying *stx*2g (serotypes O15:H16, O30:H25, O187:H28 and O149:H8 (also carrying the *stx*2a variant)). *E. coli* O15:H16 and *E. coli* O187:H28 carrying the *stx*2g gene have been described previously [8,140].

Interestingly, three of these strains (including the *E. coli* O15:H16 strain) also carried the ST1a, a heat-stable enterotoxin gene usually associated with enterotoxigenic *E. coli* (ETEC) and carried on a plasmid. STEC/ETEC hybrid strains, including serotype O15:H16, have been isolated from many sources including humans, animals, food and water [138,141], and some have been associated with diarrhoeal disease and HUS (although the strain causing HUS did possess the *stx*2a gene [141]). A total of five Scottish strains in this study were STEC/ETEC hybrid strains which is a similar proportion to that detected in clinical cases in Finland [141]. An increasing number of "hybrid" strains associated with clinical infection strains are now being reported. For example, the *E. coli* O104 strain was an STEC/enteroaggregative *E. coli* (STEC/EAEC) hybrid strain that caused an outbreak with a high percentage of HUS cases.

At the 10th International Symposium on Shiga Toxin Producing *Escherichia coli* Infections meeting in May 2018, a number of emerging STEC strains were discussed. One of these was *E. coli* O145:H28, which is an emerging pathogen in the Republic of Ireland (RoI) and was responsible for 9% of all their HUS cases in 2017. The majority of these strains were *E. coli* O145:H28 possessing the *stx*2a subtype and ST32. Similar to RoI, the first laboratory case of *E. coli* O145:H28 in Scotland was in 2004. In total, we detected 40 *E. coli* O145:H28, *stx*2a, ST32 strains in this study and this was the most common serotype detected in 2017, due to the occurrence of a number of small epidemiologically unrelated outbreaks (Table 12).

Table 12: Incidence of *E. coli* O145:H28 (2015-2017)

2015	2016	2017
6.8% (5/73)	1.7% (1/59)	15.3% (9/59)

WGS was pivotal in being able to compare with strains in England (where a small number of cases were also detected) and in distinguishing between these incidents which occurred within a similar timeframe but, despite a UK wide Problem Assessment Group (PAG) and investigation, it was not possible to determine a source of infection.

Another emerging serotype is *E. coli* O80:H2, originally reported by France [100] and elsewhere [142,143], and associated with a high rate of HUS. These strains tend to be positive for *stx*2a or *stx*2d and possess a rare intimin gene (*eae*-ξ), and *ehx*A.

Another recorded trait of this serotype is they demonstrate resistance to a number of antibiotic classes. This also appears to be a hybrid strain, possessing genes associated with the extraintestinal pathogenic *E. coli* (ExPEC) pS88 plasmid and an ability to produce invasive infection. A retrospective study of Belgian human and cattle strains showed this pathotype was present in cattle since 1987 and in humans since 2008 but was only been associated with mild disease. A 10 year retrospective study in the USA noted that this pathotype caused sporadic infection and one outbreak but there was no association with HUS. We detected three *E. coli* O80:H2 strains, two in 2013 (both *stx*2d) and one in 2014 (*stx*2a) carrying a similar virulence profile to the Belgian strains and multiple AMR genes.

E. coli O55:H7, carrying *stx*2a, emerged in England in 2014 and has caused recurrent seasonal outbreaks each year [144]. This strain appears to be associated with a high incidence of HUS, however it has not yet been isolated in Scotland. Colleagues in England also recently reported an outbreak of the rare serotype *E. coli* O117:H7 [145], among men who have sex with men. Although we detected three strains of *E. coli* O117:H7 over the study period, the timeframe of isolation suggests these cases were not linked. Since August 2017, all Scottish *E. coli* O157 and non-O157 STEC are routinely sequenced and sequences compared with those identified in England & Wales on a weekly basis so cross border incidents and outbreaks are rapidly identified and this has been one of the main benefits of introducing WGS. Previously we had to rely on the transportation of Hazard Group 3 strains which was time-consuming and could be costly.

Antimicrobial resistance (AMR) in STEC is used as a surveillance tool to monitor trends and emerging resistance rather than a guide to treatment, as antibiotics are contraindicated in the treatment of STEC patients. AMR may also provide evidence for the transmission of resistance from animals to humans [146]. We have previously demonstrated that phenotypic antimicrobial susceptibility testing (using the disk diffusion assay) correlated well with WGS predicted antimicrobial sensitivity and this correlates with what others have reported [146]. This study demonstrated that 82.6% of non-O157 STEC strains in this study lacked identifiable resistance genes.

Of all non-O157 STEC in this study, 17.6% were resistant to at least one class of antibiotic and 12.1% would be classed as MDR as they demonstrated resistance to three or more antibiotic classes. Although we do not have the corresponding predicted resistance patterns for *E. coli* O157 strains over the same time-period (as we only commenced routine WGS in August 2017), of the 319 *E. coli* O157:H7 strains that have been sequenced at the SERL to date, 44 (13.8%) were resistant to at least one class of antibiotic and 12.3% of all O157:H7 strains were resistant to more than 3 antibiotic classes.

This comprehensive collection of non-O157 STEC sequences from a single country is a valuable and unique resource, and will be further enhanced by linking the data to the clinical and epidemiological information for each patient. Health Protection Scotland (HPS) undertakes enhanced surveillance of all cases of STEC infection and collates information about potential exposures, whether each case is part of an outbreak or sporadic or whether it is a primary or secondary case. Foreign travel in the 14 days prior to onset of illness will also be assessed to identify potentially imported strains. Information is also collected on clinical presentation (e.g. bloody diarrhoea, HUS, hospitalisation). Other organisms can also cause bloody diarrhoea and similar gastrointestinal symptoms so it is not always straightforward to definitively attribute infection with the isolated STEC strain to the episode of infection. Therefore, where possible, HPS will look at co-infection with other pathogens and underlying gut pathologies for each patient, such as ulcerative colitis or Crohn’s disease.

6. Overall Conclusions

We have carried out WGS on the archived Scottish clinical culture collection of non-O157 STEC and now have a unique and comprehensive genomic database of DNA sequences for interrogation, which will be a valuable resource for surveillance, particularly during future outbreaks of infection.

Although two of the most common serotypes identified in this study (*E. coli* O26:H11 and *E. coli* O103:H2) are also two of the most common non-O157 STEC identified worldwide, we also identified a diverse range of serotypes which we were able to comprehensively characterise to a level not previously possible. Reassuringly we did not identify some of the pathogenic strains (such as O55:H7 and certain pathogenic O26:H11 clones) currently emerging in certain countries although we have identified some strains that we will continue to closely monitor, including *E. coli* O145:H28.

We were also able to perform a molecular risk assessment based on the virulence and adhesion genes carried by each strain and assign levels of risk of each strain to cause disease, based on a recently published risk assessment from the Food and Agriculture Organization of the United Nations (FAO) and the World Health Organisation (WHO) (JEMRA). The majority (59.5%) of Scottish non-O157 STEC strains were assigned to Levels 1-4, meaning potential to cause at least bloody diarrhoea, with 40.5 % of non-O157 STEC strains assigned to Level 5 with potential to cause diarrhoea.

The second phase of this project will compare the predicted disease potential (the JEMRA Level assigned, based on virulence gene combination) with the actual symptoms experienced by each patient, to determine whether this molecular risk assessment approach might support decisions on public health interventions for STEC infection in the future.

7. **References**

[1] Joint FAO/WHO expert meetings on microbiological risk assessment (JEMRA). Shiga toxin-producing *Escherichia coli* (STEC) and food: attribution, characterisation and monitoring 2018.

[2] Centers for Disease Control and Prevention (CDC). National Shiga toxin-producing *Escherichia coli* (STEC) Surveillance Annual Report, 2016. Atlanta, Georgia. 2018.

[3] Kuehne A, Bouwknwgt M, Havelaar A, Gilsdorf A, Hoyer P, Stark K, et al. Estimating true incidence of O157 and non-O157 Shiga toxin-producing *Escherichia coli* illness in Germany based on notification data of haemolytic uraemic syndrome. Epidemiol Infect 2016;144:3305–15. doi:10.1017/S0950268816001436.

[4] 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;64:371–6. doi:10.1093/cid/ciw708.

[5] Government of Canada. National Enteric Surveillance Program (NESP): Annual summary 2016. Public Health Agency Canada 2018.

[6] Terajima J, Izumiya HY, Hara-Kudo Y, Ohnishi M. Shiga Toxin (Verotoxin)-producing *Escherichia coli* and Foodborne Disease: A Review 2017;5:35–53.

[7] National Services Scotland. STEC in Scotland 2017:Enhanced Surveillance and Reference Laboratory Data. 2017.

[8] Chattaway MA, Dallman TJ, Gentle A, Wright MJ, Long SE, Ashton PM, et al. Whole Genome Sequencing for Public Health Surveillance of Shiga Toxin-Producing *Escherichia coli* Other than Serogroup O157. Front Microbiol 2016;7:258. doi:10.3389/fmicb.2016.00258.

[9] Frankel G, Phillips AD, Rosenshine I, Dougan G, Kaper JB, Knutton S. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. Mol Microbiol 1998;30:911–21.

[10] Ito K, Iida M, Yamazaki M, Moriya K, Moroishi S, Yatsuyanagi J, et al. Intimin types determined by heteroduplex mobility assay of intimin gene (eae)-positive *Escherichia coli* strains. 2007;45. doi:10.1128/JCM.01103-06.

[11] Bielaszewska M, Mellmann A, Zhang W, Köck R, Fruth A, Bauwens A, et al. Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: A microbiological study. Lancet Infect Dis 2011;11:671–6. doi:10.1016/S1473-3099(11)70165-7.

[12] Scheutz F, Teel LD, Beutin L, Pierard D, Buvens G, Karch H, et al. Multicenter Evaluation of a Sequence-Based Protocol for Subtyping Shiga Toxins and Standardizing Stx Nomenclature. J Clin Microbiol 2012;50:2951–63. doi:10.1128/JCM.00860-12.

[13] Newton HJ, Sloan J, Bulach DM, Seemann T, Allison CC, Tauschek M, et al. Shiga toxin-producing *Escherichia coli* strains negative for locus of enterocyte effacement. Emerg Infect Dis 2009;15:372–80. doi:10.3201/eid1503.080631.

[14] Scheutz F, Strockbine N. Genus I. Escherichia. The proteobacteria Part B the gammaproteobacteria. In: G.M. Garrity, D. J. Brenner, N. R. Krieg and JTS, editor. Bergey's Man. Syst. Bacteriol., New York, NY: Springer; 2005, p. 607–624.

[15] Valilis E, Ramsey A, Sidiq S, DuPont HL. Non-O157 Shiga toxin-producing *Escherichia coli*-A poorly appreciated enteric pathogen: Systematic review. Int

J Infect Dis 2018;76:82–7. doi:10.1016/j.ijid.2018.09.002.
[16] Karmali MA, Mascarenhas M, Shen S, Ziebell K, Johnson S, Reid-Smith R, et al. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J Clin Microbiol 2003;41:4930–40.
[17] European Food Safety Authority. Scientific Opinion on VTEC-seropathotype and scientific criteria regarding pathogenicity assessment. EFSA J 2013;11:3138.
[18] Scheutz F. Experiences with non-O157 STEC and implications on public health programs. Public Meet. Arlington, VA, 17 Oct., 2007.
[19] European Food Safety Authority. Scientific Opinion of the Panel on Biological Hazards (BIOHAZ) - Monitoring of verotoxigenic *Escherichia coli* (VTEC) and identification of human pathogenic VTEC types. EFSA J 2007.
[20] Locking ME, O'Brien SJ, Reilly WJ, Wright EM, Campbell DM, Coia JE, et al. Risk factors for sporadic cases of *Escherichia coli* O157 infection: the importance of contact with animal excreta. Epidemiol Infect 2001;127:215–20.
[21] Nations WHO& F and AO of the U. Shiga toxin-producing *Escherichia coli* (STEC) and food: attribution, characterization, and monitoring. 2018.
[22] Tilden J, Young W, McNamara AM, Custer C, Boesel B, Lambert-Fair MA, et al. A new route of transmission for *Escherichia coli*: infection from dry fermented salami. Am J Public Health 1996;86:1142–5.
[23] Paton AW, Ratcliff RM, Doyle RM, Seymour-Murray J, Davos D, Lanser JA, et al. Molecular microbiological investigation of an outbreak of hemolytic-uremic syndrome caused by dry fermented sausage contaminated with Shiga-like toxin-producing *Escherichia coli*. J Clin Microbiol 1996;34:1622–7.
[24] Buvens G, Possé B, De Schrijver K, De Zutter L, Lauwers S, Piérard D. 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;8:421–6. doi:10.1089/fpd.2010.0693.
[25] Hald T, Aspinall W, Devleesschauwer B, Cooke R, Corrigan T, Havelaar AH, et al. World Health Organization Estimates of the Relative Contributions of Food to the Burden of Disease Due to Selected Foodborne Hazards: A Structured Expert Elicitation. PLoS One 2016;11:e0145839. doi:10.1371/journal.pone.0145839.
[26] Canada Communicible Disease Report. Severe Outbreak of *Escherichia coli* O157:H7 in Health Care Institutions in Charlottetown, Prince Edward Island, Fall, 2002.
[27] González-Escalona N, Kase J. Virulence gene profiles and phylogeny of Shiga toxin-positive *Escherichia coli* strains isolated from FDA regulated foods during 2010-2017. PLOSONE 2019;1:e0214620.
[28] https://www.foodnavigator.com/Article/2018/05/14/Seven-sick-due-to-E.-coli-in-French-Reblochon-cheese
[29] Mylius M, Dreesman J, Pulz M, Pallasch G, Beyrer K, Claußen K, et al. Shiga toxin-producing *Escherichia coli* O103:H2 outbreak in Germany after school trip to Austria due to raw cow milk, 2017 - The important role of international collaboration for outbreak investigations. Int J Med Microbiol 2018;308:539–44. doi:10.1016/j.ijmm.2018.05.005.
[30] Centers for Disease Control and Prevention (CDC). Multistate Outbreak of

Shiga toxin-producing *Escherichia coli* Infections Linked to Flour (Final Update) 2016.
[31] Centers for Disease Control and Prevention (CDC). Multistate Outbreaks of Shiga toxin-producing *Escherichia coli* O26 Infections Linked to Chipotle Mexican Grill Restaurants (Final Update) 2015.
[32] Centers for Disease Control and Prevention (CDC). Multistate Outbreak of Shiga toxin-producing *Escherichia coli* O121 Infections Linked to Raw Clover Sprouts (Final Update) 2014.
[33] Centers for Disease Control and Prevention (CDC). Multistate Outbreak of Shiga toxin-producing *Escherichia coli* O121 Infections Linked to Farm Rich Brand Frozen Food Products (Final Update) 2013.
[34] European Food Safety Authority. Multi- country outbreak of Shiga toxin-producing *Escherichia coli* infection associated with haemolytic uraemic syndrome. 2016.
[35] National Institute of Infectious Diseases and Tuberculosis and Infectious Diseases Control Division, Ministry of Health, Labour and Welfare J. Enterohemorrhagic *Escherichia coli* infection in Japan as of April 2012. Infect Agents Surveill Rep 2012;33:115–6.
[36] European Food Safety Authority. Shiga toxin-producing E. coli (STEC) O104:H4 2011 outbreaks in Europe: Taking Stock. 2011;9:2390.
[37] Robbins A, Anand M, Nicholas DC, Egan JS, Musser KA, Giguere S, et al. Ground beef recall associated with non-O157 Shiga toxin-producing *Escherichia coli*, United States. Emerg Infect Dis 2014;20:165–7. doi:10.3201/eid2001.130915.
[38] https://www.outbreak-database.com/Home.aspx
[39] Taylor E V, Nguyen TA, Machesky KD, Koch E, Sotir MJ, Bohm SR, et al. Multistate outbreak of *Escherichia coli* O145 infections associated with romaine lettuce consumption, 2010. J Food Prot 2013;76:939–44. doi:10.4315/0362-028X.JFP-12-503.
[40] Rounds JM, Rigdon CE, Muhl LJ, Forstner M, Danzeisen GT, Koziol BS, et al. Non-O157 Shiga toxin-producing *Escherichia coli* associated with venison. Emerg Infect Dis 2012;18:279–82. doi:10.3201/eid1802.110855.
[41] Bradley KK, Williams JM, Burnsed LJ, Lytle MB, McDermott MD, Mody RK, et al. Epidemiology of a large restaurant-associated outbreak of Shiga toxin-producing *Escherichia coli* O111:NM. Epidemiol Infect 2012;140:1644–54. doi:10.1017/S0950268811002329.
[42] De Schrijver K, Buvens G, Possé B, Van den Branden D, Oosterlynck O, De Zutter L, et al. Outbreak of verocytotoxin-producing *E. coli* O145 and O26 infections associated with the consumption of ice cream produced at a farm, Belgium, 2007. Euro Surveill 2008;13.
[43] Schimmer B, Nygard K, Eriksen H-M, Lassen J, Lindstedt B-A, Brandal LT, et al. Outbreak of haemolytic uraemic syndrome in Norway caused by stx2-positive *Escherichia coli* O103:H25 traced to cured mutton sausages. BMC Infect Dis 2008;8:41. doi:10.1186/1471-2334-8-41.
[44] https://www.hps.scot.nhs.uk/web-resources-container/guidance-for-the-public-health-management-of-escherichia-coli-o157-and-other-shiga-toxin-producing-stec-infections/
[45] Advisory Committee on the Microbiological Safety of Food. Report on Verocytotoxin-producing *Escherichia coli*. 1995.

[46] The Pennington Group. Report on the circumstances leading to the 1996 outbreak of infection with E.coli O157 in Central Scotland, the implications for food safety and the lessons learned. Edinburgh: The Stationary Office, 1997.
[47] Scottish Executive Health Department/Food Standards Agency (Scotland). Report of the *E. coli* O157 Task Force. Edinburgh: The Stationary Office, 2001.
[48] Zhang H, Morrison S, Tang Y-W. Multiplex Polymerase Chain Reaction Tests for Detection of Pathogens Associated with Gastroenteritis. Clin Lab Med 2015;35:461–86. doi:10.1016/j.cll.2015.02.006.
[49] Underwood AP, Dallman T, Thomson NR, Williams M, Harker K, Perry N, et al. Public health value of next-generation DNA sequencing of enterohemorrhagic *Escherichia coli* isolates from an outbreak. J Clin Microbiol 2013;51:232–7. doi:10.1128/JCM.01696-12.
[50] Turabelidze G, Lawrence SJ, Gao H, Sodergren E, Weinstock GM, Abubucker S, et al. Precise dissection of an *Escherichia coli* O157:H7 outbreak by single nucleotide polymorphism analysis. J Clin Microbiol 2013;51:3950–4. doi:10.1128/JCM.01930-13.
[51] Rusconi B, Sanjar F, Koenig SSK, Mammel MK, Tarr PI, Eppinger M. Whole Genome Sequencing for Genomics-Guided Investigations of *Escherichia coli* O157:H7 Outbreaks. Front Microbiol 2016;7:985. doi:10.3389/fmicb.2016.00985.
[52] 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. doi:10.2807/1560-7917.ES.2019.24.4.1800346.
[53] Holmes A, Dallman TJ, Shabaan S, Hanson M, Allison L. Validation of Whole-Genome Sequencing for Identification and Characterization of Shiga Toxin-Producing *Escherichia coli* To Produce Standardized Data To Enable Data Sharing. J Clin Microbiol 2017;56. doi:10.1128/JCM.01388-17.
[54] European Food Safety Authority. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. EFSA J 2018;16:5500.
[55] Scallan E, Hoekstra RM, Angulo FJ, Tauxe R V, Widdowson M-A, Roy SL, et al. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis 2011;17:7–15. doi:10.3201/eid1701.091101p1.
[56] Pearce MC, Evans J, McKendrick IJ, Smith AW, Knight HI, Mellor DJ, et al. Prevalence and virulence factors of *Escherichia coli* serogroups O26, O103, O111, and O145 shed by cattle in Scotland. Appl Environ Microbiol 2006;72:653–9. doi:10.1128/AEM.72.1.653-659.2006.
[57] Pearce MC, Jenkins C, Vali L, Smith AW, Knight HI, Cheasty T, et al. Temporal shedding patterns and virulence factors of *Escherichia coli* serogroups O26, O103, O111, O145, and O157 in a cohort of beef calves and their dams. Appl Environ Microbiol 2004;70:1708–16.
[58] Holmes A, Allison L, Ward M, Dallman TJ, Clark R, Fawkes A, et al. Utility of Whole-Genome Sequencing of *Escherichia coli* O157 for Outbreak Detection and Epidemiological Surveillance. J Clin Microbiol 2015;53:3565–73. doi:10.1128/JCM.01066-15.
[59] Lindsey RL, Pouseele H, Chen JC, Strockbine NA, Carleton HA. Implementation of Whole Genome Sequencing (WGS) for Identification and Characterization of Shiga Toxin-Producing *Escherichia coli* (STEC) in the

United States. Front Microbiol 2016;7:766. doi:10.3389/fmicb.2016.00766.
[60] Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. Identification of *Escherichia coli* and Shigella Species from Whole-Genome Sequences. J Clin Microbiol 2017;55:616–23. doi:10.1128/JCM.01790-16.
[61] Bogaerts B, Winand R, Fu Q, Van Braekel J, Ceyssens P-J, Mattheus W, et al. Validation of a Bioinformatics Workflow for Routine Analysis of Whole-Genome Sequencing Data and Related Challenges for Pathogen Typing in a European National Reference Center: Neisseria meningitidis as a Proof-of-Concept. Front Microbiol 2019;10:362. doi:10.3389/fmicb.2019.00362.
[62] Nadon C, Van Walle I, Gerner-Smidt P, Campos J, Chinen I, Concepcion-Acevedo J, et al. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. Euro Surveill 2017;22:30544. doi:10.2807/1560-7917.ES.2017.22.23.30544.
[63] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20. doi:10.1093/bioinformatics/btu170.
[64] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25. doi:10.1186/gb-2009-10-3-r25.
[65] Tewolde R, Dallman T, Schaefer U, Sheppard CL, Ashton P, Pichon B, et al. MOST: a modified MLST typing tool based on short read sequencing. PeerJ 2016;4:e2308. doi:10.7717/peerj.2308.
[66] Ashton PM, Perry N, Ellis R, Petrovska L, Wain J, Grant KA, et al. Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. PeerJ 2015;3:e739. doi:10.7717/peerj.739.
[67] Ogura Y, Gotoh Y, Itoh T, Sato MP, Seto K, Yoshino S, et al. Population structure of *Escherichia coli* O26:H11 with recent and repeated stx2 acquisition in multiple lineages. Microb Genomics 2017;3. doi:10.1099/mgen.0.000141.
[68] Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 2010;26:589–95. doi:10.1093/bioinformatics/btp698.
[69] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9. doi:10.1093/bioinformatics/btp352.
[70] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303. doi:10.1101/gr.107524.110.
[71] Dallman T, Ashton P, Schafer U, Jironkin A, Painset A, Shaaban S, et al. SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. Bioinformatics 2018;34:3028–9. doi:10.1093/bioinformatics/bty212.
[72] https://github.com/phe-bioinformatics/PHEnix
[73] Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, et al. Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res 2012;40:e6–e6. doi:10.1093/nar/gkr928.
[74] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3. doi:10.1093/bioinformatics/btu033.

[75] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 2013;30:2725–9. doi:10.1093/molbev/mst197.
[76] DebRoy C, Fratamico PM, Yan X, Baranzoni G, Liu Y, Needleman DS, et al. Comparison of O-Antigen Gene Clusters of All O-Serogroups of *Escherichia coli* and Proposal for Adopting a New Nomenclature for O-Typing. PLoS One 2016;11:e0147434. doi:10.1371/journal.pone.0147434.
[77] Holt KE, Ingle DJ, Levine MM, Kuzevski A, Valcanis M, Robins-Browne RM, et al. In silico serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. Microb Genomics 2016;2:e000064. doi:10.1099/mgen.0.000064.
[78] Geue L, Menge C, Eichhorn I, Semmler T, Wieler LH, Pickard D, et al. Evidence for Contemporary Switching of the O-Antigen Gene Cluster between Shiga Toxin-Producing *Escherichia coli* Strains Colonizing Cattle. Front Microbiol 2017;8:424. doi:10.3389/fmicb.2017.00424.
[79] Zhang WL, Bielaszewska M, Liesegang A, Tschäpe H, Schmidt H, Bitzan M, et al. Molecular characteristics and epidemiological significance of Shiga toxin-producing *Escherichia coli* O26 strains. J Clin Microbiol 2000;38:2134–40.
[80] Bielaszewska M, Mellmann A, Bletz S, Zhang W, Köck R, Kossow A, et al. Enterohemorrhagic *Escherichia coli* O26:H11/H−: A New Virulent Clone Emerges in Europe. Clin Infect Dis 2013;56:1373–81. doi:10.1093/cid/cit055.
[81] Paton AW, Woodrow MC, Doyle RM, Lanser JA, Paton JC. Molecular characterization of a Shiga toxigenic *Escherichia coli* O113:H21 strain lacking eae responsible for a cluster of cases of hemolytic- uremic syndrome. J Clin Microbiol 1999;37:3357–61.
[82] Law D. Virulence factors of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli*. J Appl Microbiol 2000;88:729–45.
[83] Rogers TJ, Paton JC, Wang H, Talbot UM, Paton AW. Reduced virulence of an fliC mutant of Shiga-toxigenic *Escherichia coli* O113:H21. Infect Immun 2006;74:1962–6. doi:10.1128/IAI.74.3.1962-1966.2006.
[84] Nagy A, Xu Y, Bauchan GR, Shelton DR, Nou X. Aggregative adherence fimbriae I (AAF/I) mediate colonization of fresh produce and abiotic surface by Shiga toxigenic enteroaggregative *Escherichia coli* O104:H4. Int J Food Microbiol 2016;229:44–51. doi:10.1016/j.ijfoodmicro.2016.04.007.
[85] Hauser E, Bruederle M, Reich C, Bruckbauer A, Funk J, Schmidt H. Subtilase contributes to the cytotoxicity of a Shiga toxin-producing *Escherichia coli* strain encoding three different toxins. Int J Food Microbiol 2016;217:156–61. doi:10.1016/j.ijfoodmicro.2015.10.023.
[86] Aas CG, Drabløs F, Haugum K, Afset JE. Comparative Transcriptome Profiling Reveals a Potential Role of Type VI Secretion System and Fimbriae in Virulence of Non-O157 Shiga Toxin-Producing *Escherichia coli*. Front Microbiol 2018;9:1416. doi:10.3389/fmicb.2018.01416.
[87] Zheng L-L, Li Y-X, Ding J, Guo X-K, Feng K-Y, Wang Y-J, et al. A Comparison of Computational Methods for Identifying Virulence Factors. PLoS One 2012;7:e42517. doi:10.1371/journal.pone.0042517.
[88] Krause M, Barth H, Schmidt H. Toxins of locus of enterocyte effacement-negative Shiga toxin-producing *Escherichia coli*. Toxins (Basel) 2018;10:1–19. doi:10.3390/toxins10060241.

[89] Paton AW, Paton JC. *Escherichia coli* subtilase cytotoxin. Toxins (Basel) 2010;2:215–28. doi:10.3390/toxins2020215.
[90] Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, et al. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc Natl Acad Sci 2009;106:17939–44. doi:10.1073/pnas.0903585106.
[91] Alikhan N, Bachmann N, Ben Zakour N, Petty N, et.al. Multiple evolutionary trajectories for non-O157 Shiga toxigenic *Escherichia coli*. BioRxiv 2019.
[92] Ju W, Cao G, Rump L, Strain E, Luo Y, Timme R, et al. Phylogenetic Analysis of Non-O157 Shiga Toxin-Producing *Escherichia coli* Strains by Whole-Genome Sequencing. J Clin Microbiol 2012;50:4123–7. doi:10.1128/JCM.02262-12.
[93] Iguchi A, Iyoda S, Ohnishi M, EHEC Study Group. Molecular characterization reveals three distinct clonal groups among clinical shiga toxin-producing *Escherichia coli* strains of serogroup O103. J Clin Microbiol 2012;50:2894–900. doi:10.1128/JCM.00789-12.
[94] Ishijima N, Lee K, Kuwahara T, Nakayama-Imaohji H, Yoneda S, Iguchi A, et al. Identification of a New Virulent Clade in Enterohemorrhagic *Escherichia coli* O26:H11/H- Sequence Type 29. Sci Rep 2017;7:43136. doi:10.1038/srep43136.
[95] Delannoy S, Mariani-Kurkdjian P, Webb HE, Bonacorsi S, Fach P. The Mobilome; A Major Contributor to *Escherichia coli* stx2-Positive O26:H11 Strains Intra-Serotype Diversity. Front Microbiol 2017;8:1625. doi:10.3389/fmicb.2017.01625.
[96] Cooper KK, Mandrell RE, Louie JW, Korlach J, Clark TA, Parker CT, et al. Complete Genome Sequences of Two *Escherichia coli* O145:H28 Outbreak Strains of Food Origin. Genome Announc 2014;2. doi:10.1128/genomeA.00482-14.
[97] Brzuszkiewicz E, Thürmer A, Schuldes J, Leimbach A, Liesegang H, Meyer F-D, et al. Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). Arch Microbiol 2011;193:883–91. doi:10.1007/s00203-011-0725-6.
[98] Ferdous M, Zhou K, de Boer RF, Friedrich AW, Kooistra-Smid AMD, Rossen JWA. Comprehensive Characterization of *Escherichia coli* O104:H4 Isolated from Patients in the Netherlands. Front Microbiol 2015;6:1348. doi:10.3389/fmicb.2015.01348.
[99] Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, et al. Origins of the *E. coli* Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany. N Engl J Med 2011;365:709–17. doi:10.1056/NEJMoa1106920.
[100] Soysal N, Mariani-Kurkdjian P, Smail Y, Liguori S, Gouali M, Loukiadis E, et al. Enterohemorrhagic *Escherichia coli* Hybrid Pathotype O80:H2 as a New Therapeutic Challenge. Emerg Infect Dis 2016;22:1604–12. doi:10.3201/eid2209.160304.
[101] Pradel N, Bertin Y, Martin C, Livrelli V. Molecular analysis of shiga toxin-producing *Escherichia coli* strains isolated from hemolytic-uremic syndrome patients and dairy samples in France. Appl Environ Microbiol 2008;74:2118–28. doi:10.1128/AEM.02688-07.

[102] Zweifel C, Giezendanner N, Corti S, Krause G, Beutin L, Danuser J, et al. Characteristics of shiga toxin-producing *Escherichia coli* isolated from Swiss raw milk cheese within a 3-year monitoring program. J Food Prot 2010;73:88–91.
[103] Lee JB, Han D, Lee HT, Wi SM, Park JH, Jo J-W, et al. Pathogenic and phylogenetic characteristics of non-O157 Shiga toxin-producing *Escherichia coli* isolates from retail meats in South Korea. J Vet Sci 2018;19:251–9. doi:10.4142/jvs.2018.19.2.251.
[104] Werber D, Beutin L, Pichner R, Stark K, Fruth A. Shiga toxin-producing *Escherichia coli* serogroups in food and patients, Germany. Emerg Infect Dis 2008;14:1803–6. doi:10.3201/eid1411.080361.
[105] Feng PCH, Delannoy S, Lacher DW, Bosilevac JM, Fach P, Beutin L. Shiga Toxin-Producing Serogroup O91 *Escherichia coli* Strains Isolated from Food and Environmental Samples. Appl Environ Microbiol 2017;83. doi:10.1128/AEM.01231-17.
[106] Hernandez LB, Cadona JS, Christensen M, Fernández D, Padola NL, Bustamante AV, et al. Virulence genes and genetic diversity assessment of Shiga toxin-producing *Escherichia coli* O91 strains from cattle, beef and poultry products. Microb Pathog 2018;125:463–7. doi:10.1016/j.micpath.2018.10.009.
[107] van Hoek AHAM, van Veldhuizen JNJ, Friesema I, Coipan C, Rossen JWA, Bergval IL, et al. Comparative genomics reveals a lack of evidence for pigeons as a main source of stx2f-carrying *Escherichia coli* causing disease in humans and the common existence of hybrid Shiga toxin-producing and enteropathogenic *E. coli* pathotypes. BMC Genomics 2019;20:271. doi:10.1186/s12864-019-5635-z.
[108] De Rauw K, Jacobs S, Piérard D. Twenty-seven years of screening for Shiga toxin-producing *Escherichia coli* in a university hospital. Brussels, Belgium, 1987-2014 2018;13:e0199968. doi:10.1371/journal.pone.0199968.
[109] Ørskov F. On the occurrence of *E. coli* belonging to O-group 26 in cases of infantile diarrhoea and white scours. Acta Pathol Microbiol Scand 1951;29:373–8.
[110] Konowalchuk J, Speirs JI, Stavric S. Vero response to a cytotoxin of *Escherichia coli*. Infect Immun 1977;18:775–9.
[111] Fukushima H, Hoshina K, Gomyoda M. Long-term survival of shiga toxin-producing *Escherichia coli* O26, O111, and O157 in bovine feces. Appl Environ Microbiol 1999;65:5177–81.
[112] Bono et al. Identification of a New Shiga Toxin-containing O26:H11 stx2 SNP Clonal Complex in the US. 2018.
[113] Food Standards Scotland. Comparison of Human and Cattle *E. coli* O26 Isolates by PFGE. https://www.foodstandards.gov.scot/publications-and-research/publications/comparison-of-human-and-cattle-e-coli-o26-isolates-by-pulsed-field-gel-ele.
[114] Menrath A, Wieler LH, Heidemanns K, Semmler T, Fruth A, Kemper N. Shiga toxin producing *Escherichia coli*: identification of non-O157:H7-Super-Shedding cows and related risk factors. Gut Pathog 2010;2:7. doi:10.1186/1757-4749-2-7.
[115] Murphy BP, McCabe E, Murphy M, Buckley JF, Crowley D, Fanning S, et al. Longitudinal Study of Two Irish Dairy Herds: Low Numbers of Shiga Toxin-

Producing *Escherichia coli* O157 and O26 Super-Shedders Identified. Front Microbiol 2016;7:1850. doi:10.3389/fmicb.2016.01850.
[116] Food Safety and Inspection Service U. Risk Profile for Pathogenic Non-O157 Shiga Toxin-Producing *Escherichia coli* (non-O157 STEC) 2012.
[117] Kindle P, Nuesch-Inderbinen M, Cernela N, Roger S. Detection, Isolation, and Characterization of Shiga Toxin–Producing *Escherichia coli* in Flour. J Food Prot 2019;82:164–7. doi:10.4315/0362-028X.JFP-18-256.
[118] Söderlund R, Hurel J, Jinnerot T, Sekse C, Aspán A, Eriksson E, et al. Genomic comparison of *Escherichia coli* serotype O103:H2 isolates with and without verotoxin genes: implications for risk assessment of strains commonly found in ruminant reservoirs. Infect Ecol Epidemiol 2016;6:30246. doi:10.3402/iee.v6.30246.
[119] Sekse C, Sunde M, Hopp P, Bruheim T, Cudjoe KS, Kvitle B, et al. Occurrence of Potentially Human-Pathogenic *Escherichia coli* O103 in Norwegian Sheep. Appl Environ Microbiol 2013;79:7502. doi:10.1128/AEM.01825-13.
[120] Castro VS, Carvalho RCT, Conte-Junior CA, Figuiredo EES. Shiga-toxin Producing *Escherichia coli*: Pathogenicity, Supershedding, Diagnostic Methods, Occurrence, and Foodborne Outbreaks. Compr Rev Food Sci Food Saf 2017;16:1269–80. doi:10.1111/1541-4337.12302.
[121] Montero DA, Velasco J, Del Canto F, Puente JL, Padola NL, Rasko DA, et al. Locus of Adhesion and Autoaggregation (LAA), a pathogenicity island present in emerging Shiga Toxin-producing *Escherichia coli* strains. Sci Rep 2017;7:7011. doi:10.1038/s41598-017-06999-y.
[122] Paton AW, Srimanote P, Talbot UM, Wang H, Paton JC. A new family of potent AB5 cytotoxins produced by Shiga toxigenic *Escherichia coli*. J Exp Med 2004;200:35–46. doi:10.1084/jem.20040392.
[123] Zotta E, Paton AW, Ochoa F, Melendi S, Castro Parodi M, Damiano A, et al. Systemic effects of Subtilase cytotoxin produced by *Escherichia coli* O113:H21. Toxicon 2017;127:49–55. doi:10.1016/j.toxicon.2016.12.014.
[124] Melton-Celsa AR. Shiga Toxin (Stx) Classification, Structure, and Function. Microbiol Spectr 2014;2:EHEC-0024-2013. doi:10.1128/microbiolspec.EHEC-0024-2013.
[125] Weinstein DL, Jackson MP, Samuel JE, Holmes RK, O'Brien AD. Cloning and sequencing of a Shiga-like toxin type II variant from *Escherichia coli* strain responsible for edema disease of swine. J Bacteriol 1988;170:4223–30.
[126] Lo AWH, Moonens K, De Kerpel M, Brys L, Pardon E, Remaut H, et al. The molecular mechanism of Shiga toxin Stx2e neutralization by a single-domain antibody targeting the cell receptor-binding domain. J Biol Chem 2014;289:25374–81. doi:10.1074/jbc.M114.566257.
[127] Piérard D, Muyldermans G, Moriau L, Stevens D, Lauwers S. Identification of new verocytotoxin type 2 variant B-subunit genes in human and animal *Escherichia coli* isolates. J Clin Microbiol 1998;36:3317–22.
[128] Thomas A, Cheasty T, Chart H, Rowe B. Isolation of Vero cytotoxin-producing *Escherichia coli* serotypes O9ab:H- and O101:H-carrying VT2 variant gene sequences from a patient with haemolytic uraemic syndrome. Eur J Clin Microbiol Infect Dis 1994;13:1074–6.
[129] Martin A, Beutin L. Characteristics of Shiga toxin-producing *Escherichia coli* from meat and milk products of different origins and association with food producing animals as main contamination sources. Int J Food Microbiol

2011;146:99–104. doi:10.1016/j.ijfoodmicro.2011.01.041.
[130] Prager R, Fruth A, Siewert U, Strutz U, Tschäpe H. *Escherichia coli* encoding Shiga toxin 2f as an emerging human pathogen. Int J Med Microbiol 2009;299:343–53. doi:10.1016/j.ijmm.2008.10.008.
[131] Schmidt H, Scheef J, Morabito S, Caprioli A, Wieler LH, Karch H. A new Shiga toxin 2 variant (Stx2f) from *Escherichia coli* isolated from pigeons. Appl Environ Microbiol 2000;66:1205–8.
[132] Friesema I, van der Zwaluw K, Schuurman T, Kooistra-Smid M, Franz E, van Duynhoven Y, et al. Emergence of *Escherichia coli* encoding Shiga toxin 2f in human Shiga toxin-producing *E. coli* (STEC) infections in the Netherlands, January 2008 to December 2011. Euro Surveill 2014;19:26–32.
[133] Friesema IHM, Keijzer-Veen MG, Koppejan M, Schipper HS, van Griethuysen AJ, Heck MEOC, et al. Hemolytic uremic syndrome associated with *Escherichia coli* O8:H19 and Shiga toxin 2f gene. Emerg Infect Dis 2015;21:168–9. doi:10.3201/eid2101.140515.
[134] Ooka T, Seto K, Kawano K, Kobayashi H, Etoh Y, Ichihara S, et al. Clinical significance of Escherichia albertii. Emerg Infect Dis 2012;18:488–92. doi:10.3201/eid1803.111401.
[135] Brandal LT, Tunsjø HS, Ranheim TE, Løbersli I, Lange H, Wester AL. Shiga Toxin 2a in Escherichia albertii. J Clin Microbiol 2015;53:1454–5. doi:10.1128/JCM.03378-14.
[136] Gioia-Di Chiacchio RM, Cunha MPV, de Sá LRM, Davies YM, Pereira CBP, Martins FH, et al. Novel Hybrid of Typical Enteropathogenic *Escherichia coli* and Shiga-Toxin-Producing *E. coli* (tEPEC/STEC) Emerging From Pet Birds. Front Microbiol 2018;9:2975. doi:10.3389/fmicb.2018.02975.
[137] Leung PHM, Peiris JSM, Ng WWS, Robins-Browne RM, Bettelheim KA, Yam WC. A newly discovered verotoxin variant, VT2g, produced by bovine verocytotoxigenic *Escherichia coli*. Appl Environ Microbiol 2003;69:7549–53.
[138] Prager R, Fruth A, Busch U, Tietze E. Comparative analysis of virulence genes, genetic diversity, and phylogeny of Shiga toxin 2g and heat-stable enterotoxin STIa encoding *Escherichia coli* isolates from humans, animals, and environmental sources. Int J Med Microbiol 2011;301:181–91. doi:10.1016/j.ijmm.2010.06.003.
[139] Persson S, Olsen KEP, Ethelberg S, Scheutz F. Subtyping method for *Escherichia coli* shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. J Clin Microbiol 2007;45:2020–4. doi:10.1128/JCM.02591-06.
[140] Beutin L, Miko A, Krause G, Pries K, Haby S, Steege K, et al. Identification of Human-Pathogenic Strains of Shiga Toxin-Producing *Escherichia coli* from Food by a Combination of Serotyping and Molecular Typing of Shiga Toxin Genes. Appl Environ Microbiol 2007;73:4769–75. doi:10.1128/AEM.00873-07.
[141] Nyholm O, Heinikainen S, Pelkonen S, Hallanvuo S, Haukka K, Siitonen A. Hybrids of Shigatoxigenic and Enterotoxigenic *Escherichia coli* (STEC/ETEC) Among Human and Animal Isolates in Finland. Zoonoses Public Health 2015;62:518–24. doi:10.1111/zph.12177.
[142] Fierz L, Cernela N, Hauser E, Nüesch-Inderbinen M, Stephan R. Characteristics of Shigatoxin-Producing *Escherichia coli* Strains Isolated during 2010-2014 from Human Infections in Switzerland. Front Microbiol 2017;8:1471. doi:10.3389/fmicb.2017.01471.

[143] Wijnsma KL, Schijvens AM, Rossen JWA, Kooistra-Smid AMDM, Schreuder MF, van de Kar NCAJ. Unusual severe case of hemolytic uremic syndrome due to Shiga toxin 2d-producing *E. coli* O80:H2. Pediatr Nephrol 2017;32:1263–8. doi:10.1007/s00467-017-3642-3.
[144] McFarland N, Bundle N, Jenkins C, Godbole G, Mikhail A, Dallman T, et al. Recurrent seasonal outbreak of an emerging serotype of Shiga toxin-producing *Escherichia coli* (STEC O55:H7 Stx2a) in the south west of England, July 2014 to September 2015. Euro Surveill 2017;22:30610. doi:10.2807/1560-7917.ES.2017.22.36.30610.
[145] 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 Genomics 2018;4. doi:10.1099/mgen.0.000181.
[146] Day M, Doumith M, Jenkins C, Dallman TJ, Hopkins KL, Elson R, 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. doi:10.1093/jac/dkw371.

8. Appendices

Appendix 1 - Genomes used from the public database.

Strain ID	Serotype	Accession Number
93302	O26:H11	SRR3241997
93285	O26:H11	SRR3241859
93288	O26:H11	SRR3241849
93304	O26:H11	SRR3240981
194236	O26:H11	SRR3578935
131458	O26:H11	DRX066965
133042	O26:H11	DRX066979
140921	O26:H11	DRX066967
132265	O26:H11	DRX066972
141425	O26:H11	DRX066961
132777	O26:H11	DRX066959
2011C-3506	O26:H11	SRX1692204
2010C-3871	O26:H11	SRX1692174
2010C-4788	O26:H11	SRX1692143
2010C-3902	O26:H11	SRX1692172
2009C-3612	O26:H11	SRX1592478
11368	O26:H11	NC_O13661
STEC1117	O26:H11	NZ_LOFU
STEC2144	O26:H11	NZ_LOGU
STEC563	O26:H11	NZ_LODD
STEC931	O26:H11	NZ_LOFS
36084	O26:H11	NZ_LDXI
36708	O26:H11	NZ_LDXG
2011C_3493 pAA-EA11 pESBL-EA11 pG-EA11	O104:H4	NC_018658.1 CP003291.1 CP003290.1 CP003292.1
RM13514 pO145-13514 pRM13514	O145:H28	CP006027.1 CP006028.1 CP006029.1
RM13516 pO145-13516 pRM13516	O145:H28	CP006262.1 CP006263.1 CP006264.1
17-00944	O103:H2	SRR7403873

Appendix 2 - Isolate information and WGS results for the clinical isolates. The isolates were labelled (ID) according to their phylogenetic position on the tree (Figure 9).

ID	Year	Serotype	7-gene MLST	ST	*eae*	*stx* subtype	JEMRA
1	2016	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
2	2014	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
3	2016	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
4	2013	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
5	2015	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
6	2013	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
7	2016	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
8	2015	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
9	2017	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
10	2017	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
11	2016	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
12	2016	O63:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
13	2016	O125ac:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
14	2014	O125ac:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
15	2016	O125ac:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
16	2016	O125ac:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
17	2014	O125ac:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
18	2014	O125ac:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
19	2015	O125ac:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
20	2016	O125ac:H6	15,24,10,46,17,37,36	583	*eae*	*stx2f*	5
21	2014	O126:H34	86,24,19,46,23,94,36	4101	*eae*	*stx2f*	5
22	2014	O145:H34	124,24,19,13,23,50,17	722	*eae*	*stx2f*	5
23	2015	O145:H34	124,24,19,13,23,50,17	722	*eae*	*stx2f*	5
24	2015	O145:H34	124,24,19,13,23,50,17	722	*eae*	*stx2f*	5
25	2014	O145:H34	124,24,19,13,23,50,17	722	*eae*	*stx2f*	5
26	2015	O145:H34	124,24,19,13,23,50,17	722	*eae*	*stx2f*	5
27	2017	O145:H34	124,24,19,13,23,50,17	722	*eae*	*stx2f*	5
28	2017	O117:H7	76,43,9,36,404,14,10	5292	no *eae*	*stx1a*	5
29	2017	O117:H7	76,43,9,36,404,14,10	5292	no *eae*	*stx1a*	5
30	2011	O117:H7	76,43,9,36,17,14,10	504	no *eae*	*stx1a*	5
31	2013	O156:H7	76,43,9,36,17,14,10	504	no *eae*	*stx1a*	5
32	2015	O156:H7	76,43,9,36,17,14,10	504	no *eae*	*stx1a*	5
33	2005	O50/O2:H6	13,52,156,14,17,25,17	998	no *eae*	*None*	n/a
34	2006	O50/O2:H6	13,52,10,14,17,25,17	141	no *eae*	*stx2b*	5
35	2015	Ou:H5	36,44,10,13,17,10,25	1161	no *eae*	*stx2f*	5
36	2013	O113:H6	13,21,13,22,17,36,15	121	*eae*	*stx2f*	5
37	2015	O137:H6	13,14,19,22,17,506,481	6675	*eae*	*stx2f*	5

38	2015	O123/O186:H21	15,18,13,13,30,115,13	4179	*eae*	*stx2f*	5
39	2016	O146:H28	130,45,41,22,7,50,45	738	no *eae*	*stx2b*	5
40	2016	O146:H28	130,45,41,22,7,50,45	738	no *eae*	*stx2b*	5
41	2017	O146:H28	130,45,41,22,7,50,45	738	no *eae*	*stx2b*	5
42	2016	O146:H28	612,45,41,22,7,50,45	6674	no *eae*	*stx2b*	5
43	2014	O17/O77/O106:H45	120,151,46,48,35,40,38	662	no *eae*	*stx2d*	2
44	2008	O17/O77/O106:H45	120,54,46,48,532,40,38	7083	no *eae*	*stx1a*	5
45	2008	O17/O77/O106:H45	120,54,46,48,532,40,38	7083	no *eae*	*stx1a*	5
46	2016	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
47	2015	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
48	2015	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
49	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
50	2014	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
51	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2d*	2
52	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2d*	2
53	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2d*	2
54	2011	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
55	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
56	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
57	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
58	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
59	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
60	2010	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
61	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
62	2004	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
63	2008	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
64	2014	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
65	2014	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
66	2014	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
67	2014	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
68	2014	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
69	2018	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
70	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
71	2008	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
72	2015	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
73	2017	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
74	2008	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
75	2008	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
76	2010	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
77	2010	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
78	2010	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1

79	2010	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
80	2010	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
81	2005	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
82	2009	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
83	2009	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
84	2015	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
85	2011	O145:H28	19,23,730,24,21,2,16	8625	*eae*	*stx1a*	4
86	2007	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
87	2010	O145:H28	19,23,18,24,21,2,16	32	*eae*	*stx2a*	1
88	2013	O145:H28	19,23,51,24,21,5,16	6130	*eae*	*stx2a*	1
89	2012	O145:H28	19,23,51,24,21,5,16	6130	*eae*	*stx2a*	1
90	2015	O145:H28	19,23,51,24,21,2,16	137	*eae*	*stx2a*	1
91	2016	Ou:H20	179,789,225,286,15,2,2	6060	no *eae*	*stx1c*	5
92	2012	O27:H30	12,93,136,30,112,1,2	753	no *eae*	*stx2b*	5
93	2015	O105:H20	83,260,596,30,15,22,295	8638	no *eae*	*stx1d*	5
94	2013	O51:H28	83,23,186,82,159,1,2	8876	no *eae*	*stx2b*	5
95	2015	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1c*	5
96	2015	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1c*	5
97	2014	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1a stx1c*	5
98	2017	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1c*	5
99	2017	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b*	5
100	2015	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b*	5
101	2008	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1c*	5
102	2017	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1c*	5
103	2016	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1c*	5
104	2012	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1c*	5
105	2015	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b stx1c*	5
106	2012	O166:H28	87,90,231,10,1,187,138	1819	no *eae*	*stx2b*	5
107	2016	O138:H48	58,53,53,58,24,1,42	219	no *eae*	*stx2e*	5
108	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
109	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
110	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
111	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
112	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
113	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
114	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
115	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
116	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
117	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
118	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
119	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4

120	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
121	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
122	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
123	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
124	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
125	2003	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
126	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
127	2003	O26:H11	16,4,12,16,9,7,7	21	no *eae*	*stx1a*	5
128	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
129	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
130	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
131	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
132	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
133	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
134	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
135	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
136	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
137	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
138	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
139	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
140	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
141	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
142	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
143	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
144	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
145	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
146	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
147	2002	O26:H11	16,135,12,16,9,7,7	574	*eae*	*stx1a*	4
148	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
149	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
150	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
151	2009	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
152	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
153	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
154	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
155	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
156	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
157	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
158	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
159	2009	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
160	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1

161	2017	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
162	2017	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
163	2017	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
164	2017	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
165	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
166	2008	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
167	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
168	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
169	2003	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
170	2006	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
171	2006	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
172	2009	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
173	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
174	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
175	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
176	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
177	2009	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
178	2017	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
179	2009	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
180	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
181	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
182	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
183	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
184	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
185	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
186	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
187	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
188	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
189	2008	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
190	2017	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
191	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
192	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
193	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2c stx1a*	3
194	2006	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
195	2008	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
196	2005	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
197	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
198	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
199	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
200	2009	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
201	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1

202	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
203	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
204	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
205	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
206	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
207	2007	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
208	2007	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
209	2007	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
210	2007	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
211	2012	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
212	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
213	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
214	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
215	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
216	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
217	2017	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
218	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
219	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a*	1
220	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a*	1
221	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a*	1
222	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a*	1
223	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a*	1
224	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a*	1
225	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a*	1
226	2013	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
227	2017	O26:H11	6,4,12,16,9,7,7	29	*eae*	*stx2a*	1
228	2015	O26:H11	6,4,12,16,9,7,7	29	*eae*	*stx2a*	1
229	2014	O26:H11	6,4,12,16,9,7,7	29	*eae*	*stx2a stx1a*	1
230	2014	O26:H11	6,4,12,16,9,7,7	29	*eae*	*stx1a*	4
231	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
232	2014	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
233	2006	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
234	2016	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
235	2007	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
236	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
237	2015	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
238	2011	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
239	2007	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
240	2007	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
241	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
242	2008	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4

243	2010	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
244	2007	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
245	2005	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
246	2006	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx2a stx1a*	1
247	2004	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
248	2005	O26:H11	16,4,12,16,9,7,7	21	*eae*	*stx1a*	4
249	2017	O177:H11	6,4,12,16,9,7,7	29	*eae*	*stx1a*	4
250	2017	O177:H11	6,4,12,16,9,7,7	29	*eae*	*stx1a*	4
251	2012	O177:H11	6,4,12,16,9,7,7	29	*eae*	*stx1a*	4
252	2015	O177:H11	6,4,12,16,9,7,7	29	*eae*	*stx2a stx1a*	1
253	2015	O177:H11	6,4,12,16,9,7,7	29	*eae*	*stx2a stx1a*	1
254	2007	O123/O186:H11	6,4,12,16,9,360,7	4738	*eae*	*stx2a*	1
255	2008	O123/O186:H11	6,4,12,16,9,360,7	4738	*eae*	*stx2a*	1
256	2016	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx2a stx1a*	1
257	2006	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx2a stx1a*	1
258	2006	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx2a stx1a*	1
259	2013	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx1a*	4
260	2007	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx2a stx1a*	1
261	2009	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx2a stx1a*	1
262	2012	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx1a*	4
263	2014	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx1a*	4
264	2014	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx1a*	4
265	2012	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx1a*	4
266	2013	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx1a*	4
267	2013	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx1a*	4
268	2017	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx2a stx1a*	1
269	2010	O111:H8	6,4,12,16,9,7,12	16	*eae*	*stx2a stx1a*	1
270	2010	O71:H8	6,135,12,16,9,7,7	2836	*eae*	*stx2c stx1a*	3
271	2016	O71:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
272	2010	O71:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
273	2006	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
274	2018	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
275	2018	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
276	2011	O109:H32	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
277	2016	O4:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
278	2016	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
279	2006	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
280	2015	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
281	2015	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
282	2007	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
283	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4

284	2014	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
285	2014	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
286	2015	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
287	2009	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
288	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
289	2013	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
290	2009	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
291	2009	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
292	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
293	2014	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
294	2008	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
295	2016	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
296	2012	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
297	2017	O103:H2	6,4,3,17,7,7,6	17	no *eae*	*stx1a*	5
298	2011	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
299	2010	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
300	2005	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
301	2007	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
302	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
303	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
304	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
305	2014	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
306	2013	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
307	2007	O103:H2	6,4,3,962,7,7,6	8506	*eae*	*stx1a*	4
308	2013	O103:H2	91,4,3,17,7,7,6	386	*eae*	*stx1a*	4
309	2003	O103:H2	91,4,3,17,7,7,6	386	*eae*	*stx1a*	4
310	2009	O103:H2	91,4,3,17,7,7,6	386	*eae*	*stx1a*	4
311	2010	O103:H2	91,4,3,17,7,7,6	386	*eae*	*stx1a*	4
312	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
313	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
314	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
315	2014	O123/O186:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
316	2016	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
317	2017	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
318	2011	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
319	2014	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
320	2012	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
321	2012	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
322	2012	O103:H2	6,4,3,17,7,7,6	17	*eae*	*stx1a*	4
323	2004	O15:H2	6,4,3,18,7,7,6	20	*eae*	*None*	n/a
324	2016	O4:H2	6,4,3,18,7,7,6	20	*eae*	*stx2f*	5

325	2016	O45:H2	6,4,3,18,7,7,6	20	*eae*	*stx2f*	5
326	2017	O45:H2	6,4,3,18,7,7,6	20	*eae*	*stx2f*	5
327	2013	O45:H2	6,4,3,18,7,7,6	20	*eae*	*stx2f*	5
328	2014	O103:H2	6,4,3,18,7,7,6	20	*eae*	*stx2f*	5
329	2014	O103:H2	6,4,3,18,7,7,6	20	*eae*	*stx2f*	5
330	2016	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
331	2014	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx1c*	5
332	2015	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
333	2015	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
334	2006	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
335	2017	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
336	2010	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
337	2011	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
338	2010	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
339	2016	O128ab:H2	6,6,14,21,9,7,148	6265	no *eae*	*stx2b stx1c*	5
340	2016	O128ab:H2	6,6,14,21,9,7,148	6265	no *eae*	*stx2b stx1c*	5
341	2015	O128ab:H2	6,6,14,21,9,7,148	6265	no *eae*	*stx2b stx1c*	5
342	2015	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
343	2012	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx1c*	5
344	2015	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b*	5
345	2016	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b stx1c*	5
346	2011	O128ab:H2	6,6,14,21,117,7,7	811	no *eae*	*stx2b*	5
347	2010	O128ab:H2	6,6,14,21,117,7,7	811	no *eae*	*stx2b stx1c*	5
348	2016	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b*	5
349	2010	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b*	5
350	2014	O128ab:H2	6,6,14,21,9,7,7	25	no *eae*	*stx2b*	5
351	2016	O174:H21	6,95,15,18,9,8,14	677	no *eae*	*stx2c*	5
352	2012	O174:H21	6,95,15,18,9,8,14	677	no *eae*	*stx2d stx2b*	2
353	2015	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b*	5
354	2015	O146:H21	6,95,33,18,9,8,14	442	no eae	*stx2b*	5
355	2015	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b*	5
356	2011	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx1c*	5
357	2017	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b stx1c*	5
358	2010	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b*	5
359	2014	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b*	5
360	2017	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b*	5
361	2015	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b stx1c*	5
362	2014	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b stx1c*	5
363	2008	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx1c*	5
364	2007	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx1c*	5
365	2007	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx1c*	5

366	2011	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx1c*	5
367	2012	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b stx1c*	5
368	2015	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx1c*	5
369	2012	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b stx1c*	5
370	2008	O146:H21	6,95,33,983,9,8,14	8661	no *eae*	*stx2b stx1c*	5
371	2011	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b stx1c*	5
372	2015	O146:H21	6,95,33,18,9,8,14	442	no *eae*	*stx2b stx1a*	5
373	2016	Ou:H8	6,19,15,18,11,8,14	26	no *eae*	*stx2b*	5
374	2014	O11:H8	6,19,15,18,11,8,14	26	no *eae*	*stx2b*	5
375	2015	Ou:H8	6,19,15,18,11,8,14	26	no *eae*	*stx2b*	5
376	2015	Ou:H8	6,19,15,18,11,8,14	26	no *eae*	*stx2b*	5
377	2014	Ou:H16	6,4,3,88,11,8,14	3236	no *eae*	*stx1a*	5
378	2014	O76:H19	6,23,32,16,9,8,7	675	no *eae*	*stx1c*	5
379	2013	O76:H19	6,23,32,16,9,8,7	675	no *eae*	*stx1c*	5
380	2013	O76:H19	6,23,32,16,9,8,7	675	no *eae*	*stx1c*	5
381	2015	O76:H19	6,23,32,16,9,8,7	675	no *eae*	*stx1c*	5
382	2014	O76:H19	6,23,32,16,9,8,7	675	no *eae*	*stx1c*	5
383	2014	O76:H19	6,23,32,16,9,8,7	675	no *eae*	*stx1c*	5
384	2016	O76:H19	6,23,32,16,9,8,7	675	no *eae*	*stx2b stx1c*	5
385	2007	O76:H19	6,23,32,16,9,8,7	675	no *eae*	*stx1c*	5
386	2013	O153/O178:H19	63,30,3,18,43,8,6	8369	no *eae*	*stx2f*	5
387	2008	Ou:H7	6,29,14,16,9,8,14	2005	no *eae*	*stx1a*	5
388	2014	OX18:H2	6,6,5,26,9,13,98	847	no *eae*	*stx2a stx2c*	5
389	2015	Ou:H16	6,23,33,1,7,8,156	3188	no *eae*	*stx1d*	5
390	2013	O55:H12	43,41,15,18,11,7,6	101	no *eae*	*stx1a*	5
391	2015	O22:H8	6,19,3,26,11,8,6	446	no *eae*	*stx2d*	2
392	2017	O187:H28	6,4,5,26,7,8,14	200	no *eae*	*stx2g*	5
393	2007	O187:H28	6,4,5,26,746,8,14	8656	no *eae*	*stx2a*	5
394	2015	O109:H21	6,4,5,26,20,8,14	40	*eae*	*stx2f*	5
395	2016	O21:H21	6,4,4,18,24,5,14	56	no *eae*	*stx2b*	5
396	2008	O113:H21	6,4,4,18,24,8,14	223	no *eae*	*stx2a*	5
397	2013	O49:H21	6,4,4,16,24,8,14	58	no *eae*	*stx1d*	5
398	2014	O30:H25	837,4,14,16,24,8,14	8660	no *eae*	*stx2g*	5
399	2003	O113:H21	9,19,5,18,11,122,14	3695	no *eae*	*stx2d*	2
400	2009	O123/O186:H10	9,19,5,18,11,122,14	3695	no *eae*	*stx2b stx1c*	5
401	2015	O130:H11	6,65,32,26,9,8,2	297	no *eae*	*stx2a stx2c stx1a*	5
402	2011	O130:H11	6,65,32,26,9,8,2	297	no *eae*	*stx2a stx1a*	5
403	2015	O130:H11	6,65,668,26,9,8,2	7931	no *eae*	*stx2a*	5
404	2014	O130:H11	6,65,668,26,9,8,2	7931	no eae	*stx2a stx2c*	5
405	2010	O171:H25	6,65,32,26,9,8,2	297	no *eae*	*stx2a*	5
406	2006	O179:H8	6,65,32,26,9,8,2	297	no *eae*	*stx2a*	5

407	2009	O81:H21	6,4,32,26,9,8,7	737	no *eae*	*stx1c*	5
408	2008	O43:H2	6,29,5,26,24,8,6	937	no *eae*	*stx2b*	5
409	2016	O91:H10	9,6,33,131,426,8,7	8629	no *eae*	*stx2d*	2
410	2015	O91:H10	9,6,33,131,24,8,7	641	no *eae*	*stx2a*	5
411	2006	O91:H10	9,6,33,131,24,8,7	641	no *eae*	*stx2b stx2c*	5
412	2012	O181:H49	6,6,15,16,42,46,7	173	no *eae*	*stx2a*	5
413	2011	O181:H49	6,6,15,16,42,46,7	173	no *eae*	*stx2a*	5
414	2014	O153/O178:H7	9,23,64,18,11,8,6	278	no *eae*	*stx1c*	5
415	2013	O153/O178:H7	9,23,64,18,11,8,6	278	no *eae*	*stx1c*	5
416	2009	O153/O178:H7	9,23,64,18,11,8,6	278	no *eae*	*stx1c*	5
417	2018	O153/O178:H7	9,23,64,18,11,8,6	278	no *eae*	*stx1c*	5
418	2013	O153/O178:H7	9,23,64,18,11,8,6	278	no *eae*	*stx2b stx1c*	5
419	2014	O153/O178:H7	9,23,64,18,11,8,6	278	no *eae*	*stx2b stx1c*	5
420	2016	O153/O178:H7	9,ND,64,18,11,8,6	8873	no *eae*	*stx2b stx1c*	5
421	2015	O34:H4	9,23,33,18,11,8,6	642	*eae*	*stx2f*	5
422	2015	O76:H7	9,23,33,18,9,8,6	795	*eae*	*stx2a*	1
423	2005	O104:H7	9,8,5,231,11,194,7	1817	no *eae*	*stx1c*	5
424	2017	O112ab:H8	6,6,5,10,20,23,6	75	no *eae*	*stx1a*	5
425	2010	O112ab:H8	6,6,5,10,20,23,6	75	no *eae*	*stx1a*	5
426	2008	O8:H8	6,29,14,16,24,7,2	1656	no *eae*	*None*	n/a
427	2017	O8:H19	9,65,5,1,9,13,6	162	no *eae*	*stx2e*	5
428	2014	O8:H19	6,65,5,1,9,8,6	201	no *eae*	*stx2e*	5
429	2006	O112ab:H2	6,4,12,18,24,7,6	388	no *eae*	*stx1c*	5
430	2014	O112ab:H2	6,4,12,652,24,7,6	6260	no *eae*	*stx1c*	5
431	2011	O104:H4	6,6,5,136,9,7,7	678	no *eae*	*stx2a*	1
432	2016	O104:H4	6,6,5,136,9,7,7	678	no *eae*	*stx2a*	1
433	2016	O174:H8	836,6,5,9,9,8,2	8630	no *eae*	*stx2b stx1c*	5
434	2008	O174:H8	836,6,5,9,9,8,2	8630	no *eae*	*stx2b stx1c*	5
435	2010	O174:H8	836,6,5,9,9,8,2	8630	no *eae*	*stx2b stx1c*	5
436	2013	O174:H8	6,6,5,9,9,8,2	13	no *eae*	*stx2b stx1c*	5
437	2002	O174:H8	6,6,5,9,9,8,2	13	no *eae*	*stx1a stx1c*	5
438	2016	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
439	2017	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
440	2014	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
441	2016	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
442	2015	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b*	5
443	2015	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
444	2012	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b*	5
445	2006	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
446	2003	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
447	2016	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5

448	2005	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
449	2009	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
450	2015	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
451	2015	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
452	2012	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx2b stx1a*	5
453	2017	O91:H14	6,4,4,1,22,8,7	33	no *eae*	*stx1a*	5
454	2017	Ou:H14	6,23,5,16,24,18,6	1249	no *eae*	*stx1c*	5
455	2017	Ou:H14	6,23,5,16,24,18,6	1249	no *eae*	*stx1c*	5
456	2015	Ou:H14	6,23,5,16,24,18,6	1249	no *eae*	*stx1c*	5
457	2016	OX25:H11	6,4,12,941,9,2,7	8644	no *eae*	*stx2a*	5
458	2014	OX18:H8	781,4,12,10,24,18,7	8658	*eae*	*stx2a*	1
459	2016	O87:H16	6,8,4,16,7,13,2	2101	no *eae*	*stx2b*	5
460	2005	O87:H16	6,8,4,16,7,13,2	2101	no *eae*	*stx2b*	5
461	2016	O78:H4	6,279,12,1,273,12,7	3101	no *eae*	*stx1c*	5
462	2017	O78:H4	6,279,12,1,273,12,7	3101	no *eae*	*stx1c*	5
463	2009	O78:H4	6,279,12,1,273,12,7	3101	no *eae*	*stx1c*	5
464	2014	O8:H9	6,4,12,1,20,12,7	88	no *eae*	*None*	n/a
465	2015	O102:H6	6,1091,12,1,20,293,7	7926	no *eae*	*stx2b*	5
466	2016	O182:H25	77,7,72,18,53,56,7	300	*eae*	*stx2a*	1
467	2014	O182:H25	77,7,72,18,53,56,7	300	*eae*	*stx1a*	4
468	2014	O182:H25	77,7,72,18,53,56,7	300	*eae*	*stx1a*	4
469	2014	O182:H25	77,7,72,18,53,56,7	300	*eae*	*stx1a*	4
470	2006	O182:H25	77,7,72,18,53,56,7	300	*eae*	*stx1a*	4
471	2018	O182:H25	77,7,72,18,53,56,7	300	*eae*	*stx1a*	4
472	2018	O182:H25	77,7,72,18,53,56,7	300	*eae*	*stx1a*	4
473	2013	O98:H21	77,7,7,18,54,18,7	306	*eae*	*stx1a*	4
474	2012	O98:H21	77,7,7,18,54,18,7	306	*eae*	*stx1a*	4
475	2010	O84:H2	77,7,7,18,54,18,7	306	*eae*	*stx1a*	4
476	2014	O150:H2	77,7,7,18,54,18,7	306	*eae*	*stx2a stx1a*	1
477	2014	O121:H19	100,23,68,45,1,35,7	655	*eae*	*stx2a*	1
478	2008	O121:H19	100,23,68,45,1,35,7	655	*eae*	*stx1a*	4
479	2013	O121:H19	100,23,68,45,1,35,7	655	*eae*	*stx2a*	1
480	2013	O121:H19	100,23,68,45,1,35,7	655	*eae*	*stx2a*	1
481	2008	O121:H19	100,23,68,45,1,35,7	655	*eae*	*stx2a*	1
482	2008	O121:H19	100,23,68,45,1,35,7	655	*eae*	*stx2a*	1
483	2008	O49:H-	231,4,567,10,20,7,7	8264	*eae*	*stx2a*	1
484	2008	O49:H-	231,4,567,10,20,7,7	8264	*eae*	*stx2a*	1
485	2014	O165:H25	48,46,43,45,11,34,35	119	*eae*	*stx2a stx2c stx1a*	1
486	2015	O165:H25	48,46,43,45,11,34,35	119	*eae*	*stx2a stx1a*	1
487	2016	O165:H25	48,46,43,45,11,34,35	119	*eae*	*stx2a stx1a*	1
488	2017	O165:H25	48,46,43,45,11,34,35	119	*eae*	*stx2a stx1a*	1

489	2014	O172:H25	121,46,43,45,11,34,35	660	*eae*	*stx2a*	1
490	2008	O172:H25	121,46,43,45,11,34,35	660	*eae*	*stx2a*	1
491	2014	O177:H25	1,46,69,1,20,34,7	342	*eae*	*stx2c*	3
492	2006	O177:H25	1,46,69,1,20,34,7	342	*eae*	*stx2c*	3
493	2009	O5:H9	1,46,69,1,20,34,7	342	*eae*	*stx1a*	4
494	2009	O5:H9	1,46,69,1,20,34,7	342	*eae*	*stx1a*	4
495	2009	O5:H9	1,46,69,1,20,34,7	342	*eae*	*stx1a*	4
496	2011	O5:H9	1,46,69,1,20,34,7	342	*eae*	*stx1a*	4
497	2003	O177:H25	1,46,123,1,20,34,7	659	*eae*	*stx2c*	3
498	2016	O113:H4	10,11,4,8,8,8,2	10	no *eae*	*stx2d*	2
499	2015	O113:H4	10,11,4,8,8,8,2	10	no *eae*	*stx2b stx1c*	5
500	2014	O113:H4	10,11,4,8,8,8,2	10	no *eae*	*stx2b stx1c*	5
501	2016	O113:H4	10,11,4,8,8,8,2	10	no *eae*	*stx2b stx1c*	5
502	2014	O113:H4	10,11,4,8,8,8,2	10	no *eae*	*stx2b stx1c*	5
503	2017	O113:H4	10,11,4,8,8,8,2	10	no *eae*	*stx2b stx1c*	5
504	2017	O113:H4	10,11,4,8,8,8,2	10	no *eae*	*stx2b stx1c*	5
505	2006	O113:H4	10,11,4,8,8,8,2	10	no *eae*	*stx2b stx1c*	5
506	2010	O38:H26	10,11,4,8,8,8,2	10	no *eae*	*stx2b stx1c*	5
507	2005	O118/O151:H12	10,11,4,8,8,8,2	10	no *eae*	*stx2b*	5
508	2009	O118/O151:H12	10,11,4,8,8,8,2	10	no *eae*	*stx2b*	5
509	2013	O49:H10	6,7,5,1,8,18,2	206	*eae*	*stx1a*	4
510	2013	O80:H2	78,27,5,10,12,8,2	301	*eae*	*stx2d*	2
511	2013	O80:H2	78,27,5,10,12,8,2	301	*eae*	*stx2d*	2
512	2014	O80:H2	78,27,5,10,12,8,2	301	*eae*	*stx2a*	1
513	2013	O180:H2	78,27,5,10,12,8,2	301	*eae*	*stx2a*	1
514	2016	Ou:H45	10,7,4,8,8,8,2	656	no *eae*	*stx2b*	5
515	2015	Ou:H45	10,7,4,8,8,8,2	656	no *eae*	*stx2b*	5
516	2014	O15:H16	6,11,5,1,8,8,6	325	no *eae*	*stx2g*	5
517	2017	O149:H8	64,92,5,83,24,8,6	344	no *eae*	*stx2a stx2g*	5
518	2005	Ou:H6	62,100,17,31,5,5,4	362	no *eae*	*stx2a*	5
519	2012	O17/O44:H18	21,35,27,6,5,5,4	69	no *eae*	*stx2d*	2
520	2012	O17/O44:H18	21,35,27,6,5,5,4	69	no *eae*	*stx2d*	2
521	2012	O17/O44:H18	21,35,27,6,5,5,4	69	no *eae*	*stx2d*	2
522	2004	Ou:H2	92,4,87,96,70,58,2	648	no *eae*	*None*	n/a

Appendix 3 - Acquired antimicrobial resistance genes and profiles

ID	Serotype	AMR Genes	AMR Classes
28	O117:H7	*aadA5,mph(A),sul1,dfrA17*	Aminoglycoside,Macrolide,Sulphonamide,Trimethoprim
29	O117:H7	*strA,strB,sul1,sul2,dfrA5*	Aminoglycoside,Sulphonamide,Trimethoprim
30	O117:H7	*strA,strB,blaTEM-1B,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
31	O156:H7	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
32	O156:H7	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
33	O50/O2:H6	*aadA1,blaSHV-48, sul1,tet(D)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
66	O145:H28	*strA,strB,blaTEM-1B,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
76	O145:H28	*strA,strB,blaTEM-1B,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
77	O145:H28	*strA,strB,blaTEM-1B,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
78	O145:H28	*strA,strB,blaTEM-1B,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
79	O145:H28	*strA,strB,blaTEM-1C,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
80	O145:H28	*strA,strB,blaTEM-1B,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
81	O145:H28	*strA,strB,blaTEM-1C,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
86	O145:H28	*strA,strB,aadA1,blaTEM-1B, sul1,sul2,tet(A),tet(B),dfrA1*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline,Trimethoprim
121	O26:H11	*blaTEM-1A*	Beta-lactamase
127	O26:H11	*strA,strB*	Aminoglycoside
128	O26:H11	*strA,strB,aadA1,sul1,sul2,tet(A),dfrA1*	Aminoglycoside,Sulphonamide,Tetracycline,Trimethoprim
129	O26:H11	*strA,strB,blaTEM-1C,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
130	O26:H11	*strA,strB,blaTEM-1B,catA1,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Phenicol,Sulphonamide,Tetracycline
131	O26:H11	*strA,strB,blaTEM-30*	Aminoglycoside,Beta-lactamase
136	O26:H11	*strA,strB,floR,sul2*	Aminoglycoside,Phenicol,Sulphonamide
141	O26:H11	*strA,strB,blaTEM-1B,floR,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Phenicol,Sulphonamide,Tetracycline
142	O26:H11	*strA,strB,blaTEM-1B,floR,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Phenicol,Sulphonamide,Tetracycline
151	O26:H11	*strA,strB,aadA1,mph(B),floR,sul1,sul2,tet(A),dfrA1*	Aminoglycoside,Macrolide,Phenicol,Sulphonamide,Tetracycline,Trimethoprim
152	O26:H11	*strA,strB,floR,sul2*	Aminoglycoside,Phenicol,Sulphonamide
159	O26:H11	*strA,strB*	Aminoglycoside
160	O26:H11	*strA,strB,blaTEM-1C,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
170	O26:H11	*blaTEM-1B*	Beta-lactamase
185	O26:H11	*aadA12,sul1,tet(A)*	Aminoglycoside,Sulphonamide,Tetracycline
190	O26:H11	*strA, strB,aph(3')-Ia,blaTEM-1B, sul1,sul2,tet(A),dfrA7*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline,Trimethoprim
195	O26:H11	*strA,strB,blaTEM-1B,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
196	O26:H11	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
203	O26:H11	*strA,strB,blaTEM-1B,sul2*	Aminoglycoside,Beta-lactamase,Sulphonamide
207	O26:H11	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
208	O26:H11	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
209	O26:H11	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
212	O26:H11	*blaTEM-1A*	Beta-lactamase
213	O26:H11	*blaTEM-1A*	Beta-lactamase
214	O26:H11	*strA,strB,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline

223	O26:H11	*strA,strB,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
224	O26:H11	*strA,strB,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
233	O26:H11	*strA*	Aminoglycoside
234	O26:H11	*strA,strB,blaTEM-1B,sul2,tet(B),dfrA8*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline,Trimethoprim
237	O26:H11	*aph(3')-Ia,tet(A)*	Aminoglycoside,Tetracycline
238	O26:H11	*strA,strB,blaTEM-1B,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
245	O26:H11	*strB*	Aminoglycoside
246	O26:H11	*strA,sul2,tet(A)*	Aminoglycoside,Sulphonamide,Tetracycline
255	O123/O186:H11	*strA,strB,aph(3')-Ia,blaTEM-1B, sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
256	O111:H8	*tet(A)*	Tetracycline
257	O111:H8	*strA,strB,aph(3')-Ia,blaTEM-1B,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
258	O111:H8	*strA,strB,aph(3')-Ia,blaTEM-1B, sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
259	O111:H8	*strA,strB,sul2,tet(A),tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
261	O111:H8	*strA,strB,aadA1,blaTEM-1B, mph(B),sul1,sul2,tet(A),dfrA1*	Aminoglycoside,Beta-lactamase,Macrolide,Sulphonamide,Tetracycline,Trimethoprim
263	O111:H8	*strA,strB,aph(3')-Ia,blaTEM-1B, sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
264	O111:H8	*strA,strB,blaTEM-1B,sul2,tet(A),dfrA5*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline,Trimethoprim
265	O111:H8	*strA,strB,blaTEM-1B,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
268	O111:H8	*strA,strB,aph(3')-Ia,blaTEM-1B,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
270	O71:H8	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
271	O71:H2	*strA,strB,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
272	O71:H2	*strA,strB,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
273	O103:H2	*strA,strB,sul2,tet(A)*	Aminoglycoside,Sulphonamide,Tetracycline
292	O103:H2	*strA, strB,blaTEM-1B,floR,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Phenicol,Sulphonamide,Tetracycline
302	O103:H2	*aadA1,sul1*	Aminoglycoside,Sulphonamide
303	O103:H2	*aadA1,sul1*	Aminoglycoside,Sulphonamide
304	O103:H2	*aadA1,sul1*	Aminoglycoside,Sulphonamide
308	O103:H2	*strA,strB,blaTEM-1B,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
310	O103:H2	*strA,strB,blaTEM-30,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline
313	O103:H2	*strA,strB,aadA1,blaTEM-1C, mph(B),sul1,sul2,tet(A),dfrA1*	Aminoglycoside,Beta-lactamase,Macrolide,Sulphonamide,Tetracycline,Trimethoprim
314	O103:H2	*strA,strB,aadA1,sul1,sul2,dfrA1*	Aminoglycoside,Sulphonamide,Trimethoprim
316	O103:H2	*strA,strB,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
317	O103:H2	*strA,strB,aph(3')-Ia,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
318	O103:H2	*strA,strB,aph(3')-Ia,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
326	O45:H2	*aadA5,sul2,tet(A),dfrA17*	Aminoglycoside,Sulphonamide,Tetracycline,Trimethoprim
344	O128ab:H2	*strA,strB,floR,sul2*	Aminoglycoside,Phenicol,Sulphonamide
345	O128ab:H2	*blaTEM-1A*	Beta-lactamase
348	O128ab:H2	*strA,strB,aadA1,blaTEM-1B, mph(B),sul1,sul2,tet(A),dfrA1*	Aminoglycoside,Beta-lactamase,Macrolide,Sulphonamide,Tetracycline,Trimethoprim
361	O146:H21	*strA,strB,sul2,dfrA14*	Aminoglycoside,Sulphonamide,Trimethoprim
362	O146:H21	*aadA12,sul1,tet(A)*	Aminoglycoside,Sulphonamide,Tetracycline

407	O81:H21	*strA,strB*	Aminoglycoside
431	O104:H4	*strA,strB,blaCTX-M-15,blaTEM-1B,sul1,sul2,tet(A),dfrA7*	Aminoglycoside,Beta-lactamase,Sulphonamide,tetracycline,Trimethoprim
432	O104:H4	*aadA5,aac(3)-IId,blaTEM-1B, mph(A),sul1,dfrA17*	Aminoglycoside,Beta-lactamase,Macrolide,Sulphonamide,Trimethoprim
446	O91:H14	*strB*	Aminoglycoside
449	O91:H14	*strA,strB*	Aminoglycoside
491	O177:H25	*strA,strB,sul2,tet(B)*	Aminoglycoside,Sulphonamide,Tetracycline
493	O5:H9	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
494	O5:H9	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
495	O5:H9	*strA,strB,sul2*	Aminoglycoside,Sulphonamide
496	O5:H9	*strA,strB,aph(4)-Ia,aac(3)-Iva, sul2,tet(31)*	Aminoglycoside,Sulphonamide,Tetracycline
510	O80:H2	*strA,strB,aadA1,aadB,aph(3')-Ia, blaTEM-1B,catA1,floR,sul1,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Phenicol,Sulphonamide,Tetracycline
511	O80:H2	*strA,strB,aac(3)-Iva,aph(3')-Ia,aph(4)-Ia, blaTEM-1B,sul2,tet(A),dfrA5*	Aminoglycoside,Beta-lactamase,Sulphonamide,Tetracycline,Trimethoprim
512	O80:H2	*strA,strB,aadA1,aadB,aph(3')-Ia, blaTEM-1B, catA1,floR,sul1,sul2,tet(A)*	Aminoglycoside,Beta-lactamase,Phenicol,Sulphonamide,Tetracycline
515	Ou:H45	*blaTEM-1B*	Beta-lactamase

Appendix 4. Virulence genes listed in order of frequency detected among the strains.

Gene	Name	No. strains carrying gene	Category
FimH	Type 1 fimbrial protein	518	Adhesin
gad	Glutamate decarboxylase	512	Acid Resistance
iss	Increased serum survival	424	Other
ehxA	Enterohaemolysin	399	Toxin
hlyD	Haemolysin D	398	Toxin
lpfA	Long polar fimbriae	339	Adhesin
eae	Intimin	336	Adhesin
espA	Type III secretion system	336	Secretion system
tir	Translocated intimin receptor protein	336	Adhesin
nleB	Non-LEE encoded effector B	319	Secretion system
iha	Adherence protein	318	Adhesin
astA	Heat-stable enterotoxin 1	316	Toxin
espF	Type III secretion system	312	Secretion system
espJ	Prophage-encoded type III secretion system effector	302	Secretion system
cif	Type III secreted effector	298	Secretion system
espB	Secreted protein B	272	Secretion system
nleA	Non-LEE encoded effector A	272	Secretion system
espP	serine protease	242	SPATE
efa1	EHEC factor for adherence	231	Adhesin
nleC	Non-LEE encoded effector C	225	Secretion system
katP	Plasmid-encoded catalase peroxidase	211	Other
PAI(MalX)	Pathogenicity island	209	Other
toxB	Toxin B	189	Toxin
fyuA	Ferric yersiniabactin receptor	186	Siderophore
celb	Endonuclease colicin E2	176	Colicin
TraT	Outer membrane lipoprotein	165	Other
cba	Colicin B	163	Colicin
tccP	Tir-cytoskeleton coupling protein	136	Secretion system
subA	Subtilase toxin subunit	136	Toxin
ireA	Siderophore receptor	120	Other
mchF	ABC transporter protein MchF	112	Microcin
mchB	Microcin H47 part of colicin H	102	Microcin
mchC	MchC protein	102	Microcin
espI	serine protease	88	SPATE

senB	Plasmid-encoded enterotoxin	77	Toxin
etpD	Type II secretion protein	58	Secretion system
cma	Colicin M	56	Colicin
epeA	serine protease	32	SPATE
ibeA	Invasion of brain endothelium protein A	31	Other
hra	heat-resistant agglutinin	31	Other
iutA	Ferric aerobactin receptor	28	Siderophore
espC	serine protease	27	SPATE
mcmA	Microcin M part of colicin H	27	Microcin
eilA	Salmonella HilA homolog	25	Other
air	Enteroaggregative immunoglobulin repeat protein	25	Other
iroN	Enterobactin siderophore receptor protein	19	Other
pic	Serine protease	17	SPATE
cvi-cvaC	Transporter accessory protein- Colicin V immunity protein	16	Other
usp	Uropathogenic-specific protein gene	15	Other
saa	Auto agglutinating adhesin	14	Adhesin
hlyF	Haemolysin F	11	Toxin
OmpT	Outer membrane protease	10	Other
capU	Hexosyltransferase homolog	9	Other
vat	serine protease	8	SPATE
sfa/foc	S and F1C fimbriae	8	Adhesin
kpsMK5II	Capsule synthesis proteins	7	Other
kpsMK1	Capsule synthesis protein	7	Other
sigA	Secretory immunoglobin A	7	SPATE
cdtB	Cytolethal distending toxin B	5	Toxin
sta1	Heat-stabile enterotoxin ST-Ia	5	Toxin
hlyA	Haemolysin A	4	Toxin
aaiC	a*ggR*-activated island C	4	Secretion system
sepA	Serine protease	3	SPATE
papC	P Fimbriae	2	Adhesin
papG-allele-II	P Adhesin	2	Adhesin
aap	dispersin	2	Adhesin
aar	AggR-activated regulator	2	Adhesin
aatA	Dispersin transporter protein	2	Adhesin
aggR	AraC transcriptional activator	2	Adhesin
ORF3	Isoprenoid Biosynthesis	2	Other
ORF4	Putative isopentenyl-diphosphate delta-isomerase	2	Other
clbB	Marker for the 5' region of the pks island	2	Other
clbN	Marker for the 3' region of the pks island	2	Other
papEF	P Adhesin	2	Adhesin
bmaE	M-agglutinin	2	Other
rfc	O antigen polymerase	2	Other

agg3A	AAF/III major fimbrial subunit	1	Adhesin
agg3B	AAF/III minor adhesin.	1	Adhesin
agg3C	Usher, AAF/III assembly unit	1	Adhesin
agg3D	Chaperone, AAF/III assembly unit	1	Adhesin
cnf1	Cytotoxic necrotizing factor	1	Toxin
FocG	F1C Fimbrial adhesin	1	Adhesin
sfaS	S-fimbriae minor subunit	1	Adhesin
f17A	F17 fimbrial protein	1	Adhesin
f17G	F17 fimbrial protein	1	Adhesin
aggA	AAF/I major fimbrial subunit	1	Adhesin
aggB	AAF/I minor adhesin	1	Adhesin
aggC	Usher, AAF/I assembly unit	1	Adhesin
aggD	Chaperone, AAF/I assembly unit	1	Adhesin

[1] SPATE Serine Protease Autotransporters of *Enterobacteriaceae*

[2] Bold denotes genes only detected in *eae*-ve strains.