Extended-spectrum beta-lactamase (ESBL)-producing and non-ESBL-producing *Escherichia coli* isolates causing bacteremia in the Netherlands (2014 – 2016) differ in ST distribution, antimicrobial resistance gene and virulence gene content

Denise van Hout*[1], Tess D. Verschuuren[1], Patricia C.J. Bruijning-Verhagen[1,2], Thijs Bosch[2], Anita C. Schürch[3], Rob J.L. Willems[3], Marc J.M. Bonten[1-3], Jan A.J.W. Kluytmans[1,4]

Author affiliations:

[1] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands

[2] Center for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[3] Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, University Utrecht, The Netherlands

[4] Microvida Laboratory for Medical Microbiology and Department of Infection Control, Amphia Hospital, Breda, The Netherlands

*** Corresponding author**:
Denise van Hout
UMC Utrecht; Julius Center
D.vanHout-3@umcutrecht.nl

ABSTRACT

Background: Knowledge on the molecular epidemiology of *Escherichia coli* causing *E. coli* bacteremia (ECB) in the Netherlands is almost exclusively based on extended-spectrum beta-lactamase producing *E. coli* isolates (ESBL-PEc) that are responsible for about 10% of all ECB episodes. We determined clonal differences and differences in resistance and virulence gene content between non-ESBL-producing *E. coli* (non-ESBL-PEc) and ESBL-PEc isolates with different epidemiological characteristics.

Materials/methods: A random selection of non-ESBL-PEc isolates as well as all available ESBL-PEc blood isolates were obtained from two Dutch hospitals between January 2014 and December 2016. For comparative analysis, whole genome sequencing was performed of all isolates to infer the sequence types (STs), serotypes, antibiotic resistance by either acquired genes or chromosomal mutations and virulence gene (VG) scores, based on presence of 49 predefined putative pathogenic VG.

Results: ST73 was most prevalent among the 212 non-ESBL-PEc (N = 26, 12.3%) and ST131 among the 69 ESBL-PEc (N = 30, 43.5%). Prevalence of ST131 among non-ESBL-PEc was 10.4% (N = 22, *P* value < 0.001 compared to ESBL-PEc). O25:H4 was the most common serotype in non-ESBL-PEc and ESBL-PEc. Median total resistance gene counts were 42 (IQR 39 – 45) and 46 (IQR 43 – 50) for non-ESBL-PEc and ESBL-PEc, respectively (*P* value < 0.001). Median acquired resistance gene counts were 1 (IQR 1 – 6) and 7 (IQR 4 – 9) for non-ESBL-PEc and ESBL-PEc, respectively (*P* value < 0.001). Median VG scores were 13 (IQR 9 – 20) and 12 (IQR 8 – 14) for non-ESBL-PEc and ESBL-PEc isolates, respectively (*P* value = 0.002). Resistance gene and VG content varied between different *E. coli* STs.

Conclusions: We observed differences between non-ESBL-PEc and ESBL-PEc blood isolates in ST distribution, resistance gene and virulence gene content.

INTRODUCTION

Despite advances in medical healthcare and in contrast to the decline in other infectious diseases, the annual incidence of Gram-negative bacteremia in Europe is increasing [1–4]. *Escherichia coli* is the leading causative pathogen in Gram-negative bacteremia and is associated with 30-day mortality up to 18% [1,4–6]. Antibiotic treatment options of *E. coli* bacteremia (ECB) are getting compromised by the pandemic presence of extended-spectrum beta-lactamases (ESBLs) [1–4]; enzymes that confer resistance to antibiotics commonly used for ECB treatment such as third-generation cephalosporins. ESBLs can be exchanged between strains by horizontal gene transfer, such as through carry-over of mobile genetic elements. In some European countries, the incidence of ECB with antibiotic-resistant strains seems to increase faster than ECB caused by susceptible strains [1–4]. The individual patient and financial burden is increased for ECB episodes that are caused by resistant *E. coli*. Yet, ECB due to susceptible strains is far more common and therefore determines the major part of the *E. coli* bacteremia disease burden in the population [1–4]. The majority of ECBs is of community onset and is preceded by an infection in the urinary tract, but other sources, such as the hepatic-biliary tract, also comprise important primary foci [4,7]. These clinical characteristics of ECB episodes are important because they can indicate different target populations for prevention. More insight in the molecular epidemiology of ESBL-negative as well as ESBL-positive ECB with different clinical characteristics is needed to help identify key targets for the development of future preventive strategies such as *E. coli* vaccines, which are currently being developed [8]. Up to now, the molecular epidemiology of ECB in the Netherlands has been mainly described in single-center studies [9] and among antimicrobial resistant isolates only [10,11]. Dutch studies combining patient characteristics with high-resolution genetic data of *E. coli* isolates are limited, specifically for ECB, with its potential severe clinical consequences.

In this study, we aimed to analyze the current population structure of ECB in the Netherlands, with special attention to differences in antimicrobial resistance and virulence gene content and serotype distribution between isolates with different clinical epidemiological characteristics and between non-ESBL-producing *E. coli* (non-ESBL-PEc) and ESBL-producing *E. coli* (ESBL-PEc) blood isolates.

METHODS

Study design

Details of the study design, epidemiological data collection and laboratory methods (i.e. phenotypic ESBL detection) are described elsewhere [12]. In short, patients with ECB were retrospectively identified from medical microbiological records in the University Medical Center Utrecht (UMCU), a 1,042-bed tertiary care center and the Amphia Hospital in Breda, an 837-bed teaching hospital. In each hospital, we selected a random sample of 40 isolates of unique patients per year for the years 2014, 2015 and 2016, comprising ~24% of all first bacteremic *E. coli* isolates in a year. In addition to this random sample, all ESBL-PEc blood isolates from 2014 – 2016 were selected from the two hospitals. Whole genome sequencing (WGS) was performed by The Netherlands National Institute for Public Health and the Environment (RIVM) using the Illumina HiSeq 2500 (BaseClear, Leiden, the Netherlands). De novo assembly was performed using SPAdes genome assembler v.3.6.2 and the quality of assembles was assessed using QUAST [13]. Only genomes with an estimated genome size between 3 MB and 6 MB and number of contigs not exceeding 1,000 were included in further analyses. Baseline clinical epidemiological characteristics were compared between the non-ESBL-PEc and ESBL-PEc ECB episodes. ESBL-production was defined as confirmed phenotypic ESBL-positivity, unless described otherwise [12]. Baseline characteristics were compared by the Fisher's Exact or Pearson χ^2 test for categorical variables and by Mann-Whitney U test for continuous variables when applicable. A two-tailed *P* value <0.05 was considered statistically significant.

This study 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) and informed consent was not obtained. All statistical analyses were performed with Statistical Package for Social Sciences V.25.0 (SPSS, Chicago, Illinois, USA) and R Version 3.4.1. Boxplots were made with R packages *ggplot2* and *ggpubr* and bar charts were made with Graphpad Prism Version 8.0.1.

Multi-locus sequence types (MLST)

Multi-locus sequence types (STs) were determined using mlst2.0 (https://github.com/tseemann/mlst) by scanning contig files against the *E. coli* PubMLST typing scheme (updated May 12th, 2018). ST (i.e. clonal) distribution was presented stratified for non-ESBL-PEc and ESBL-PEc isolates and by epidemiological subgroup (i.e. community versus hospital onset; different primary foci of ECB). Genotype (ST) diversity was analysed by Simpson's diversity index [14]. A core-genome (cg) neighbour-joining (NJ) phylogenetic tree was constructed in Seqsphere with the *Escherichia/Shigella* cgMLST v1 scheme developed by Enterobase (https://enterobase.warwick.ac.uk/species/index/ecoli), containing 2,513 target genes and visualised using the free online available web-tool Microreact (https://microreact.org/showcase) [15]. The cgNJ method reconstructs phylogeny by using a distance matrix that contains the genetic distance between each pair of sequences.

Serotyping

We assigned serotypes by using the web-tool SerotypeFinder 2.0 from the Center for Genomic Epidemiology at the Danish Technical University, Lyngby, Denmark (https://cge.cbs.dtu.dk/services/SerotypeFinder) [16]. Simpson's index for serotype diversity was calculated for non-ESBL-PEc and ESBL-PEc isolates. Serotype distribution among non-ESBL-PEc and ESBL-PEc was compared to two current *E. coli* vaccine candidates [8,17],

123 excluding isolates in which no definitive serotype could be defined and the occurrence of
124 serotypes was described by primary focus of ECB

125 **Antimicrobial resistance genes**

126 Abricate (https://github.com/tseemann/abricate) version 0.8.13 was used for mass screening of
127 contigs for antimicrobial resistance genes using the ResFinder 3.1.0 database (acquired
128 resistance genes), date of download 24 January 2019, and the Comprehensive Antibiotic
129 Resistance Database (CARD) (all resistance genes), date of download 1 March 2019 [18,19].
130 The thresholds for coverage length and sequence identity were 80% and 95%, respectively. A
131 resistance gene count using each of the databases was made per isolate, which was defined as
132 the total number of resistance genes (using CARD) and the total number of acquired resistance
133 genes (using ResFinder) identified, respectively. In case of double detection of identical
134 resistance genes within a single isolate, they were only counted once. The resistance gene
135 scores were compared between non-ESBL-PEc and ESBL-PEc with the non-parametric
136 Wilcoxon rank sum test (for this comparison only, the scores of the ESBL-PEc isolates were
137 corrected for presence of the ESBL gene). Resistance gene scores were then analysed for non-
138 ESBL-PEc and ESBL-PEc separately and were compared between isolates with different
139 epidemiological characteristics and different STs using Kruskal-Wallis one-way ANOVA. In case
140 of an overall ANOVA *P* value <0.05, post-hoc pairwise comparisons were made and the Holm-
141 Bonferroni *P* value correction was applied to account for multiple testing. For pairwise
142 comparisons, the non-parametric Wilcoxon rank sum test was used.

143 **Virulence genes**

144 The presence of putative virulence factor genes (VG) was identified using abricate version
145 0.8.13 for BLAST against the VFDB database (http://www.mgc.ac.cn/VFs), date of download 8
146 February 2019, with minimal coverage length and sequence identity 80% and 95% [20]. We
147 searched for 49 putative VG that were previously described as extra-intestinal pathogenic *E. coli*

(ExPEC)-associated VG [21–25]. If any of the predefined VG were not included in VFDB, BLAST against the ecoli_VF_collection database was performed (date 8 February 2019), a repository that contains known VG from VFDB supplemented with additional *E. coli* VG that have been reported in literature [26]. The *kpsM*, *afa/dra* and *sfa/foc* operons were considered present if any of the corresponding genes or allelic variants were identified. A virulence score was made per isolate and was defined as the total number of pre specified VG, adjusted for multiple detection of the *afa/dra* (Afa/Dr adhesins), *pap* (P fimbrial adhesins), *sfa/foc* (S and F1C fimbrial adhesins) and *kpsM* (group 2 and III capsule) operons, as described previously [23]. If a VG was detected multiple times within a single isolate (i.e. with different quality measures), it was only counted once. These virulence scores were then compared between isolates with different epidemiological characteristics and between different STs using Kruskal-Wallis one-way ANOVA. In case of an overall ANOVA *P* value <0.05, post-hoc pairwise comparisons were made with the non-parametric Wilcoxon rank sum test and the Holm-Bonferroni *P* value correction was applied to account for multiple testing.

RESULTS

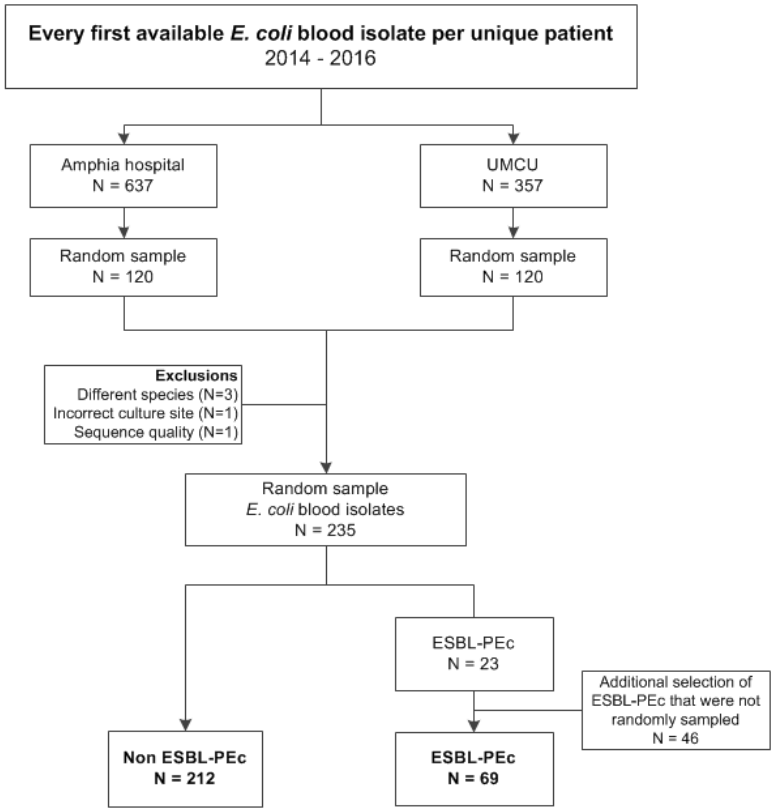
Patient characteristics

The isolate collection consisted of 212 phenotypic non-ESBL-PEc and 69 ESBL-PEc blood isolates (Fig. 1). Distribution of age, sex, onset of infection and primary foci were comparable between non-ESBL-PEc and ESBL-PEc bacteremia episodes (Table 1). As compared to non-ESBL-PEc, ECB episodes with ESBL-PEc were less often of community onset (63.8% versus 81.1%, *P* value = 0.003). Crude 30-day and 1-year mortality was higher in ECB episodes caused by ESBL-PEc (27.5% and 50.7%, respectively) in comparison with ECB episodes caused by non-ESBL-PEc (11.3% and 29.2%, respectively).

Clonal distribution

Among non-ESBL-PEc, ST73 was the most frequently observed ST (12.3%), followed by ST131 (10.4%). Isolates of ST73, 95, 127, 141, 80 and 1193 were solely identified among non-ESBL-PEc (Fig. 2). ST131 was dominant among ESBL-PEc (N = 30, 43.5%) and prevalence was higher than among non-ESBL-PEc (*P* value < 0.001). Simpson's index for ST diversity was 95.6% (95% CI 94.4% – 96.8%) and 80.6% (95% CI 70.9% – 90.4%) for non-ESBL-PEc and ESBL-PEc, respectively. The occurrence of different STs did not differ between nosocomial and community onset ECB (S1 Figure and S2 Table). ST131 was the dominant ST among ESBL-positive ECB episodes with a primary urinary (63%) and hepatic-biliary focus (57%), which was higher as compared to other primary foci of ESBL-positive ECB (i.e. 21% among primary hepatic-biliary focus, see S3 Figure and S4 Table). The NJ-phylogenetic tree of all isolates can be found in the Supporting Information (S5 Figure).

Figure 1. Flowchart of selection of *E. coli* blood isolates

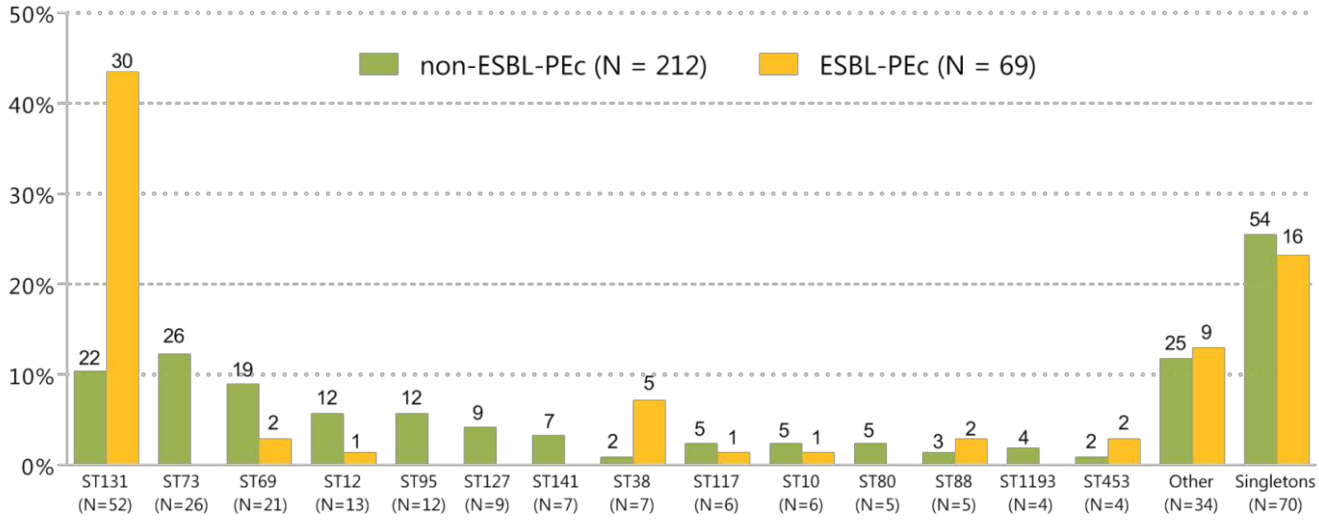


186 **Table 1.** Baseline epidemiological characteristics of *E. coli* bacteremia episodes

	Non-ESBL-PEc[a] N = 212		ESBL-PEc[a] N = 69		*P* value[b]
Median age, years (IQR)	69	(59 – 77)	69	(56 – 76)	0.802
Female sex (%)	102	(48.1)	32	(46.4)	0.802
Community onset (%)	172	(81.1)	44	(63.8)	*0.003*
Primary focus of ECB (%)					
Urinary tract	103	(48.6)	30	(43.5)	0.785
Hepatic-biliary	46	(21.7)	14	(20.3)	
Gastro-intestinal	23	(10.8)	7	(10.1)	
Other	10	(4.7)	5	(7.2)	
Unknown	30	(14.2)	13	(18.8)	
Urinary catheter (%)	69	(32.5)	28	(40.6)	0.223
Ward (%)					
Non-ICU	182	(85.8)	58	(84.1)	0.714
ICU	30	(14.2)	11	(15.9)	
Mortality (%)					
30-day	24	(11.3)	19	(27.5)	*0.001*
1-year	62	(29.2)	35	(50.7)	*0.001*

ECB, *E. coli* bacteremia; ESBL, extended-spectrum beta-lactamase; ESBL-PEc, ESBL-producing *E. coli*; ICU, intensive care unit; IQR, interquartile range; non-ESBL-PEc, non-ESBL-producing *E. coli*
[a]ESBL-positivity based on phenotype.
[b]*P* value of comparison between non-ESBL-PEc versus ESBL-PEc, calculated with Pearson's χ^2, Fisher's exact, or Mann-Whitney U test when applicable. *P* values in italic represent *P* values <0.05.

187 **Figure 2.** ST distribution among non-ESBL-PEc versus ESBL-PEc[a] in order of frequency[b]



188
189 ESBL-PEc, ESBL-producing *E. coli*; non-ESBL-PEc, non-ESBL-producing *E. coli*; ST, sequence type
190 [a]ESBL-positivity based on phenotype.
191 [b]Missing STs and STs that occurred ≤3 times are grouped in "Other". STs that only occurred once are grouped in
192 "Singletons". The height of each individual bars represents the proportion of the ST within the group of non-ESBL-
193 PEc and ESBL-PEc, respectively. The numbers represent the absolute numbers of occurrence.

Serotypes

The most common serotype was O25:H4 and was identified in 19 (9.0%) non-ESBL-PEc and 24 (34.8%) ESBL-PEc isolates, which largely reflected the prevalence of ST131 in each group (Table 2). Multiple serotypes only occurred among non-ESBL-PEc, such as O6:H1 and O6:H31. ST73 was most often of serotype O6:H1 (16 / 26, 61.5%). Simpson's index for serotype diversity was 96.7% (95% CI 95.8% – 97.6%) and 83.8% (95% CI 76.9% – 90.6%) for non-ESBL-PEc and ESBL-PEc, respectively. The distribution of O:serotypes per primary focus of ECB was mostly in line with the distribution of the most important STs per primary focus (i.e. ST131, ST73) (S6 Table).

53 (25.0%) non-ESBL-PEc and 25 (36.2%) ESBL-PEc isolates belonged to either O1, O2, O6 or O25, the serotypes of the 4-valent *E. coli* vaccine that has reached phase 2 development stage [8,28], whereas the majority of non-ESBL-PEc (N = 113; 53.3%) and ESBL-PEc isolates (N = 35; 50.7%) belonged to one of the O:serotypes of the new 10-valent conjugant *E. coli* vaccine (ExPEC-10V) that is currently in development [17].

Antimicrobial resistance genes

In total, 110 unique resistance genes were identified with CARD and 69 unique acquired resistance genes were identified with ResFinder 3.1.0 (see S7 Table). ESBL-genes were detected in 65 (94.2%) of 69 *E. coli* blood isolates with phenotypic ESBL production. $bla_{\text{CTX-M-15}}$ was the most prevalent ESBL gene (N = 28, 43.1%), followed by $bla_{\text{CTX-M-9}}$ (N = 14, 21.5%) and $bla_{\text{CTX-M-27}}$ (N = 9, 13.8%). Assemblies of the phenotypic ESBL-PEc isolates in which no ESBL-gene was identified (N = 4) were individually uploaded on the DTU Resfinder 3.1.0 website (date 11 March 2019, thresholds for coverage length 80% and sequence identity 95%); these isolates remained genotypically ESBL-negative. One of these isolates was positive for $bla_{\text{CMY-2}}$ (AmpC gene).

The median total resistance gene count was 42 (IQR 39 – 45) for non-ESBL-PEc and 46 (IQR 43 – 50) for ESBL-PEc iolates (*P* value < 0.001). The median acquired resistance gene count for non-ESBL-PEc versus ESBL-PEc was 1 (IQR 1 – 6) versus 7 (IQR 4 – 9) (*P* value < 0.001). Among non-ESBL-PEc, total and acquired resistance gene counts were not different between community and hospital-onset ECB episodes (S8 Figure and S9 Table). Among non-ESBL-PEc, there were statistically significant differences in resistance gene count for different primary foci of ECB, but absolute differences were small (S8 Figure). The median acquired resistance gene count of non-ESBL-PEc isolates from ECB with a primary hepatic-biliary focus was 1 (IQR 1 – 1), whereas for a primary urinary focus this was 2 (IQR 1 – 6) (*P* value ≤0.001), primary gastro-intestinal focus this was 4 (IQR 1 – 8) (*P* value ≤0.01) and unknown primary focus this was 1 (IQR 1 – 6) (*P* value ≤0.0001) (S9 Table). Among ESBL-PEc isolates, there were no statistical significant differences in total and acquired resistance gene counts between community and hospital-onset ECB or different primary foci of ECB (S8 Figure and S9 Table). Among non-ESBL-PEc, there was heterogeneity in total resistance gene count between almost all dominant STs; this was not the case for acquired resistance gene count (Fig. 3 and S10 Table). No statistically significant differences were observed among ESBL-PEc isolates of different STs (S10 Table).

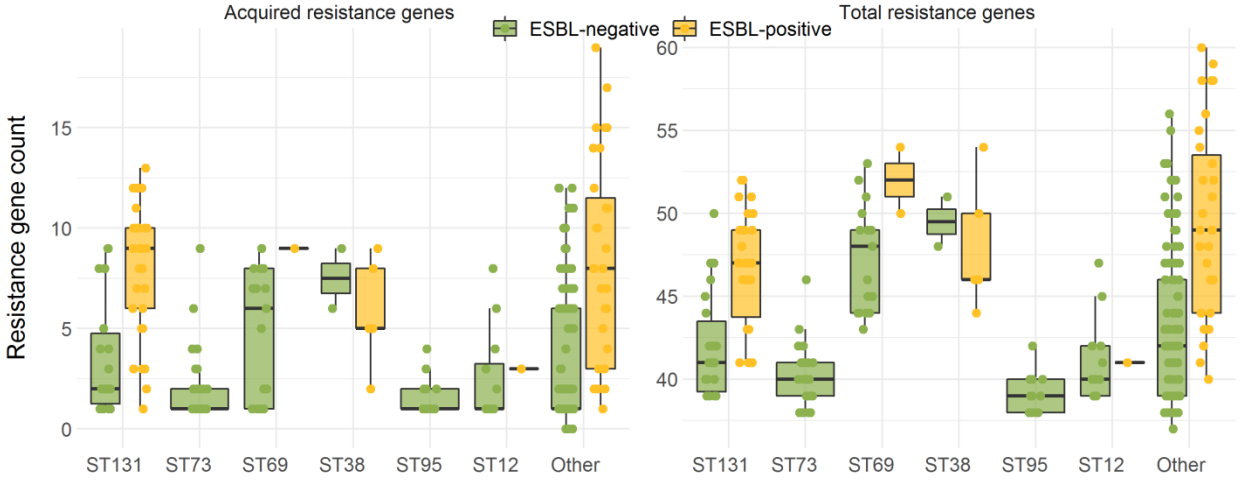
235 **Table 2.** Serotype distribution among *E. coli* blood isolates, stratified for ESBL-positivity

	Non-ESBL-PEc N = 212 (%)	**ESBL-PEc[a] N = 69** (%)
O25:H4 (%)	19 (9.0)	24 (34.8)
O6:H1 (%)	16 (7.5)	-
O2/O50:H6 (%)	10 (4.7)	-
O6:H31 (%)	9 (4.2)	-
O15:H18 (%)	7 (3.3)	2 (2.9)
O17/O44/O77:H18 (%)	8 (3.8)	-
O4:H5 (%)	7 (3.3)	1 (1.4)
O75:H5 (%)	8 (3.8)	-
O8:H9	5 (2.4)	2 (2.9)
O16:H5 (%)	3 (1.4)	3 (4.3)
O86:H18	1 (0.5)	4 (5.8)
O4:H1 (%)	5 (2.4)	-
O1:H7	4 (1.9)	-
O117:H4	4 (1.9)	-
O2/O50:H1	4 (1.9)	-
O23:H16	2 (0.9)	2 (2.9)
O25:H1	4 (1.9)	-
O18/O18ac:H7	3 (1.4)	-
O2/O50:H7	3 (1.4)	-
O45:H7	3 (1.4)	-
O75:H7	3 (1.4)	-
O8:H17	3 (1.4)	-
O9:H17	-	2 (2.9)
O9/O104:H9	-	2 (2.9)
O13/O135:H4	2 (0.9)	-
O18:H1	2 (0.9)	-
O18:H5	2 (0.9)	-
O22:H1	2 (0.9)	-
O24:H4	2 (0.9)	-
O8:H10	2 (0.9)	-
O8:H25	2 (0.9)	-
O8:H30	2 (0.9)	-
Singletons	45 (21.2)	13 (18.8)
Unknown	20 (9.4)	14 (20.3)

ESBL, extended-spectrum beta-lactamase; ESBL-PEc, ESBL-producing *E. coli*, non-ESBL-PEc, non-ESBL-producing *E. coli*

236

Figure 3. Resistance gene count per ST, stratified for ESBL-positivity[a]



ESBL, extended-spectrum beta-lactamase; ST, sequence type
[a]ESBL-positivity based on phenotype.
Boxplots display median resistance gene count and inter quartile range (IQR); and every dot represents a single isolate. Results of the CARD database are depicted in the left graph and included all resistance genes. The results of the ResFinder 3.1.0 database are depicted in the right graph and only included acquired resistance genes. Only STs that occurred >5% within ESBLs or non-ESBLs were grouped into main groups, the rest was categorized as "Other". Results of the pairwise comparisons can be found in S10 Table.

Virulence genes

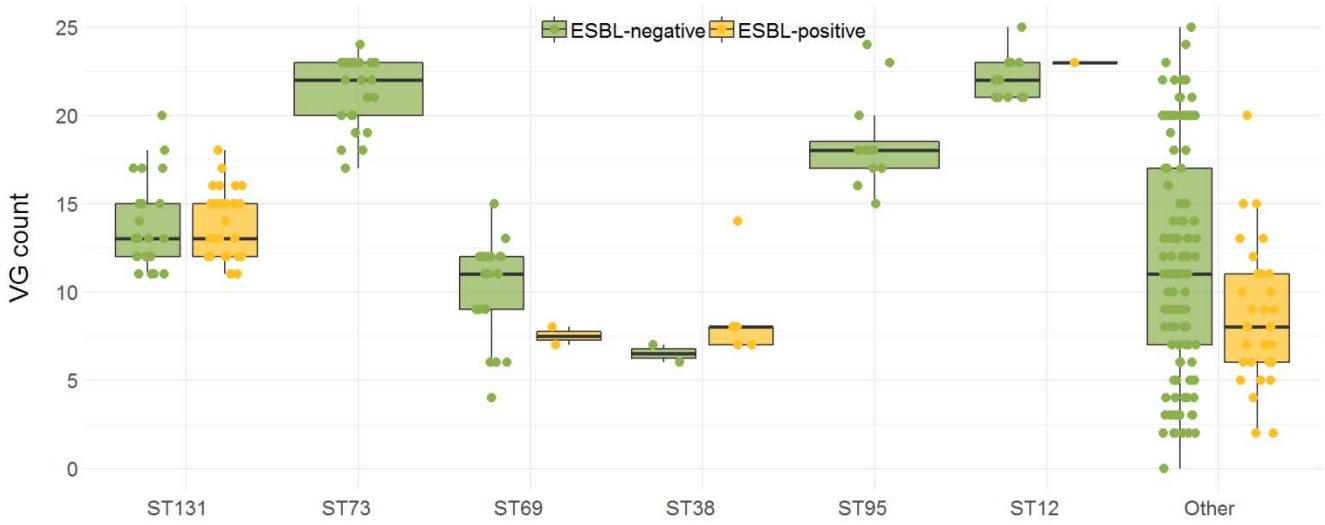
Of the 49 predefined ExPEC-associated VG, 44 (89.8%) were detected in at least one *E. coli* blood isolate (S11 Table). The median VG score was 13 (IQR 9 – 20) for non-ESBL-PEc and 12 (IQR 8 – 14) for ESBL-PEc blood isolates (*P* value = 0.002). In one non-ESBL-PEc isolate no predefined ExPEC-associated VG was detected, while a maximum VG score of 25 was found in two non-ESBL-PEc isolates.

For non-ESBL-PEc and ESBL-PEc isolates, there was no significant difference in the VG score between isolates that caused community or hospital onset ECB (S13 Table). Non-ESBL-PEc isolates that caused ECB with a primary gastro-intestinal focus (median 10, IQR 6 – 13) and hepatic-biliary focus (median 11, IQR 5 – 18) had lower VG scores as compared to isolates with a primary focus in the urinary tract (median 15, IQR 11 – 21) (*P* value = 0.007 and *P* value = 0.036, respectively, see S12 Figure and S13 Table). Among non-ESBL-PEc and ESBL-PEc, there were no statistical significant differences in VG scores between isolates of patients without

versus with a urinary catheter, between patients alive or deceased after 30 days or between patients admitted to the intensive care unit (ICU) versus a non-ICU ward (data not shown).

There was heterogeneity in VG scores between non-ESBL-PEc of different STs, this was less pronounced for ESBL-PEc isolates (Fig. 4 and S14 Table). ESBL-negative ST38 had the lowest average VG score (median 7, IQR 6 – 7) and ESBL-positive ST12 had the highest VG score (median 23, IQR 23 – 23). Median VG score of both ESBL-negative and ESBL-positive ST131 isolates was 13 (IQR 12 – 15). All pairwise comparisons between ESBL-negative STs yielded Holm-Bonferroni adjusted *P* values < 0.05, except for the comparison ST12 versus ST73 and all pairwise comparisons that included ST38.

Figure 4. ExPEC-associated VG score in different STs, stratified for ESBL-positivity[a]



ESBL, extended-spectrum beta-lactamase; ST, sequence type
[a]ESBL-positivity based on phenotype.
Boxplots display median VG score and inter quartile range (IQR); and every dot represents a single isolate. Only STs that occurred >5% within ESBLs or non-ESBLs were grouped into main groups, the rest was categorized as "Other". Results of pairwise comparisons can be found in S14 Table.

DISCUSSION

In this study, we found that ESBL-producing *E. coli* blood isolates were different from non-ESBL-producing *E. coli* causing bacteraemia in terms of clonal distribution, serotype distribution, antimicrobial resistance gene count and VG scores.

The ST distribution among ESBL-PEc blood isolates was less diverse as compared to non-ESBL-PEc. This was mainly caused by the predominance of ST131 within ESBL-PEc, as has been described before [29,30]. In contrast, ST73, a ST that is known for its susceptible antibiotic profile [29], was only identified among non-ESBL-PEc blood isolates. The association between ESBL phenotype and STs in *E. coli*, which is repeatedly found, implies that the molecular backbone of strains can increase (or decrease) its propensity to acquire and subsequently maintain plasmids carrying ESBL genes. A recent large-scale study that compared the pan-genomes of invasive *E. coli* isolates, including ST131 and ST73, suggested that due to ongoing adaptation to long term human intestinal colonisation and consequent evolutionary gene selection, ST131 might have become able to reduce the fitness costs of long term plasmid maintenance [31,32]. Interestingly, in our study, isolates that belonged to ST73 had low resistance gene content but relatively high VG score as compared to other STs. Furthermore, the average VG score among non-ESBL-PEc was slightly higher in comparison to ESBL-PEc blood isolates, which demonstrates that ESBL-positivity in *E. coli* is not necessarily related to an increased VG content. In case we would assume that VG content is associated with virulent potential, i.e. the ability of a strain to cause invasive disease, then these findings do not support the theory that increased virulence of resistant strains causes the increased incidence of resistant ECB as compared to sensitive ECB. This theory has been suggested for other pathogens, such as MRSA [1,33,34]. Still, results of the current study show that molecular characteristics of ESBL-PEc cannot be merely generalized to non-ESBL-PEc blood isolates, highlighting the importance of not preselecting on ESBL-positivity when investigating the molecular epidemiology of ECB.

One of our hypotheses was that the distributions of STs, resistance gene and VG content would differ between ECB episodes of community and hospital onset and between different primary foci, as a possible result of different levels of sub specialization of intestinal *E. coli* strains. However, we did not observe large differences when assessing epidemiological

subgroups, but found that these differences in molecular content mostly depended on phenotypic ESBL-production and STs. This confirms the findings from a recent study that was performed in Scotland [35]. In that study, there were combinations of virulence genes as well as a particular accessory gene composition that differentiated between STs rather than between epidemiological factors. The association between ST69 and community onset ECB, as found in the Scottish study, was not identified in the current study. Other differences were the large proportion of *E. coli* isolates from ECB episodes that were deemed hospital-acquired (62%) as compared to our study (18.4% for ESBL-negative and 36.2% for ESBL-positive ECB) and in that study, analyses were not stratified for ESBL-positivity. More studies that combine clinical characteristics with molecular characteristics of ECB are important, because these data help to further elucidate the role of host-specific factors versus strain-specific factors in the pathogenesis of ECB. Since different determinants of ECB might indicate different targets for surveillance or infection-prevention, a thorough understanding of the molecular epidemiology is needed to reduce the occurrence of this invasive infectious disease with potential severe clinical consequences.

We identified serotype O25:H4 as the most prevalent serotype causing ESBL-negative as well as ESBL-positive ECB in the Netherlands, followed by O6:H1. The serotype distribution among non-ESBL-PEc was more heterogeneous as compared to ESBL-PEc, similar to the differences in clonal diversity between these two groups. A large recent European surveillance study that included 1,110 *E. coli* blood isolates from adults between 2011 and 2017 showed that there is heterogeneity in serotype distribution among different countries, which highlights the need for country specific data [17]. Furthermore, we showed that the coverage of the new potential 10-valent vaccine was higher as compared to the 4-valent vaccine and was actually doubled for non-ESBL-PEc bacteremia. The findings of the current study can be used for future studies and can help further evaluation and implementation of *E. coli* vaccines.

Strengths of the current study are the multicenter design and combination of epidemiological characteristics and highly discriminatory genetic data. There are also important limitations. Firstly, *E. coli* is a heterogeneous species, of which the seven MLST genes only constitute a small proportion of the entire gene content. Because we also only investigated presence of a small fraction of the genes that are commonly part of the accessory genome, such as virulence and acquired resistance genes, but did not assess the entire accessory gene pool, we could have missed genomic differences between isolates that are reflected in the accessory gene pool only. Secondly, we selected *E. coli* isolates from a tertiary care center and teaching hospital from the Netherlands from two different regions, which we considered to be representative of the Netherlands. The description of strains that were identified here might not be entirely generalizable to other countries since there could be differences between circulating *E. coli* strains, dependent on local population characteristics and antimicrobial resistance levels. Thirdly, many pairwise comparisons between subgroups were performed, which increases the risk of false-positive findings (i.e. type I errors). Even though we applied a strict p-value correction for multiple testing, this naturally does not eliminate the risk of false-positive findings. The analyses on resistance gene and VG content should therefore be viewed as hypothesis generating.

In conclusion, there are molecular differences between non-ESBL-PEc and ESBL-PEc blood isolates that reach beyond their phenotypic ESBL positivity. Future genomic research of *E. coli* should preferably focus on *E. coli* without preselection on ESBL-positivity, to limit the risk of inferring characteristics of resistant *E. coli* to the *E. coli* population as a whole. Furthermore, more studies are needed to better understand repeatedly found associations between gene content and STs, which could aid the development of targeted preventive interventions.

FUNDING

This study was internally funded by the Netherlands National Institute for Public Health and the Environment (RIVM) and the University Medical Center Utrecht (UMCU) and the authors received no specific funding for this work.

ACKNOWLEDGMENTS

We would sincerely like to thank Kim van der Zwaluw, Carlo Verhulst and Judith Vlooswijk for their contributions in the laboratory execution of the study. We would also like to thank Janetta Top for her contributions to the core genome phylogenetic analyses. Preliminary results of this paper were presented at the 29th European Congress of Clinical Microbiology & Infectious Diseases, Amsterdam, the Netherlands (P1450).

AUTHOR CONTRIBUTIONS

Author	Roles
D. van Hout	Conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing – original draft preparation, writing – review & editing
T.D. Verschuuren	Conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, writing – review & editing
P.C.J. Bruijning-Verhagen	Conceptualization, formal analysis, methodology, supervision, validation, writing – review & editing
T. Bosch	Conceptualization, data curation, formal analysis, methodology, investigation, methodology, funding acquisition, resources, software, supervision, writing – review & editing
A.C. Schürch	Methodology, software, resources, supervision, validation, writing – review & editing
R.J.L. Willems	Methodology, software, resources, supervision, writing – review & editing
M.J.M. Bonten	Conceptualization, formal analysis, funding acquisition, resources, supervision, writing – review & editing, writing – review & editing
J.A.J.W. Kluytmans	Conceptualization, formal analysis, resources, supervision, writing – review & editing

REFERENCES

1. de Kraker MEA, Jarlier V, Monen JCM, Heuer OE, van de Sande N, Grundmann H. The changing

epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. Clin Microbiol Infect. 2013;19: 860–868. doi:10.1111/1469-0691.12028

2. Van Der Steen M, Leenstra T, Kluytmans JAJW, Van Der Bij AK. Trends in expanded-spectrum cephalosporin-resistant Escherichia coli and Klebsiella pneumoniae among Dutch clinical isolates, from 2008 to 2012. PLoS One. 2015;10: e0138088. doi:10.1371/journal.pone.0138088

3. Schlackow I, Stoesser N, Walker AS, Crook DW, Peto TEA, Wyllie DH. Increasing incidence of Escherichia coli bacteraemia is driven by an increase in antibiotic-resistant isolates: electronic database study in Oxfordshire 1999-2011. J Antimicrob Chemother. 2012;67: 1514–1524. doi:10.1093/jac/dks082

4. Vihta K-D, Stoesser N, Llewelyn MJ, Quan TP, Davies T, Fawcett NJ, et al. Trends over time in Escherichia coli bloodstream infections, urinary tract infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998-2016: a study of electronic health records. Lancet Infect Dis. 2018;18: 1138–1149. doi:10.1016/S1473-3099(18)30353-0

5. Abernethy JK, Johnson AP, Guy R, Hinton N, Sheridan EA, Hope RJ. Thirty day all-cause mortality in patients with Escherichia coli bacteraemia in England. Clin Microbiol Infect. 2015;21: 251.e1-8. doi:10.1016/j.cmi.2015.01.001

6. Fitzpatrick JM, Biswas JS, Edgeworth JD, Islam J, Jenkins N, Judge R, et al. Gram-negative bacteraemia; a multi-centre prospective evaluation of empiric antibiotic therapy and outcome in English acute hospitals. Clin Microbiol Infect. 2016;22: 244–251. doi:10.1016/j.cmi.2015.10.034

7. Abernethy J, Guy R, Sheridan EA, Hopkins S, Kiernan M, Wilcox MH, et al. Epidemiology of Escherichia coli bacteraemia in England: results of an enhanced sentinel surveillance programme. J Hosp Infect. 2017;95: 365–375. doi:10.1016/j.jhin.2016.12.008

8. Huttner A, Hatz C, van den Dobbelsteen G, Abbanat D, Hornacek A, Frolich R, et al. Safety, immunogenicity, and preliminary clinical efficacy of a vaccine against extraintestinal pathogenic Escherichia coli in women with a history of recurrent urinary tract infection: a randomised, single-blind, placebo-controlled phase 1b trial. Lancet Infect Dis. 2017;17: 528–537. doi:10.1016/S1473-3099(17)30108-1

9. Overdevest ITMA, Bergmans AMC, Verweij JJ, Vissers J, Bax N, Snelders E, et al. Prevalence of phylogroups and O25/ST131 in susceptible and extended-spectrum β-lactamase-producing Escherichia coli isolates, the Netherlands. Clin Microbiol Infect. 2015;21: 570.e1-e4. doi:10.1016/j.cmi.2015.02.020

10. Zhou X, Garcia-Cobos S, Ruijs GJHM, Kampinga GA, Arends JP, Borst DM, et al. Epidemiology of Extended-Spectrum beta-lactamase-Producing E. coli and Vancomycin-Resistant Enterococci in the Northern Dutch-German Cross-Border Region. Front Microbiol. 2017;8: 1914. doi:10.3389/fmicb.2017.01914

11. Voor In 't Holt AF, Wattel AA, Boers SA, Jansen R, Hays JP, Goessens WHF, et al. Detection of Healthcare-Related Extended-Spectrum Beta-Lactamase-Producing Escherichia coli Transmission Events Using Combined Genetic and Phenotypic Epidemiology. PLoS One. 2016;11: e0160156. doi:10.1371/journal.pone.0160156

12. van Hout D, Verschuuren TD., Bruijning-Verhagen PCJ., Bosch T., Reuland EA., Fluit AC., et al. Design of the EPIGENEC Study: Assessing the EPIdemiology and GENetics of Escherichia coli in the Netherlands. Preprints. 2019; 2019020066. doi:10.20944/preprints201902.0066.v1

13. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29: 1072–1075. doi:10.1093/bioinformatics/btt086

14. Simpson EH. Measurement of diversity. Nature. 1949;163.

15. Argimon S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb genomics. 2016;2: e000093. doi:10.1099/mgen.0.000093

16. Joensen KG, Tetzschner AMM, 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: 2410–2426. doi:10.1128/JCM.00008-15

17. Geurtsen J, Weerdenburg E, Davies T, Go O, Spiessens B, Geet G Van, et al. Extraintestinal pathogenic Escherichia coli surveillance study to determine O-serotype prevalence and antibiotic resistance in blood isolates collected in Europe, 2011–2017. ECCMID Conf #P1451.

18. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 2017;45: D566–D573. doi:10.1093/nar/gkw1004

19. 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: 2640–2644. doi:10.1093/jac/dks261

20. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 2019;47: D687–D692. doi:10.1093/nar/gky1080

21. Johnson JR, Stell AL. Extended Virulence Genotypes of Escherichia coli Strains from Patients with Urosepsis in Relation to Phylogeny and Host Compromise. J Infect Dis. 2000;181: 261–72. doi:10.1086/315217

22. Johnson JR, Johnston BD, Porter S, Thuras P, Aziz M, Price LB. Accessory Traits and Phylogenetic Background Predict Escherichia coli Extraintestinal Virulence Better Than Does Ecological Source. J Infect Dis. 2019;219: 121–132. doi:10.1093/infdis/jiy459

23. Johnson JR, Porter S, Johnston B, Kuskowski MA, Spurbeck RR, Mobley HLT, et al. Host Characteristics and Bacterial Traits Predict Experimental Virulence for Escherichia coli Bloodstream Isolates From Patients With Urosepsis. Open Forum Infect Dis. 2015;2: ofv083. doi:10.1093/ofid/ofv083

24. Dale AP, Woodford N. Extra-intestinal pathogenic Escherichia coli (ExPEC): Disease, carriage and clones. J Infect. 2015;71: 615–626. doi:10.1016/j.jinf.2015.09.009

25. Dale AP, Pandey AK, Hesp RJ, Belogiannis K, Laver JR, Shone CC, et al. Genomes of Escherichia coli bacteraemia isolates originating from urinary tract foci contain more virulence-associated genes than those from non-urinary foci and neutropaenic hosts. J Infect. 2018;77: 534–543. doi:10.1016/j.jinf.2018.10.011

26. Leimbach A. ecoli_VF_collection: V.0.1. Zenodo. [Internet]. 2016. Available: http://dx.doi.org/10.5281/zenodo.56686.

27. Severiano A, Pinto FR, Ramirez M, Carrico JA. Adjusted Wallace coefficient as a measure of congruence between typing methods. J Clin Microbiol. 2011;49: 3997–4000. doi:10.1128/JCM.00624-11

28. Frenck RW, Ervin J, Chu L, Abbanat D, Spiessens B, Go O, et al. Safety and immunogenicity of a vaccine for extra-intestinal pathogenic Escherichia coli (ESTELLA): a phase 2 randomised controlled trial. Lancet Infect Dis. 2019;19: 631–640. doi:10.1016/S1473-3099(18)30803-X

29. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, et al. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res. 2017;27: 1437–

1449. doi:10.1101/gr.216606.116

30. Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. Escherichia coli ST131, an intriguing clonal group. Clin Microbiol Rev. 2014;27: 543–574. doi:10.1128/CMR.00125-13

31. McNally A, Oren Y, Kelly D, Pascoe B, Dunn S, Sreecharan T, et al. Combined Analysis of Variation in Core, Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. PLoS Genet. 2016;12: e1006280. doi:10.1371/journal.pgen.1006280

32. McNally A, Kallonen T, Connor C, Abudahab K, Aanensen DM, Horner C, et al. Diversification of Colonization Factors in a Multidrug-Resistant Escherichia coli Lineage Evolving under Negative Frequency-Dependent Selection. MBio. 2019;10: e00644-19. doi:10.1128/mBio.00644-19

33. Laupland KB, Lyytikäinen O, Sgaard M, Kennedy KJ, Knudsen JD, Ostergaard C, et al. The changing epidemiology of Staphylococcus aureus bloodstream infection: a multinational population-based surveillance study. Clin Microbiol Infect. 2013;19: 465–471. doi:10.1111/j.1469-0691.2012.03903.x

34. Mostofsky E, Lipsitch M, Regev-Yochay G. Is methicillin-resistant Staphylococcus aureus replacing methicillin-susceptible S. aureus? J Antimicrob Chemother. 2011;66: 2199–2214. doi:10.1093/jac/dkr278

35. Goswami C, Fox S, Holden M, Connor M, Leanord A, Evans TJ. Genetic analysis of invasive Escherichia coli in Scotland reveals determinants of healthcare-associated versus community-acquired infections. Microb genomics. 2018;4. doi:10.1099/mgen.0.000190