International Journal of Medical Microbiology 314 (2024) 151610

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.com/locate/ijmm

Genomic surveillance of STEC/EHEC infections in Germany 2020 to 2022 permits insight into virulence gene profiles and novel O-antigen gene clusters

Angelika Fruth, Christina Lang, Tobias Größl, Thomas Garn, Antje Flieger *

Division of Enteropathogenic Bacteria and Legionella (FG11) and National Reference Centre for Salmonella and other Bacterial Enteric Pathogens, Robert Koch Institute, Wernigerode, Germany

ARTICLE INFO

Keywords:
STEC
EHEC
WGS
Molecular surveillance
Virulence
Outbreak detection
Emerging pathogens

ABSTRACT

Shiga toxin-producing *E. coli* (STEC), including the subgroup of enterohemorrhagic *E. coli* (EHEC), are important bacterial pathogens which cause diarrhea and the severe clinical manifestation hemolytic uremic syndrome (HUS). Genomic surveillance of STEC/EHEC is a state-of-the-art tool to identify infection clusters and to extract markers of circulating clinical strains, such as their virulence and resistance profile for risk assessment and implementation of infection prevention measures. The aim of the study was characterization of the clinical STEC population in Germany for establishment of a reference data set. To that end, from 2020 to 2022 1257 STEC isolates, including 39 of known HUS association, were analyzed and lead to a classification of 30.4 % into 129 infection clusters. Major serogroups in all clinical STEC analyzed were O26, O146, O91, O157, O103, and O145; and in HUS-associated strains were O26, O145, O157, O111, and O80. *stx1* was less frequently and *stx2* or a combination of *stx, eaeA* and *ehxA* were more frequently found in HUS-associated strains. Predominant *stx* gene subtypes in all STEC strains were *stx*1a (24 %) and *stx*2a (21 %) and in HUS-associated strains were mainly *stx*2a (69 %) and the combination of *stx*1a and *stx*2a (12.8 %). Furthermore, two novel O-antigen gene clusters (RKI6 and RKI7) and strains of serovars O45:H2 and O80:H2 showing multidrug resistance were detected. In conclusion, the implemented surveillance tools now allow to comprehensively define the population of clinical STEC strains including those associated with the severe disease manifestation HUS reaching a new surveillance level in Germany.

1. Introduction

The zoonotic pathogen Shiga toxin-producing *Escherichia coli* (STEC) includes the subgroup of enterohemorrhagic *E. coli* (EHEC) and represents a multifaceted family the members thereof share Shiga toxin (Stx) but otherwise are characterized by a diverse set of virulence factors and by metabolic heterogeneity. STEC cause a range of symptoms from mild watery or bloody diarrhea, and hemolytic colitis to the severe manifestation of disease hemolytic uremic syndrome (HUS) where fatalities may occur. The prime STEC virulence factor, Stx, is an AB5 toxin which shows N-glycosidase activity in its A subunit and CD77/Gb3 (globotriaosylceramide 3) receptor binding of the B subunit. Stx acts on ribosomal RNA, thereby halts protein synthesis and causes cell death. It is found in two toxin families; Stx1 encoded by the *stx*1 subunit A and B genes, *stx1A* and *stx1B*, respectively, and Stx2 encoded by *stx2A* and *stx2B*. Currently, three subtypes of Stx1 (Stx1a, Stx1c, Stx1d) and 15 subtypes of Stx2 (Stx2a-o) are known (Bai et al., 2021; Gill et al., 2022; Harada et al., 2023; Yang et al., 2022). Stx subtypes Stx2a, Stx2c, and Stx2d are predominantly associated with development of HUS (Matussek et al., 2023; Scheutz, 2014). The level of STEC pathogenicity is further determined by the combination of *stx* with other virulence genes, especially those for attachment to the intestinal mucosa, such as the attaching and effacing genes (*eae*) or aggregation-mediating factors (e.g. *saa*) (Lang et al., 2018).

Although a multitude of STEC serotypes is known, O157:H7 strains worldwide predominate in severe disease and in outbreaks. Nevertheless, so called non-O157-STEC are increasingly coming into focus because of their association with HUS and/or with outbreaks, as for example STEC O104:H4 (Frank et al., 2011), O145:H28 (Taylor et al., 2013), O26:H11 (Jones et al., 2019; Minary et al., 2022), O103:H2

* Correspondence to: Robert Koch Institute, Burgstrasse 37, 38855 Wernigerode, Germany.
E-mail address: fliegera@rki.de (A. Flieger).

https://doi.org/10.1016/j.ijmm.2024.151610
Received 6 November 2023; Received in revised form 17 January 2024; Accepted 22 January 2024
Available online 1 February 2024

(Mylius et al., 2018), or O111:H8 (Centers for Disease C and Prevention, 2012). For risk assessment in clinical and public health context and in cooperation with food control authorities, it is crucial to know whether a pathogen detected in a patient or food shows a gene panel characteristic of highly virulent and HUS-associated strains (Scheutz, 2014).

In Germany, detection of STEC/EHEC infection and HUS as clinical syndrome is notifiable according to the Infection Protection Act in two separate categories, respectively. The diagnosis of a disease is made by primary diagnostics laboratories. This is based on the detection of the Stx protein or the *stx* genes. Culture for further typing is usually not performed. Stx-positive samples are sent to specialized laboratories for further confirmation and subtyping (Fig. 1). The NRC receives ~ 60 % of samples from reported cases (Fruth et al., 2023; Lang et al., 2018). STEC national surveillance data have been available from the NRC since 2001 and are based on phenotypic methods (i.e. serotyping) and molecular methods (i.e. PCR, PFGE). For Germany encompassing a population of 83.2 million inhabitants, the incidence of STEC/EHEC infections in 2019 was 2.3 per 100,000. Since 2015, molecular surveillance by whole genome sequencing (WGS) has been gradually established, initially focusing on STEC belonging to the HUSEC group (Taylor et al., 2013). From 2020, WGS analysis of all STEC isolates has been performed at NRC. The resulting data form the basis for infection cluster analysis and for outbreak detection, identification of emerging pathogens, and the monitoring of trends in STEC phylogenetic lineages. They thus feed into the development of national guidelines and risk assessment, such as recommendations for STEC colonized children with respect to re-admission to community facilities (Pörtner et al., 2019).

After introduction of WGS for all STEC isolates at NRC in 2020, the aim of the here presented study was to characterize the clinical STEC population in Germany to establish a reference data set for STEC/EHEC molecular surveillance. To that end, a comprehensive analysis of the samples received within Germany-wide surveillance activities from 2020 to 2022 was undertaken to study their virulence markers, their antibiotic resistance profile, and perform outbreak detection.

2. Methods

2.1. Bacterial strains

From 2020 to 2022, the National Reference Centre for *Salmonella* and other enteric bacterial pathogens (NRC) analyzed 2903 clinical samples (2020: 961, 2021: 945, 2022: 997) received for STEC analysis from primary diagnostics laboratories and federal state laboratories of public health authorities in Germany. 2225 of these samples were confirmed as *stx* gene positive (2020: 777, 2021: 720, 2022: 728). 1257 unique STEC isolates (2020: 244, 2021: 539, 2022: 474) were recovered from the samples and subjected to further analysis and genome sequencing (Fig. 1, Tab. S1). Samples were grown on nutrient agar (Oxoid GmbH, Germany) or in tryptic soy broth (TSB; BD-BBL, Germany), if not stated otherwise. Testing for enterohemolysin production was performed on enterohemolysin agar (Sifin GmbH, Germany) (Taylor et al., 2013).

2.2. PCR-based virulence gene analysis

All strains were tested for presence of *stx* and *eae*A (encoding adhesin intimin) genes and genotypes and *ehx*A gene (enterohemolysin) using PCR as described earlier (Lang et al., 2019). We determined the O and H antigen genes by PCR (panel of 12 genes characteristic for specific O-antigen gene clusters (OAGC) and 12 different *fli*C type genes characteristic for H-antigen; 19) and after subjection of the strains to genome sequencing, we predicted the type of the OAGC and H antigen gene clusters (HAGC) (see below).

2.3. Antibiotics susceptibility testing

The strains were tested for susceptibility to 15 standard antibiotics

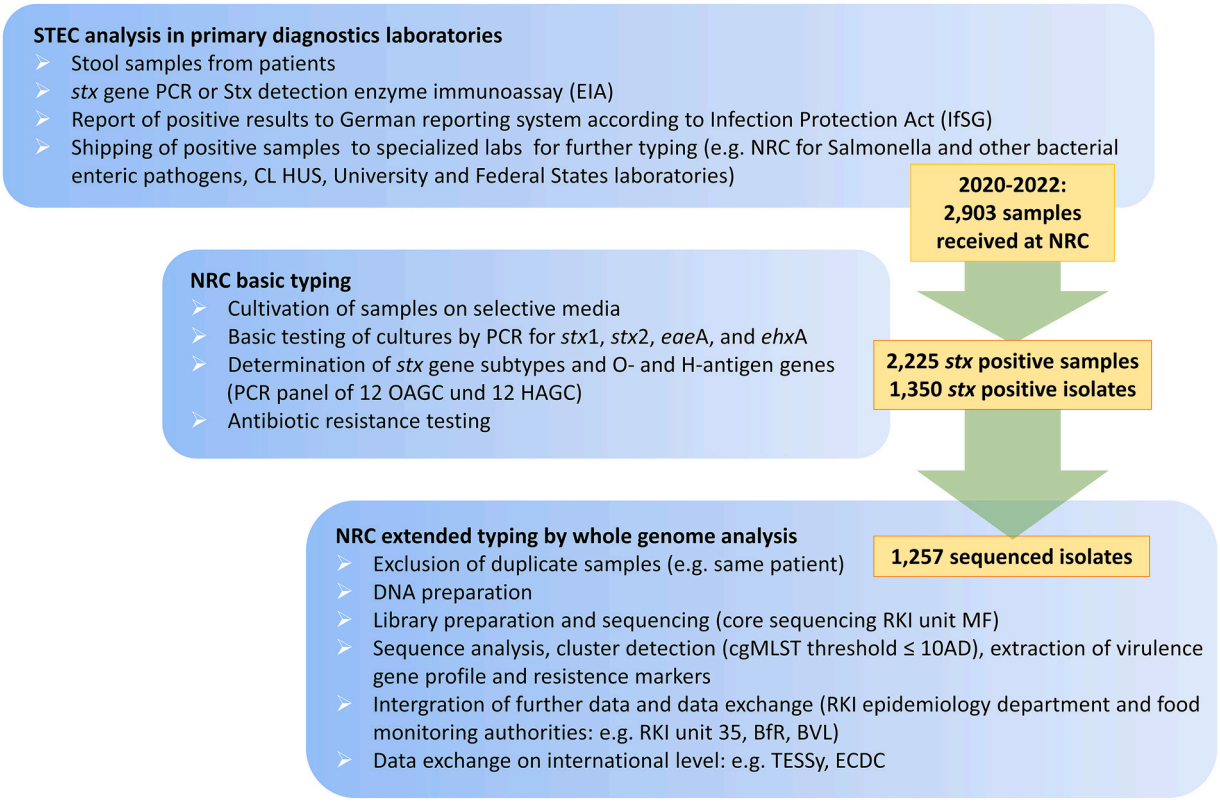

Fig. 1. Workflow and sample number of STEC analysis involving the steps in primary diagnostics laboratories, basic typing and extended typing by means of whole genome analysis at NRC, 2020–2022.

Table 1
Selection of HUS-associated STEC/EHEC strains, 2020–2022: serotype (OAGC, HAGC), MLST ST, *stx*- and *eae*A-subtype and further virulence gene markers, phenotypic antibiotic resistance profile* (further HUS-associated strains see Tab. S1).

RKI No.	OAGC	HAGC	MLST ST type Warwick	*stx* gene subtype	*eae*A gene subtype	*ehx*A	EAST1	*esp*P	*fyu*A	*iha*	*irp*2	*iuc*A	*kat*P	*lpf*O26	*lpf*O104	*lpf*O113	*saa*	*sfp*A	*sub*AB	*ter*A	*aai*C	Resistance profile	Reference
20-03804	O26	H11	21	1a	beta 1	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	n.d.	this study
21-01439	O26	H11	21	1a, 2a	beta 1	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	AMP, CHL	this study
22-07323	O26	H11	29	2a	beta 1	+	+	-	+	+	+	-	-	+	+	+	-	-	-	+	-	susceptible	this study
20-03221	O178	H19	205	2a	-	+	-	+	-	+	-	-	-	-	+	+	+	-	+	-	-	susceptible	this study
22-04972	O128	H2	297	2b	-	+	-	-	-	+	-	+	-	+	+	+	-	-	+	-	-	susceptible	this study
21-06651	O71	H2	17	2a	epsilon 1	+	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	TCY	this study
21-02230	O45	H2	301	2a	xi	+	-	+	-	+	-	-	-	-	-	-	-	-	-	+	-	AMP, CHL, KAN, NAL, SXT, TCY, TMP	this study
20-04319	O80	H2	301	2a	xi	+	-	+	-	-	+	+	-	-	-	-	-	-	-	+	-	CIP, CHL, GEN, KAN, SXT, NAL	this study
20-00797	O80	H2	301	2d	xi	+	-	+	-	+	-	+	-	-	-	-	-	-	-	-	-	AMP, GEN, KAN, NAL, SXT, STR, TCY	this study
20-04287	O177	H25	659	2a	beta 1	+	-	+	-	+	-	+	+	-	+	+	-	-	-	+	-	susceptible	this study
20-05217	O145	H28	32	2a	gamma 1	+	+	+	-	+	-	-	-	-	-	-	-	-	-	+	-	susceptible	this study
21-03553	O128ac	H34	5278	2a	epsilon 6	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	susceptible	this study
21-00958	O157	H7	11	2a	gamma 1	+	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	susceptible	this study
21-04780	O157	H7	11	2c	gamma 1	+	+	+	-	+	-	-	+	-	-	-	-	-	-	+	-	susceptible	this study
22-04577	O111	H8	16	1a, 2a	theta 2	+	+	-	-	+	-	+	-	+	+	+	-	-	-	+	-	AMP, KAN, TCY	this study
20-05756	O111	H8	16	1a	theta2	+	+	-	-	+	-	-	-	+	+	+	-	-	-	+	-	susceptible	this study

* ampicillin (AMP), chloramphenicol (CHL), ciprofloxacin (CIP), gentamicin (GEN), kanamicin (KAN), nalidixic acid (NAL), streptomycin (STR), trimethoprim/sulfamerazin (SXT), tetracylines (TCY), trimethoprim (TMP), n.d.= not determined

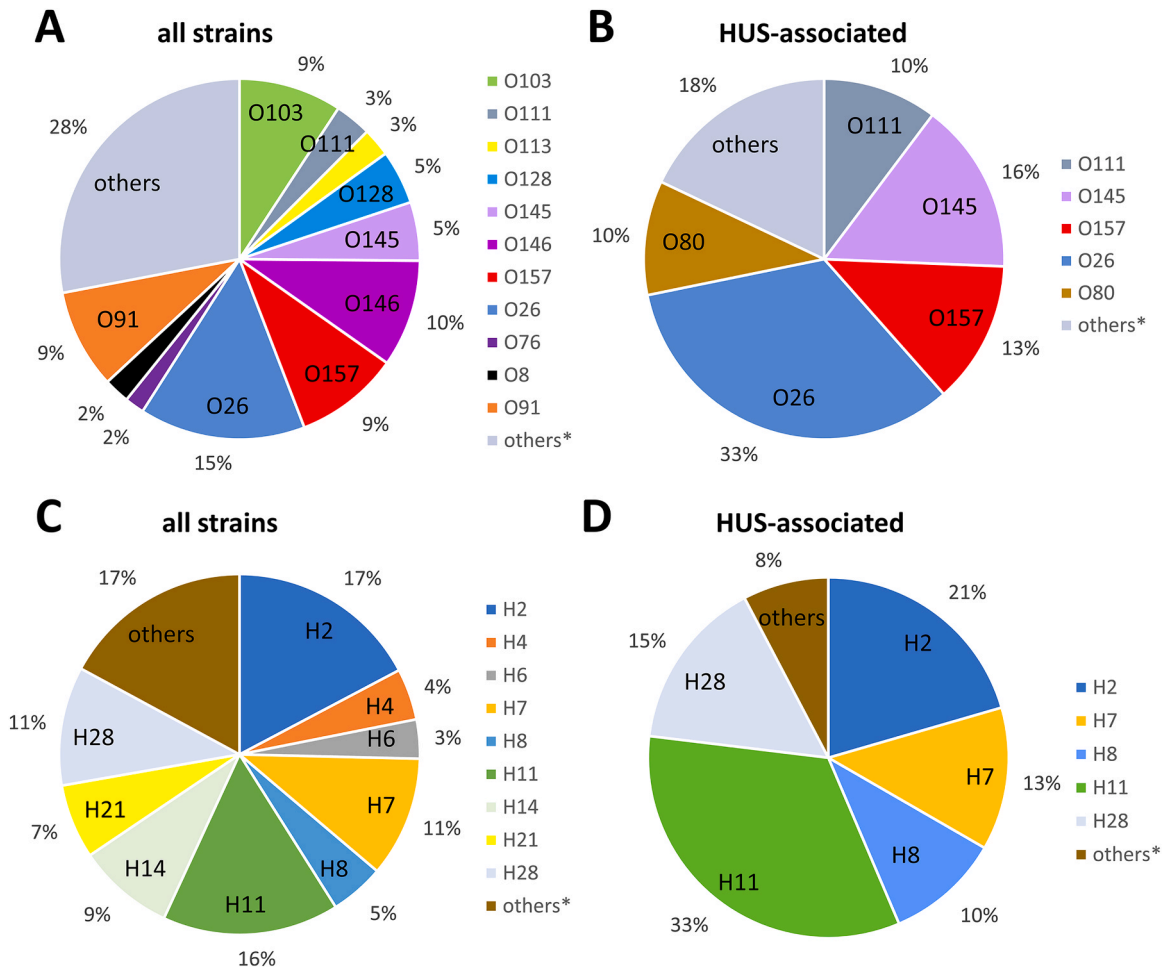

Fig. 2. O-antigen gene cluster and H-antigen gene cluster analysis for all STEC versus HUS-associated strains, 2020–2022. O-antigen gene cluster groups (A, B) and H-antigen gene clusters (C, D). Gene clusters were extracted from WGS data for all STEC strains, n = 1257 (A, C) or from HUS-associated strains, n = 39 (B, D). Please also refer to Table 1 and S1.

according to EUCAST recommendations for *E. coli* by a broth microdilution assay (http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Disk_test_documents/2019_manuals/Reading_guide_BMD_v_1.0_2019.pdf). Specific antibiotics were ampicillin (AMP), azithromycin (AZM), cefotaxime (CTX), ceftazidime (CAZ), cefoxitin (COX), chloramphenicol (CHL), ciprofloxacin (CIP), gentamicin (GEN), kanamicin (KAN), meropenem (MEP), nalidixic acid (NAL), streptomycin (STR), trimethoprim/sulfamerazin (SXT), tetracylines (TCY), and trimethoprim (TMP).

2.4. WGS

Whole genome sequencing (WGS) was accomplished using short-read paired-end sequencing with the MiSeq (2 × 300 bp), HiSeq 1500 (2 × 250 bp) and NextSeq 2000 instruments (Illumina, San Diego, CA). DNA from the *E. coli* strains was isolated by glass bead extraction (Koser et al., 2014) and 1 ng of DNA was used to generate libraries by using the Nextera XT DNA - or NEBNext Ultra II DNA library prep Kit according to the manufacturer's instructions (Illumina, San Diego, CA; NEB, Frankfurt am Main).

2.5. Bioinformatics analyses

Ridom SeqSphere+(at least version 8.5.1) was used for analysis of the sequences (Junemann et al., 2013). This includes: 1) running FastQC for read data quality control (Andrews, 2010), 2) assembling with SKESA (Souvorov et al., 2018), 3) scanning for 2,513 targets belonging to the cgMLST of Enterobase (Zhou et al., 2020), 4) determining the ST for *E. coli* MLST Warwick (Wirth et al., 2006), 5) running CGE SerotypeFinder for O- and H-antigen determination (Joensen et al., 2015), 6) using *E. coli* VFDB for classification for bacterial virulence factors (Liu et al., 2022) and finally 7) *E. coli* NCBI AMRFinderPlus a tool that identifies AMR genes and resistance associated point mutations (Feldgarden et al., 2021). Analysis for surveillance and cluster detection with the common cgMLST scheme of Enterobase for *E. coli* was performed by using Ridom SeqSphere+ software by pairwise analysis, ignoring missing values. Sequence data were uploaded in database miGenomesurv.

In parallel serotype and virulence genes of interest were called by mapping against the respective reference sequence using standard Geneious mapper (settings: medium sensitivity; none finetuning, trim sequences before mapping; at least Geneious prime version 2021.2.2; Biomatters Ltd.). Requirements for positive matches were 100 % coverage of the reference sequence, 90 % identity with the reference sequence, and high quality for 90 % bases in the sequence. Reference sequences for serotype determination and for virulence marker genes were downloaded from the Center for Genomic Epidemiology (CGE; DTU, Denmark; SerotypeFinder, VirulenceFinder; https://cge.cbs.dtu.dk/services/data.php). Further reference sequences for serotyping were obtained from NCBI (Lang et al., 2019). For analyzing of potentially novel O- antigen loci the reads were *de novo* assembled as described elsewhere (Lang et al., 2019).

Table 2
Dominant STEC/EHEC strain types analyzed at NRC, 2020–2022: associated symptoms, serotype (OAGC, HAGC), MLST ST, *stx*- and *eae*A-subtype, further virulence gene markers, phenotypic antibiotic resistance profile* (summary of all strains analyzed see Tab. S1).

rank	associated symptoms	OAGC	HAGC	MLST ST type Warwick	*stx* gene subtype	*eae*A gene subtype	*ehx*A	EAST1	*esp*P	*fyu*A	*iha*	*irp*2	*iuc*A	*kat*P	*lpf*O26	*lpf*O104	*lpf*O113	*saa*	*sfp*A	*sub*AB	*ter*A	*aai*C	Resistance profile
1	(bloody) diarrhea	O26	H11	21	2a	beta 1	+	+	+	+	+	+	-	+	+	+	+	-	-	-	+	-	susceptible
	(bloody) diarrhea			29	2a	beta 1	+	+	-	+	+	+	+	-	+	+	+	-	-	-	+	-	susceptible
2	diarrhea	O146	H21	442	1c, 2b	-	+	-	-	-	+	-	+	-	+	+	+	-	-	+	+	-	susceptible
	diarrhea		H28	738	2b	-	-	-	-	-	+	-	-	-	-	+	+	-	-	+	+	-	susceptible
3	diarrhea	O91	H14	33	1a, 2b	-	+	-	-	-	+	-	+	-	+	+	+	-	-	+	-	-	susceptible
	diarrhea				1a	-	+	-	-	-	+	-	+	-	+	+	+	-	-	-	-	-	susceptible
4	(bloody) diarrhea	O157	H7	11	2a	gamma 1	+	+	+	-	+	-	-	+	-	-	-	-	-	-	+	-	susceptible
	(bloody) diarrhea				1a, 2c	gamma 1	+	+	+	-	+	-	-	+	-	-	-	-	-	-	+	-	AMP, CHL, GEN, SXT, TMP
5	diarrhea	O145	H28	32	2a	gamma 1	+	+	+	-	+	-	-	-	-	-	-	-	-	-	+	-	susceptible
6	diarrhea	O103	H2	17	1a	epsilon 1	+	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	susceptible

Table 3
STEC/EHEC strains showing novel O-antigen (OAGC) gene cluster analyzed at NRC, 2020–2022: associated symptoms, serotype, MLST ST type, *stx*- and *eae*A-subtype, further virulence gene markers, phenotypic antibiotic resistance profile* .

RKI No.	associated symptoms	OAGC	HAGC	MLST ST type Warwick	*stx* gene subtype	*eae*A gene subtype	*ehx*ABCD	EAST1	*esp*P	*fyu*A	*iha*	*irp*2	*iuc*A	*kat*P	*lpf*O26	*lpf*O104	*lpf*O113	*saa*	*sfp*A	*sub*AB	*ter*A	*aai*C	Resistance profile	Reference
22-03186	n.a.[#]	RKI6	H4	10	1a, 2c	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	AMP, CHL, NAL, SXT, TCY, TMP	this study
20-05749	n.a.[#]	RKI7	H6	28	2f	beta 1	-	+	-	-	-	+	-	-	-	-	-	-	+	-	-	-	susceptible	this study

n.a. = data not available

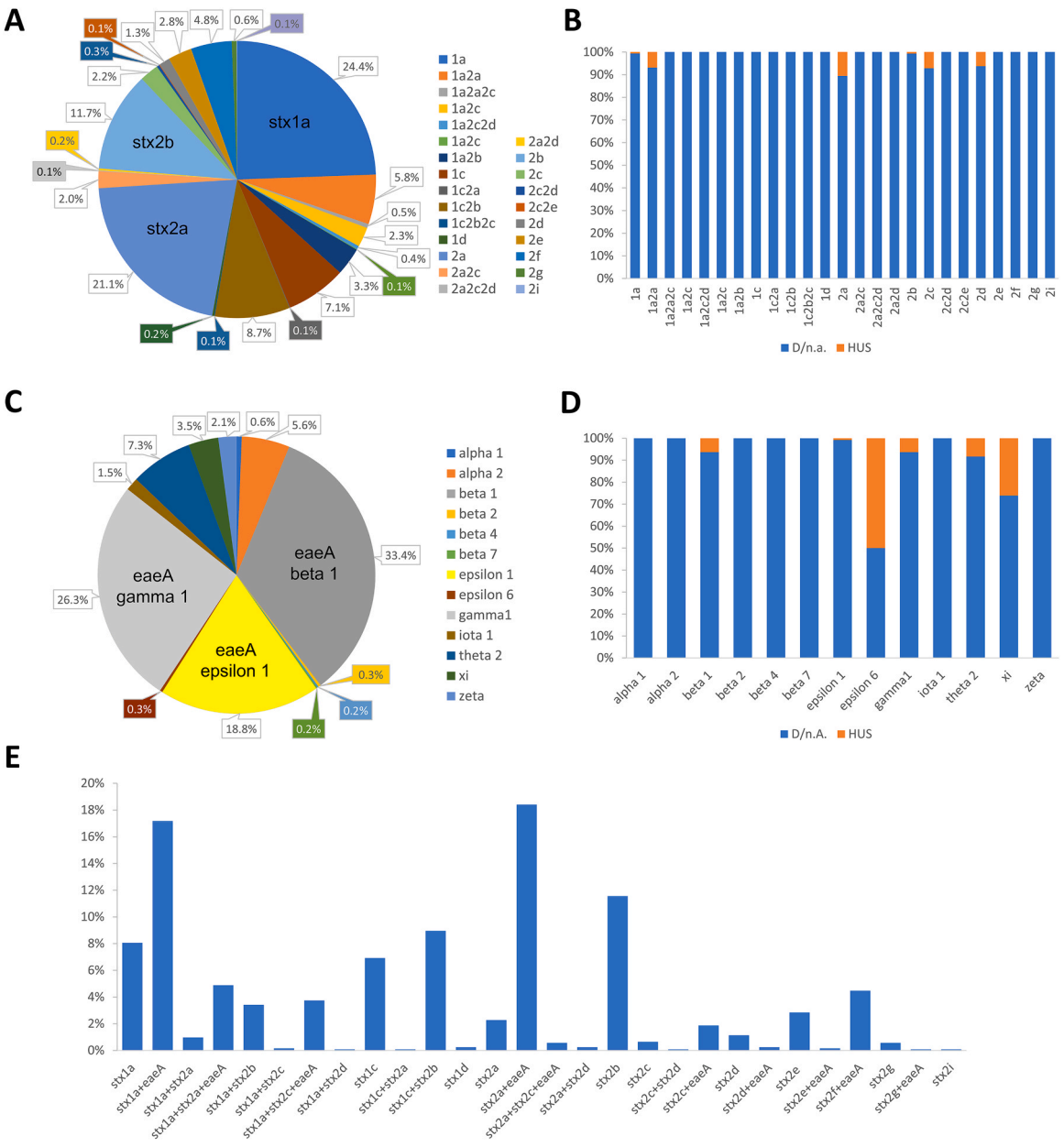

Fig. 3. *stx* and *eaeA* subtype analysis for all STEC and HUS-associated strains, 2020–2022. Stx (A, B) and eaeA subtypes (C, D) extracted from genomes for total samples analyzed at NRC 2020–22 (A, C, E) and percentage of HUS-associated strains from specific subtypes (B, D). Combinations of stx and eaeA subtypes are shown in Fig. 3E. Please also refer to Table 1 and S1.

3. Results

3.1. *stx1* is less frequently and *stx2* or a combination of *stx*, *eaeA* and *ehxA* are more frequently found in HUS-associated strains

From 2020 to 2022, 2903 clinical samples were received at NRC from Germany-wide STEC surveillance activities. This represents about 60 % of reported STEC cases (2020:1377, 2021:1611, 2022:1828). 2225 of those were tested positive for *stx* presence and 1350 STEC isolates were obtained from the *stx*-positive samples. Out of these, 1257 isolates were sequenced after exclusion of multiple submission samples (2020: 244, 2021: 539, 2022: 474). This revealed an STEC isolation rate of 46.5 % of the original samples (Fig. 1, Tab. S1). Median age of the patients was 34.1 years (range 0.0–94.4). Children $\leq$5 years of age were represented by 58.5 %, and the distribution females/males was 54.1 versus 45.9 %. Information of disease progression into HUS was available for 39 (3.1 %) of the isolate-associated cases (Table 1). The strains were characterized for primary virulence genes, such as *stx*, *eae*A, and *ehx*A. 32 % of all isolates showed *stx*1, 47 % *stx*2, and 21% *stx*1 and *stx*2. 53 % of all strains in addition to *stx* contained *eae*A, 73 % *ehx*A, and 45 % showed all three virulence markers. 5 % of HUS-associated strains showed *stx*1, 82 % *stx*2 %, and 13 % *stx*1 combined with *stx*2. 95 % of HUS-associated strains contained *eae*A, 95 % *ehx*A, and 89 % all three markers. This confirmed, as described before, that among all STEC strains analyzed, the presence of *stx*1 was less prevalent and *stx*2 more prevalent in HUS-associated strains. Furthermore, all three virulence markers were found in a higher frequency in HUS than in all isolates (Matussek et al., 2023; Pörtner et al., 2019).

3.2. O26:H11 was most frequently detected among all STEC strains and HUS-associated strains. Two novel STEC O-antigen gene clusters were defined

In *E. coli,* 192 O-antigen (OAG) types and 53 H-antigen (HAG) types are known (Andrews, 2010). We found a high diversity of OAG types and most frequent types were O26 (15 %), O146 (10 %), O91, O103, O157 (each 9 %), and O145 (5 %) (Fig. 2A, Table 2). Within the HUS-associated isolates, O26 were most common (33 %) followed by O145 (16 %) and O157 (13 %) (Fig. 2B, Table 1). Most common HAG types for all strains were H2 (17 %) and for the HUS strains H11 (33 %) (Fig. 2C, 2D, Table 2). A vast number of serotypes then results from the combination of the OAG and HAG types. Dominant serotypes and those commonly associated with human disease were O26:H11 (15 %), O146:H21/H28 (10 %), O91:H14 (9 %) and O157:H7 (9 %); for HUS O26:H11 (33 %), O145:H28 (16 %) and O157:H7 (13 %). Very rare serotypes were alike detected, such as O2:H6, O156:H25, and O182:H25. For 30 (0.2 %) of the strains, the OAG could only be correctly classified by genome analysis (otherwise ONT), including new types recently described (Lang et al., 2019; Iguchi et al., 2015) and two strains which did not match with so far known OAG loci. Therefore, these were classified as novel OAG gene clusters and were designated as RKI6 and RKI7 (Table 3).

3.3. Most of the strains harbored stx2 and HUS-associated strains mostly contained stx2a, stx2c, stx2d. A timely shift from stx1 to stx2 gene presence was recently observed for O26:H11

Different virulence associated genes (VGAs) and combinations were extracted from the genome sequences (Tab. S1). As mentioned above, about 30 % of the isolates contained *stx*1-genes (71 % *stx*1a, 27 % *stx*1c, 2 % *stx*1d) and 70 % were positive for *stx*2-genes (44 % *stx*2a, 36 % *stx*2b, 9 % *stx*2c, 1 % *stx*2d, 6 % *stx*2e, 3 % *stx*2f, 1 % *stx*2g) (Fig. 3A). HUS-associated strains mostly contained *stx*2a, *stx*2c, *stx*2d or the combination of *stx*1a and *stx*2a which are known to be predominantly associated with HUS (Matussek et al., 2023; Scheutz, 2014), Fig. 3B). A small fraction of HUS-associated strains revealed *stx*1a and *stx*2b (Fig. 3B).

Over the last years, a timely shift in *stx* gene subtypes has been observed in some STEC serotypes. For example, in the period considered here, 32.7 % of the isolates of serotype O26:H11 were positive for *stx*1a (in comparison to NRC data from 2001: 83.2 %), 11.7 % were positive for *stx*1a and *stx*2a (2001: 4 %), but 55.7 % were positive for *stx*2a (2001: 12.8 %) (Fig. S1). Further, 59.8 % of *stx*1-only positive isolates of serotype O91:H14 were *stx*1a positive (2001: 88.5 %), whereas 32.1 % were *stx*1a and *stx*2b positive (2001: 10.3 %) and 8 % were exclusively *stx*2b positive (2001: 1.1 %).

In addition, several novel combinations of *stx* subtype and rare serotypes were detected. For example, strains of serotype O80:H2 which usually harbor *stx*2a were also detected with both *stx*2a and/or *stx*2d. These occurred in 0.9 % and 0.2 %, respectively, primarily in Southern Germany. Further, variant *stx*1d was detected in 3 (0.2 %) isolates belonging to serotypes O9:H10 and OgN14:H23. *stx*2g was detected in 7 (0.5 %) isolates of serotype O187:H28 and, for the first time, *stx*2i was detected in an isolate of serotype O8:H9 in Germany.

As a marker for the pathogenicity island LEE, *eae*A subtype was identified based on the genome sequences. (Tab. S1). Dominant *eae*A gene subtypes beta 1 (33 %), gamma 1 (26 %) and epsilon 1 (19 %) correlate with the corresponding serotypes with O26:H11, O157:H7, O145:H28, and O103:H2 (see Fig. 3C and D). Combination of *stx* and *eae*A was most frequently detected in strains harboring *stx*2a (18.4 %) and *stx*1a (17.2 %). A similar pattern was observed for HUS-associated strains (Fig. 3E, Table 1 and S1). We also identified from the WGS data a panel of different virulence markers which are listed for the specific strains in table S1. Of special note, in 3 (0.2 %) isolates of serotype OX18:H21 *aai*C was found as another VGA in addition to *stx*1c (Tab. S1).

3.4. Strains of serovars O45:H2 and O80:H2 showed multidrug resistance

76 % of the analyzed STEC isolates were completely susceptible towards the test set of 16 antibiotics. 18 % were resistant towards 1 to 4 antibiotics, and only 6 % were multidrug resistant (> 4 antibiotics). Among the isolates resistant towards multiple antibiotics, strains of serotypes O45:H2 and O80:H2 (ST301) were predominant (Table 1). All of these ST301 strains showed multi drug resistance (Tab. S1). ESBL resistance was detected only very occasionally, e.g. in an isolate of serotype O26:H11 with a CTX-M-1 beta-lactamase (Tab.S1).

3.5. Phylogenetic analysis identified 129 infection clusters

In the mentioned period, 383 isolates (30.4 %) of the 1257 isolates were assigned to 129 clusters (threshold allelic distance > 10 AD) including two to ten isolates. The majority of the strains in these clusters were from individuals of the same household and three outbreaks were further analyzed. In 2020, a total of seven STEC OgN3:H25 cases (*stx*1-, *stx*2a+, *eae*A-, *sub*AB+, ST 11013) were detected in Northwest of Germany. The cases were on average 33.1 years old (range 2–77 years) and had mild diarrhea. No source of infection was determined. Another outbreak in 2020 occurred in Northwest-Mecklenburg. Here, several daycare centers were affected, which were supplied by the same caterer. 31 cases of mild and bloody diarrhea were registered, as well as 7 asymptomatic cases. The average age was 13 years (range 1–86 years). The outbreak strain was an STEC O26:H11 (*stx*1a+, *stx*2-, *eae*A+, *ehx*A+, ST29). No food was identified as source of infection. The third outbreak involved several daycare centers in Bavaria, which was most likely caused by food from a caterer. STEC O111:H8 (*stx*1a+, *stx*2a+, *eae*A+, *ehx*A+, ST16) was found associated with the cases. More than 30 cases of illness were recorded. The majority showed diarrhea and vomiting and at least eight cases developed HUS. At NRC, ten samples were processed for this purpose and genome-based phylogenetic analysis identified an infection cluster showing allele distances of 0 alleles.

4. Discussion

Molecular subtyping of STEC for strain characterization, the detection of pathogen diversity and assignment to possible outbreaks has been carried out at NRC since 2001. The methods used previously were classical serotyping, PCR for *stx*, *eae*A and *ehx*A gene detection and PFGE or MLVA and MLST for phylogenetic analysis. Pre-screening in the laboratories of primary diagnostics was based on the detection of Shiga toxin by means of enzyme immunoassay (EIA), without pre-selection of certain pathogen types, as for example O157 strains. This procedure was world-wide unique compared to laboratories abroad which focused at the time on STEC O157 detection. The *stx* centered analysis led for example to the rapid detection of the STEC O104:H4 2011 outbreak strain in Germany (Bai et al., 2021). In the meantime, the panel of these detection methods has been expanded to include real-time PCR in routine diagnostics laboratories and subtyping, including WGS, in the work of the NRC. A wide range of open source tools is available for this purpose, which enable fast and reliable work.

The results presented in this manuscript are now reaching a new surveillance level in Germany because a large portion of isolates from reported cases underwent comprehensive strain subtyping including genome analysis. This allows us to define the population of disease-associated strains and to extract the characteristics of strains associated with the severe disease manifestation HUS. For example, we confirmed that *stx*1 is less frequently and *stx*2 more frequently found in HUS-associated strains. Here, HUS-associated strains mostly contained *stx*2a, *stx*2c, *stx*2d as shown before (Matussek et al., 2023; Scheutz, 2014). We also highlighted that a wide variety of serotypes was detected among STEC strains and that O26:H11 was most frequently found

among all strains and the HUS-associated strains. We also defined two novel STEC OAG gene clusters and our data revealed that strains of serovars O45:H2 and O80:H2 showed multidrug resistance. These observations are in line with the data collected from other European countries or internationally. Specifically, the increased occurrence of multi-resistant strains of serotype O80:H2 in association with HUS has been reported mainly from France and Switzerland (Cointe et al., 2020, 2018; Nuesch-Inderbinen et al., 2023). Further, a shift from *stx*1 to *stx*2 gene presence was observed for O26:H11. A change in the *stx* gene population in dominant clones of serotype O26:H11 from the Czech Republic, Italy and France has already been reported (Jones et al., 2019; Karnisova et al., 2018). This shows that there are common lines of STEC evolution also seen in an international context.

Using WGS for outbreak detection allowed more sensitive analysis across different health sectors, which in 2015 led for the first time both to the detection of an outbreak with STEC of serotype O103:H2 in patients from Germany and identification of the food source in Austria (Matussek et al., 2023). WGS has been standard practice since then and so far, food or other sources of infection were often not identified (Minary et al., 2022; Jenkins et al., 2019; Rodwell et al., 2022, 2023, 2021). But recent expansion of WGS will increase chances for source identification. In exploratory surveys, raw milk and uncooked meat products or contact with STEC colonized animals and humans were identified as common causes of infection. But recently, plant-based foods have been increasingly suspected as infection sources. In addition to ready-to-eat salads, flour and flour products are getting more often into focus, as in France and Belgium in an outbreak due to frozen pizza (ECDC-EPIS UI: 2022-FWD-00017).

5. Conclusion

In recent years, genome analysis was successfully implemented to accomplish virulence gene typing and serovar prediction for detection of novel und highly virulent STEC variants. Further sensitive phylogenetic analysis allows improved strain discrimination and cluster detection in an unprecedented manner. Therefore, the integration of epidemiological data and data of the competent food authorities on national and international level leads to highly efficient control strategies positively impacting public health.

Funding

This work was supported by the Federal Ministry of Health of Germany (grant D81959).

CRediT authorship contribution statement

Lang Christina: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Fruth Angelika:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Flieger Antje:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Garn Thomas:** Formal analysis, Investigation, Methodology. **Größl Tobias:** Formal analysis, Investigation, Methodology.

Acknowledgements

We highly appreciate our outgoing head of department Martin Mielke for his constant interest and intensive support in the implementation of the genomic STEC surveillance project. Furthermore, we want to thank the colleagues of unit MFI Method Development, Research Infrastructure and Information Technology at Robert Koch Institute, Berlin, especially Andrea Thürmer, Stephan Fuchs, Torsten Semmler and the technical staff for genome sequencing.

The authors thank RKI unit FG35 Gastrointestinal Infections, Zoonoses and Tropical Infections of Department for Infectious Epidemiology, Klaus Stark, Christina Frank and Gerd Falkenhorst for support and fruitful discussions regarding cluster definition and outbreak analysis. For technical assistance in cultivation and typing of strains, we would like to thank members of technical staff of our unit Jenny Gabrisch, Sabrina Diederich, Steffen Schneider. In particular, we would like to thank CL HUS, University Münster, Alexander Mellmann, Barbara Middendorf-Bauchart for supportive cooperation, all laboratories and cooperation partners (German STEC surveillance network) for providing samples and isolates.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ijmm.2024.151610.

References

Andrews S.: FastQC: A Quality Control Tool for High Throughput Sequence Data. ⟨http://www.bioinformatics.babraham.ac.uk/projects/fastqc/⟩, 2010.
Bai, X., Zhang, J., Hua, Y., et al., 2021. Genomic insights into clinical shiga toxin-producing escherichia coli strains: a 15-year period survey in Jonkoping, Sweden. Front Microbiol 12, 627861.
Centers for Disease C and Prevention, 2012. Outbreak of Shiga toxin-producing Escherichia coli O111 infections associated with a correctional facility dairy - Colorado, 2010. MMWR Morb. Mortal. Wkly Rep. 61, 149–152.
Cointe, A., Birgy, A., Bridier-Nahmias, A., et al., 2020. Escherichia coli O80 hybrid pathotype strains producing Shiga toxin and ESBL: molecular characterization and potential therapeutic options. J. Antimicrob. Chemother. 75, 537–542.
Cointe, A., Birgy, A., Mariani-Kurkdjian, P., et al., 2018. Emerging multidrug-resistant hybrid pathotype shiga toxin-producing Escherichia coli O80 and related strains of clonal complex 165, Europe. Emerg. Infect. Dis. 24, 2262–2269.
Feldgarden, M., Brover, V., Gonzalez-Escalona, N., et al., 2021. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci. Rep. 11, 12728.
Frank, C., Werber, D., Cramer, J.P., et al., 2011. Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. N. Engl. J. Med 365, 1771–1780.
Fruth, A., Simon, S., Halbedel, S., Banerji, S., Flieger, A., 2023. COVID-19-Pandemie führte zu starkem Rückgang von darmpathogenen Erregern – Ergebnisse der integrierten molekularen Surveillance. Epid Bull 5, 3–9.
Gill, A., Dussault, F., McMahon, T., et al., 2022. Characterization of atypical shiga toxin gene sequences and description of Stx2j, a new subtype. J. Clin. Microbiol 60, e0222921.
Harada, T., Wakabayashi, Y., Seto, K., Lee, K., Iyoda, S., Kawatsu, K., 2023. Real-time PCR assays to detect 10 Shiga toxin subtype (Stx1a, Stx1c, Stx1d, Stx2a, Stx2b, Stx2c, Stx2d, Stx2e, Stx2f, and Stx2g) genes. Diagn. Microbiol Infect. Dis. 105, 115874.
Iguchi, A., Iyoda, S., Kikuchi, T., et al., 2015. A complete view of the genetic diversity of the Escherichia coli O-antigen biosynthesis gene cluster. DNA Res 22, 101–107.
Jenkins, C., Dallman, T.J., Grant, K.A., 2019. Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing Escherichia coli serogroup O157:H7, England, 2013 to 2017. Eur. Surveill. 24.
Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M., Scheutz, F., 2015. Rapid and easy in silico serotyping of Escherichia coli isolates by use of whole-genome sequencing data. J. Clin. Microbiol 53, 2410–2426.
Jones, G., Lefevre, S., Donguy, M.P., et al., 2019. Outbreak of Shiga toxin-producing Escherichia coli (STEC) O26 paediatric haemolytic uraemic syndrome (HUS) cases associated with the consumption of soft raw cow's milk cheeses, France, March to May 2019. Eur. Surveill. 24.
Junemann, S., Sedlazeck, F.J., Prior, K., et al., 2013. Updating benchtop sequencing performance comparison. Nat. Biotechnol. 31, 294–296.
Karnisova, L., Marejkova, M., Hrbackova, H., et al., 2018. Attack of the clones: whole genome-based characterization of two closely related enterohemorrhagic Escherichia coli O26 epidemic lineages. BMC Genom. 19, 647.
Koser, C.U., Bryant, J.M., Comas, I., et al., 2014. Comment on: characterization of the embB gene in Mycobacterium tuberculosis isolates from Barcelona and rapid detection of main mutations related to ethambutol resistance using a low-density DNA array. J. Antimicrob. Chemother. 69, 2298–2299.
Lang, C., Fruth, A., Holland, G., et al., 2018. Novel type of pilus associated with a Shiga-toxigenic E. coli hybrid pathovar conveys aggregative adherence and bacterial virulence. Emerg. Microbes Infect. 7, 203.
Lang, C., Hiller, M., Konrad, R., Fruth, A., Flieger, A., 2019. Whole-Genome-Based Public Health Surveillance of Less Common Shiga Toxin-Producing Escherichia coli

Serovars and Untypeable Strains Identifies Four Novel O Genotypes. J. Clin. Microbiol 57.
Liu, B., Zheng, D., Zhou, S., Chen, L., Yang, J., 2022. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res 50, D912–D917.
Matussek, A., Mernelius, S., Chromek, M., et al., 2023. Genome-wide association study of hemolytic uremic syndrome causing Shiga toxin-producing Escherichia coli from Sweden, 1994-2018. Eur. J. Clin. Microbiol Infect. Dis. 42, 771–779.
Minary, K., Tanne, C., Kwon, T., et al., 2022. Outbreak of hemolytic uremic syndrome with unusually severe clinical presentation caused by Shiga toxin-producing Escherichia coli O26:H11 in France. Arch. Pedia 29, 448–452.
Mylius, M., Dreesman, J., Pulz, M., et al., 2018. Shiga toxin-producing Escherichia coli O103:H2 outbreak in Germany after school trip to Austria due to raw cow milk, 2017 - the important role of international collaboration for outbreak investigations. Int. J. Med Microbiol 308, 539–544.
Nuesch-Inderbinen, M., Treier, A., Stevens, M.J.A., Stephan, R., 2023. Whole genome sequence-based characterisation of Shiga toxin-producing Escherichia coli isolated from game meat originating from several European countries. Sci. Rep. 13, 3247.
Pörtner, K., Fruth, A., Flieger, A., Middendorf-Bauchart, B., Mellmann, A., Falkenhorst, G., 2019. Überarbeitung der RKI Empfehlungen für die Wiederzulassung zu Gemeinschaftseinrichtungen gemäß § 34 IfSG nach EHEC Infektion. Epid Bull. 47, 506–509.
Rodwell, E.V., Chan, Y.W., Sawyer, C., et al., 2022. Shiga toxin-producing Escherichia coli clonal complex 32, including serotype O145:H28, in the UK and Ireland. J. Med. Microbiol. 71.
Rodwell, E.V., Simpson, A., Chan, Y.W., Godbole, G., McCarthy, N.D., Jenkins, C., 2023. The epidemiology of Shiga toxin-producing Escherichia coli O26:H11 (clonal complex 29) in England, 2014-2021. J. Infect. 86, 552–562.
Rodwell, E.V., Vishram, B., Smith, R., et al., 2021. Epidemiology and genomic analysis of Shiga toxin-producing Escherichia coli clonal complex 165 in the UK. J. Med. Microbiol. 70.
Scheutz, F., 2014. Taxonomy meets public health: the case of shiga toxin-producing Escherichia coli. Microbiol Spectr. 2.
Souvorov, A., Agarwala, R., Lipman, D.J., 2018. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol. 19, 153.
Taylor, E.V., Nguyen, T.A., Machesky, K.D., et al., 2013. Multistate outbreak of Escherichia coli O145 infections associated with romaine lettuce consumption, 2010. J. Food Prot. 76, 939–944.
Wirth, T., Falush, D., Lan, R., et al., 2006. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol. Microbiol 60, 1136–1151.
Yang, X., Liu, Q., Sun, H., Xiong, Y., Matussek, A., Bai, X., 2022. Genomic characterization of Escherichia coli O8 strains producing shiga toxin 2l subtype. Microorganisms 10.
Zhou, Z., Alikhan, N.F., Mohamed, K., Fan, Y., Agama Study G and Achtman M, 2020. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res. 30, 138–152.