Review Article

Enteroaggregative E. coli O104 from an outbreak of HUS in Germany 2011, could it happen again?

Marie Anne Chattaway¹, Tim Dallman¹, Iruke N. Okeke², John Wain¹

¹Laboratory of Gastrointestinal Pathogens, Health Protection Agency, London, England
²Molecular Microbiology, Haverford College, Pennsylvania, USA

Abstract

Enterohaemorrhagic E. coli (EHEC) particularly O157:H7 (Sequence type 11 complex), is the best documented and most well-known of E. coli that cause diarrhoea. The importance of EHEC lies in the severity of disease. Outbreaks can infect thousands of people causing bloody diarrhoea and haemolytic uremic syndrome (HUS) that in turn can result in protracted illness or even death. The ability of EHEC to colonise the human gut is normally associated with the presence of genes from another group of diarrhoeagenic E. coli, the enteropathogenic E. coli (EPEC), via the locus of enterocyte effacement. However, the massive outbreak in Germany was caused by an EHEC which had acquired virulence genes from yet another group of diarrhoeagenic E. coli, the enteroaggregative E. coli (EAEC). In reality EAEC is probably the most common bacterial cause of diarrhoea but is not identified in most diagnostic laboratories. This outbreak emphasises the importance of being able to detect all diarrhoeagenic E. coli and not to focus on E. coli O157:H7 alone. Routine surveillance systems for EAEC, a once ignored global pathogen, would go a long way to reaching this goal. This review describes methods for identifying non-O157 EHEC and describes the key genetic features of EHEC and EAEC. Our aim is to provide information for laboratories and policy makers which enables them to make informed decisions about the best methods available for detecting newly emergent strains of diarrhoeagenic E. coli.

Key words: EAEC; EHEC; HUS; Outbreak O104:H4; ST678


(Received 01 July 2011—Accepted 02 July 2011)

Introduction

Beginning in early May 2011, an unusually high number of haemolytic uraemic syndrome (HUS) cases were reported in Germany. The outbreak was caused by an enterohaemorrhagic E. coli (EHEC) which had characteristics of both a verotoxigenic E. coli (VTEC) (for a guide to nomenclature see Table 1) and of the less well-known diarrhoeagenic E. coli, enteroaggregative E. coli (EAEC). There is considerable expertise in diarrhoeagenic E. coli in Germany but even with support from the reference laboratory in Rome, the pathogen responsible for the outbreak proved challenging to characterize. Within most diagnostic laboratories the current methodology for VTEC detection is aimed at detecting sorbitol negative VTEC O157:H7 and for most European countries, the sorbitol positive outbreak strain O104:H4 could not be detected. It is therefore important that we examine the methods used by diagnostic and public health microbiology laboratories to characterise VTEC isolates and begin the process of global standardisation. A universal approach based on genomic features would be more generally applicable and transportable than current methods.

The medical care provision required to manage thousands of patients with haemolytic uraemic syndrome (HUS) was a major challenge. Even the well-funded hospitals in Northern Germany were forced to loan dialysis and other medical equipment to manage the unexpected case load. Boosting diagnostic and epidemiological apparatus to improve source attribution during outbreaks is imperative to reduce the burden on already overstretched health care facilities. It is now clear that the ability to isolate and identify novel, e.g. non-O157:H7 VTECs, as well as known diarrhoeagenic E. coli (DEC) must be considered.

Strains with combinations of virulence factors from different E. coli pathotypes have been described before but it is the size and severity of the outbreak in Germany which has highlighted the importance and unpredictability of the consequences of genetic exchange amongst gut bacteria. This review will present what is currently known about the outbreak strain and discuss the preliminary genomic analysis
in the context of what is known about other isolates from the EAEC and VTEC pathogroups. It will also highlight an important lesson learned from this outbreak – the importance of a global epidemiological capacity, encompassing the developing world, to detect novel and emerging pathogens in addition to well-known ones.

The outbreak strain

Initial testing by German laboratories showed that the strain associated with the outbreak was of sequence type (ST) 678, serotype O104:H4, and contained genetic elements found in both EHEC (vtx) and EAEC (aggR). ST678 also contains the EAEC (55989) sequenced strain and the HUS causing O104:H4 VTEC (deposited on the public MLST database by Karch in 2001). The most closely related sequence type (the ST25 group) (Figure 1) is a VTEC O128:H2. The O128:H2 serotype (although no sequence type data is available) has been previously seen in sheep [1] and also isolated from infantile EPEC infections [2]. The group of E. coli most closely related to the outbreak strain are therefore a mixture of pathotypes: EAECs, EHECs and EPECs.

Although serotyping data in the MLST database is incomplete, it is clear that E. coli within the O104 serogroup occurs in several different unrelated STs. Therefore, this serogroup does not represent a related group of organisms and comparison of the outbreak strain with other O104 isolates might be redundant. However, including the flagella type, O104:H4 seems to describe a very closely related group of isolates, all within ST678. Members of this sequence type can belong to either the VTEC or the EAEC pathotypes (Figure 1) which can be associated with HUS in humans [3,4] but have not been commonly isolated in Europe. Possibly because detection techniques are optimised for VTEC O157:H7 in diagnostic laboratories. Methods traditionally used to detect the commonly known VTEC O157 were not successful (O157 agglutination negative, sorbitol positive) and a combination of phenotypic and genotypic methods were necessary. The following sections describe the methods used at the reference laboratories.

Figure 1. Spanning Tree of MLST data for EHEC and EAEC
supplemented by published methods, and describe the potential for using genomic data.

**Enrichment and isolation**

Isolation directly from faecal specimens, stored at 4-8°C, was performed as soon as possible as viability of the organism decreases each day and plasmid loss may occur. An enrichment broth of Modified Tryptone Soya was inoculated with mixed faecal matter. The faeces were directly plated onto selective cefixime tellurite sorbitol MacConkey agar (CT-SMAC) and the EHEC ST678 (O104:H4) colonies grew very well producing a creamy pink morphology. The broth and agar plates were incubated aerobically at 37°C overnight. If CT-SMAC selective plates are unobtainable, other methods can be used such as exploiting the antibiotic resistant properties of this strain and using MacConkey agar supplemented with streptomycin (20 mg/ml) and/or tetracycline (10 mg/ml) [5]; commercial media is now available for this outbreak strain. Although useful for the O104:H4 outbreak strain, it is possible that other outbreaks will occur in which this supplemented media may not be appropriate.
Typing methods

Microbial typing relied on the isolation of the outbreak strain from faeces and this was straightforward because positive cases grew as an almost pure culture on CT-SMAC plates. Single colonies were picked and tested against O104:H7 antigens. Somatic antibodies (including O104) are available commercially. Molecular serotyping was performed using the O104 antigen-associated gene (wzxO104) and the gene encoding the H4 flagellar antigen (fliCH4) [6,7]. Antibodies raised against the O104 antigen are also positive with the K9 capsular antigen; therefore, O8:K9 and O9:K9 antigens can also be positive. Separate O8, O9 and O9a specific PCR has been carried out in other studies to rule out these other serotypes [8].

For serological typing of unknown isolates, screening against the whole panel of at least 185 somatic and 56 flagella E. coli antigens might be necessary. This is likely to be conducted by regional or international reference laboratories; the local testing laboratories can then purchase the antibodies for testing.

Alternatively, the genes that encode the specific O antigens in E. coli are clustered in the genome [9] and DNA sequencing can be used to predict the serotype [10].

Virulence detection

For detection of virulence factors by PCR, extraction of DNA was performed from the enrichment broth. Targets recommended for detecting the outbreak strain (Table 2 and Figure 2) include EHEC targets such as the rarely found vtx1, the commonly found vtx2 and intimin (eae), an adhesion factor responsible for the attaching and effacing (A/E) lesions found in EHEC and EPEC [11]. The outbreak strain was vtx2 positive and vtx1/eae negative.

Although a sub-typing scheme is available for vtx1 and vtx2 [12] and described for detecting the outbreak sub-type vtx2a, the variation at the nucleotide level is difficult to detect by PCR and needs careful optimisation. Detection using the generic vtx2 primers and the presence of the EAEC plasmid with the absence of intimin was considered to be sufficient for the screening of the outbreak strain.

The PCR targets described for EAEC are not as stable as EHEC possibly because most are plasmid encoded; plasmids are variable and sometimes they may be lost completely during culture in the laboratory. Targets used for detecting the outbreak strain included a regulator (aggR) [13] of multiple EAEC virulence factors including an anti-aggregator transporter gene (aat) [14] and a dispersing protein (aap) that coats the bacterial surface [15], although this marker has also been found in other E. coli [16]. These gene products are linked in that they all play a role in the EAEC colonisation of the gut by aiding the translocation of dispersin across the membrane [17] and are usually found together.

Multilocus sequence typing

Multilocus sequence typing (MLST [http://www.mlst.net]) of seven gene loci (adk, fumC, gyrB, mdh, purA & recA) [18], define the outbreak strain as a member of ST 678. The combination of MLST and virulence marker targets (vtx1, vtx2, eae, aggR etc) is a robust and accessible test that can accurately identify strains of all E. coli, including unusual EHEC’s. Sequence type profiles should ideally be submitted to the public database so that we can start to gain an understanding of the E. coli pathotypes causing disease globally and enabling the assignment of new alleles and STs.

Detection of other EHEC strains

Focusing on the serotype of EHEC outbreak strains has led to a bias in laboratory testing for the detection of O157:H7. Recent research studies in both developing and developed countries have shown that non-O157 EHEC strains are prevalent and can be more dominant than O157:H7 in some geographical areas [19-22]. However, front-line laboratories still test only for VTEC O157 and so the true burden of non-O157:H7 EHEC is not known. For example, many O157 EHEC are sorbitol-negative, but the strain in this outbreak was not, nor did it react with the common antisera for EHEC, such as those recognizing O157, O26, O111 and H7 antigens.

A multiplex PCR has recently been described to specifically detect the most common toxin producing VTEC serogroups (O157, O103, O91, O113, O145, O111, and O26) [10] yet O104 had not been recognised as a potential pathogen. A microarray has also been designed to detect the most clinically relevant EHEC with additional targets for O104, O121, O118, O45 and O55 included [6]; however, microarray technology is not feasible to implement in most laboratories. A comprehensive selection of EHEC flagella antigens was included in these studies but they were selected from a historical prospective
### Table 1. Nomenclature of Diarrhoeagenic *E.coli* and Enterohaemorrhagic *E. coli* toxin sub-types

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEC</td>
<td>Diarrhoeagenic <em>Escherichia coli</em></td>
<td>Any defined group of <em>E. coli</em> which has been associated with the ability to cause diarrhoea</td>
</tr>
<tr>
<td>DEAC</td>
<td>Diffusely-adherent <em>Escherichia coli</em></td>
<td>A group of <em>E. coli</em> which has been associated with the ability to cause diarrhoea defined by a specific pattern of adherence using the HEp-2 cell assay</td>
</tr>
<tr>
<td>EAEC</td>
<td>Enteroaggregative <em>Escherichia coli</em></td>
<td>A group of <em>E. coli</em> which has been associated with the ability to cause diarrhoea defined by a specific pattern of aggregation using the HEp-2 cell assay</td>
</tr>
<tr>
<td>EIEC</td>
<td>Enteroinvasive <em>Escherichia coli</em></td>
<td>A group of <em>E. coli</em> which has been associated with the ability to cause diarrhoea defined by the presence of invasion genes also found in <em>Shigella</em>.</td>
</tr>
<tr>
<td>EPEC</td>
<td>Enteropathogenic <em>Escherichia coli</em></td>
<td>A group of <em>E. coli</em> which has been associated with the ability to cause diarrhoea originally defined as specific serotypes and by a specific pattern of adherence using the HEp-2 cell assay but now by the presence of certain virulence factors including the locus of enterocyte effacement and associated effectors</td>
</tr>
<tr>
<td>ETEC</td>
<td>Enterotoxigenic <em>Escherichia coli</em></td>
<td>A group of <em>E. coli</em> which has been associated with the ability to cause diarrhoea defined by the presence of heat stable or heat labile toxins</td>
</tr>
<tr>
<td>VTEC</td>
<td>Verocytotoxic <em>Escherichia coli</em></td>
<td>A group of <em>E. coli</em> which has been associated with the ability to cause diarrhoea defined by the presence of a toxin gene, <em>vtx</em>, which has activity against cultured vero cells</td>
</tr>
<tr>
<td>STEC</td>
<td>Shiga Toxin-Producing <em>Escherichia coli</em></td>
<td>A group of <em>E. coli</em> which has been associated with the ability to cause diarrhoea defined by the presence of a toxin gene, <em>stx</em>, because of genetic similarity with the toxin of <em>Shigella dysenteriae</em>.</td>
</tr>
</tbody>
</table>

The terms *stx/vtx* or STEC/VTEC are entirely interchangeable and here we follow the European reference laboratories guidance and use *vtx* and VTEC.

For *E. coli* these two gene names are synonymous – only in *Shigella dysenteriae* type 1 is *stx* used exclusively. The discussion about which name should be used revolves around the scientifically agreed use of the same gene name for genes which show homology (shared ancestry); *vtx*1 and *stx* are homologues but for *vtx2/stx* this may not be true.

*stx1/vtx1* | Toxin gene type 1 | Several genetic variants including *vtx1a, vtx1c, vtx1d* |
<p>|<em>stx2/vtx2</em> | Toxin gene type 2 | Several genetic variants including <em>vtx2a, vtx2b, vtx2c, vtx2d, vtx2e, vtx2f and vtx2g</em>. |</p>
<table>
<thead>
<tr>
<th>Target</th>
<th>Function</th>
<th>Pathotype</th>
<th>Primer Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>aat</td>
<td>Anti-aggregator transporter gene</td>
<td>EAEC</td>
<td>Forward: CTG GCG AAA GAC TGT ATC AT Reverse: CAA TGT ATA GAA ATC CGC TGT T</td>
<td>[58]</td>
</tr>
<tr>
<td>aggR</td>
<td>Regulator multiple EAEC virulence factors</td>
<td>EAEC</td>
<td>Forward: CTA ATT GTA CAA TCG ATG TA Reverse: AGA GTC CAT CTC TTT GAT AAG</td>
<td>[59]</td>
</tr>
<tr>
<td>aap</td>
<td>Anti-aggregation protein (dispersin)</td>
<td>EAEC</td>
<td>Forward: CTT GGG TAT CAG CTC GAA TG Reverse: AAC CCA TTC GGT TAG AGC AC</td>
<td>[59]</td>
</tr>
<tr>
<td>elitB</td>
<td>Heat Liable Toxin</td>
<td>ETEC</td>
<td>Forward: CTA ATT GTA CAA TCG ATG TA Reverse: AGA GTC CAT CTC TTT GAT AAG</td>
<td>[60]</td>
</tr>
<tr>
<td>estA</td>
<td>Heat Stable Toxin</td>
<td>ETEC</td>
<td>Forward: CTA ATT GTA CAA TCG ATG TA Reverse: AGA GTC CAT CTC TTT GAT AAG</td>
<td>[61]</td>
</tr>
<tr>
<td>Ipa-H</td>
<td>Invasion plasmid</td>
<td>EIEC</td>
<td>Forward: GTT CCT TGA CCG CCT TTC CTA CGT C Reverse: GCC GGT CAG CCA CCC TCT GAG AGT AC</td>
<td>[62]</td>
</tr>
<tr>
<td>EAF</td>
<td>EPEC adherence factor</td>
<td>EPEC</td>
<td>Forward: CAG GGT AAA AGA AAG ATG ATA A Reverse: TAT GGG GAC CAT GTA TTA TCA</td>
<td>[63]</td>
</tr>
<tr>
<td>eaeA</td>
<td>etaching and effacing gene</td>
<td>EPEC/EHEC</td>
<td>Forward: CTG AAC GGC GAT TAC GCG AA Reverse: CCA GAC GAT ACG ATC CAG</td>
<td>[64]</td>
</tr>
<tr>
<td>afa</td>
<td>Afrimbrial adhesion</td>
<td>DAEC</td>
<td>Forward: GCT GGG CAG CAA ACT GAT AAC TCT Reverse: CAT CAA GCT GTT TGT TCG TCC GCC G</td>
<td>[65]</td>
</tr>
<tr>
<td>Stx1</td>
<td>Shiga toxin 1</td>
<td>EHEC</td>
<td>Forward: CCCCAGTTCAW GTRAGRTCMACRTC Reverse: GCC GGT CAG CCA CCC TCT GAG AGT AC Probe: Cy5-CTGGATGATGCTCAGTGGCGTTCTTATGTAA-BHQ</td>
<td>[66]</td>
</tr>
<tr>
<td>eae</td>
<td>intimin</td>
<td>EHEC</td>
<td>Forward: CAT TGA TCA GGA TTT TTC TCG GGA TAG Reverse: CTC ATG CCG AAA TAG CCG TTA Probe: Yak-ATAGTCTCGCCAGTATTCGCCACCAATACC-BHQ</td>
<td>[67]</td>
</tr>
<tr>
<td>Stx2</td>
<td>Shiga toxin 2</td>
<td>EHEC</td>
<td>Forward: CTG AACTTTAATTGAGATGTGCTCACA Reverse: CCAGTCAACGCTAAGGATGTTCAAC Probe: Yak-TACTGCACTGGACATCGTGAATACGCTCC-BHQ</td>
<td>[66]</td>
</tr>
</tbody>
</table>

**Additional Specific ST 678 (O104:H4) Outbreak Targets**

<table>
<thead>
<tr>
<th>Target</th>
<th>Function</th>
<th>Pathotype</th>
<th>Primer Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>flfCH4</td>
<td>Gene encoding H4 flagella antigen</td>
<td>E. coli</td>
<td>Forward: GCTGGGGGCTAACAAGTCAAC Reverse: CATGTCCCGAAGGATGTTCAAC</td>
<td>[7]</td>
</tr>
<tr>
<td>Stx2a</td>
<td>Sub-type of toxin 2a</td>
<td>EHEC</td>
<td>Forward: GGCAACCTTACTGTAATGCGTGC Reverse: 2: GGCCACCTTACTGTAATGCGTGC</td>
<td>[12]</td>
</tr>
</tbody>
</table>

*These are examples of published primers used, ideally primers should be self-designed and optimised to keep up with sequence variation.

EPEC: enteropathogenic E. coli; EHEC: enterohaemorrhagic E. coli; ETEC: enterotoxigenic E. coli; EIEC: enteroinvasive E. coli; DAEC: diffusely-adherent E. coli; EAEC: enteroaggregative E. coli.*
and so emergence of unforeseen serological profiles such as H4 was unexpected. VTEC isolates can be of 60 O/H types [22] and to include all O/H types for detection on the front line is impractical. Relying on serological typing of common EHEC antigens has led to insufficient systems within the front-line laboratories to detect emerging EHEC outbreaks. The switch to molecular serotyping will overcome this problem but will always be problematic for new serotypes.

The importance of designing globally relevant tests to detect the virulence and background of circulating strains is clear and this outbreak has shown the adaptability and ability of E. coli to accumulate virulence genes. Therefore, several pathotypes of E. coli should now be considered when screening for EHEC strains to help identify emerging hybrid strains. To do this for verocytotoxin producing E. coli strains belonging to different serotype requires toxin assays or molecular identification of the toxin genes.

The best option for the majority of laboratories is DNA-based diagnostics for multiple DEC genes (Figure 2). Multiple genes must be sought since different strains can harbour different combinations of known virulence loci, especially in heterogeneous groups such as EAEC. Thus isolates such as the recent ST678 (O104:H4) outbreak strain can only be reliably identified using molecular methods. Molecular methods can also be used to track virulence genes in specimens or suspected sources that may no longer contain live organisms, an important feature for outbreak analyses. Moreover, the versatility of these methods means that they can be adapted when new strains appear, which is important because we cannot predict when or where a new hypervirulent E. coli strain will appear.

For detecting future EHEC strains, the same methodology used for the detection of the outbreak strain could be employed but with the addition of other DEC virulent targets: the invasive gene (ipaH) for enteroinvasive E. coli (EIEC); heat labile (eltB) and heat stable toxin (estaA) for enterotoxigenic E. coli (ETEC); the enteropathogenic E. coli (EPEC) adherence factor (EAF) and the afimbrial adhesion gene (afa) for diffusely adherent E. coli (DAEC) (Table 2 and Figure 2).

**O104:H4 outbreak strain – genetic content**

The O104:H4 outbreak was a verocytotoxin producing E. coli strain containing vtx2; however, the strain is different from many VTEC strains because it lacks both the locus for the enterocyte effacement (LEE) pathogenicity island and the EHEC virulence plasmid. The strain tested positive in initial screens for aggR, which encodes a transcriptional regulator of aggregative adherence genes and is located on the virulence plasmid of many EAEC strains. The strain has since been shown to possess an aggregative adherence plasmid and to demonstrate aggregative adherence; it carries an aggregative adherence plasmid as well as the verocytotoxin gene, the two elements that were of most interest to clinical microbiologists. However, there are also other multiple prophages, transposons and a number of horizontally-acquired antimicrobial resistance genes.

**Enteraggregative E. coli (EAEC)**

Enteraggregative E. coli (EAEC) is a large, diverse pathogroup of diarrhoeegenic E. coli (DEC) which was defined in 1987 when it was observed that some non-toxigenic strains of E. coli from cases of diarrhoea were not adhering to HEp-2 cells in the localised pattern typical of classical enteropathogenic E. coli (EPEC) but aggregated in a stacked brick formation [23,24]. Early research on EAEC linked these strains to persistent diarrhoea in children in developing countries but EAEC have since been shown to be an important cause of acute diarrhoea as well, and to be important in the etiology of intestinal infections in industrialized countries [25].

EAEC are known for their heterogeneity and although there are serotypes associated with this group, such as O44:H18, O111:H12, O125, and O126:H7 [25-29], they are not unique to EAEC. Studies have shown a wide selection of EAEC serotypes and many are untypeable [30-32]; therefore, serotyping is not a useful tool in distinguishing this problematic group.

The group contains organisms of multiple lineages [24] which harbour a virulence plasmid; because the HEp-2 assay is difficult to perform and interpret, it is detection of the virulence plasmid which forms the mainstay for identification and so diagnosis of the disease. The following are problems associated with the use of plasmid markers: plasmids have variable gene content; plasmids may be lost on sub-culture; and the plasmid may transfer and be detected in entirely unrelated bacteria which are not actually able to cause diarrhoea.

This group is a main cause of health costs in the developing world but its variable pathogenicity means that funding has not been a priority and comparative pangenome analysis of EAEC in relation
to other *E. coli* pathotypes and commensals has not been extensively conducted. It has therefore not been possible to define unique stable chromosomal markers for identification. One chromosomal maker (also known to be plasmid encoded) is the *pic* gene which is present in the sequenced 042 strain [33]. This gene is a multi-functional secreted protease but is not unique in the Enterobacteriaceae. Flanking sequence around this gene in EAEC and *Shigella* is different, suggesting that this gene has been acquired by horizontal transfer [34].

These problems with diagnostics have resulted in a poor understanding of this heterogeneous pathotype, which has in turn led to a lack of knowledge of its true burden and impact on human health. Despite ample evidence that EAEC is the most common DEC [35-39], it remains less well-known compared to EPEC, EIEC AND ETEC (Table 1).

**Outbreaks of EAEC**

There have been some reports of this organism being associated with outbreaks, the largest of which was in Japan in 1993 when 2,697 schoolchildren became ill after eating food contaminated with EAEC with their lunch [40]. Although evidence pointed to white radish sprouts in the stir-fried vegetables, the bacteria were never isolated from the most likely food source. Multiple outbreaks in association with EAEC have been reported in the United Kingdom in association with public functions such as restaurants, hotels and conference centres [41]. EAEC has also caused outbreaks in hospitals (in Serbia 19 babies were infected in a neonatal ward [42]); from well water (in India 20 cases were reported including multiple age groups [43]); and from food (seen in 24 cases in an Italian holiday resort associated with cheese made with unpasteurized sheep milk). Furthermore, they may be an animal reservoir for some EAEC strains [44]. There is no common serotype associated with EAEC outbreaks.

**Genomics of EAEC**

At least four EAEC genomes have been completed, are nearly completed, or are in progress (Table 3). EAEC strain 042 produced diarrhoea in three of five adult volunteers in a challenge study in which other EAEC strains tested did not produce symptoms [45]. Strain 101-1 was responsible for the largest documented EAEC outbreak prior to 2011 [40]. Interestingly, although the 101-1 did harbour some virulence genes seen in strain 042, its presumed hypervirulence as suggested by the outbreak remained enigmatic for several years. Recent data demonstrates that in addition to multiple horizontally acquired virulence genes, strain 101-1 harbours a pathoadaptive mutation [46]. Like *Shigella* and some EHEC lineages, it has lost the lysine decarboxylase or *cad* genes. Inserting these genes onto the chromosome of 101-1 attenuates the strain. [46]. The genome of strain 55989 (source) is also in progress; of the four fully or partially sequenced EAEC genomes begun in June, it is this strain that shares the most genomic sequence with the ST 678 (O104:H4) outbreak isolate whose draft genome sequence was completed in June.

**Enterohemorrhagic *E. coli* (EHEC)**

Enterohemorrhagic *E. coli* (EHEC) causes haemorrhage of the intestinal tract of humans. The mechanism for this is complex but, for EHEC infection, always involves a toxin [47] called verotoxin (*vtx*) or shigatoxin (*stx*) (Table 1). Originally described as a rare *E. coli* serotype in 1983 [48] causing hemorrhagic colitis, O157:H7 VTEC, a cow-adapted *E. coli*, has since expanded in the bovine population and spill-over into humans, associated with disease, is such that it is currently the

---

**Table 3. Table of sequenced strains for Enteroaggregative *E. coli* and Enterohaemorrhagic *E. coli***

<table>
<thead>
<tr>
<th>Strain Name</th>
<th>Accession Number</th>
<th>Genome size (MB)</th>
<th>Serotype</th>
<th>ST</th>
<th>AAF</th>
<th>Phylotype</th>
<th>VT</th>
</tr>
</thead>
<tbody>
<tr>
<td>042</td>
<td>FN554766</td>
<td>5.35</td>
<td>O44:H18</td>
<td>414</td>
<td>II</td>
<td>D</td>
<td>N/A</td>
</tr>
<tr>
<td>101-1</td>
<td>AAMK000000000</td>
<td>4.98</td>
<td>O111:H10</td>
<td>493</td>
<td>II</td>
<td>B2</td>
<td>N/A</td>
</tr>
<tr>
<td>55989</td>
<td>CU928145.2</td>
<td>5.15</td>
<td>Unknown</td>
<td>678</td>
<td>III</td>
<td>B1</td>
<td>N/A</td>
</tr>
<tr>
<td>H112180280</td>
<td>AP010953.1</td>
<td>5.5</td>
<td>O104:H4</td>
<td>678</td>
<td>I</td>
<td>B1</td>
<td>stx2</td>
</tr>
<tr>
<td>O111</td>
<td>AP010960.1</td>
<td>5.80</td>
<td>O111:H-</td>
<td>16</td>
<td>N/A</td>
<td>B1</td>
<td>stx1/stx2</td>
</tr>
<tr>
<td>O26</td>
<td>AP010953.1</td>
<td>5.86</td>
<td>O26:H11</td>
<td>21</td>
<td>N/A</td>
<td>B1</td>
<td>stx2</td>
</tr>
<tr>
<td>O103</td>
<td>AP010958.1</td>
<td>5.48</td>
<td>O103:H2</td>
<td>17</td>
<td>N/A</td>
<td>B1</td>
<td>stx1/stx2</td>
</tr>
<tr>
<td>Sakai</td>
<td>BA000007.2</td>
<td>5.60</td>
<td>O157:H7</td>
<td>11</td>
<td>N/A</td>
<td>E</td>
<td>stx1/stx2</td>
</tr>
</tbody>
</table>
most commonly isolated EHEC (ST11 complex). Another commonly isolated sub-type is the ST21 complex EHECs which are predominantly serotype O26:H11 but may also be O111:H- or O111:H8. The outbreak EHEC ST678 strain clusters away from these “common” EHECs but clusters, as a double locus variant, with EHECs of serotype O128:H2. It is clear that EHEC, as with EAEC, represent a diverse group of *E. coli* which have acquired virulence genes on several different occasions (Figure 1). It is not just the virulence genes, but also the background into which the virulence genes are acquired, which results in the ability of a strain to cause disease and spread; adherence and toxin production have both been implicated for EHEC. Cases of EHEC infection normally present to health facilities as bloody diarrhoea although more severe complications can occur. The frequency of these complications is dependent on the toxin encoded; the presence of *vtx2a* has been shown to be associated with a more virulent infection [49] partly due to increased expression [50]. The ability to adhere to intestinal cells has also been shown to be associated with virulence and although EHEC, as with EPEC, normally adhere using the LEE [47], there are other mechanisms of attachment within *E. coli* and outbreaks have been caused by several different lineages of EHEC using non-LEE mediated attachment. For the outbreak strain, adherence is presumably mediated via the acquired EAEC virulence factors. It is possible that this adherence is more effective than LEE mediated adherence and so may explain why the outbreak strain caused such a virulent infection.

**Outbreaks of EHEC**

EHEC outbreaks are more often reported from industrialised countries than from developing countries because surveillance and reporting systems are in place. The most common type from outbreaks is O157:H7 which was responsible for one of the largest outbreaks which included 106 HUS cases from 2,764 confirmed infections in Japan in 1996 [51]. Outbreaks caused by non-O157 EHEC have for several years been highlighted as a potential risk [52] and are well documented again in some industrialised countries [53]. One of the common non-O157 VTECs in the USA is O111:H8 and one of the largest outbreaks was caused by an EHEC O111 (ST and H group not given) in the USA in 2008 causing 341 illnesses [54]. Another strain of EHEC O111:H2 (unknown ST) caused an outbreak in 1998 [55] and had features very similar to those of the German outbreak strain: *eae* negative, EAEC aggregative adherence, and associated with HUS. In 2007 there was an outbreak of EHEC in which five children were infected by two serotypes (O145 and O26) from consumption of ice-cream produced from a Belgium farm [56], perhaps showing the widespread nature of non-O157 EHECs and emphasising their potential to contaminate food handled by people. Thus highly virulent non-O157 *E. coli* has been circulating for some time but the potential impact may not be fully appreciated.

**Genomics of EHEC**

It is clear that EHEC, as with EAEC, is a heterogeneous group of DECs defined by a virulence factor (*vtx*). The best studied, single locus variants of ST11 share the serotype O157:H7 and show a

---

**Table 4. List of Useful Links**

<table>
<thead>
<tr>
<th>Links to other resources</th>
<th>Useful links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robert Koch Institute</td>
<td><a href="http://www.rki.de/EN/Home/homepage_node.html">http://www.rki.de/EN/Home/homepage_node.html</a></td>
</tr>
<tr>
<td>Centers for disease control and prevention</td>
<td><a href="http://www.cdc.gov/index.htm">http://www.cdc.gov/index.htm</a></td>
</tr>
<tr>
<td>github repository for the “crowdsharing” efforts</td>
<td><a href="https://github.com/ehec-outbreak-crowdsourced/">https://github.com/ehec-outbreak-crowdsourced/</a></td>
</tr>
</tbody>
</table>
conserved genome containing around 1.5 Mb of horizontally acquired DNA which includes a type III secretion system and effectors, the LEE. The LEE contains around 30 coding sequences in 5 operons and encodes the ability of both EPEC and EHEC to attach to the gut and cause disease. Several (currently 24) non-LEE effectors have been described [47] for which the cellular function is being investigated. There is some variability within the EHEC ST11 (O157:H7) group in toxins (Table 1) and in the other accessory genes. This suggests that acquisition has occurred on several occasions and that the genomes of these closely related bacteria are dynamically exchanging DNA with other gut flora. It is believed that O157:H7 as a group evolved from the O55 ancestor, after the horizontal acquisition of genes encoding the O157 antigen, and then branched into two lineages O157:H7 and O157:H-. There has been little radiation in human isolates since this occurred and comparison with the cattle strains (the normal host for ST11 VTEC) suggests that it is a limited subset of cattle-adapted strains which cause infection in humans. This may be due to the source-sink nature of the population dynamics. The source is cattle which support the majority of the bacterial population whilst spill-over into the human population occurs with a restricted set of strains that have the ability to shed in high numbers from cattle and/or to amplify in the environment as well as the ability to colonise and cause disease in humans.

There are several non-O157 EHECs now described and there is sequence data available for ST/serotype: ST21/O26:H11, ST16/O111:H- and ST17/O103:H2 (Table 3). These non-O157 EHECs are from different lineages and yet contain a set of relatively conserved accessory genes [57]. Although analysis of the accessory genome suggests that selective forces within the same environment have led to the acquisition and maintenance of a similar accessory gene content (parallel evolution), there is high level clustering of several of the non-O157 EHECs suggesting a common ancestry. It seems likely that some *E. coli* lineages acquire genetic material via horizontal exchange more often than others; however, whether this is driven by a pathogenic lifestyle [18] or whether pathogens have emerged from strains with a commensal lifestyle within such lineages is not clear. What is clear is that there are many diverse *E. coli* in which vtx genes have been found but it is only those that can also adhere to the intestine which will remain in the *E. coli* population and come to our notice as a cause of infectious disease in humans or animals. The latest of these emergent *E. coli* caused the massive outbreak of HUS in Germany (see the useful links in Table 4).

**The emergence of new pathogenic *E. coli***

The distribution of EAEC and EHEC across the tree drawn from MLST (Figure 1) suggests that both pathotypes have arisen on several occasions from several ancestral strains. However, there are clear patterns in the ancestry; the majority of VTECs are clustered around two STs: ST11 (O157:H7) and ST21 (O26:H11). The ST678 (O104:H4) strain from the outbreak clusters with other non-O157 VTECs, possibly around ST25, but most closely with the strain 55989 (also ST678 and also an EAEC). However, this is away from most EAEC isolates suggesting that the acquisition of the plasmid encoding the EAEC phenotype has occurred independently into the ST678 lineage and is not a previously widespread EAEC strain. It seems likely that the emergence of “new” VTECs will be from lineages of *E. coli* which have the ability to adhere to the gut of an animal host, which may be human, either by the mechanisms classically shown by EPEC (the LEE) or by the virulence plasmid of EAEC.

**Conclusion**

Although EHEC is the best documented and most severe of the DEC it is not the most common cause of *E. coli* diarrhoea. The importance of EAEC as the most common causative agent is now being realised. This may be due to improved testing rather than a recent increase in identified cases, in which case as testing improves further so will estimates of the burden of EAEC disease. The recent combination of EAEC and EHEC virulence factors in a single outbreak strain causing such severe disease emphasises the importance of being able to detect all DEC using appropriate genetic methods and not just focus on *E. coli* O157:H7. Routine surveillance systems for EAEC, a once ignored global pathogen, would go a long way to reaching this goal.

**Acknowledgements**

Thank you to all members of the Gastrointestinal Infections Reference Unit at Colindale and Flemming Scheutz for valuable discussion and to the scientists who released sequence data on the outbreak strain into the public domain. Thank you to the authors of published MLST data [18,24,32] which was used for the analysis in this paper.
References


27. Chart H, Spencer J, Smith HR, Rowe B (1997) Magnesium ions are required for HEp-2 cell adhesion by


Corresponding Author
Dr John Wain
Laboratory of Gastrointestinal Pathogens
Health Protection Agency, Colindale
61, Colindale Avenue
London NW9 5HT
United Kingdom
John.wain@hpa.org.uk

Conflict of Interests: No conflict of interests is declared.