Design of the EPIGENEC study: assessing the EPIdemiology and GENetics of *Escherichia coli* in the Netherlands

Denise van Hout1*, Tess D. Verschuuren1, Patricia C.J. Bruining-Verhagen1,2, Thijs Bosch2, E. Ascelijn Reuland2,3,4, Ad C. Fluit4, Anita C. Schürch4, Rob J.L. Willems4, Sabine C. de Greeff2, Annemarie van ’t Veen3,4, Jan A.J.W. Kluymans1,5, Marc J.M. Bonten1,2,4

1. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
2. Center for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
3. Saltro Diagnostic Center for Primary Care, Utrecht, The Netherlands
4. Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands
5. Department of Infection Control, Amphia Hospital, Breda, The Netherlands

* Corresponding author: Denise van Hout, D.vanHout-3@umcutrecht.nl

ABSTRACT: Background: Infections caused by *E. coli* cause considerable disease burden and range from frequently occurring and relatively innocent urinary tract infection (UTI) to severe bloodstream infection (BSI). The incidence of infections caused by ESBL-producing *E. coli* (ESBL-PEc) is increasing, justifying surveillance and development of preventive strategies in several domains. Faecal carriage is universal and believed to be the most important reservoir for *E. coli* from which infections can originate. It is currently unknown to what extent Dutch *E. coli* carriage strains in the community reflect isolates causing disease. In this study, we will perform comparative genomics to infer the population structures of human-derived ESBL-PEc from community- and hospital-acquired infections and from community-based faecal carriage samples in the Netherlands. Furthermore, we will describe the molecular epidemiology of *E. coli* isolates causing invasive disease (BSI). Methods: This study uses four different microbiological data sources: 1) ESBL-PEc from patients with community-acquired UTI tested in primary care between May and November 2017, 2) ESBL-PEc from urine cultures obtained from patients hospitalized between January 2014 and December 2016, 3) *E. coli* from blood cultures obtained from patients hospitalized between January 2014 and December 2016, and 4) ESBL-PEc from faecal samples collected in a national population-prevalence study performed between January 2014 and January 2017. Clinical epidemiological data was collected from all patients and all isolates were subjected to whole genome sequencing. Discussion: The EPIGENEC study (EPIdemiology and GENetics of *E. coli*) will describe the molecular epidemiology of *E. coli* BSI and assess the genomic population structure of ESBL-PEc strains from community-acquired and nosocomial infections, and of ESBL-PEc reflecting community-based faecal carriage. Information from these studies may assist in optimizing surveillance strategies and determining targets and potential impact of future new preventive measures.

Keywords: *Escherichia coli*, antimicrobial resistance, infection, molecular epidemiology
Escherichia coli (E. coli) is commonly found as a gut commensal in humans. Besides its commensal lifestyle E. coli is also an important pathogen in humans, as it can establish disease in tissues other than the gastrointestinal tract. These so-called extra-intestinal pathogenic E. coli (ExPEC) can cause a wide spectrum of diseases, from uncomplicated cystitis to bloodstream infections (BSI) with 30-day mortality up to 18% 1–3. E. coli is a very heterogeneous species, only 20% of the genes in a typical E. coli genome is usually shared among all strains 4. E. coli is known to easily acquire antimicrobial resistance. Molecular characterization studies have shown that E. coli strains predominantly become resistant through the exchange of mobile genetic elements carrying resistance genes, such as those encoding for extended-spectrum beta-lactamases (ESBL) 5. ESBL-producing E. coli (ESBL-PEc) are often co-resistant to other classes of antibiotics 6. Infections caused by antibiotic-resistant E. coli strains occur with increasing frequency, which potentially increase the total overall E. coli disease burden 3,7,8. Furthermore, in a recent modelling study, ESBL-PEc was found to be responsible for approximately a third of the estimated 33,000 antibiotic-resistance related deaths in Europe in 2015 9. The increasing availability of whole genome sequencing (WGS) has allowed a more detailed insight into the genetics of E. coli virulence and resistance and provided further insight into the distribution of acquired virulence and resistance genes in pathogenic and commensal E. coli strains of different genetic backgrounds 10–12.

Intestinal carriage is believed to be the most important human reservoir for ESBL-PEc from which infections can originate 13. The estimated prevalence of ESBL-PEc faecal carriage in Dutch community-dwelling inhabitants ranges from 5.2% in the general population 14 to 10.1% in urbanized areas 15, and from 5.0% 16 to 6.1% 17 in hospitalized patients. Surveillance of the molecular epidemiology of antibiotic resistance in the community reservoir is important to identify trends in resistance development. Yet, such surveillance is labour-intensive and costly,
and, therefore, not regularly performed. It is currently unknown to what extent the molecular epidemiology of these ESBL-PEc strains present in the Dutch community relates to the molecular epidemiology of ESBL-PEc strains causing community-acquired and nosocomial *E. coli* infections. In case of good correlations, urine *E. coli* isolates from primary care patients or from hospitalized patients could be used for surveillance of the molecular epidemiology of antibiotic-resistant *E. coli* in the community in the Netherlands.

Information on to what extent *E. coli* strains from different niches and patient populations in the Netherlands differ genomically, is scarce. Possibly, there is also a difference in pathogenic potential within invasive *E. coli* isolates, reflected for example by molecular differences at the genome level in strains that have caused community-acquired BSI as compared to strains that cause BSI in a population that is already vulnerable to infection. Such information is critical for informing strategies around surveillance, prevention and treatment of this important pathogen. In particular for *E. coli* BSI, which is characterized by high morbidity and mortality, more insight in the clinical as well as molecular epidemiology in the Netherlands is needed to help identify targets and potential impact of future preventive strategies such as *E. coli* vaccines, of which one is currently being developed.

Here, the rationale and study design of the EPIGENEC Study (EPIdemiology and GENetics of *E. coli*) is described, which aims to 1) Investigate the association between the molecular epidemiology of faecal ESBL-PEc isolates obtained from carriage in the community and from clinical cultures (UTI and BSI) in the Netherlands, and 2) Characterize *E. coli* causing BSI in the Netherlands by evaluating clinical characteristics and mapping the genomic population structure of *E. coli* causing BSI based on WGS.
METHODS

Primary objectives

1. To compare the genomic population structure of ESBL-PEc isolates from community-acquired and nosocomial infections with ESBL-PEc isolates found in community faecal samples based on WGS.

2. To describe the clinical and molecular epidemiology of E. coli isolates causing BSI, and in particular, to assess the association between different clinical and molecular characteristics.

Study design and population

This observational study consists of a prospective as well as a retrospective part. Four sources of data and samples will be obtained from clinical care and the community (see Figure 1).

1. Community-acquired UTI

Patients with a community-acquired UTI caused by ESBL-PEc were identified prospectively by a positive urine culture result at Saltro, a medical laboratory providing service to primary care practices, primarily in the Utrecht (city) region. Urine samples were either inoculated in enrichment broth (Isobouillon with tobramycin, vancomycin, and nystatin) if specifically requested on ESBL or identified by elevated MIC for cephalosporins. Screening for ESBL-producing Enterobacteriaceae was performed by inoculation onto a selective screening agar, the Brilliance ESBL screening agar (Oxoid, Basingstoke, United Kingdom). All broths and plated were incubated overnight at 36°C. Species identification and antibiotic susceptibility testing of colonies growing on the Brilliance ESBL plates were performed with respectively the MALDITOF-MS (Bruker, Bremen, Germany) and the Vitek 2 system (Vitek AST, bioMérieux, Marcy-l'Étoile, France). The MIC breakpoints used for interpreting the results were according to the criteria of the EUCAST. Phenotypic confirmation of ESBL was performed by combination disk diffusion test, as recommended by the Dutch national guideline for laboratory ESBL
All ESBL-PEc isolates from positive urine cultures between May 2017 and November 2017 were stored at Saltro, at -80°C.

2. **Nosocomial UTI**

Patients with nosocomial UTI caused by ESBL-PEc were retrospectively identified from medical microbiological records in two participating hospitals: 1) University Medical Center Utrecht (UMCU), and 2) Amphia Hospital in Breda. The UMCU is a 1,042-bed tertiary hospital, providing care to the Utrecht (province) region and serves as a regional referral center. The Amphia Hospital is an 837-bed teaching hospital that provides service to a region of approximately ~400,000 residents. Sample inoculation and confirmation of phenotypic ESBL production was performed as described for the community UTI samples, except that CHROMagar ESBL plates were used (CHROMagar, Paris, France). In both hospitals, every first ESBL-PEc isolate per patient is routinely stored and frozen at -80°C by the medical microbiology department. For this study, we selected all ESBL-PEc isolates from nosocomial UTIs (sample taken >2 days after hospital admission) during the years 2014, 2015 and 2016.

3. **BSI**

In the same two hospitals, patients with *E. coli* BSI, both ESBL-producing and non-ESBL-producing, were retrospectively identified from medical microbiological records by growth of *E. coli* in blood cultures. In these hospitals, *E. coli* isolates from blood cultures are routinely stored at -80°C. For the years 2014, 2015 and 2016, a random sample of 40 isolates per year, comprising ~25% of all bacteraemic *E. coli* isolates in a year, was drawn from each hospital. In addition to the random sample, all ESBL-PEc isolates from 2014-2016 were selected for WGS. Consequently, this set of ESBL-PEc, together with the random sample of the bacteraemic *E. coli* strains, comprises the total blood isolate collection for the current study (see Figure 1).
All ESBL-PEc isolates obtained from the clinical samples (community UTI, nosocomial UTI and BSI) and the random sample of *E. coli* BSI isolates were included for further molecular analyses at The Netherlands National Institute for Public Health and the Environment (RIVM). Because we expected follow-up cultures to often grow the same *E. coli* isolate as the first culture, and for efficiency reasons, we selected only the first available *E. coli* isolate for each patient (all ages), irrespective of time between cultures.

4. Community-based intestinal carriage

The fourth dataset consists of ESBL-PEc isolates collected from faecal samples of a national population study for ESBL-producing Enterobacteriaceae, performed between November 2014 and November 2016. In this cross-sectional study, every month a random sample of ~2,000 residents of the Netherlands was drawn from Dutch municipalities (covering the entire population of the Netherlands). One person per household was invited to fill in a web-based questionnaire, and upon completion of the questionnaire, the participant was asked to provide a faecal sample. ESBL-producing *Escherichia coli*, *Klebsiella pneumoniae*, and the *Enterobacter cloacae* complex were isolated using MacConkey agar with 1 mg/L cefotaxime or after enrichment 2 mL of LB with 1 mg/mL cefotaxime. Up to five colonies with different morphologies were selected. Species identification was performed using MALDI-TOF-MS (Bruker, Bremen, Germany). ESBL-encoding genes were identified by PCR and isolates negative in the PCR were tested for the presence of other ESBL encoding genes by the Check MDR CT-101 microarray (Check-points, Wageningen, the Netherlands). The genes were identified by conventional sequencing. PCR-based Replicon Typing (PBRT) was performed to identify the plasmid type that encoded the ESBL. All ESBL-producing Enterobacteriaceae were stored at -80°C in the UMCU and were subjected to WGS (see Genotyping). Further details of the study design can be found elsewhere. For the current study, only genetic data of the first sampled faecal ESBL-PEc isolate of a patient were collected. No age restrictions were used.
Figure 1. Data sources

Legend Figure 1: this figure depicts the sources and time-span of the different *E. coli* collections that are used in this study.

**E. COLI**

**ESBL-POSITIVE E. COLI**

- Blood
  - Random sample
  - Retrospective collection
  - 2 hospitals
  - 2014-2016

- Nosocomial UTI
  - Retrospective collection
  - 2 hospitals
  - 2014-2016

- Community UTI
  - GP
  - Prospective collection in primary care and laboratory
  - 2017

- Faecal community carriage
  - Population study
  - 2014-2017

**Epidemiological variables**

The following information was collected from all patients: age, sex, postal code, type of infection (community UTI, nosocomial UTI, BSI), date of sample collection, and community or nosocomial (i.e. sample taken >2 days after hospital admission) onset of infection. In addition, for UTIs it was recorded whether the urine sample was a catheter sample. For patients with *E. coli* BSI, additional information regarding presence of a urinary catheter, hospital ward (ICU versus non-ICU), 30-day and 1-year mortality and the primary focus of BSI was obtained from electronic medical records. Possible primary foci were: urinary tract (i.e. pyelonephritis, prostatitis), gastrointestinal (i.e. diverticulitis, bacterial translocation), hepatic-biliary (i.e. cholangitis), respiratory,
gynaecological, other (i.e. meningitis, venous catheter), and unknown. The primary focus of BSI
(portal of entry) was defined on the basis of clinical and/or radiologic features and the isolation
of \textit{E. coli} from the presumed source of infection. If \textit{E. coli} was not isolated from the presumed
primary focus (i.e. because of previous antimicrobial treatment or invasive procedure that was
needed to isolate \textit{E. coli} from primary source), the presumed primary focus was based on a firm
clinical suspicion (given that all other possible sources of infection were excluded). In case of
multiple possible primary foci, consensus was reached by discussion by DH and TV.

**Genotyping**

All \textit{E. coli} isolates that were selected for the current study were inoculated on (non-selective)
blood agar and species confirmation was performed by MALDITOF-MS prior to WGS, which
was performed at the RIVM. All \textit{E. coli} strains, except the strains from the external dataset, were
subjected to WGS using the Illumina HiSeq 2500 (BaseClear, Leiden, the Netherlands). For
this, cell pellets were made from 1500\(\mu\)l of overnight culture in Brain Heart Infusion (BHI) broth
(Tritium Microbiologie BV, Eindhoven, the Netherlands). The pellets were washed in 500\(\mu\)l
saline and subsequently re-suspended in 200\(\mu\)l DNA/RNA Shield (Zymo Research, Irvine CA,
U.S.A.). From the pellets, nucleic acid extraction was performed by BaseClear (Leiden, the
Netherlands) using an in-house protocol. Library preparation and sequencing of the bacterial
genomes was performed using the Illumina Nextera XT kit and the HiSeq 2500 with a paired-
end 100 cycles protocol. The ESBL-PEc from the fourth dataset, i.e. ESBL-PEc recovered from
faeces of healthy humans, were sequenced using the MiSeq or NextSeq platforms (Illumina,
San Diego, CA). Contigs of all four datasets were assembled with SPAdes genome assembler
v.3.6.2. The resulting WGS data were used to determine the multi-locus sequence type (MLST),
virulence gene and resistance gene content using the mlst2.0 (https://github.com/tseemann/mlst) to scan contig files against the \textit{E. coli} PubMLST typing
scheme (updated May 12th, 2018), and Abricate (https://github.com/tseemann/abricate) for
mass screening of contigs for antimicrobial resistance and virulence genes. Abricate comes bundled with multiple resistance gene and virulence gene databases. For this study, the ResFinder and VFDB databases were used. Serotypes were assigned by using the web-tool SerotypeFinder 2.0 from the Center for Genomic Epidemiology at the Danish Technical University, Lyngby, Denmark (http://www.genomicepidemiology.org). This tool uses presence of O- and H-antigen-processing genes to predict *E. coli* serotypes

**Planned analyses**

**Primary objective 1**

The population structure of ESBL-PEc from the clinical and faecal samples will be compared on three levels. Firstly, the core genome will be assessed with MLST, a core genome phylogeny based on SNP and allelic profile variation using SeqSphere, and the ESBL-PEc populations will be partitioned in sequence clusters. For this, different methods are available like hierarchical Bayesian Analysis of Population Structure (BAPS) or PopPUNK. Secondly, the accessory genome will be assessed by comparing acquired resistance genes in the ESBL-PEc populations using Resfinder, and the plasmid composition will be predicted using the recently developed mlplasmid© algorithm. Lastly, a pan-genome analysis will be performed using PANINI, to assess if the total gene content differs per different ESBL-PEc population.

**Primary objective 2**

To assess the association between epidemiological characteristics and molecular characteristics of *E. coli* blood isolates, MLST, virulence and antimicrobial resistance gene content will be described according to the different epidemiological subgroups. A core-genome tree will be constructed with the same method as mentioned above. A virulence score will be made per isolate and will be defined as the total number of virulence genes present in that strain. These virulence scores will then be compared between isolates with different epidemiological characteristics and between ST131 and non-ST131 isolates, respectively.
Serotype distribution of the bacteraemia population will be compared to current *E. coli* vaccine candidates. Furthermore, a genome-wide association approach will be used to see whether any epidemiological characteristics are associated with certain molecular traits.

**Ethics**

This study is conducted according to the principles of the Declaration of Helsinki (World Medical Association, 2013) and does not fall under the scope of the Medical Research Involving Human Subjects Act, the Medical Research Ethics Committee of the UMCU has therefore waived the need for official approval by the Ethics Committee (IRB number 18/056). The study uses pseudonymised data and informed consent is not obtained from study participants. Patients that participated in the open population study (ESBLAT study, IRB number 14/219-C) have provided informed consent for the use of clinical data and faecal samples in future studies such as the current study. In this study, in case of age <13 years, parents provided informed consent. In case of age 13-17 years, both the child and parents provided informed consent.

**DISCUSSION**

The EPIGENEC study aims to assess the genomic population structure of ESBL-PEc strains from community and nosocomial infections and ESBL-PEc strains representing community faecal carriage. It will also carefully describe the clinical epidemiology and genomic population structure of *E. coli* BSI, which is important in determining the targets and impact of possible new preventive measures.

**Strengths**

One of the key aspects of the current study is the combined use of epidemiological data and detailed whole genome sequence data of strains from several different domains in order to obtain a more complete picture of the current molecular epidemiology of (ESBL-producing) *E. coli* in the Netherlands. Furthermore, the use of WGS techniques allows us to map the
population structure of *E. coli* and the association of the genomic make-up of strains with their isolation source with high resolution and discriminatory power. Also, all strains were uniformly assembled and analysed, reducing the risk of information bias.

**Limitations**

This study also has limitations. Guidelines for Dutch primary care physicians recommend to only send in urine cultures for microbiological testing for patients with complicated UTI (i.e. symptoms accompanied with fever, or in case of male patients with UTI symptoms), clinical treatment failure, recurrent UTIs, or a possibly resistant infection, which implies selection of patients with community UTI. However for our study, we do not consider this to cause selection bias, since we are particularly interested in the molecular epidemiology of ESBL-PEc from urine samples in the way they are currently being performed, so as according to clinical practice. Also, ideally we would be able to pick up time-trends in the change in molecular epidemiology of community faecal carriage of ESBL-PEc and assess whether these trends are reflected in the molecular epidemiology of clinical cultures, for example from community or nosocomial UTI. One could imagine using such results to assess the possible value of ESBL-PEc isolates from clinical cultures as a proxy of changes in the molecular epidemiology of community faecal carriage. However, considering the heterogeneity in the *E. coli* species and the limited amount of years of which we have faecal samples, this will prohibit us to draw hard conclusions. We still believe this comparison will provide us valuable information and will guide future research on the possible use of routine clinical samples in the assessment of the molecular epidemiology of ESBL-PEc.
DECLARATIONS

Acknowledgements

We sincerely thank Kim van der Zwaluw, Judith Vlooswijk, Carlo Verhulst and Jelle Scharringa for their contribution in the preparation of this study.

Conflicts of interest

None to declare.

Study funding

This study is internally funded by the Netherlands National Institute for Public Health and the Environment (RIVM) and the UMCU and has not received external funding or assistance from any commercial organization.
REFERENCES


