Taxonomy Meets Public Health: The Case of Shiga Toxin-Producing Escherichia coli

FLEMMING SCHEUTZ1

1WHO Collaborating Centre for Reference and Research on Escherichia and Klebsiella, Department of Microbiology and Infection Control, Statens Serum Institut, DK-2300 Copenhagen S, Denmark

ABSTRACT To help assess the clinical and public health risks associated with different Shiga toxin-producing Escherichia coli (STEC) strains, an empirical classification scheme was used to classify STEC into five "seropathotypes" (seropathotype A [high risk] to seropathotypes D and E [minimal risk]). This definition is of considerable value in cases of human infection but is also problematic because not all STEC infections are fully characterized and coupled to reliable clinical information. Outbreaks with emerging hybrid strains continuously challenge our understanding of virulence potential and may result in incorrect classification of specific pathotypes; an example is the hybrid strain that caused the 2011 outbreak in Germany, STEC/EAggEC O104:H4, which may deserve an alternative seropathotype designation. The integration of mobile virulence factors in the stepwise and parallel evolution of pathogenic lineages of STEC collides with the requirements of a good taxonomy, which separates elements of each group into subgroups that are mutually exclusive, unambiguous, and, together, include all possibilities. The concept of (sero)-pathotypes is therefore challenged, and the need to identify factors of STEC that absolutely predict the potential to cause human disease is obvious. Because the definition of hemolytic-uremic syndrome (HUS) is distinct, a basic and primary definition of HUS-associated E. coli (HUSEC) for first-line public health action is proposed: stx2 in a background of an eae- or aggR-positive E. coli followed by a second-line subtyping of stx genes that refines the definition of HUSEC to include only stx2a and stx2d. All other STEC strains are considered "low-risk" STEC.

PATHOTYPES AND TAXONOMY

The term enteropathogenic Escherichia coli was originally used to refer to strains belonging to a limited number of O groups epidemiologically associated with infantile diarrhea (1). Subsequently, E. coli strains isolated from intestinal diseases have been grouped into at least six main categories on the basis of epidemiological evidence, phenotypic traits, clinical features of the disease they produce, and specific virulence factors. The well-described intestinal pathotypes or categories of diarrheagenic E. coli groups are enteropathogenic E. coli (EPEC), Shiga toxin-producing E. coli (STEC) or verocytotoxin-producing E. coli (VTEC) (including enterohemorrhagic E. coli [EHEC]), enterotoxigenic E. coli (ETEC), enteraggregative E. coli (EAEC), enteroinvasive E. coli, and diffusely adherent E. coli. The general definition of an E. coli pathotype as "a group of strains of a single species that cause a common disease using a common set of virulence factors" (2) has been further refined for STEC to help assess the clinical and public health risks associated with different STEC strains (3). An empirical classification scheme was used to classify STEC serotypes into five "seropathotypes" (A through E) according to the reported association of serotypes with human intestinal disease, outbreaks, and hemolytic-uremic syndrome (HUS) (3). This classification system uses a gradient...
ranging from seropathotype A (high risk) to seropathotypes D and E (minimal risk). This definition has been of considerable value in cases of human infection but is also problematic because the majority of isolates from STEC infections are not fully characterized and coupled to reliable clinical information. Although the definition of HUS is distinct, the spectrum of diarrheal disease varies considerably and may include a range of symptoms from nonbloody to scanty blood to true hemorrhagic colitis. Additionally, the use of A through E adds confusion because Shiga toxin subtypes are also named alphabetically. Most importantly, the concept of (sero)-pathotypes collides with the requirements of a good taxonomy, which separates elements of each group into subgroups that are mutually exclusive, unambiguous, and, together, include all possibilities. In practice, a good taxonomy should be simple to apply, easy to remember, and easy to use. The need to define human pathogenic STEC and to identify factors of STEC that absolutely predict the potential to cause human disease is obvious in terms of clinical management, supportive or antibiotic treatment, quarantine measurements, risk assessment, surveillance, and outbreak investigations and management. This chapter presents a brief history of the concept of pathotypes and describes the possible alternatives for categorizing STEC based on phenotypic or molecular typing.

PATHOTYPES

First discovered in 1977 (4), verocytotoxin was found to be biologically and structurally similar to Shiga toxin produced by Shigella dysenteriae Type 1 (5, 6). It was soon realized that antigenically distinct cytotoxins could be found in different E. coli serotypes (7, 8). STEC or VTEC strains are characterized by their ability to produce either one or both of these cytotoxins, referred to as Stx1 or VT1 (first described as Shiga-like toxin I, SLTI) and Stx2 or VT2 (first described as Shiga-like toxin II, SLTII). The cytotoxin production is usually bacteriophage-mediated (9–12), and the diversity of this toxin family has since become clear.

The public health significance of STEC was first recognized in 1982, when two outbreaks in the United States affected at least 47 people in Oregon and Michigan. Nine of 12 stool cultures yielded a rare E. coli serotype, O157:H7, that was also isolated from a beef patty from a suspected lot of meat in Michigan (13). This strain was designated EDL933 and has since been used as the prototype STEC strain by researchers worldwide. Distinct clinical features of hemorrhagic colitis (HC) included abdominal cramps, copious bloody diarrhea described as “all blood and no stool,” unaccompanied by fecal leukocytes, and no fever. Duration of illness was 2 to 9 days, and there were no deaths, complications, or sequelae in any of the cases.

In an outbreak of HC in November 1982 at a Canadian institution for elderly patients, sorbitol-negative E. coli O157:H7 was shown to produce verocytotoxin. Two of six sporadic cytotoxic O157:H7 strains were associated with HC, and 70% of 78 cytotoxic serotypes isolated from sporadic cases of diarrhea during 1978 to 1982 were E. coli O26:H11 (14). The cytotoxicity of an O26 strain (H30, described as a verocytotoxin producer by Konowalchuk et al. [4]), two of the E. coli O157:H7 strains from the U.S. cases of HC, and the beef patty isolate EDL933 from this outbreak could be neutralized by rabbit antiserum to purified Shiga toxin from S. dysenteriae Type 1, substantiating the premise that these cytotoxins were the same and that they played an important role in E. coli diarrheal diseases (15). In 1983, 11 of 15 sporadic cases of enteropathic HUS were shown to have evidence of infection with VTEC, indicating an association between sporadic cases of HUS and cytotoxin-producing E. coli strains (16). Verocytotoxin was proposed as having direct etiological importance in the pathogenesis of both HUS and HC (17).

The term EHEC, coined to refer to strains such as O157:H7 that manifest the above-mentioned clinical, epidemiological, and pathogenic features (18), further defined a pathogenic subgroup of STEC strains based on their association with disease in humans and their ability to hybridize to a DNA probe (CVD419) derived from a large plasmid present in most O157:H7 strains and the majority of other STEC strains isolated from cases of HC (19). Many of the other STEC strains belonged to classical EPEC O groups or serotypes such as O26, O111, O114, O125, O126, and O128 (4, 5, 12, 19–21), but non-EPEC O groups were also found to produce Stx: O1, O2, O4, O5, O6, O18, O45, O50, O68, O91, O103, O113, O121, and O145 (4, 5, 19). Furthermore, Stx production was also described in O groups O138, O139, and O141 isolated from weaned pigs with edema disease (12). Among the more than 472 STEC/VTEC serotypes, and apart from O157:H– and O157:H7, those in O groups O26, O103, O111, and O145 are most commonly isolated from humans worldwide (22). They, along with strains that have caused outbreaks, are clearly recognized as pathogens.

Consequently, serotypes and their association with diseases of varying severity in humans and with sporadic
disease or outbreaks have been used to classify STEC into five seropathotypes (A through E) according to the reported occurrence of serotypes in human disease, in outbreaks, and in HUS (3). Seropathotype A consists of O157:H7 and O157:NM, considered to be the most virulent. Seropathotype B originally consisted of five serotypes that were similar to seropathotype A in causing severe disease and outbreaks but occurred at lower frequency, but in the United States, seropathotype B has been extended to include 13 STEC serotypes: O26:H11 and NM; O45:H2 and NM; O103:H2, H11, H25, and NM; O111:H8 and NM; O121:H19 and H7; and O145:NM (23). Seropathotype C includes serotypes infrequently implicated in sporadic HUS but not typically with outbreaks and includes O5:NM, O91:H21, O104: H21, O113:H21, O121:NM, and O165:H25. Seropathotype D is composed of 12 serotypes that have been implicated in sporadic cases of diarrhea but not with outbreaks or HUS, and seropathotype E is composed of at least 14 animal serotypes that have not been implicated in disease in humans (3).

This approach has been of considerable value in defining pathogenic STEC serotypes of importance in cases of human infection and also for STEC isolates from ruminants. However, classification of strains based on the criteria above is also problematic because the majority of isolates from STEC infections are not fully serotyped nor characterized for the presence of virulence factors. A recent Belgian study of STEC added 14 serotypes to seropathotype C (including four serotypes associated with HUS) and 54 serotypes to seropathotype D, demonstrating how versatile the definition of seropathotypes can be (24). Outbreaks with emerging or new hybrid strains with hitherto unknown virulence factors may also continuously challenge our understanding and appreciation of virulence potential. The limitation to “relevant” serotypes may therefore result in the omission or incorrect classification of specific pathotypes. Strain O104:H4 is such an example, because until the May–June 2011 outbreak in Germany this highly virulent strain would have been classified as seropathotype D on the basis of its sporadic occurrence, its lack of association with outbreaks, and its limited association with HUS (25). In fact, other STEC serotypes such as O104:H21 and O113:H21 strains lacking the eae gene were responsible for an outbreak and a cluster of three HUS cases in the United States and Australia, respectively (26, 27).

It has been suggested that these unusual and emerging types should have their own seropathotype designation (28). Using data from the European Surveillance System (TESSy data) as provided by the European Centre for Disease Prevention and Control (ECDC) and data available in the European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2011 (29), the European Food Safety Authority Panel on Biological Hazards (BIOHAZ Panel) concluded that the seropathotype classification by Karmali et al. (3) does not define pathogenic VTEC nor does it provide an exhaustive list of pathogenic serotypes. Eighty-five percent of 13,545 confirmed cases of human VTEC infection from 2007 to 2010 were not fully serotyped and could not be classified by using the seropathotype concept, and about 27% of the cases could not be assigned to a seropathotype group as they were not listed by Karmali et al. (25). However, about half of the isolates with missing H type were from cases of O157 infection (5,610 cases) and would most likely have been typed as O157:H7 or O157:H– and therefore assigned to seropathotype A. This would have expanded this group to 6,657 cases, or 87% of cases, 78% of fatal cases, 91% of the hospitalizations, 91% of the HUS cases, and 95% of the cases with bloody diarrhea (25).

Reporting of detailed clinical data is often incomplete. Of the reported confirmed VTEC cases in the European Union between 2007 and 2010, the health outcome was reported for 53% of diarrheal cases and 59% of HUS cases (25). Most patients (ca. 64%) presented with only diarrhea. Clinical information is not obtained according to standardized guidelines and definitions, and detailed information on the individual clinical course of disease is generally absent. Although the definition of HUS is distinct, the spectrum of diarrheal disease varies considerably and may include symptoms ranging from nonbloody to scanty blood to true hemorrhagic colitis. A specific STEC pathotype may even be associated with clinical presentations ranging from asymptomatic carriage to life-threatening HUS and death, as observed during outbreaks and in person-to-person transmission, whereby an index patient may experience only mild symptoms whereas secondary cases may develop into severe complications (30).

The BIOHAZ Panel concluded that “pathogenicity can neither be excluded nor confirmed for a given STEC serogroup or serotype based on the seropathotype concept or analysis of the public health surveillance data” (25). In addition, even though the clinical manifestations of non-O157 STEC infection may differ considerably from those of O157:H7 (31), STEC O157:H7 has also been isolated from stools of healthy individuals. Many studies have tried to correlate the presence of specific virulence factors with disease or severity of disease.
(see discussion in reference 32). The combination of the locus of enterocyte effacement (LEE)-encoded eae gene for intimin and stx2 is significantly more frequent in isolates from serotypes found in humans and is most strongly associated with disease in humans, particularly with severe disease (32–34). The reverse is true for stx1, which is found more frequently in serotypes not found in humans (32). Enterohemolysin, a plasmid-encoded toxin expressed by the ehxA gene that readily causes the hemolysis of washed sheep erythrocytes and liberates hemoglobin from the red blood cells during infection, has been linked to severe disease symptoms (35, 36).

The taxonomy of EPEC and VTEC is intimately intertwined in that many VTEC types share specific virulence factors such as pathogenicity islands, e.g., LEE (including the eae gene), O island (OI) 122, and plasmids with EPEC. Non-LEE (nle)-encoded genes have also been found in both EPEC and STEC isolates, such as the EspN/NleA effector protein encoded on a prophage (37). Several studies have indicated that the major difference between certain EPEC and STEC isolates is the absence or presence of the bacteriophage encoding Stx. However, the presence of LEE does not seem to be essential for full virulence, as a wide number of LEE-negative STEC strains have been associated with sporadic cases and small outbreaks of HC and HUS (38).

How are environmental, food, and veterinary STEC types, which by nature cannot be isolated from human cases of either HC or HUS, classified? Animals often carry types that are referred to as EHEC or could be classified in the A through E pathotype scheme without any clinical symptoms. The classification based on the clinical course of disease in humans is clearly host associated, and the term EHEC or the A through E classification of nonhuman isolates could be misleading.

**PHYLOGENY AND DISEASE**

An *E. coli* genome contains between 4,200 and 5,500 genes, with fewer than 2,000 genes conserved among all strains of the species (the core genome). Comparison of 61 *E. coli* and *Shigella* spp. sequenced genomes has shown that the genetic repertoire or pan-genome comprises 15,741 gene families and that only 993 (6%) of the families are represented in every genome (39). The variable or “accessory” genes thus make up more than 90% of the pan-genome and about 80% of a typical genome; some of these variable genes tend to be collocated on genomic islands. Continuous gene flux has occurred during *E. coli* divergence, mainly as a result of horizontal gene transfers and deletions. This genetic plasticity accelerates the adaptation of *E. coli* to varied environments and lifestyles, as it allows multiple gene combinations that result in phenotypic diversification and the emergence of new hypervirulent strains such as the hybrid STEC and EAggEC O104:H4-B1-ST678 strain that combine resistance and virulence genes, which in classical pathogenic *E. coli* strains traditionally have been mutually exclusive (40). More than describing STEC stains as a separate group of *E. coli*, STEC represents many, if not the majority, of phylogenetic lineages of *E. coli* in general. The sequence-based method targeting housekeeping genes (using distinct sets of genes), multi-locus-sequence typing (MLST), generally has less discriminatory power than pulsed-field gel electrophoresis (PFGE) but has been considered the most reliable method to determine the genetic relatedness of epidemiologically unrelated isolates. However, when in silico MLST is performed on whole-genome sequences, many of the strains appear jumbled and less well resolved (39). *E. coli* strains are assigned by MLST to different sequence types (STs), and within each ST diverse clusters can be observed by PFGE. Three distinct sequence-based methods targeting housekeeping genes exist for *E. coli*. As of October 2013, Institut Pasteur’s MLST scheme using *dinB, icda, pabB, polB, putP, trpA, trpB,* and *uidA* genes lists 599 unique STs ([http://www.pasteur.fr/recherche/genopole/PS8/mlst/EColi.html](http://www.pasteur.fr/recherche/genopole/PS8/mlst/EColi.html)). Mark Achtman’s MLST scheme using *adk, fumC, gyrB, icda, mdb, purA,* and *recA* lists 5,873 isolates belonging to 3,874 STs, of which 1,485 are found in 54 ST complexes and 4,388 isolates belonging to 3,562 unique STs have not been assigned an ST complex ([mlst.warwick.ac.uk](http://www.mlst.warwick.ac.uk)). T. Whittam’s MLST scheme, the *EcMLST*, lists allele sequences and allele profiles for 679 *E. coli* strains and uses different combinations of 15 housekeeping genes, i.e., internal fragments of the seven *aspC, clpX, fadD, icda, lysP, mdb,* and *uidA* genes where the associated sequence type is defined as st7 or further characterized by using internal fragments of the eight additional *arcA, aroE, cyoA, dnaG, grpE, mtlD, mutS,* and *rpoS* genes where the associated sequence type is defined as st15. Some isolates are only characterized by using internal fragments of two of the seven housekeeping genes, *mdb* and *uidA,* where the associated sequence type is defined as st2 ([www.shigatox.net](http://www.shigatox.net)). A comparison study is ongoing to establish correspondence between the two former MLST schemes.

There is increasing evidence that within serotype O157:H7 there are differences in the clinical outcome and association with HUS. Molecular methods have identified different genetic lineages of *E. coli* O157:H7.
Octamer-based genome scanning and microarray comparative genomic hybridization were first to identify three lineages designated I, II, and I/II (41–43). Nucleotide polymorphism–derived genotyping and phylogenetic analyses identified eight major STEC O157 lineages (44). Seven lineages are typically found in cattle, including one that does not associate with human disease and may be evolving away from human virulence and two other lineages accounting for a majority of human disease (44). The 30 sorbitol-fermenting O157 strains belong to a lineage VIII exclusively isolated from humans (44). In a study of 528 O157 strains primarily from Michigan patients but also including strains from Argentina, Australia, Canada, Germany, Japan, and the United Kingdom from 1982 to 2006, Manning et al. identified 39 single nucleotide polymorphism (SNP) genotypes that differed at 20% of SNP loci and separated them into nine distinct clades (45). The outbreak strain TW14359, implicated in a multistate outbreak associated with the consumption of bagged spinach in North America in 2006 (46, 47), was shown to be a member of clade 8, which was significantly associated with younger age (0 to 18 years) and patients with HUS, who were seven times more likely to be infected with clade 8 strains than patients with strains from clades 1 to 7 combined (45). The study revealed substantial genomic differences between clades, suggesting that an emergent subpopulation of the clade 8 lineage has acquired critical factors that contribute to more severe disease. Comparison of the phylogenetically divergent O157:H7 outbreak strains TW14359 and RIMD0509952, which caused the largest O157:H7 outbreak to date in Sakai, Japan, showed that these two strains vary in their ability to colonize or initiate the disease process (48). Interestingly, most LEE genes, the stx2 genes, and several pO157-encoded genes that promote adherence, including type II secretion genes and their effectors stcE and adfO, are upregulated in the spinach outbreak strain, whereas flagellar and chemotaxis genes are primarily upregulated in the Sakai strain (48).

Evolutionary analyses of STEC by multilocus enzyme electrophoresis (49) and partial sequencing of 13 housekeeping genes (50) have identified two distantly related clonal groups classified as EHEC 1, including serotype O157:H7 and its inferred ancestor O55:H7, and EHEC 2, represented by several O groups (O26, O111, O118, etc.). These two clonal groups differ in their virulence and global distribution. Although several fully annotated genomic sequences exist for strains of serotype O157:H7, much less is known about the genomic composition of EHEC 2. Analysis of 24 clinical EHEC 2 strains representing serotypes O26:H11, O111:H8/H11, O118:H16, O153:H11, and O15:H11 from humans and animals by comparative genomic hybridization supports the hypothesis that extensive modular shuffling of mobile DNA elements has occurred among STEC strains, and the gene content variation of phage-related genes in EHEC 2 seems to indicate that EHEC 2 is a multiform pathogenic clonal complex, characterized by substantial intraserotype genetic variation. The heterogeneous distribution of mobile elements is especially seen in O26:H11 more than in other EHEC 2 serotypes (51). Comparative analysis of whole-genome phylogeny and of type III secretion system effectors of 114 LEE+ E. coli isolates shows that attaching and effacing E. coli is divided into five distinct genomic lineages and that the LEE+stx+/bfp− genomes are primarily divided into two genomic lineages, the O157/O55 EHEC1 and non-O157 EHEC2 (52). Most importantly, phylogenetic relatedness was independent of the presence or absence of stx-encoding phages, highlighting the close relation between LEE+ EPEC and STEC lineages (52). In this study of 138 whole genomes of which 114 were LEE+, stx genes were only found in phylogroups B1, including EHEC2, EPEC2, and unclassified attaching and effacing E. coli, and in E, represented by O157 EHEC1 (52). Even less is known about non-LEE STEC, but whole-genome comparative analysis of nine non-LEE genomes revealed that phage-encoded genes, including non-LEE-encoded effectors, were absent from all nine STEC genomes. Several plasmid-encoded virulence factors reportedly identified in LEE-negative STEC isolates were identified in only a subset of the nine LEE-negative isolates, further confirming the diversity of this group. Characterization of the lambdoid stx-encoding phages showed that although the integrase gene sequence corresponded with genomic location, it was not correlated with stx subtype, highlighting the mosaic nature of these phages. A wide range of basal and induced expression of the Shiga toxin genes, stx1 and stx2, and the Q genes was observed (53).

**TOXIN AND DISEASE**

The Shiga toxin family is divided into two branches, Stx1 (almost identical to Stx from S. dysenteriae Type 1) and Stx2. Subtypes, denoted by Arabic letters that follow the main type name, may exhibit significant differences in biologic activity, including serologic reactivity, receptor binding, and the capacity to be activated by elastase in intestinal mucus. Variants have been defined by relatedness of sequence within a subtype that
differs by one or more AAs from the prototype (54). The phylogenetic relationship of variants has been analyzed (54–60), but not all variants have been examined for all classical phenotypic differences, biologic activity, and hybridization properties. The variants are designated by toxin subtype, O group if the host strain is E. coli and generic name of the host bacterium if the host strain is not E. coli, followed by the strain name or number from which that toxin was described (54). At present, 107 variants have been identified: subtypes of Stx1 include 9 variants of Stx1a (including Shiga toxin from S. dysenteriae), 4 of Stx1c, and 1 of Stx1d, and subtypes of Stx2 include 21 variants of Stx2a, 16 of Stx2b, 18 of Stx2c, 18 of Stx2d, 14 of Stx2e, 2 of Stx2f, and 4 of Stx2g (Fig. 1).

Stx subtypes, and maybe specific variants, are clinically relevant. Stx2a (with or without Stx2c) seems to be highly associated with HUS (30, 61–64). The combination of eae and stx2 especially has been associated with the development of HUS and bloody diarrhea (34, 62, 63, 65, 66), and an unspecified synergism between the adhesin intimin encoded by eae and Stx2 has been suggested (32). A German study of 922 patients with STEC infection found that 81 of 107 patients (76%) with sorbitol-fermenting O157 had HUS (54). This particular STEC strain is typically positive for both stx2a and eae. Sporadic adult cases of HUS in France were seen in patients infected with six different serotypes, all vtx2 (vtx1 and eae negative) (67). In Australia, a VTEC O113:H21 strain lacking eae was responsible for a cluster of three cases of HUS (27). In Finland, O174:H21 (68) and O:rough:H49 (69), both eae negative and vtx2 positive, were isolated from two separate cases of HUS. In Germany, a large outbreak in 2011 of a hybrid STEC-EAggEC strain O104: H4, stx2a and eae negative but positive for many EAggEC-associated genes, affected 3,167 patients without HUS (16 deaths) and 908 with HUS (34 deaths) (70).

In Germany, an association between the activatable subtype vtx2d in eae-negative STEC strains and HUS has been found (61). The median age of 21 years in patients with vtx2d was considerably higher than in other patients with STEC. Stx2 subtypes Stx2a and Stx2d studied in vitro with Vero monkey kidney cells and primary human renal proximal tubule epithelial cells were at least 25 times more potent than Stx2b and Stx2c. The in vivo potency of Stx2b and Stx2c in mice was similar to that of Stx1, whereas Stx2a and Stx2d were 40 to 400 times more potent than Stx1 (71). It has been suggested that disease outbreaks select for producers of high levels of Stx2a among E. coli O157: H7 strains shed by animals and that Stx1 expression is unlikely to be significant in human outbreaks (72). Nearly all lineage I strains carry stx2a, whereas all lineage II strains carry stx2c, and 4 of 14 lineage I/II strains have copies of both stx2a and stx2c (73). Real-time PCR and enzyme-linked immunosorbent assay have demonstrated that lineage I and I/II strains produce significantly more stx2a mRNA and Stx2a than lineage II strains. However, among lineage I strains significantly more Stx2a is also produced by strains from humans than from cattle. Therefore, lineage-associated differences among E. coli O157: H7 strains, such as prophage content, toxin type, and toxin expression, may contribute to host isolation bias. However, the level of Stx2 production alone may also play an important role in the within-lineage association of O157: H7 strains with human clinical disease. Indeed, clade 8 strains associated with HUS overexpress Stx2a when compared to strains from clades 1 to 3, and SNPs, which may affect Stx2a expression and could be useful in the genetic differentiation of highly virulent strains, have been described (74).

Analyses of the 2006 outbreak of O157: H7 clade 8 strains in spinach suggested the presence of stx2a and an stx2c variant (75), but not all clade 8 strains have both stx2a and stx2c, and none of the strains has only...
The presence and presumable production of the Stx2c variant alone cannot be solely responsible for the enhanced virulence attributed to the clade 8 lineage. This also is true for the production of Stx2a, because it was detected in nearly every strain representing all nine clades (45).

In a Danish outbreak with an O157:H7 strain with a toxin gene subtype profile consisting of eae, stx1a, and stx2a, HUS developed in 62% of the patients. HUS had been previously found in two of six (33%) patients infected with STEC. Infections with STEC O157 containing the vtx2a toxin profile were associated with a higher number of HUS cases (24 to 30%), whereas HUS developed in only 1 of 31 (3%) patients infected with O157:H7 eae + stx2c-positive strains and did not develop in 93 patients with eae + stx1a with (85 patients) or without (8 patients) stx2c (76). The association between stx2a and severity of disease has also been demonstrated in two animal models for clinical genotype (CG) strains (carrying stx2a with or without Stx2c), which induced more severe clinical symptoms, earlier and higher mortality, and more severe histopathologic lesions compared to bovine-biased genotype (BBG) strains (carrying stx2c only) (72). Purified Stx2a has also been shown to be more potent than Stx2c against primary human kidney cell lines and in mouse models (71). It is possible that carriage and expression of Stx2a alone are sufficient to confer increased virulence in animal models and increased expression of human disease, but alternatively, these phenotypes may result in whole or in part in other genetic factors that are correlated with Stx2a.

Stx2c has occasionally been associated with HUS but with a significantly lower risk (62, 64). The majority of 210 patients in Japan infected with Stx2c strains presented no or mild symptoms, except for 3 patients with bloody diarrhea (78). Also in Japan, 169 strains carrying only stx2cβα (now stx2c) (54) were probably less virulent and caused bloody diarrhea less frequently (79). In an Austrian study of 201 STEC strains collected from patients and environmental sources, the stx2a and stx2c alleles were associated with high virulence and the ability to cause HUS, whereas stx2d, stx2e, stx1a, and stx1c occurred in milder or asymptomatic infections (80). However, caution is warranted for infections by Stx2c O157 strains, in addition to Stx2a O157 strains.

Other subtypes or variants of Stx1 and Stx2 are primarily associated with a milder course of disease (61–63). Except for the few individual, unusual cases mentioned below, HUS did not develop in a total of 825 Danish patients as follows: 426 with vtx1a STEC (313 eae positive and 113 eae negative), 65 with stx1c (5 eae positive and 60 eae negative), 13 stx1d (all eae negative), 87 with stx2b (1 eae positive and 86 eae negative), 12 with stx2e (1 eae positive and 11 eae negative), 37 with stx2f (all eae positive), 5 with stx2g (eae negative), and 180 with seven different combinations of stx1 and stx2 subtypes, excluding stx2a and stx2d (70 eae positive and 110 eae negative) (unpublished Danish surveillance data). In France, sporadic cases of STEC infection complicated with HUS have been described in patients infected with a clone of O103:H2, virulence type vtx1 and eae (81). Among infections with VTEC strains with only the vtx1 gene, one case of HUS has been described; a strain with the serotype O115:H18 (82) and 8 of 169 non-O157 HUS-associated strains representing 42 different ST types have been subtyped as vtx1-only strains (54). Data from the European Surveillance System (TESSy data), as provided by the ECDC on virulence characteristics of reported confirmed VTEC cases in 2007–2010, including all cases, hospitalized cases only, and HUS cases only, show that most HUS cases (89.2%), for which information was reported on virulence factors, were either eae, vtx2 positive or eae, vtx1, vtx2 positive and that an additional 5.9% were either vtx2 positive or vtx1, vtx2 positive but without the eae gene. Only 2.3% were positive for stx1 (1.6% eae, stx1 and 0.7% stx1) (23). None of the 124 reported infections due to O103:H2, of which all but one were eae, vtx1 positive, caused HUS (25). Surveillance data rarely include detailed clinical descriptions of individual HUS cases. Follow-up on clinical data on Danish patients with unusual or rare association between HUS and the virulence profile eae, vtx1 revealed possible complicating factors such as antibiotics during the acute infection, underlying nephrotic syndrome, association with outbreak cases infected with an HUS-associated outbreak O157:H7 strain, or a double infection with O157:H7 (eae + vtx1 + vtx2). HUS also developed in one patient infected with O53:H12 vtx1 only who had received six different antibiotics during the acute phase and in one patient with O13,073:K1:H18 vtx2d (unpublished data). The association of specific virulence factors with HUS should be carefully examined for underlying disease, epidemiologic relation, and antibiotic treatment during the acute phase of illness.
from patients with diarrhea and HUS. These O groups in STEC strains are not usually associated with edema disease (84). Thus, STEC without both stx2 and eae is, with a few exceptions, only sporadically associated with HUS, and toxin subtyping can be useful in identifying high-risk or HUSEC strains.

**Stx-ASSOCIATED BACTERIOPHAGE INSERTION**

In *E. coli*, the capacity to produce Shiga toxin is encoded by genes on bacteriophages, and it has been suggested that rather than describing the taxonomy of STEC bacteria, the biology and taxonomy of the bacteriophages that are hosted by certain lineages of *E. coli* could serve as the basis for refining this pathogenic group.

Several common stx phage insertion sites have been reported in LEE-positive STEC genomes. These insertion sites include wrbA, yecE, torS/T, sbcB, yehV, argW, ssrA, and prfC (85, 86). In LEE-negative strains, additional insertion sites are often different and include patC, yciD, ynhH, serU, mlrA, and yjbM (53) (Fig. 2). However, a number of sites have not been determined for some of the most common serotypes like O157, O26, O111, and O103 (85). Insertion site occupancy by stx phages depends on the host strain and on the availability of the preferred locus in the host strain. For the most part, yebV is occupied by stx1 encoding phages in LEE-positive O157 strains whereas wrbA or argW is preferentially selected by the stx2a phages and sbcB by stx2c phages (87). Phages preferentially use one insertion site, but if this primary insertion site is unavailable, then a secondary insertion site is selected (88, 89). Molecular epidemiological studies using Stx-associated bacteriophage insertion sites for strain differentiation have shown that specific bacteriophage-associated genetic factors underlie the differential virulence of CG and BBG, where genotypes that are significantly overrepresented among human isolates, as compared to cattle isolates, are referred as CG and genotypes among cattle isolates that show significantly overrepresentation or similar representation, as compared to human isolates, are referred as BBG. STEC O157 can be further classified into several genotypes (90). In a study of a panel of 419 STEC O157 strains from geographically and temporally diverse cattle- and human-origin isolates, the presence of Stx2a-associated bacteriophage sequences detected adjacent to either wrbA or argW and the detection of stx2a were significantly associated with CG compared to BBG strains, of which many (42.9%) had Stx2a-associated bacteriophage sequences adjacent to wrbA or argW but lacked detectable stx2a. All 281 CG strains had Stx2a-associated bacteriophage sequences inserted in wrbA or argW and carried stx2a. In contrast, the presence of Stx2c-associated bacteriophage sequences inserted in wrbA or argW and carried stx2a. In contrast, the presence of Stx2c-associated bacteriophage sequences inserted in sbcB and the presence of stx2c were significantly more common in BBG compared to CG isolates. All 107 BBG isolates had both traits, whereas only 11.8% of CG isolates belonged to genotypes with Stx2c-associated bacteriophage sequences inserted in sbcB, and only 7.1% of CG isolates carried stx2c (87). Deep sequencing of pooled STEC O157 DNAs from human clinical cases (n = 91) and cattle (n = 102) identified 42 genotypes that could be tagged by a minimal set of
32 polymorphisms. Phylogenetic trees of these genotypes are also divided into clades that represent strains of cattle origin, or cattle and human origin (91), thus confirming that certain O157 lineages are more associated with the bovine reservoir and do not turn up in human disease.

The lineage-associated differences among STEC O157:H7 strains such as prophage content, toxin type, and toxin expression may contribute to the observed host isolation bias. However, the level of Stx2 production alone may also play an important role in the within-lineage association of STEC O157:H7 strains with human clinical disease (73).

**TOXIN EXPRESSION AND Q GENES**

The biology of the Stx-encoding phages contributes greatly to the production of Stx, and many research results during the past decade have contradicted the prevailing assumption that phages serve merely as agents for virulence gene transfer.

In a majority of STEC strains, expression of stx genes within lambdoid phages is believed to be largely under the control of the late phage promoter, pR, and the Q antiterminator protein (92). In STEC lysogenic for Stx-converting bacteriophages, expression of the stx genes is usually repressed and production of Stx is preceded by prophage induction (93, 94). Variations in the Q gene have been proposed to influence the quantitative expression of Stx (95). The Q gene transcription is increased under inducing conditions allowing for increased transcription of the stx genes that are downstream of the Q-binding site (92). Inducing factors include nutritional stress, oxidative stress, UV radiation, antibiotics (mitomycin C, quinolones, β-lactams, etc.), EDTA and other chelating agents, hydrogen peroxide, heat shock, and quorum sensing (96), but Stx1 production can also be induced through low-iron induction of pStx1 (97). Different Q variants have been described. Q933 was found in the Stx2a-producing strain EDL933 (98). Q21 was found in phage 21, which does not contain any stx genes (99), but phage 2851 encoding vtx2c contains a Q gene with 96.9% identity to Q21 and has frequently been detected in Stx2c-producing O157 strains, indicating that phages related to 2851 are associated with Stx2c production in O157 strains from different locations and time periods (100), although it can be found in other Stx subtypes (101). The Q0111:H− was found in a Japanese O111:H− isolate, 11128, in a study of O157:H− clinical isolates related to HUS (86). Other Q genes found in Stx1- and Stx2a-producing phages show high identity to the Q933, i.e., the Stx1 phage in H19-B, which shows 96.5% identity to Q933 (98); the Stx2a-producing phage, VT2-Sakai, which is identical to Q933 (102); the Stx1-producing phage, VT1-Sakai, which shows 97.4% identity with Q933 (103); and the Stx1-producing phage in strain Morioka, which shows 97% identity to Q933 (104). Using the above-mentioned Q gene sequences as references, Jacobsen (105) recently queried 287 E. coli genome sequences of varying quality: Two genome sequences were from Los Alamos National Laboratory and 285 genome sequences were publicly available online (106 from three sequencing projects: Escherichia group, E. coli O104:H4, and E. coli antibiotic resistance) from the Broad Institute of Harvard and Massachusetts Institute of Technology (www.broadinstitute.org) and 179 from National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). Information about serotype was available for 131 (45.6%) strains and resulted in data on 57 different serotypes with a bias toward O157:H7 and O104:H4. Jacobsen identified 474 antiterminator Q genes in 235 of 287 genome sequences, of which 80 Q genes were found upstream of the stx operon in 62 genomes of at least 14 different serotypes. On the basis of their amino acid sequence, these 80 Q genes could be classified into five main groups (I to V) and further into 10 variants whereby variants within the main groups are distinguished by one amino acid difference. Thirteen Q genes are classified as variant QI, which is homolog to Q2851 and Q21 and, in accordance with previous findings, primarily found in stx2c phages although one stx2a phage was also identified. Main group QII is represented by six Q genes, is homolog to Q0111:H− found in stx2d phages (one) and stx2e (four) phages, respectively. QIII is the smallest group with only two Q genes in stx2e phages, whereas QIV has as many as 59 Q genes, with homologs to QVT1-Sakai, QH-19B, QMorioka, and QVT1-Sakai. QIV can be divided into six variants, none of which is found in stx2c, stx2d, or stx2e phages, but QIVa is found in stx2a and stx2e, QIVb in stx1a, QIVc in stx1c, QIVd in stx2a, QIVe in stx1a, and QIVf in stx2a phages (103). Group V contained only two sequences that were not used in Jacobsen’s study. Among all Q variants, 23 of 169 sites were conserved sites, and 32 were functionally conserved. There were 13 different Q-vtx combinations based on Q variants and vtx subtypes, but some strains carried two VT phages, resulting in 17 different Q-vtx genotypes. These phylogenetic analyses of the associated stx genes thus revealed six different vtx subtypes, vtx1a,
vtx1c, vtx2a, vtx2c, vtx2d, and vtx2e. Q genes associated with the stx2b and stx2g phages in isolates EH250 and 7V, respectively, are apparently associated with another phage in the isolate, indicating that Q expression might have a contribution from additional Q genes in non-stx phages (53).

The genetic architecture does not appear to be the only factor affecting Stx expression. The inferred phylogeny based on the alignment of stx bacteriophages indicates a broad phylogenetic diversity also observed with whole-genome phylogeny (53). However, taken together, the present data seem to suggest that the primary sequence of Q may play a major role in the regulation and released amount of Shiga toxin. In O157, DNA sequencing of the genes flanking the stx2c gene from a lineage II strain, EC970520, has revealed that the Q gene is replaced by a pphA (serine/threonine phosphatase) homolog in this and all other lineage I strains, which appears to be a useful molecular marker to distinguish among Stx2a- and Stx2c-encoding phages in O157 strains (73).

The great variation in Q genes compared to stx genes indicates that the stx genes have been taken up by lambdoid phages on several occasions and that there is a high genetic selection for this gene combination and that all the Q genes, functionally expressing their lysis genes, can take up the stx genes under certain favorable conditions.

### Genomic Islands

Sequencing of STEC O157:H7 EDL933 revealed O157: H7-specific DNA organized in genomic O islands (106), some of which are referred to as pathogenicity islands because they can contain a flexible pool of virulence-associated genes. One such pathogenicity island is LEE, which encodes a type III secretion system involved in the formation attaching and effacing lesions. The most important protein is intimin encoded by the gene eae, which is responsible for both the intimate adhesion of bacteria to the intestinal epithelium and the attaching and effacing lesion. Analysis of the variable C-terminal encoding sequence of eae defines at least 29 distinct intimin types (α1, α2, β1, β2, β3, γ1, δ1, δ, ε1, ε2, ε3, ε4, ε5, ζ1, ζ3, η1, η2, t1-A, t1-B, t1-C, ι2, ι1, ι3, ι4, ι5, ι6, ι7, ι8, ι9, ι10, ι11, ι12, ι13, ι14, ι15, ι16, ι17, ι18, ι19, ι20, ι21, ι22, ι23, ι24, ι25, ι26, ι27, ι28, ι29) (107–109) that have been associated with tissue tropism. Intimin binds to the cell receptor Tir, which is translocated by bacteria to the enterocyte through a type III secretion system (110). One SNP, 255 T>A in tir, has been shown to predict the propensity of STEC O157 isolates to cause human clinical disease (111). The overrepresentation of the tir 255 T>A T allele in human-derived isolates versus the tir 255 T>A A allele suggests that these isolates have a higher propensity to cause disease. The high frequency of bovine isolates with the A allele suggests a possible bovine ecological niche for this STEC O157 subset. A Norwegian study of 167 human and nonhuman E. coli O157:H7/NM (nonmotile) isolates with respect to the 255 T AT SNP in the tir gene was able to differentiate STEC O157 into distinct virulence clades (1 to 3 and 8) and found an overrepresentation of the T allele among human strains compared to nonhuman strains, including five of six HUS cases (112).

OI-122 is another 23-kb pathogenicity island with at least six genes with significant homology to known virulence genes: pagC of Salmonella enterica serovar Typhimurium; sen of Shigella flexneri; two non-LEE effector (nle) genes nleB and nleE of Citrobacter rodentium; and the EHEC factor for adherence gene cluster efa1 and efa2 found in STEC O157:H7. A modular arrangement of OI-122 genes based on their association with each other across HUS-associated non-O157 VTEC strains was proposed: module 1 contains Z4318, pagC, and Z4332; module 2 contains Z4323, sen, nleB, and nleE; and module 3 contains the efa gene cluster, also referred to as lifA (lymphocyte activation inhibitor) (113, 114). A progressive decrease in the prevalence of OI-122 genes in non-O157 VTEC strains belonging to seropathotype B through D, with a concomitant increase in the frequencies of incomplete and absent OI-122. The variable virulence profiles of the non-O157 serotypes indicated that complete OI-122 was more frequently present in isolates associated with HUS and that the individual genes stx2, eae, espP, as well as the OI-122-associated genes sen, nleB, nleE, and the efa/lifA gene cluster were significantly more often present in non-O157 STEC associated with HUS, and that the combined virulence profile vtx2–nleE–efal/lifA showed the strongest association with HUS (24). The presence of putative transposases in OI-122 has led to the hypothesis that its elements are acquired or lost in a modular manner. Although pagC, Z4322, sen, nleB, nleE, and efa1/lifA individually are more prevalent in non-O157 VTEC associated with HUS, the simultaneous presence of all of these genes strengthens the association with serious disease. Thus, after acquiring OI-122 modules from STEC O157:H7 through
horizontal transfer, less pathogenic non-O157 strains could cross a virulence threshold, resulting in sufficient pathogenicity to cause HUS (115).

**TAXONOMY AND PUBLIC HEALTH**

More than 472 O:H serotypes of STEC have been described (28). Some seropathotypes seem to be more closely related to specific reservoirs and specific diseases, and their number and characteristics are ever changing. However, there seems to be a much closer relationship between the virulence profile, i.e., stx subtype, additional “cocktail” genes, and the clinical course of disease than to the serotype itself—even for O157 (76). Many of these genes are acquired by horizontal gene transfer, and it is therefore possible (and plausible) that certain types of *E. coli* that have never been associated with disease or outbreaks could acquire mobile genetic elements such as the Stx encoding bacteriophages through parallel evolution, which in turn could convert them into virulent strains capable of causing human infection and outbreaks. Understanding this complex biology of host-pathogen interaction is evidently a high-priority research subject, and a better understanding of the real genetic basis behind our classification of seropathotypes is needed. The presence of LEE and O1-122 in many of the A and B seropathotypes has been very useful but has also been challenged by the serious O104:H4 outbreak in Germany in 2011 as well as by other non-LEE STEC outbreaks and HUS cases. Furthermore, the genome sequences for some of the common non-O157 O groups, O26, O103, and O111, have revealed that these pathogenic groups have arisen multiple times by acquisition of mobile genetic elements harboring virulence genes. Lambda prophages and other integrative elements have played a major role in the stepwise evolution of pathogenic lineages of STEC, especially the acquisition of type 3 secretion effectors, i.e., non-LEE-encoding *nle* genes, and some of these otherwise unlinked effectors can work in concert with one another to produce a desired effect on the host cell (116, 117). More whole-genome sequence information on non-O157 STEC and STEC-related *E. coli* strains is needed to better understand and further categorize the clinically relevant pathotypes for the necessary public health response and action.

In public health the primary goal is to (i) prevent onset of disease, (ii) minimize the risk of progression of the disease in individuals or transmission of illness, and (iii) provide rehabilitation to prevent the worsening of an individual’s health. The methods and techniques to categorize an STEC strain isolated from a patient at the present time require extensive analyses that are usually not available in the primary diagnostic line where public health action and management of STEC-infected patients begins. In the first-line diagnosis, three primary issues need immediate attention:

- Risk of severe disease and prognosis
- Quarantine of infected individuals
- Treatment with antibiotics, especially with long-term otherwise healthy carriers

In Denmark, the observed and well-documented association between Stx2 and eae or ability to colonize and persist in the gut, such as the EAEC hybrid strains, has resulted in a basic and primary definition of HUSEC for first-line public health action: *stx2* in a background of *eae*- or *aggR*-positive *E. coli* is HUS-associated and is kept “on hold” until *stx* is subtyped and further characterized. All other STEC strains with the virulence profiles of *vtx1*, *vtx1* and *vtx2*, *vtx1* and *vtx2*, respectively, of serotype, are considered “low-risk” STEC. If this HUSEC definition had been applied, 71% (1,454 of 2,046) of Danish patients would have been informed that they had a “low-risk” STEC infection during the years 1983 to 2012. Twenty-nine percent (552 of 2,046) would have been informed that they might have an HUSEC infection. If *vtx2f* is diagnosed as unspecified *vtx2*, an additional 55 (3%) patients would have been added, but as *stx2f* requires a different and specific set of primers, this would not have been relevant in most cases. Further refining of the definition of HUSEC by *stx* subtyping to include only *stx2a* and *stx2d* as HUSEC would reduce the number of patients with HUSEC to 11% (224 of 2,062). For specific, mainly socioeconomic reasons, antibiotic treatment of patients infected with VTEC is also considered according to specific criteria, which include the isolation of identical “low-risk” types of STEC found in two separate stool specimens, no isolation of HUSEC, and a detailed characterization of the STEC strain’s virulence profile and serotype. In the first line this includes the detection of the *eae*, *stx1*, *stx2*, *aggR*, and *aaiC* genes followed by a more detailed second-line subtyping of the *stx* genes. This simplified approach has been generally accepted by the primary diagnostic clinical microbiology laboratories and public health officers in Denmark. Though not complete in the characterization of each STEC isolate, this approach is practical and easy to use in an operational environment, which is expected to quickly (i) evaluate the risk of progression of the disease in individuals, (ii) minimize
transmission of HUSEC, (iii) rehabilitate individuals to prevent the worsening of an individual’s health, and (iv) reduce the socioeconomic impact on their families. However, such a simple approach should be applied with prudence because each individual case is unique, and each patient should be carefully evaluated with regard to predisposing factors, general clinical condition, contact with other STEC-infected individuals, possible link to an outbreak, and so forth. The identification of “low-risk” STEC does not per se exclude the risk of progression to severe disease, dehydration, or HUS, and patients with bloody stools and/or affected kidney function must be carefully monitored.

The versatile integration of mobile virulence factors in the stepwise and parallel evolution of pathogenic lineages of STEC collides with the requirements of a good taxonomy and complicates how public health goals are met in a timely manner. The intertwining of the traditional taxonomy concept and the association of (sero)-pathotypes to epidemiological evidence, phenotypic traits, clinical features of the disease, and specific virulence factors is therefore an ongoing challenge, and the need to identify factors of STEC that absolutely predict the potential to cause human disease is obvious in terms of clinical management, supportive or antibiotic treatment, quarantine measurements, risk assessment, surveillance, and outbreak investigations and management.

However, recent sequence information has provided advanced analytical tools that will reduce the clinical divide between the STEC types that are associated with severe disease from the seemingly benign STEC types.

ACKNOWLEDGMENT
I declare no conflicts of interest with regard to the manuscript.

REFERENCES


Taxonomy Meets Public Health: The Case of STEC


