Clinical Microbiology and Infection 22 (2016) 642.e1–642.e9

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Original article

Molecular characterization and phylogeny of Shiga toxin–producing *Escherichia coli* isolates obtained from two Dutch regions using whole genome sequencing

M. Ferdous [1], A.W. Friedrich [1], H. Grundmann [1], R.F. de Boer [2], P.D. Croughs [4], M.A. Islam [3], M.F.Q. Kluytmans-van den Bergh [5, 6], A.M.D. Kooistra-Smid [1, 2], J.W.A. Rossen [1, *]

[1] *Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands*
[2] *Department of Medical Microbiology, Certe Laboratory for Infectious Diseases, Groningen, The Netherlands*
[3] *Department of Psychiatry, Rob Giel Research Center, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands*
[4] *Star-MDC, Rotterdam, The Netherlands*
[5] *Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht, The Netherlands*
[6] *Amphia Academy Infectious Disease Foundation, Amphia Hospital, Breda, The Netherlands*

ARTICLE INFO

Article history:
Received 12 February 2016
Received in revised form
17 March 2016
Accepted 25 March 2016
Available online 4 April 2016

Editor: E. Bottieau

Keywords:
Core genome MLST
Disease outcome
Genetic diversity
Multilocus sequence typing (MLST)
Phylogenetic analysis
Shiga toxin–producing *Escherichia coli* (STEC)
Whole genome sequencing (WGS)

ABSTRACT

Shiga toxin–producing *Escherichia coli* (STEC) is one of the major causes of human gastrointestinal disease and has been implicated in sporadic cases and outbreaks of diarrhoea, haemorrhagic colitis and haemolytic uremic syndrome worldwide. In this study, we determined the molecular characteristics and phylogenetic relationship of STEC isolates, and their genetic diversity was compared to that of other *E. coli* populations. Whole genome sequencing was performed on 132 clinical STEC isolates obtained from the faeces of 129 Dutch patients with gastrointestinal complaints. STEC isolates of this study belonged to 44 different sequence types (STs), 42 serogenotypes and 14 *stx* subtype combinations. Antibiotic resistance genes were more frequently present in *stx*1-positive isolates compared to *stx*2 and *stx*1 + *stx*2 –positive isolates. The *iha, mchB, mchC, mchF, subA, ireA, senB, saa* and *sigA* genes were significantly more frequently present in *eae*-negative than in *eae*-positive STEC isolates. Presence of virulence genes encoding type III secretion proteins and adhesins was associated with isolates obtained from patients with bloody diarrhoea. Core genome phylogenetic analysis showed that isolates clustered according to their ST or serogenotypes irrespective of *stx* subtypes. Isolates obtained from patients with bloody diarrhoea were from diverse phylogenetic backgrounds. Some STEC isolates shared common ancestors with non-STEC isolates. Whole genome sequencing is a powerful tool for clinical microbiology, allowing high-resolution molecular typing, population structure analysis and detailed molecular characterization of strains. STEC isolates of a substantial genetic diversity and of distinct phylogenetic groups were observed in this study. **M. Ferdous, CMI 2016;22:642.e1–642.e9**

Introduction

Shiga toxin–producing *Escherichia coli* (STEC) is a pathogen of significant public health concern associated with both outbreaks and sporadic cases of human gastrointestinal illness worldwide [1]. Enterohaemorrhagic *E. coli*, a subpopulation of STEC, can cause bloody diarrhoea in humans, and some can cause haemolytic-uremic syndrome [2]. The ability of STEC to cause disease is associated with the production of Shiga-like toxins (Stx), which are classified into two major types, Stx1 and Stx2 (encoded by the *stx*1 and *stx*2 genes). Stx1 and Stx2 are further categorized into several subtypes; according to the new classification proposed by Scheutz *et al.* [3], Stx1 consists of three variants, Stx1a, Stx1c and Stx1d, whereas Stx2 is a diverse group composed of seven distinct variants,

* Corresponding author. J.W.A. Rossen, Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, HPC EB 80, Hanzeplein 1, 9713 GZ Groningen, The Netherlands
E-mail address: j.rossen@umcg.nl (J.W.A. Rossen).

http://dx.doi.org/10.1016/j.cmi.2016.03.028
1198-743X/

namely Stx2a, Stx2b, Stx2c, Stx2d, Stx2e, Stx2f and Stx2g. The production of Stx might not be solely responsible for pathogenesis of STEC [4]; several mobile genetic elements, e.g. plasmids, transposons, phages and pathogenicity islands, also play a role in disease outcome [4,5]. One of the major virulence factors is the outer membrane protein intimin, encoded by *eae* gene, which is part of a pathogenicity island named the locus of enterocyte effacement (LEE) [6]. Intimin is thought to be the genetic determinant of the formation of attaching and effacing (A/E) lesions, which is characterized by the intimate attachment of the bacteria to the enterocyte membrane and by the effacement of the microvilli of the enterocyte [6].

More than 100 O serotypes (based on the somatic antigen) of STEC have been associated with human disease [7], with STEC O157:H7 being the predominant serotype implicated in foodborne infections worldwide and the cause of outbreaks in many countries, including Canada, Japan, the United Kingdom and the United States [8]. However, several non-O157 STEC types have also been associated with sporadic cases and outbreaks. Six O serotypes (O26, O111, O103, O121, O45 and O145) have been reported by the US Centers for Disease Control and Prevention as the cause of 71% of non-O157 STEC infection, and these serotypes are considered as the top six STEC [7]. Moreover, in Europe, infections caused by non-O157 STEC strains are more common than those caused by O157:H7 strains [8].

Different methods have been used to classify STEC. Karmali *et al.* [9] introduced seropathotypes to assess the pathogenic potential of STEC on the basis of their reported frequencies in human illness. Subsequently, the classification was modified by the European Food Safety Authority on the basis of the health outcome of reported confirmed human verotoxigenic *E. coli* cases. For epidemiologic purposes, various genetic fingerprinting methods have been developed to identify, trace and prevent dissemination of STEC [10]. Among the sequence-based methods, multilocus sequence typing (MLST) is a reliable method to determine genetic relatedness of epidemiologically unrelated isolates, but it has limited discriminatory power [11]. Currently whole genome sequencing (WGS) of bacterial genomes is an accessible and affordable method [12,13]. An obvious application for WGS is epidemiologic typing to detect and support outbreak investigations, to define transmission pathways of pathogens and to reveal laboratory cross-contamination [14]. Because of the public health importance of STEC infections, epidemiologic and molecular surveillance systems are essential for early outbreak detection and to differentiate STEC strains based on their potential to cause severe illness in humans [15,16].

This study was performed to determine the molecular characteristics, phylogenetic relationship and diversity of STEC isolates from faeces of patients obtained from two regions in the Netherlands and to reveal the relation between molecular determinants and disease outcome.

Methods

Collection of isolates

A multicentre prospective study, STEC-ID-net, was performed during the period from April 2013 to March 2014 in the Dutch regions of Groningen and Rotterdam. Stool samples from patients with suspected infectious gastroenteritis were screened using quantitative PCR targeting the *stx*1/*stx*2/*escV* genes, and positive samples were further processed to obtain pure *stx*-positive isolates [17].

Phenotypic antibiotic resistance pattern

Antibiotic resistance patterns of the STEC isolates were determined using VITEK2 (bioMerieux, Marcy l'Etoile, France) following European Committee on Antimicrobial Susceptibility Testing guidelines.

Whole genome sequencing

To perform the WGS of *stx*-positive isolates, DNA was extracted using the UltraClean microbial DNA isolation kit (MO BIO Laboratories, Carlsbad, CA, USA) according to the manufacturer's protocol. A DNA library was prepared using the Nextera XT kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions, then run on a MiSeq (Illumina) for generating paired-end 250 bp reads, aiming at a coverage of at least 60-fold.

Data analysis and molecular characterization

De novo assembly was performed using CLC Genomics Workbench 7.0.3 (CLC Bio, Aarhus, Denmark) after quality trimming (Qs $\geq$ 28) with optimal word sizes based on the maximum N50 value [18]. The average N50 value of the sequenced 132 STEC isolates was 144 166 (range, 23 341–387 919), and the average number of contigs was 178 (range, 59–497). Gene annotation was performed by uploading the assembled genome onto the RAST server version 2.0 [19]. The sequence types (STs) and O and H serogenotypes were identified by uploading the assembled genomes to the MLST Finder 1.7 [20] and SerotypeFinder 1.1 tool [21], respectively, of the Center for Genomic Epidemiology (CGE) website. The virulence genes and *stx* subtypes were determined using CGE VirulenceFinder 1.2 [22], and antibiotic resistance genes were determined by CGE ResFinder 2.1 [23]. For the CGE server, the threshold of ID was set to 85% and the percentage of minimum overlapping gene length to 60%. The sequences assigned unknown ST by CGE MLST finder were submitted to the EnteroBase database of the University of Warwick.

This whole-genome shotgun project has been deposited in National Center for Biotechnology Information (NCBI) under BioProject PRJNA285020. The GenBank accession numbers are listed in Supplementary Table S1.

Statistical analysis

To determine the association of virulence genes with different patient groups and to compare the presence of virulence genes in *eae*-positive and -negative STEC isolates, the Pearson chi-square test was used. To observe the effect of the virulence genes on patient groups, univariate binary logistic regression was performed. All analyses were done using two-tailed tests at a 5% significance level. The statistical analyses were performed using IBM SPSS Statistics 22 (IBM SPSS, Chicago, IL, USA).

Phylogenetic analysis of STEC isolates

To determine the phylogenetic relationship of the isolates, a gene-by-gene approach was performed by SeqSphere$^+$ 3.0 (Ridom, Münster, Germany). Briefly, an *ad hoc* core genome MLST (cgMLST) scheme was developed using the genome of *E. coli* O157:H7 strain Sakai (accession no. NC_002695) as the reference genome and an additional ten *E. coli* as query genomes (Supplementary File S2) to extract open reading frames (ORFs) from the genome of each isolate using MLST+ Target Definer 2.1.0 of SeqSphere$^+$. Only the ORFs without premature stop codons and ambiguous nucleotides from contigs of assembled genomes were included. The genes shared by all isolates analysed were defined as the core genome for phylogenetic analysis [18]. A neighbour-joining (NJ) tree was constructed on the basis of a distance matrix among the isolates, depending on the core genome of all isolates. To compare the

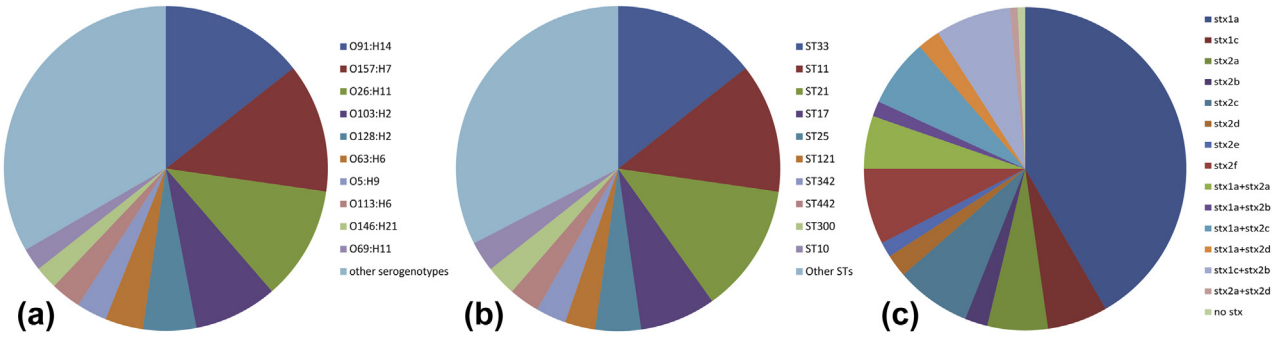

Fig. 1. Distribution of different serogenotypes (a), sequence types (STs) (b) and *stx* subtype combinations (c) among STEC isolates. *Other serogenotypes and STs of all isolates are summarized in Supplementary Table S1.

isolates with previously reported ones, five additional genomes of STEC (*E. coli* strain Sakai, strain EDL933, strain 11368, strain 2011c-3493 and strain 12009) available on NCBI were included in the phylogenetic analysis. To see the allele differences among the isolates, a minimum spanning tree (MST) was constructed by SeqSphere^{+} based on the numerical allele type for each isolates according to the sequence identity of each gene [24]. An additional NJ tree based on the accessory genome of the isolates was also constructed.

Phylogenetic comparison of STEC isolates with diarrhoeagenic E. coli (DEC) reference collection

The phylogenetic relationship of the STEC isolates of this study with isolates of the DEC reference collection (n = 76) [25] was determined by cgMLST. The DEC consists of predominant clones of diarrhoeagenic *E. coli,* including 27 STEC, 25 enteropathogenic *E. coli* (EPEC) and 24 non-STEC/EPEC *E. coli.*

Genetic diversity of STEC isolates

To reveal the genetic diversity of the STEC isolates of this study, median pairwise distance (MPD) was calculated [26] and the MPD was compared to that of two other independent strain collections, i.e. 76 isolates of the DEC collection and 131 isolates that were randomly selected from a collection of extended-spectrum β-lactamase–producing Enterobacteriaceae (ESBL-E) from a multicentre study on the epidemiology of ESBL-E in Dutch hospitalized patients (SoM study, unpublished data). Because neither of the underlying population followed normal distribution pattern, the median of the pairwise distances and interquartile ranges (IQR) were calculated in all three populations. To identify the group differences on MPD of three *E. coli* isolates, the Kruskal-Wallis test was used. Additionally, to compare the genetic diversity within the STEC isolates with that of other *E. coli* populations, the Mann-Whitney *U* test was performed, and the statistical significance level was corrected by Bonferroni correction for multiple testing.

Results

STEC isolates and patient groups

From 425 *stx*-positive faecal samples, 132 STEC isolates (Groningen, n = 70; Rotterdam, n = 62) from 129 patients were obtained. From three patients, two different types of STEC were isolated. Characteristics of the isolates are summarized in Supplementary Table S1. On the basis of clinical outcomes (available for 110 patients), patients were classified into groups with bloody diarrhoea (including one haemolytic-uremic syndrome patient) (n = 26), nonbloody diarrhoea (n = 64) and no diarrhoea (n = 20) but having other clinical symptoms, such as abdominal pain, nausea and malaise (Supplementary Table S1).

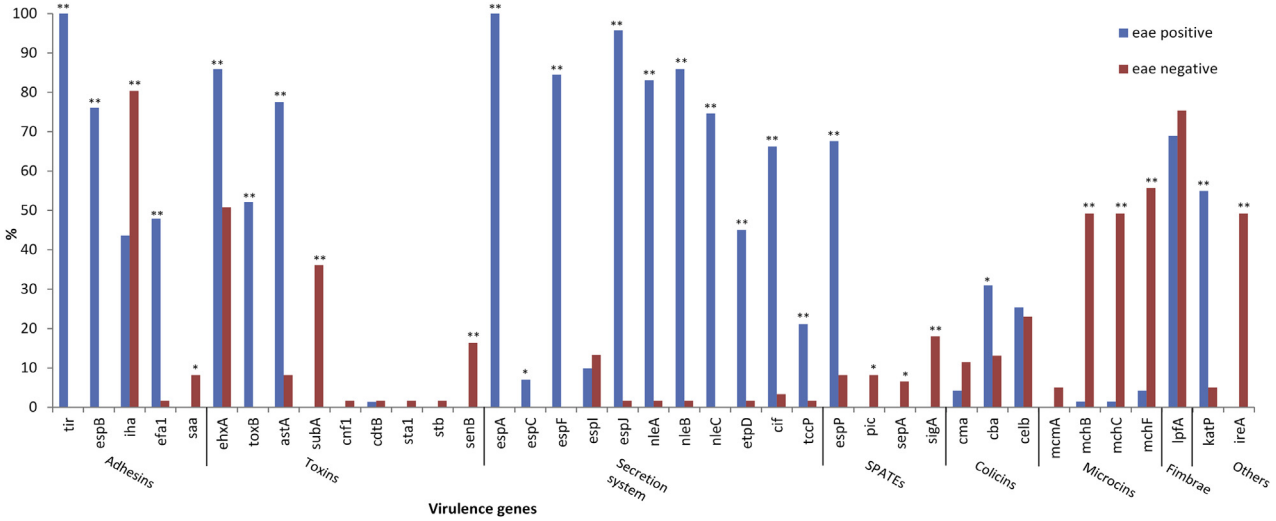

Fig. 2. Comparison of virulence genes in *eae*-positive and *eae*-negative STEC isolates. Blue and red bars represent frequency of virulence genes in *eae*-positive and *eae*-negative isolates, respectively. *p <0.05, **p <0.001.

Serotypes, STs and stx subtypes

Forty-two different serogenotypes and 44 different STs were found among the 132 isolates. The most predominant serogenotypes were O91:H14 ($n = 19$), O157:H7 ($n = 17$) and O26:H11 ($n = 15$) (Fig. 1a). The most predominant STs were ST33 ($n = 19$), ST11 ($n = 17$) and ST21 ($n = 17$) (Fig. 1b). For six of the isolates, we found new STs assigned by EnteroBase database (Supplementary Table S1). Fourteen different *stx* subtypes (combinations) were found among the isolates, with *stx*1a being most predominant ($n = 55$, 41.6%) (Fig. 1c). Almost all O91:H14 isolates contained only the *stx*1a gene except two, of which one (STEC 2620) contained the *stx*1a + *stx*2b, and the other (STEC 2110-1) contained only the *stx*2a gene (Supplementary Table S1). One *stx*2-positive isolate (STEC 2826) had both *stx*2d (encoding Stx A subunit) and *stx*2a (encoding Stx B subunit) genes.

Presence of virulence genes

Among the *eae*-positive isolates ($n = 71$), 33 (46%) contained *stx*1, 21 (30%) contained *stx*2, 16 (23%) contained *stx*1 + *stx*2 and one isolate was found without the *stx* gene. Among the *eae*-negative isolates ($n = 61$), the prevalence of *stx*1, *stx*2 and *stx*1 + *stx*2−positive isolates was 49% ($n = 30$), 26% ($n = 16$) and 24.5% ($n = 15$), respectively. In *eae*-positive isolates, the virulence genes *tir, espA, espB, espF, espJ, espP, nleA, nleB, nleC, etpD, katP, toxB, efa1* and *cif* were significantly more often present, whereas *iha, mchB, mchC, mchF, subA, ireA, sepA, senB, saa* and *sigA* genes were significantly more often found in *eae*-negative isolates (Fig. 2). The associations and the effects (odds ratios and their confidence intervals) of the presence of virulence genes with bloody diarrhoea and diarrhoea are presented in Table 1.

Antibiotic resistance patterns and presence of resistance genes

Phenotypic antibiotic resistance patterns of the isolates are shown in Supplementary Table S3. In 33 STEC isolates (25%), at least one antibiotic resistance gene was found. Although not statistically significant, the presence of resistance genes was higher in only *stx*1-positive isolates (35.5%) compared to *stx*2 (16.5%) (p 0.063)- and *stx*1 + *stx*2 (15%) (p 0.054)-positive isolates. Antibiotic resistance genes found in STEC isolates are listed in Table 2. Two isolates

Table 1
Distribution of virulence genes (other than *stx*) among isolates obtained from three patient groups

Virulence gene category	Virulence gene	Patient group, *n* (%)			p[a]	OR[b] (95% CI)	
		BD ($n = 26$)	NBD ($n = 64$)	ND ($n = 20$)		BD vs. NBD	BD vs. ND
	eae	21 (80)	34 (53)	7 (35)	0.006*	3.7 (1.24–11.04)	7.8 (2.04–29.7)
	tir	21 (80)	34 (53)	7 (35)	0.006*	3.7 (1.2–11.04)	7.8 (2.04–29.7)
	espB	18 (69)	24 (37.5)	6 (30)	0.009*	3.7 (1.4–9.9)	5.2 (1.4–18.6)
	iha	14 (54)	36 (56)	15 (75)	0.272	0.9 (0.36–2.2)	0.39 (0.11–1.38)
	efa1	11 (42)	18 (28)	4 (20)	0.231	1.87 (0.72–4.8)	2.93 (0.76–11.2)
	saa	1 (4)	2 (3)	1 (5)	0.924	1.24 (0.10–14.3)	0.76 (0.04–12.9)
Toxin	*ehxA*	24 (92)	41 (64)	13 (65)	0.023*	6.7 (1.4–31)	6.4 (1.1–35.7)
	toxB	12 (46)	15 (23)	4 (20)	0.063	2.8 (1.06–7.3)	3.4 (0.89–13)
	astA	17 (65)	26 (41)	7 (35)	0.059	2.7 (1.06–7.13)	3.5 (1.03–11.9)
	subA	2 (8)	10 (15.6)	5 (25)	0.273	0.45 (0.92–2.21)	0.25 (0.04–1.45)
	cnf1	0 (0)	1 (1.6)	0 (0)	0.696	Undefined	Undefined
	cdtB	0 (0)	1 (1.6)	1 (5)	0.441	Undefined	Undefined
	senB	1 (4)	5 (8)	2 (10)	0.704	0.47 (0.05–4.2)	0.36 (0.03–4.28)
Secretion system	*espA*	21 (80)	33 (51)	7 (35)	0.005*	3.9 (1.3–11.7)	7.8 (2.04–29.7)
	espC	1 (3.8)	3 (4.7)	0 (0)	0.619	0.81 (0.08–8.19)	Undefined
	espF	19 (73)	27 (42)	5 (25)	0.003*	3.7 (1.3–10)	8.1 (2.1–30.8)
	espI	3 (11.5)	8 (12.5)	2 (10)	0.954	0.91 (0.22–3.75)	1.17 (0.18–7.8)
	espJ	21 (80)	33 (51)	6 (30)	0.002*	3.9 (1.3–11.7)	9.8 (2.5–38.4)
	nleA	20 (77)	27 (42)	6 (30)	0.002*	4.5 (1.6–12.9)	7.7 (2.07–29.1)
	nleB	19 (73)	30 (47)	6 (30)	0.011*	3.07 (1.1–8.3)	6.3 (1.7–23)
	nleC	16 (61.5)	22 (34)	6 (30)	0.035*	3.05 (1.1–7.8)	3.7 (1.08–12.9)
	etpD	12 (46)	15 (43)	1 (5)	0.005*	2.8 (1.06–7.3)	16.2 (1.8–140)
	cif	12 (46)	26 (40.6)	6 (30)	0.534	1.25 (0.50–3.13)	2.00 (0.58–6.83)
	tccP	2 (8)	13 (20)	1 (5)	0.125	0.33 (0.06–1.56)	1.58 (0.13–18.8)
SPATE	*espP*	16 (61.5)	23 (36)	6 (30)	0.045*	2.8 (1.1–7.3)	3.7 (1.08–12.9)
	pic	1 (4)	4 (6)	0 (0)	0.494	0.60 (0.06–5.63)	Undefined
	sepA	1 (4)	2 (3)	0 (0)	0.69	1.24 (0.10–14.2)	Undefined
	sigA	1 (4)	5 (8)	3 (15)	0.387	0.47 (0.05–4.24)	0.22 (0.02–2.36)
Colicin	*cma*	3 (11.5)	3 (4.7)	1 (5)	0.465	2.65 (0.49–14)	2.47 (0.23–25.8)
	cba	7 (27)	12 (19)	5 (25)	0.648	1.6 (0.54–4.65)	1.10 (0.29–4.19)
	celb	6 (23)	17 (26.6)	5 (25)	0.941	0.83 (0.28–2.41)	0.9 (0.23–3.51)
Microcin	*mch*B	4 (15)	13 (20)	8 (40)	0.110	0.71 (0.21–2.43)	0.27 (0.06–1.1)
	mchC	4 (15)	13 (20)	8 (40)	0.110	0.71 (0.21–2.43)	0.27 (0.06–1.1)
	mchF	6 (23)	17 (26)	8 (40)	0.407	0.83 (0.28–2.41)	0.45 (0.12–1.61)
	mcmA	1 (4)	1 (1.6)	1 (5)	0.657	2.5 (0.15–41.8)	0.76 (0.04–12.9)
Fimbriae	*lpfA*	20 (77)	44 (69)	17 (85)	0.323	1.51 (0.53–4.34)	0.59 (0.12–2.71)
Other	*ireA*	3 (11.5)	14 (22)	8 (40)	0.071	0.46 (0.12–1.78)	0.19 (.04-.87)
	katP	10 (38.5)	22 (34)	4 (20)	0.379	1.12 (0.46–3.06)	2.5 (0.64–9.6)

BD, bloody diarrhoea; CI, confidence interval; NBD, nonbloody diarrhoea; ND, no diarrhoea; OR, odds ratio.
*Statistically significant p values.
[a] p values were obtained by chi-square test comparing three groups.
[b] OR obtained from binary logistic regression. OR was also calculated for group D vs. ND but was found to be not significant and therefore is not mentioned in this table.

(STEC 381-1 and STEC 1255) of serotype O104:H4 and O5:H9 were ESBL producers and contained the *blaCTX-M-15* and *blaCTX-M-1* genes, respectively.

Phylogenetic analysis

The core genome phylogenetic tree was constructed including those 2069 genes (defined as core genome) shared by all 137 isolates, including five reference STEC isolates (Fig. 3). To describe the isolates from the tree, the NJ tree in Fig. 3 is arbitrarily divided into eight groups (groups 1 to 8). Isolates of the same serogenotype or ST are clustered together, irrespective of their *stx* subtypes and isolation region (Groningen or Rotterdam). Group 2, 4, 5 and 7 contained isolates of serogenotypes O157:H7, O26:H11, O91:H14 and O103:H2, respectively, which were the most prevalent serogenotypes in this study. In group 7, isolates of serotype O128:H2 formed a subcluster. Groups 3 and 6 contained heterogeneous isolates of different STs and serogenotypes. All but one (STEC 309, in group 7) of the *stx* 2f-positive isolates clustered together (in group 1) and belonged to O63:H6 and O113:H6 serogenotypes (Fig. 3).

Several subclusters were observed in a single ST cluster. In group 4, isolates of serogenotype O26:H11 formed different subclusters within ST21; notably, one of the subclusters contained three O69:H11 isolates which had a 290 minimum allele difference (data not shown) from their closest O26:H11 isolate STEC 380. In some cases, isolates from the same O serogroups, e.g. O5:H9 and O5:H19, were found to be scattered at distinct positions in the tree. On the other hand, most of the cases—isolates of the same H type, irrespective of their O serotypes—shared a common ancestor, e.g. serogroups O103:H2 and O128:H2; serogroups O174:H21, O146:H21 and O91:H21; and serogroups O113:H6 and O63:H6.

To reveal the genetic relatedness of the isolates based on their virulence genes and other mobile genetic elements, an additional NJ tree was constructed. This tree was based on the accessory genome containing 2586 genes that were present in at least one isolate but not in all of the 137 isolates (Supplementary Fig. S1). The tree was almost identical to the core genome phylogenetic tree, and no clusters based on disease severity were found.

Distribution of isolates with clinical manifestation, geographical location and epidemiologic link

The isolates obtained from the patients with bloody diarrhoea belonged to different serogenotypes and STs (Fig. 3). No significant geographical distribution was observed among the STEC isolates in relation to their genotypes. Two exceptions should be mentioned within group 6, in which all four isolates belonging to ST442 were obtained from patients in the Rotterdam region. In addition, in group 1, seven of nine *stx*2f isolates were obtained from patients in the Groningen region. All other closely related subclusters contained isolates from both regions. Isolates of two clusters containing three (STEC 338, STEC 381-1 and STEC 381-4 in group 6) and two (STEC 690 and STEC 757 in group 4) epidemiologically related isolates were also closely genetically related. In Fig. 3, we also highlight isolates having almost similar core genomes, e.g. STEC 563 and STEC 709 (three-allele difference), STEC 479 and STEC 487 (no allele difference), STEC 2174 and STEC 2363 (one-allele difference), STEC 384 and STEC 464 (no allele difference), and STEC 299 and STEC 2441 (two-allele difference); for those isolates, no

Table 2
Presence of antibiotic resistance genes in STEC isolates

Isolate ID	Serogenotype	*stx* subtype	Presence of antibiotic resistance genes
STEC 168	O91:H14	*stx*1a	*dfrA1, strA-B, sul2*
STEC 169	O6:H10	*stx*1c	*aadA1, blaTEM-1B, dfrA1, mphB, strA-B, sul1–2, tetA*
STEC 196	O91:H14	*stx*1a	*aadA1, sul1, tetA*
STEC 200	O174:H21	*stx*2c	*strA-B, sul2, tetB*
STEC 299	O5:H9	*stx*1a	*strA-B, sul2*
STEC 329	O91:H14	*stx*1a	*aadA5, catB3, dfrA1, sul1, tetA*
STEC 338	O104:H4	*stx*2a	*blaTEM-1B, dfrA7, sul1*
STEC 370	O111:H8	*stx*1a	*aadA1, aph(3′)-Ia, catA, dfrA1, mphB, strA-B, sul1, tetA, tetM*
STEC 381-1	O104:H4	*stx*2a	*bla CTX-M-15, blaTEM-1B, dfrA7, sul1*
STEC 381-4	O104:H4	*stx*2a	*blaTEM-1B, dfrA7, sul1*
STEC 479	O26:H11	*stx*1a	*aadA1, blaTEM-1B, dfrA1, mphB, strA-B, sul1–2, tetA*
STEC 487	O26:H11	*stx*1a	*aadA1, blaTEM-1B, dfrA1, mphB, strA-B, sul1–2, tetA*
STEC 690	O69:H11	*stx*1a	*aac(3)-IIa, aadA1, aph(3′)-Ic, blaTEM-1A, catA1, dfrA1, mphB, strA-B, sul1–2, tetA-B*
STEC 691	O69:H11	*stx*1a	*aac(3)-IIa, aadA1, aph(3′)-Ic, blaTEM-1A, catA1, dfrA1, mphB, strA-B, sul1–2, tetA-B*
STEC 757	O69:H11	*stx*1a	*aac(3)-IIa, aadA1, aph(3′)-Ic, blaTEM-1A, catA1, dfrA1, mphB, strA-B, sul1–2, tetA-B*
STEC 1255	O5:H9	*stx*1a+ *stx*2a	*blaCTX-M-1, mphA*
STEC 1500	O76:H19	*stx*1c	*strA-B*
STEC 1585	O91:H14	*stx*1a	*aadA1, sul1, tetA*
STEC 2193	O103:H2	*stx*1a	*blaTEM-1C, strA-B, sul2*
STEC 2236	O113:H4	*stx*2d	*strA-B, sul2, tetB*
STEC 2359	O55:H12	*stx*1a	*blaTEM-1B, strA-B, sul2, tetA*
STEC 2441	O5:H9	*stx*1a	*strA-B, sul2*
STEC 2564	O117:H7	*stx*1a	*dfrA14, strA-B, sul2, tetA*
STEC 2573	O112:H19	*stx*1a+ *stx*2d	*blaTEM-1B, catA, floR, strA-B,sul2, tetA*
STEC 2633	O146:H10	*stx*1a	*aadA1, sul1, tetA*
STEC 2743	O103:H2	*stx*1a	*aadA1, blaTEM-1B, dfrA1, strA-B, sul1–2, tetA*
STEC 2770	O157:H7	*stx*1a+ *stx*2c	*blaTEM-1B, strA-B, sul2, tetA*
STEC 2797	O100:H30	*stx*2e	*aadA1, blaTEM-1B, dfrA1, strA-B, sul1–2, tetA*
STEC 2820	O157:H7	*stx*1a+ *stx*2c	*aadA1, aph(3′)-Ia, blaTEM-1B, dfrA8, strA-B, sul2, tetA*
STEC 2821	O157:H7	*stx*1a+ *stx*2c	*blaTEM-1B, strA-B, sul2, tetA*
STEC 3084	O91:H14	*stx*1a	*aadA5, catB, dfrA1, sul1, tetA*
STEC 3087	O91:H14	*stx*1a	*aadA5, catB, dfrA1, sul1, tetA*
STEC 3106	O91:H14	*stx*1a	*tetA*

STEC, Shiga toxin–producing *Escherichia coli.*

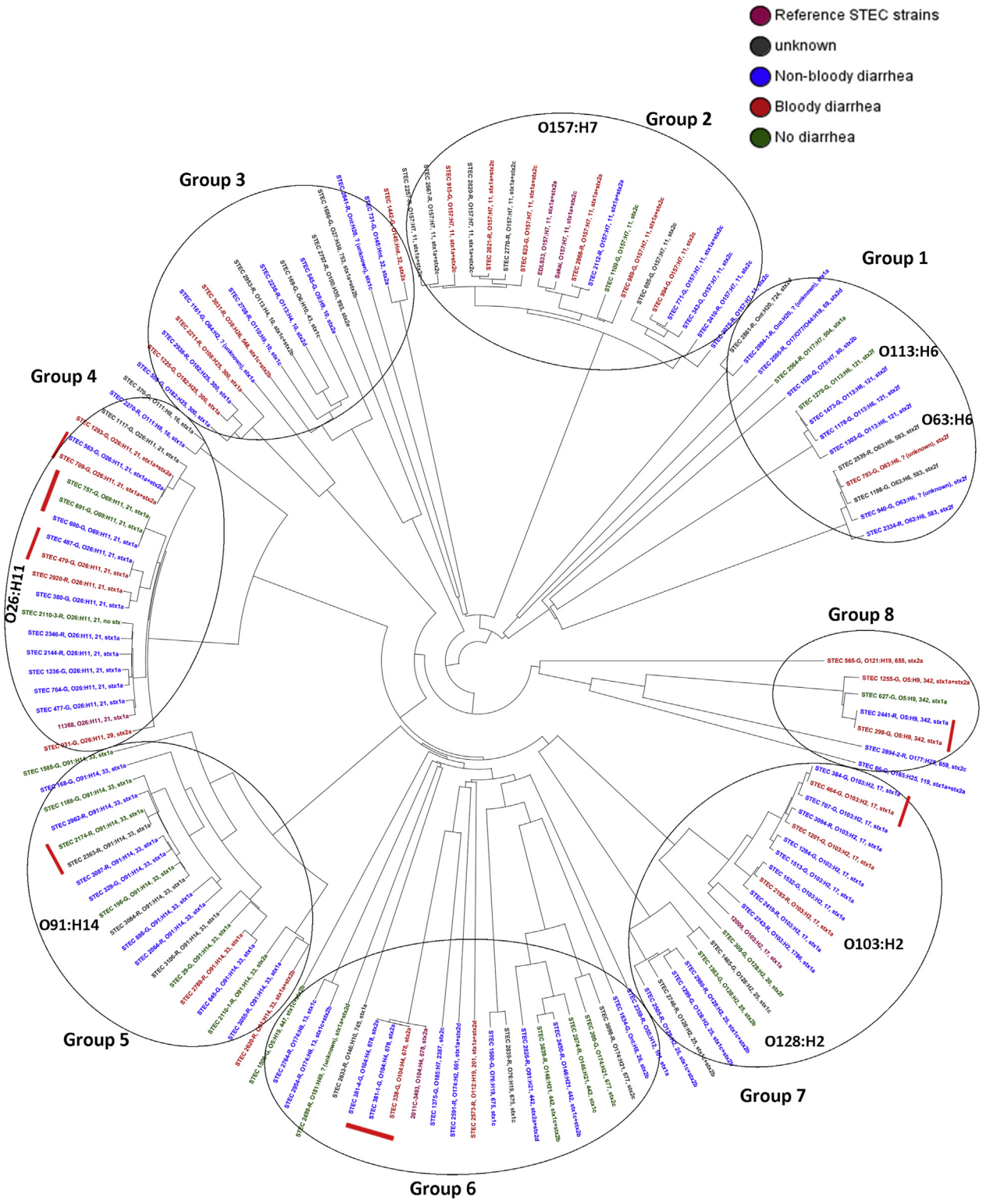

Fig. 3. Neighbour-joining (NJ) phylogenetic tree of STEC isolates based on 2069 genes (defined as core genome) shared by 137 STEC isolates (including five reference STEC). Whole tree is arbitrarily divided into eight groups, indicated by circles (groups 1 to 8). Predominant serogenotypes are mentioned in circles. Each isolate ID is followed by isolation region (G for Groningen and R for Rotterdam), serogenotype, sequence type and *stx* subtype. Closely related isolates with allele difference <5 (obtained from minimum spanning tree analysis) are highlighted with red lines behind them. New sequence types were not updated in Seqsphere server during analysis and therefore were left as '?' as unknown ST. ST, sequence type; STEC, Shiga toxin–producing *Escherichia coli*.

Fig. 4. Neighbour-joining (NJ) phylogenetic tree of STEC isolates and DEC reference isolates based on 1231 genes (defined as core genome) shared by 208 isolates. Non-STEC isolates are marked with red asterisk. For each isolate, serogenotypes and sequence types are provided behind isolate name. DEC, diarrhoeagenic *Escherichia coli;* STEC, Shiga toxin–producing *E. coli*.

epidemiologic data were available. The number of allele differences was obtained from MST analyses (data not shown).

Phylogenetic comparison of STEC isolates with DEC reference collection

Figure 4 shows a NJ tree representing all the STEC isolates from this study and *E. coli* isolates from the DEC collection. STEC isolates of this study represented a diverse collection and did not form a separate cluster but were interspersed among the DEC isolates. The phylogenetic tree shows that STEC serogroups O157:H7 evolved from *E. coli* O55:H7 isolates (DEC5) and O26:H11 clustered with DEC9 and DEC10 isolates in a separate lineage. In both cases, clusters contained only STEC isolates. In contrast, STEC O103:H2 shared a common ancestor with EPEC O111:H2 (DEC12) and O128:H2 (DEC11A–D). STEC O63:H6 and O113:H6, both belonging to the *stx*2f subtype, shared a common ancestor with EPEC O55:H6 (DEC1 and DEC2). Among the non-STEC/EPEC isolates, DEC13, DEC14 and DEC15 isolates shared a common ancestor with several of our STEC isolates of heterogeneous serotypes, and DEC6 (O111:H21) isolates clustered together with our STEC ST10 isolates.

Genetic diversity of STEC isolates compared to DEC isolates and ESBL-producing E. coli isolates

The MPD of the STEC, DEC and ESBL-producing *E. coli* isolates was 0.86 (IQR 0.24), 0.93 (IQR 0.24) and 0.97 (IQR 0.10), respectively. The result from the Kruskal-Wallis test indicated that the MPD of the three populations (STEC, DEC and ESBL) was

significantly different ($p < 0.001$). Additionally, a Mann-Whitney *U* test showed that there was a highly significant difference ($p < 0.001$) between the MPD of STEC and ESBL, and DEC and ESBL ($p < 0.001$), but not between the MPD of STEC and DEC (p 0.137).

Discussion

In this study, STEC isolates were collected from faecal samples of patients with gastrointestinal complaints from two regions of the Netherlands. Molecular characterization and high-resolution typing of the isolates was performed using WGS. Serogenotyping, MLST and subtyping of *stx* genes revealed a diverse group of STEC in the study population. Twenty-five per cent of the isolates carried one or more antibiotic resistance genes, including ESBL genes. Resistance genes were mostly found in *stx*1-positive isolates belonging to O serotypes other than the big six known to be most frequently involved in severe human infection. These isolates often originate from food-producing animals, which are regularly treated with antibiotics that may lead to STEC becoming resistant. In addition, *stx*1-positive bacteria often cause only mild symptoms without bloody diarrhoea, and patients may receive antibiotic therapy if no detailed diagnostics is performed [27–29]. Recently, several reports, including that of on the 2011 German *E. coli* O104:H4 outbreak, have described the association of STEC isolates with ESBL genes [30]. Antibiotic resistance genes are mainly carried on mobile genetic elements that can be transferred from one bacterium to the other while subjected to selective pressure, e.g. by exposure to antibiotics. Transferring of resistance genes into clinically significant bacteria for human could make their treatment option complicated [31].

The presence of several virulence genes (*iha, mchB, mchC, mchF, subA, ireA, senB, saa, sigA*) was significantly higher in *eae*-negative isolates compared to *eae*-positive ones, similar to that reported by previous studies in which these genes were described as additional virulence factors in *eae*-negative strains [32,33]. The presence of virulence genes (*eae, tir, espA, espF, espJ*) associated with the LEE pathogenicity island and the non-LEE-encoded effector (nle) that encodes translocated substrates of the type III secretion system was more frequent in isolates obtained from patients with bloody diarrhoea. These genetic determinants were also described to be associated with highly pathogenic STEC and therefore with severe disease [15,34], although a wide number of STEC LEE-negative strains also have been associated with sporadic cases and outbreaks [4]. There was no correlation between the serogenotype or STs and disease outcome. Thus, isolates obtained from bloody diarrhoea did not belong to a specific phylogenetic cluster but were scattered throughout the phylogenetic tree. This finding supports the idea that STEC from different phylogenetic backgrounds could be responsible for severe disease outcome in human by acquiring virulence factors contributing to their pathogenicity [15]. Furthermore, clustering of isolates according to their STs and serogenotype pattern irrespective of *stx* subtypes suggests that Stx-converting phages are carried by a genetically diverse group of *E. coli* [35]. In addition, the phylogenetic tree based on the accessory genome of the isolates also showed no correlation with disease severity. Clearly, disease outcome is multifactorial and does not only depend on the genetic contents or virulence factors of the isolates but also on host susceptibility factors and several Stx phage-related factors [36].

MLST provides an adequate tool for producing genetic profiles for a vast number of isolates, especially in nonepidemic circumstances, i.e. for national reference services or when comparing large international strain collections, but it has a low discriminatory power. We found several subclusters within the same ST and serogenotype clusters. In addition, isolates highly similar in their core genome were also identified by cgMLST. Five historical isolates obtained from outbreaks or sporadic cases were included in the phylogeny, and we found that two of our O157:H7 isolates clustered with two O157:H7 isolates associated with a previous outbreak [18] and that three epidemiologically related O104:H4 isolates clustered with the 2011 *E. coli* O104:H4 outbreak isolate 2011c-3493 [24]. Therefore, the detailed gene-by-gene phylogenetic approach using cgMLST enabled us to discriminate among isolates within the same STs and helped us to identify potentially more virulent clones, thereby improving risk assessment and outbreak management. Isolates belonging to different serogroups clustered with each other, suggesting that using just serogroups may cause misleading conclusions about the phylogenetic relatedness between STEC strains and their health risks [37]. However, the two regions (Groningen and Rotterdam) from which the isolates were obtained are only approximately 250 km away, and both regions probably share many food sources. This may be an explanation for why no geographical distribution was observed among the STEC isolates of the different regions. In some cases, isolates of the same O serogroup were located in different phylogenetic clusters. On the other hand, isolates of the same H type, irrespective of their O serotype, shared a common ancestor. This finding is in concordance with previous findings where H serogroups were described as monophyletic, whereas O serogroups were described as polyphyletic [37,38].

Comparing STEC isolates of this study with DEC isolates revealed that the STEC isolates represent a heterogeneous group. Some of the serogroups formed distinct branches containing only STEC isolates. However, some serogroups shared a common ancestor with EPEC and other *stx*/*eae*-negative DEC isolates. This supports the hypothesis that Stx converting bacteriophages can integrate into different *E. coli* pathogroups, thereby converting them into a more pathogenic variants [33]. Moreover, STEC isolates of this study had a similar diversity pattern compared to DEC isolates but were less diverse than ESBL-producing *E. coli* isolates. This result supports the idea that Stx converting bacteriophages may have a preference in host selection [36].

Our study has several limitations. Firstly, it was not possible to link all the isolate characteristics with patient disease outcome and epidemiology as a result of the lack of patient information. Therefore, an unknown epidemiologic linkage may exist among isolates, which might have influenced the reported diversity. Moreover, as some of the faecal samples were also positive for parasites, and other bacteria causing gastroenteritis and all samples were not tested for gastroenteric viruses, it was not possible to elucidate the exact aetiology of the disease outcome.

In conclusion, STEC isolates of a substantial genetic diversity and of distinct phylogenetic groups were observed in two regions of the Netherlands. WGS serves perfectly well for detailed characterization of STEC strains compared to serotyping and MLST, which have less discriminatory power and which do not provide any information on virulence and resistance genes. WGS of STEC could be useful for outbreak tracing within a clinical outbreak, but because STEC strains are diverse, it may not always be suitable for comparing different outbreaks. WGS may not always be useful to find common ancestors of STEC because of its great heterogeneity and incorporation of mobile genetic elements, but so far, it is the best available method. There was no clear correlation between serogenotype, *stx* subtype or ST and disease outcome, as it is also influenced by several factors in addition to virulence factors or a specific pathotype.

Acknowledgements

We thank all the project members of the STEC-ID-net study. This work was done in collaboration with the European Society of

Clinical Microbiology and Infectious Diseases Study Group on Molecular Diagnostics, Basel, Switzerland.

Transparency Declaration

All authors report no conflicts of interest relevant to this article.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cmi.2016.03.028.

References

[1] Paton JC, Paton AW. Pathogenesis and diagnosis of Shiga toxin–producing *Escherichia coli* infections. Clin Microbiol Rev 1998;11:450–79.

[2] Mellmann A, Bielaszewska M, Köck R, Friedrich AW, Fruth A, Middendorf B, et al. Analysis of collection of hemolytic uremic syndrome–associated enterohemorrhagic *Escherichia coli*. Emerg Infect Dis 2008;14:1287–90.

[3] Scheutz F, Teel LD, Beutin L, Piérard 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.

[4] Farfan MJ, Torres AG. Molecular mechanisms that mediate colonization of Shiga toxin–producing *Escherichia coli* strains. Infect Immun 2012;80:903–13.

[5] Torres AG, Vazquez-Juarez RC, Tutt CB, Garcia-Gallegos JG. Pathoadaptive mutation that mediates adherence of Shiga toxin–producing *Escherichia coli* O111. Infect Immun 2005;73:4766–76.

[6] Elliott SJ, Sperandio V, Girón JA, Shin S, Mellies JL, Wainwright L, et al. The locus of enterocyte effacement (LEE)-encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic *Escherichia coli*. Infect Immun 2000;68:6115–26.

[7] Brooks JT, Sowers EG, Wells JG, Greene KD, Griffin PM, Hoekstra RM, et al. Non-O157 Shiga toxin–producing *Escherichia coli* infections in the United States, 1983–2002. J Infect Dis 2005;15(192):1422–9.

[8] Blanco JE, Blanco M, Alonso MP, Mora A, Dahbi G, Coira MA, et al. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: prevalence in Lugo, Spain, from 1992 through 1999. J Clin Microbiol 2004;42:311–9.

[9] 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.

[10] Heir E, Lindstedt BA, Vardund T, Wasteson Y, Kapperud G. Genomic fingerprinting of shigatoxin-producing *Escherichia coli* (STEC) strains: comparison of pulsed-field gel electrophoresis (PFGE) and fluorescent amplified-fragment-length polymorphism (FAFLP). Epidemiol Infect 2000;125:537–48.

[11] European Food Safety Authority. Urgent advice on the public health risk of Shiga-toxin producing *Escherichia coli* in fresh vegetables. EFSA J 2011;9:2274.

[12] Fricke WF, Rasko DA. Bacterial genome sequencing in the clinic: bioinformatics challenges and solutions. Nat Rev Genet 2014;15:49–55.

[13] Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. Pathology 2015;47:199–210.

[14] Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog 2012;8. e1002824.

[15] Haugum K, Johansen J, Gabrielsen C, Brandal LT, Bergh K, Ussery DW, et al. Comparative genomics to delineate pathogenic potential in non-O157 Shiga toxin–producing *Escherichia coli* (STEC) from patients with and without haemolytic uremic syndrome (HUS) in Norway. PLoS One 2014;9. e111788.

[16] Berenger BM, Berry C, Peterson T, Fach P, Delannoy S, Li V, et al. The utility of multiple molecular methods including whole genome sequencing as tools to differentiate *Escherichia coli* O157:H7 outbreaks. Euro Surveill 2015;20.

[17] de Boer RF, Ferdous M, Ott A, Scheper HR, Wisselink GJ, Heck ME, et al. Assessing the public health risk of Shiga toxin–producing *Escherichia coli* by use of a rapid diagnostic screening algorithm. J Clin Microbiol 2015;53:1588–98.

[18] Ferdous M, Zhou K, Mellmann A, Morabito S, Croughs PD, de Boer RF, et al. Is Shiga toxin–negative *Escherichia coli* O157:H7 enteropathogenic or enterohemorrhagic *Escherichia coli*? Comprehensive molecular analysis using whole-genome sequencing. J Clin Microbiol 2015;53:3530–8.

[19] Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics 2008;9:75.

[20] Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 2012;50:1355–61.

[21] Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. . Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J Clin Microbiol 2015;53:2410–26.

[22] Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol 2014;52:1501–10.

[23] Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 2012;67:2640–4.

[24] 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.

[25] Hazen TH, Sahl JW, Redman JC, Morris CR, Daugherty SC, Chibucos MC, et al. Draft genome sequences of the diarrheagenic *Escherichia coli* collection. J Bacteriol 2012;194:3026–7.

[26] Tsirogiannis C, Sandel B. Computing the skewness of the phylogenetic mean pairwise distance in linear time. Algorithms Mol Biol 2014;9:15.

[27] Iweriebor BC, Iwu CJ, Obi LC, Nwodo UU, Okoh AI. Multiple antibiotic resistances among Shiga toxin producing *Escherichia coli* O157 in feces of dairy cattle farms in Eastern Cape of South Africa. BMC Microbiol 2015;15:213.

[28] Park HJ, Yoon JW, Heo EJ, Ko EK, Kim KY, Kim YJ, et al. Antibiotic resistance and virulence potentials of Shiga toxin–producing *Escherichia coli* isolates from raw meats of slaughterhouses and retail markets in Korea. J Microbiol Biotechnol 2015;25:1460–6.

[29] Ishii Y, Kimura S, Alba J, Shiroto K, Otsuka M, Hashizume N, et al. Extended-spectrum beta-lactamase-producing Shiga toxin gene (*Stx*1)-positive *Escherichia coli* O26:H11: a new concern. J Clin Microbiol 2005;43:1072–5.

[30] Mandakini R, Dutta TK, Chingtham S, Roychoudhury P, Samanta I, Joardar SN, et al. ESBL-producing Shiga-toxigenic *E. coli* (STEC) associated with piglet diarrhoea in India. Trop Anim Health Prod 2015;47:377–81.

[31] Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br J Pharmacol 2008;153(Suppl. 1):S347–57.

[32] Franz E, van Hoek AH, Wuite M, van der Wal FJ, de Boer AG, Bouw EI, et al. Molecular hazard identification of non-O157 Shiga toxin–producing *Escherichia coli* (STEC). PLoS One 2015;10:e0120353.

[33] Steyert SR, Sahl JW, Fraser CM, Teel LD, Scheutz F, Rasko DA. Comparative genomics and *stx* phage characterization of LEE-negative Shiga toxin–producing *Escherichia coli*. Front Cell Infect Microbiol 2012;2:133.

[34] Delannoy S, Beutin L, Fach P. Discrimination of enterohemorrhagic *Escherichia coli* (EHEC) from non-EHEC strains based on detection of various combinations of type III effector genes. J Clin Microbiol 2013;51:3257–62.

[35] Tozzoli R, Grande L, Michelacci V, Ranieri P, Maugliani A, Caprioli A, et al. Shiga toxin–converting phages and the emergence of new pathogenic *Escherichia coli*: a world in motion. Front Cell Infect Microbiol 2014;4:80.

[36] Gamage SD, Patton AK, Hanson JF, Weiss AA. Diversity and host range of Shiga toxin encoding phage. Infect Immun 2004;72:7131–9.

[37] 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.

[38] Yan X, Fratamico PM, Bono JL, Baranzoni GM, Chen CY. Genome sequencing and comparative genomics provides insights on the evolutionary dynamics and pathogenic potential of different H-serotypes of Shiga toxin–producing *Escherichia coli* O104. BMC Microbiol 2015;15:83.