Journal of Antimicrobial Chemotherapy

Comparative genomics of ESBL-producing *Escherichia coli* (ESBL-Ec) reveals a similar distribution of the 10 most prevalent ESBL-Ec clones and ESBL genes among human community faecal and extra-intestinal infection isolates in the Netherlands (2014–17)

T. D. Verschuuren[1]*, D. van Hout[1], S. Arredondo-Alonso[2], A. C. Fluit[2], E. A. Reuland[2,3,4], J. Top[2], A. C. Schürch[2], T. Bosch[4], M. J. M. Bonten[1,2], J. A. J. W. Kluytmans[1] and R. J. L. Willems[2]

[1]Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, The Netherlands; [2]Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, The Netherlands; [3]Saltro Diagnostic Center for Primary Care, Utrecht, The Netherlands; [4]The Netherlands Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

*Corresponding author. E-mail: t.d.verschuuren@gmail.com

Introduction: The human gut microbiota is an important reservoir of ESBL-producing *Escherichia coli* (ESBL-Ec). Community surveillance studies of ESBL-Ec to monitor circulating clones and ESBL genes are logistically challenging and costly.

Objectives: To evaluate if isolates obtained in routine clinical practice can be used as an alternative to monitor the distribution of clones and ESBL genes circulating in the community.

Methods: WGS was performed on 451 Dutch ESBL-Ec isolates (2014–17), including 162 community faeces and 289 urine and blood isolates. We compared proportions of 10 most frequently identified STs, PopPUNK-based sequence clusters (SCs) and ESBL gene subtypes and the degree of similarity using Czekanowski's proportional similarity index (PSI).

Results: Nine out of 10 most prevalent STs and SCs and 8/10 most prevalent ESBL genes in clinical ESBL-Ec were also the most common types in community faeces. The proportions of ST131 (39% versus 23%) and SC131 (40% versus 25%) were higher in clinical isolates than in community faeces ($P < 0.01$). Within ST131, H30Rx (C2) subclade was more prevalent among clinical isolates (55% versus 26%, $P < 0.01$). The proportion of ESBL gene $bla_{CTX-M-1}$ was lower in clinical isolates (5% versus 18%, $P < 0.01$). Czekanowski's PSI confirmed that the differences in ESBL-Ec from community faeces and clinical isolates were limited.

Conclusions: Distributions of the 10 most prevalent clones and ESBL genes from ESBL-Ec community gut colonization and extra-intestinal infection overlapped in majority, indicating that isolates from routine clinical practice could be used to monitor ESBL-Ec clones and ESBL genes in the community.

Introduction

In Europe, the number of bloodstream infections with *Escherichia coli* is rising, mainly driven by an increase in community onset infections.[1,2] In Europe, *E. coli* is the most frequent cause of bloodstream and urinary tract infections, and an increasing proportion is caused by ESBL-producing *E. coli* (ESBL-Ec).[3]

The main human reservoir of ESBL-Ec is the gut of community dwelling individuals.[4–8] In the Netherlands, the prevalence of intestinal ESBL-Ec carriage in the open population is approximately 5%.[9–12] ST131 and ESBL genes of the bla_{CTX-M} type currently dominate this human ESBL-Ec reservoir, replacing the TEM and SHV gene variants that dominated in the 1990s.[4,9,13–16] With possible new variants likely to arise in time, molecular surveillance of the ESBL-Ec human reservoir is fundamental to track temporal changes and to allow early detection of important antibiotic-resistant strains.[4]

Previous surveillance studies in the Netherlands that assessed ESBL-Ec carriage in the community provided valuable insight into the prevalence and population structure of the human community ESBL-Ec reservoir.[9–12] Unfortunately, such studies are logistically challenging and costly and, therefore, not performed on a regular basis. Clinical isolates that are routinely obtained in primary or secondary healthcare settings could potentially serve as an alternative to monitor clones and ESBL genes in the community ESBL-Ec reservoir as proposed by Coque *et al.*[4] in 2008. Here, we determined the genomic relatedness of human community faecal and clinical ESBL-Ec isolates using WGS, in order to determine whether ESBL-Ec isolates obtained in routine clinical practice could be used to monitor the clones and ESBL genes in the community gut reservoir.

Methods

Study design and population

Sample collection was fully described previously and included: (i) faecal ESBL-Ec isolates that originated from a Dutch cross-sectional open-population study performed between 2014 and 2016 ($n = 162$), and (ii) clinical ESBL-Ec isolates ($n = 289$).[9,17] Clinical isolates were obtained from: (i) patients with community acquired (CA) urinary tract infection, prospectively collected in primary care in 2017 ($n = 175$); (ii) hospitalized patients with nosocomial urinary tract infection between 2014 and 2016 [further referred to as hospital-acquired (HA) urine isolates], retrospectively collected ($n = 49$); and (iii) hospitalized patients with a positive blood culture between 2014 and 2016, also retrospectively collected ($n = 65$). Participating centres were Saltro, a medical laboratory providing services to primary care practices, primarily in the Utrecht region, the University Medical Center Utrecht and the Amphia Hospital in Breda. Only the first available isolate per patient was included in the current study [Figure S1 and Table S1 (available as Supplementary data at *JAC* Online)].

Ethics

Individual informed consent was given by subjects participating in the surveillance study providing community faeces isolates (IRB number 14/219-C).[9] For the use of clinical isolates the ethics review board of the University Medical Center Utrecht judged this study to be outside the scope of the Medical Research Involving Human Subjects Act and waived the need for official approval (IRB number 18/056). Based on the 'Code of conduct for health research' informed consent was not obtained.[18]

Genotyping

WGS was performed on all isolates using Illumina HiSeq 2500, MiSeq, or NextSeq platforms. All generated raw reads are available in the European Nucleotide Archive (ENA) of the European Bioinformatics Institute (EBI) under the study accession numbers PRJEB35000 and PRJEB40007. *De novo* assembly was performed using SPAdes (v3.6.2).[19] The quality of the assemblies was assessed using Quality Assessment Tool for Genome Assemblies (QUAST), using default settings.[20] STs were inferred *in silico* with MLST (v2.0) using the Achtman scheme with tseemann/mlst (v2.15.1) (https://github.com/tseemann/mlst).[21] Acquired ESBL genes were determined with a search against the ResFinder database (v3.1.0) using a minimal length of 80% and a minimal identity of 95% as cut-offs, using abricate (version 0.8.7) (https://github.com/tseemann/abricate).[22] If the minimal length or identity of an acquired ESBL gene was above the before-mentioned threshold, but below 100%, we repeated the search with ResFinder (v3.2 webserver) using raw reads to confirm the ESBL gene type. ST131-clades were based on *fimH*-type, defined with FimTyper (version 1.0).[23] As described

previously, isolates with *fimH41* are grouped to ST131-clade A, with *fimH22* grouped to ST131-clade B and with *fimH30* grouped to ST131-clade C. H30R (clade C1) is defined based on the presence of fluoroquinolone resistance (combined *gyrA/parC* mutations defined with ResFinder), while H30Rx (clade C2) is defined based on fluoroquinolone resistance and presence of ESBL gene $bla_{CTX-M-15}$.[24,25] Other observed *fimH* types were classified as either ST131-clade A, B or C(1)(2) based on their fluoroquinolone resistance, ESBL gene and position in the phylogenetic tree (Figure 1). Isolates in clade C1 carrying $bla_{CTX-M-27}$ are described separately (Table 1).[26]

Phylogeny and partitioning in whole-genome-based sequence clusters (SCs)

Genomic relatedness of ESBL-Ec isolates was determined using PopPUNK (v1.1.3), using default parameters. PopPUNK calculated a relative core and accessory distance for each pair in the dataset based on k-mer comparisons.[27] The distance matrix produced by PopPUNK was used to infer phylogeny and assign strains or sequence clusters, in this article further referred to as SCs, representing sets of isolates similar in both their core and accessory genomes relative to the rest of the population.[27] SCs were named after the most prevalent ST within the cluster. The adjusted Rand index was used to calculate the congruence between STs and SCs, where identical population partitioning was one, and completely different population partitioning was zero.[28] A core genome neighbour-joining (NJ) tree was constructed with PopPUNK and an accessory genome NJ tree was constructed using fastcluster (v1.1.25) in R, with the distance matrix produced by PopPUNK. All trees were visualized with microreact (v5.111.0).[29]

Statistical analysis

Proportions with 95% CI of the 10 most occurring STs, SCs, ESBL genes and most common ST131 clades were compared between community faecal and clinical isolates, using a two-proportion *z*-test. Czekanowski's proportional similarity index (PSI) was used to calculate the aggregate proportion of overlap between community faeces and clinical isolates. The PSI was calculated by: $PSI = 1 - 0.5 \sum k |p(faeces)k - p(clinical)k|$, where p was the proportion of a observed subtype within ST- (e.g. ST131..STn), SC- (e.g. SC131..SCn), or ESBL gene-level (e.g. $bla_{CTX-M-15}..bla_n$), respectively. Ninety-five percent CIs were calculated using 5000 bootstrap iterations.[30–32] The observed PSIs were tested against the expected PSIs under the null hypothesis that there was no difference between community faecal and clinical isolates using a permutation test.[32–34] Isolates were randomly relabelled as having a faecal or clinical source, creating a permutation distribution from 5000 iterations. In the simulated permutation distribution, ST, SC and ESBL gene assignment was independent from sample origin. The *P* value was the probability of the observed PSI (PSI_{obs}) under the null hypothesis. A *P* value < 0.05 was considered statistically significant.

Subgroup analyses were performed for the different types of clinical samples (CA-urine, HA-urine and blood) to explore if a certain sample group could be used as proxy for molecular surveillance. Furthermore, a *post hoc* analysis, repeating the analysis without ST131 was performed to assess if ST131 was the sole explaining factor for the observed difference between community faecal and clinical isolates.

All calculations were performed in RStudio Version 1.1.456.[35]

Results

Distribution of STs

In total, 108 different STs were identified among the 451 isolates. The three most common STs were ST131 (34%), ST38 (10%) and ST1193 (4%) (Figure S2a). Twenty-six STs were present in both community faecal and clinical isolates, together these accounted for 82% of all isolates. The remaining STs consisted mostly of
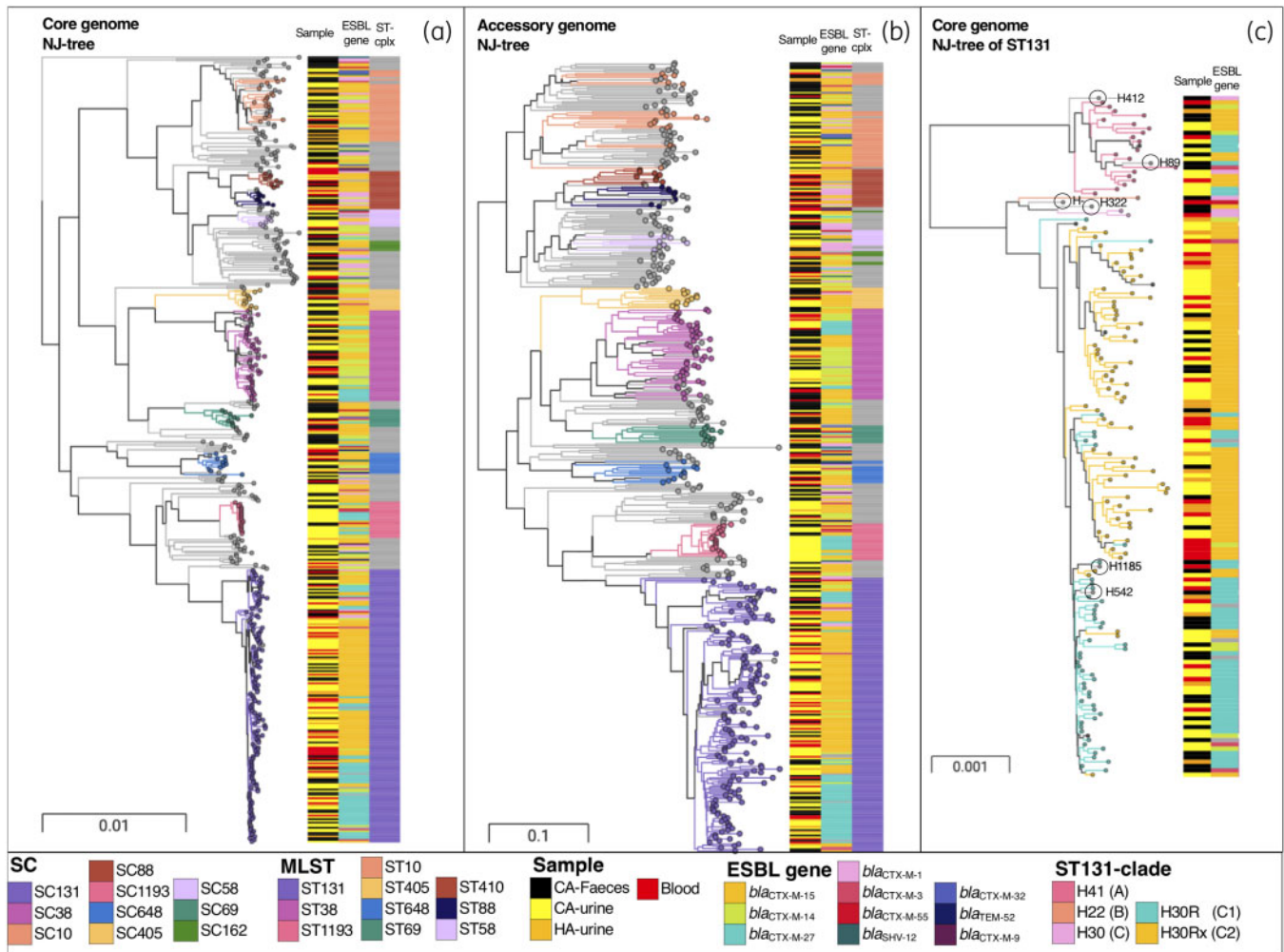
Figure 1. Neighbour-joining trees. (a) Core genome – nodes coloured according to ST (10 most frequent), (b) accessory genome – nodes coloured according to ST (10 most frequent) and (c) core genome ST131 – nodes coloured according to clade (nodes with a singleton *fimH* type indicated separately). Constructed with PopPUNK. Online view core tree (https://microreact.org/project/Vmycsy2gY/938965ce), accessory tree (https://microreact.org/project/f9Iums0yo/11f42f48), core genome ST131-subtree (https://microreact.org/project/Vmycsy2gY/107f9879). Sample: community urine (CA), nosocomial urine (HA).

singletons (Table S2). Of the 10 most frequently occurring STs in clinical isolates, nine also belonged to the 10 most frequently occurring STs in community faeces isolates (Figure 2a). Only ST95 was not found among community faeces isolates. These top 10 STs represent 54% of faecal isolates and 67% of clinical isolates. ST131 was significantly more often observed in clinical isolates (39%) than community faecal isolates (23%, $P < 0.01$) (Figure 2a). Among ST131 isolates, 48% represented the H30Rx (clade C2) type and were significantly more frequently observed among clinical isolates (55% versus 26%, $P < 0.01$) (Table 1).

Distribution of SCs
Seventy-five different SCs were assigned by PopPUNK, which were congruent with ST assignment (adjusted Rand index 0.93). The three most prevalent SCs corresponded to SC131 (35%), SC38 (12%) and SC10 (8%). Twenty-four SCs were found both in community faecal and clinical isolates, accounting for 82% of

community faecal and 95% of clinical isolates (Table S3). Of the 10 most frequently occurring SCs in clinical isolates, nine also belonged to the 10 most frequently occurring SCs in community faeces isolates, representing 64% of faecal isolates (Figure 2b). The exception was SC95 ($n = 5$, clinical isolates). SC131 was significantly more frequently observed among clinical isolates (40%) than in community faecal isolates (25%) (Figure 2b).

Distribution of ESBL genes
In total, 453 ESBL genes were identified, representing 16 different ESBL genes, of which 97%, belonged to 12 variants of the bla_{CTX-M} family. The remaining four genes were bla_{SHV-12} (1.8%, $n = 8$), bla_{TEM-52} (0.7%, $n = 3$) and two genes belonging to the bla_{TEM} family (0.4%, $n = 2$). The three most prevalent ESBL genes were $bla_{CTX-M-15}$ (48%), $bla_{CTX-M-14}$ (17%) and $bla_{CTX-M-27}$ (16%). Nine ESBL genes were found in both community faecal and clinical isolates, accounting for 98% of all isolates (Table S4). Two urine

Table 1. Subdivision of ST131 in subclades

$N = 152$	Clinical, $N = 114$, n (%)	Faeces, $N = 38$, n (%)	P value[a]
H41[b] (A), ($N = 20$)	14 (12)	6 (16)	0.8
H22[c] (B), ($N = 1$)	0 (0)	1 (3)	0.56
H30[d] (C), ($N = 125$)	98 (86)	27 (71)	0.07
H30 (C), ($N = 2$)	1 (1)	1 (3)	1
H30R (C1), ($N = 50$)	34 (30)	16 (42)	0.23
H30R (C1 with $bla_{CTX-M-27}$), ($n = 39$)	26 (23)	13 (34)	0.24
H30Rx (C2), ($N = 73$)	63 (55)	10 (26)	<0.01
Other[e]			
H412 (A)	0 (0)	1 (3)	1
H89 (A)	0 (0)	1 (3)	1
H322 (B/C)	0 (0)	1 (3)	1
H1185 (C1)	1 (0.9)	0 (0)	1
H542 (C1)	0 (0)	1 (3)	1
no *fimH* (B/C), ($N = 1$)	1 (0.9)	0 (0)	1

[a]P value derived from χ^2 statistic.
[b]*fimH41* corresponds with clade A.
[c]*fimH22* corresponds with clade B.
[d]*fimH30* corresponds with clade C. H30 subgroups: H30, no combined *gyrA/parC* point mutations; H30R (clade C1), combined *gyrA/parC* point mutations and ESBL gene other than $bla_{CTX-M-15}$; H30R (C1 with $bla_{CTX-M-27}$) is described as a subgroup of H30R (C1) group ($N = 50$); H30Rx (clade C2), combined *gyrA/parC* point mutations and ESBL gene $bla_{CTX-M-15}$.
[e]Other *fimH* types: *fimH* type singletons are classified based on fluoroquinolone resistance, presence of ESBL gene type and position in phylogenetic tree (Figure 1). *fimH89*, fluoroquinolone susceptible, presence of $bla_{CTX-M-27}$, assigned to clade A; *fimH322*, fluoroquinolone susceptible, presence of $bla_{CTX-M-15}$, assigned to clade B/C; *fimH412*, fluoroquinolone resistant, presence of $bla_{CTX-M-1}$, assigned to clade A; *fimH1185*, fluoroquinolone resistant, presence of $bla_{CTX-M-27}$, assigned to clade C1; *fimH542*, fluoroquinolone resistant, presence of $bla_{CTX-M-27}$, assigned to clade C1; no *fimH* type, fluoroquinolone susceptible, presence of $bla_{CTX-M-55}$, assigned to clade B/C.[23,24]

isolates each harboured two ESBL genes ($bla_{CTX-M-27}$ and $bla_{CTX-M-55}$). Of the 10 most frequently occurring ESBL genes in clinical isolates, eight also belonged to the most frequently occurring ESBL genes in community faeces, representing 98% of faecal isolates (Figure 2c). The exceptions were $bla_{CTX-M-9}$ (three clinical isolates) and $bla_{CTX-M-2}$ (two clinical isolates) (Figure 2c). The prevalence of $bla_{CTX-M-1}$ was lower in clinical (5%) than in community faecal isolates (18%, $P < 0.01$) (Figure 2c).

Core and accessory phylogenies
To determine the genomic relatedness of ESBL-Ec community faecal and clinical isolates, a core genome and accessory genome NJ tree was constructed based on core and accessory genome distance matrices generated by PopPUNK (Figure 1). Both core genome and accessory genome-based trees showed that the community faecal and clinical ESBL-Ec populations were diverse with no distinct clustering of community faecal and clinical isolates. This indicated that in this dataset the faecal and clinical

ESBL-Ec isolates did not constitute two distinct subpopulations based on evolutionary origin or genetic repertoire. Of the four most prevalent ESBL genes, $bla_{CTX-M-15}$ was observed throughout the trees, while $bla_{CTX-M-27}$ concentrated in ST131 (SC131), ST38 (SC38) and ST1193 (SC1193). The $bla_{CTX-M-14}$ gene concentrated mostly in ST38 (SC38), and $bla_{CTX-M-1}$ concentrated in ST58 (SC58) and ST88 (SC88). We generated a subtree based on the core genome of only ST131 isolates (Figure 1c) and plotted the *fimH* subclade typing. A separate clustering was observed for H41 (clade A), H22 (clade B) and H30 (clade C). However, H30R (clade C1) and H30Rx (clade C2) occurred alternatingly. Within H30R (clade C1), the majority of isolates carried $bla_{CTX-M-27}$. Among H30Rx, community faecal isolates were underrepresented.

PSI
To quantify the degree of similarity between the frequency distributions of STs, SCs and ESBL genes among community faeces and clinical isolates, the PSI was calculated, which is interpreted as a proportion of overlap between the two sample groups. For the frequency distributions of STs, the PSI_{obs} was 0.55, while the expected PSI (PSI_{exp}) under the null hypothesis was 0.67 ($P < 0.01$) (Figure 3). A similar result was seen for frequency distributions of SCs (PSI_{obs} 0.68 versus PSI_{exp} 0.76; $P < 0.01$) and ESBL genes (PSI_{obs} 0.81 versus PSI_{exp} 0.91; $P < 0.01$), respectively (Figure 3). These PSI differences of 0.12 for STs, 0.08 for SCs and 0.10 for ESBL genes can be interpreted as limited within the possible range of 0.00–1.00.

Subgroup and sensitivity analyses
Subgroup analyses of the clinical samples separately revealed that the similarity of community faeces and community urine was equal to that of community faeces and all clinical isolates combined (PSI_{obs} ST 0.54; SC 0.65; ESBL gene 0.78) (Figure 4 and Figures S3–S5). The sensitivity analysis excluding ST131 from the dataset did not eliminate the observed difference in frequency distributions of STs, SCs and ESBL genes, expressed as PSI, between human faecal and clinical ESBL-Ec isolates (Figures S6–S9).

Discussion
In this study, we used WGS on 451 Dutch ESBL-Ec isolates to assess the degree of similarity of human community faecal and human clinical isolates. The distribution of the 10 most frequently found STs, SCs and ESBL genes for the two groups was very similar. Nine out of 10 most prevalent STs and SCs in clinical isolates were also the most common types in community faeces. These nine STs made up more than half of all community faeces isolates. Eight out of the 10 most prevalent ESBL genes in clinical isolates were also the most common types in community faeces, which represented virtually all (98%) community faeces isolates. Furthermore, phylogenetic inferences did not reveal distinct clustering based on sample group.

The absence of distinct phylogenetic clustering of *E. coli* isolates based on source group is in line with earlier research,[36,37] as well as the observed overlap of the 10 most common STs and ESBL genes in community faeces and clinical isolates in ESBL-Ec.[6,8] Also, the observed higher prevalence of ST131, particularly the higher prevalence of ST131 clade C2 H30Rx (55% versus 26% of ST131 isolates)
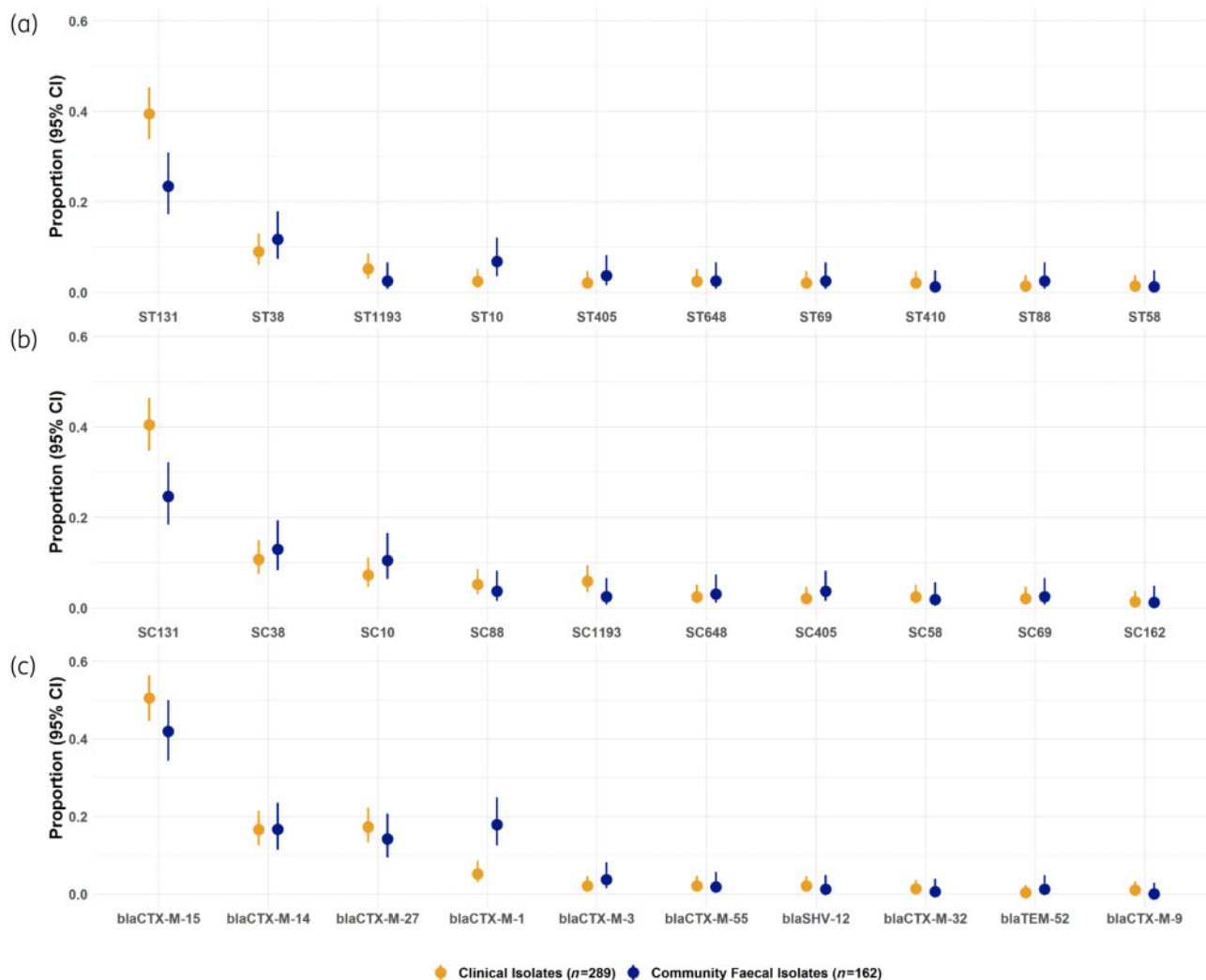
Figure 2. Proportions of the 10 most frequent genetic subtypes in clinical isolates for (a) STs, (b) SCs and (c) ESBL gene types. Proportions of ST131, SC131 and $bla_{CTX-M-1}$ differed between clinical and community faecal isolates; $P < 0.01$. P value derived from χ^2 statistic. All other proportions did not differ significantly. This figure appears in colour in the online version of *JAC* and in black and white in the print version of *JAC*.

in clinical isolates is in line with earlier findings.[6,8,11,38] It has been postulated that due to multiple evolutionary events, such as acquisition of adaptive elements, ST131 has greater pathogenic potential than other STs.[14,16,24,25,39–43] Our study shows that ST131 was indeed more prevalent among extra-intestinal infection, but that this ST was also the dominant ST in community faecal carriage of ESBL-Ec. A recent epidemiological surveillance study by Day *et al.*[8] in the UK also found a higher prevalence of ST131 in ESBL-Ec blood isolates (64%) compared with ESBL-Ec from faeces (36%). Notably, the absolute prevalence of ST131 in faecal isolates was considerably higher than in our study, which is possibly related to differences in the local epidemiology and sample collection. Faecal isolates included in the study by Day *et al.*[8] were recovered from faeces samples that were collected for specific diagnostic purposes, such as occult blood screening (a screening method for colon cancer), or the detection of intestinal pathogens, while in our study, a random sample of the Dutch open population was invited to provide a faecal sample. A lower prevalence of $bla_{CTX-M-1}$ in

clinical isolates is also in line with earlier research. The $bla_{CTX-M-1}$ gene was previously described as an important ESBL gene in intestinal carriage and non-human reservoirs.[6,8,10,11,44–46] Taking this into account, we hypothesize that this ESBL gene is more often accompanied by strains of lower virulence for humans. For our study this implies that the observed higher prevalence of ST131 and SC131, the lower prevalence of $bla_{CTX-M-1}$, and subtle differences in other clone/ESBL gene type distributions, expressed in the PSI, could be the reflection of a relatively higher prevalence of certain, possibly more virulent, strains in our clinical sample collection (Table S1).

To our knowledge, this is the most in depth comparative genomic assessment of ESBL-Ec found in community gut colonization and extra-intestinal infection to date, in a set of samples taken from a confined geographical region (the Netherlands) and from the same time period (2014–17). We used the Czekanowki's PSI to quantify the degree of similarity based on the distribution of frequencies of STs, core and accessory genome-based SCs, and ESBL
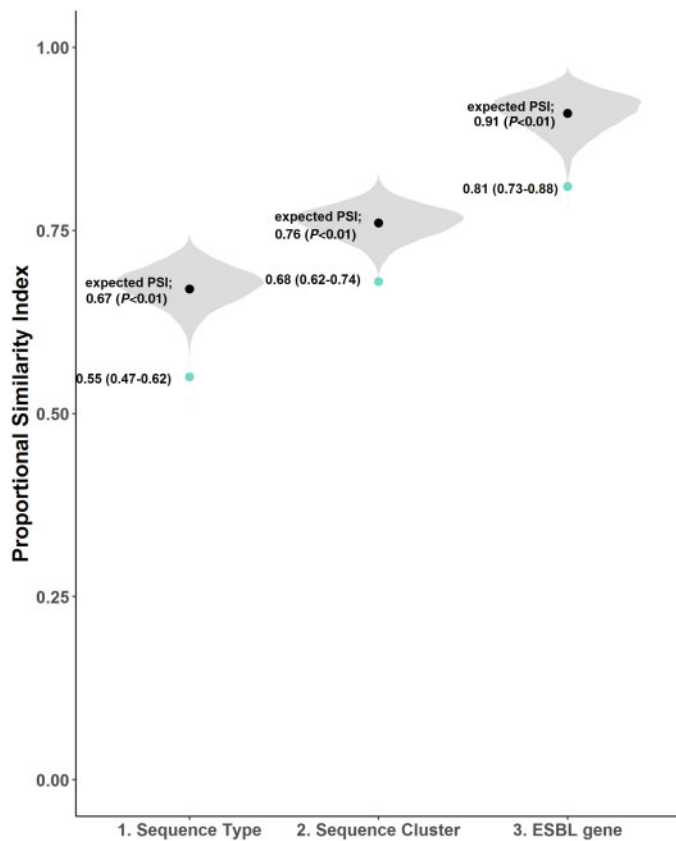
Figure 3. Mean observed PSI (PSI_{obs}), interpreted as the proportion of overlap between community faeces with clinical isolates; PSI_{obs} (green) with 95% CI calculated with 5000 bootstrap iterations. Permutation distribution with mean expected PSI (PSI_{exp}) under the null hypothesis (i.e. no difference between groups); PSI_{exp} (grey), calculated with 5000 permutations. *P* value permutation test; chance of the observed PSI under the null hypothesis (i.e. no difference between community faeces and clinical isolates). This figure appears in colour in the online version of *JAC* and in black and white in the print version of *JAC*.

Figure 4. Mean (95% CI) PSI, interpreted as the aggregate proportion of overlap between community faeces and clinical subgroups on ST-, SC- and ESBL gene-level. Each cell is coloured according to the PSI level, with a colour gradient from 0 (light) to 1 (dark). This figure appears in colour in the online version of *JAC* and in black and white in the print version of *JAC*.

genes between community gut colonization and extra-intestinal infection of ESBL-Ec. This measure, originating from ecology, has been used to express similarity between populations of several bacterial species.[31,47–49] This analysis revealed that the genomic make-up of community and clinical isolates did not entirely overlap; however, the difference in PSI_{obs} and PSI_{exp} ranged from 0.08 to 0.12, which could be interpreted as limited. Dorado-García *et al.*[31] used the PSI to assess overlap in different reservoirs for ESBL genes and plasmid replicon types. That study found an PSI (human general population versus extra-intestinal infection) of 0.7 for ESBL genes (which was comparable to the PSI of 0.8 found in our study). While, for example, the PSI for ESBL genes in the human general population versus chicken meat at the slaughterhouse was 0.3. In a *post hoc* analysis we excluded the hypothesis that the difference in proportion of ST131 among faecal and clinical isolates was the sole factor that contributed to the observed difference in genomic make-up between the two ESBL-Ec sample groups. It was a deliberate choice not to include plasmid replicon types in this article, as comparing only replicon types would not reveal the full degree of

similarity of complete ESBL gene carrying plasmids between the two ESBL-Ec populations.

Due to the cross-sectional nature of this study, our sample collection did not allow an analysis of temporal changes in circulating clones among community faecal and clinical ESBL-Ec. Furthermore, the subgroup sizes limited the subgroup analyses of the different sample types; in particular the number of ESBL-Ec isolates from nosocomial urine was small, leading to low precision. However, primary care urine was found to be equally similar to community faeces as all clinical isolates combined, indicating that primary care urine alone could be a good source for molecular surveillance. Culture indications may vary per country, particularly urine cultures in primary care. This may limit generalizability of our results outside the Netherlands. Furthermore, we did not have information on what proportion of primary care urines was healthcare associated, e.g. from patients with a recent hospitalization. All in all, the findings in this study indicate that primary care urine, nosocomial urine and blood collected in routine clinical practice provided a reliable overview of the most common circulating clones and ESBL genes within the human community ESBL-Ec reservoir. We propose molecular surveillance of the human ESBL-Ec reservoir to be implemented in the following way: (i) continuous monitoring of trends of the most frequent clones and ESBL genes using primary care urine isolates, and (ii) when the results from primary care urine demonstrate large shifts in clonal/ESBL gene distribution, proceed with conducting a point prevalence measurement of community colonization to assess clone/ESBL gene distribution and ESBL-Ec prevalence of community gut colonization, and to confirm the findings in primary care urine.

To conclude, our findings indicate that in the Netherlands the distribution of the 10 most prevalent clones and ESBL genes in community gut colonization and extra-intestinal disease causing ESBL-Ec are predominantly the same. Based on this, we postulate that clinical isolates collected in routine practice are suitable to monitor the most important clones and ESBL genes in the ESBL-Ec reservoir in the human community.

Transparency declarations

None to declare.

Author contributions

T.D.V.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing – original draft preparation, writing – review and editing. D.v.H.: conceptualization, data curation, investigation, methodology, project administration, software, validation, writing – review and editing. S.A.-A.: formal analysis, writing – review and editing. A.C.F.: data curation, validation, writing – review and editing. E.A.R.: conceptualization, data curation, validation, writing – review and editing. J.T.: formal analysis, writing – review and editing. A.C.S.: conceptualization, data curation, formal analysis, investigation, methodology, software, supervision, validation, writing – review and editing. T.B.: conceptualization, data curation, formal analysis, methodology, investigation, methodology, funding acquisition, resources, software, supervision, writing – review and editing. M.J.M.B.: conceptualization, formal analysis, methodology, investigation, methodology, funding acquisition, resources, supervision, writing – review and editing. J.A.J.W.K.: conceptualization, formal analysis, methodology, investigation, methodology, resources, supervision, writing – review and editing. R.J.L.W.: conceptualization, formal analysis, methodology, investigation, methodology, funding acquisition, software, supervision, writing – review and editing.

Supplementary data

Tables S1 to S4 and Figures S1 to S9 are available as Supplementary data at *JAC* Online.

References

1 PHE 2019. Annual epidemiological commentary: bacteraemia, MSSA bacteraemia and C. difficile infections, up to and including financial year April 2018 to March 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/843870/Annual_epidemiological_commentary_April_2018-March_2019.pdf.

2 Vihta K, Stoesser N, Llewelyn MJ *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**: 1138–49.

3 ECDC 2019. Surveillance of antimicrobial resistance in Europe 2018. https://www.ecdc.europa.eu/sites/default/files/documents/surveillance-antimicrobial-resistance-Europe-2018.pdf.

4 Coque TM, Baquero F, Canton R. Increasing prevalence of ESBL-producing Enterobacteriaceae in Europe. *Euro Surveill* 2008; **13**: pii=19044.

5 Rodríguez-Baño J, López-Cerero L, Navarro MD *et al*. Faecal carriage of extended-spectrum β-lactamase-producing *Escherichia coli*: prevalence, risk factors and molecular epidemiology. *J Antimicrob Chemother* 2008; **62**: 1142–9.

6 Mughini-Gras L, Dorado-García A, van Duijkeren E *et al*. Attributable sources of community-acquired carriage of *Escherichia coli* containing β-lactam antibiotic resistance genes: a population-based modelling study. *Lancet Planet Health* 2019; **3**: e357–69.

7 Isendahl J, Giske CG, Hammar U *et al*. Temporal dynamics and risk factors for bloodstream infection with extended-spectrum β-lactamase-producing bacteria in previously-colonized individuals: national population-based cohort study. *Clin Infect Dis* 2019; **68**: 641–9.

8 Day MJ, Hopkins KL, Wareham DW *et al*. Extended-spectrum β-lactamase-producing *Escherichia coli* in human-derived and foodchain-derived samples from England, Wales, and Scotland: an epidemiological surveillance and typing study. *Lancet Infect Dis* 2019; **19**: 1325–35.

9 van den Bunt G, van Pelt W, Hidalgo L *et al*. Prevalence, risk factors and genetic characterisation of extended-spectrum β-lactamase and carbapenemase-producing Enterobacteriaceae (ESBL-E and CPE): a community-based cross-sectional study, the Netherlands, 2014 to 2016. *Euro Surveill* 2019; **24**: pii=1800594.

10 Wielders CCH, van Hoek AHAM, Hengeveld PD *et al*. Extended-spectrum β-lactamase- and pAmpC-producing Enterobacteriaceae among the general population in a livestock-dense area. *Clin Microbiol Infect* 2017; **23**: e1–120.

11 van Hoek AHAM, Schouls L, van Santen MG *et al*. Molecular characteristics of extended-spectrum cephalosporin-resistant Enterobacteriaceae from humans in the community. *PLoS One* 2015; **10**: 1–12.

12 Reuland EA, Overdevest ITMA, Al Naiemi N *et al*. High prevalence of ESBL-producing Enterobacteriaceae carriage in Dutch community patients with gastrointestinal complaints. *Clin Microbiol Infect* 2013; **19**: 542–9.

13 Livermore DM, Canton R, Gniadkowski M *et al*. CTX-M: changing the face of ESBLs in Europe. *J Antimicrob Chemother* 2007; **59**: 165–74.

14 Kallonen T, Brodrick HJ, Harris SR *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**: 1437–49.

15 van der Bij AK, Peirano G, Goessens WH *et al*. Clinical and molecular characteristics of extended-spectrum-β-lactamase-producing *Escherichia coli* causing bacteraemia in the Rotterdam area, the Netherlands. *Antimicrob Agents Chemother* 2011; **55**: 3576–8.

16 Manges AR, Geum HM, Guo A *et al*. Global Extraintestinal Pathogenic *Escherichia coli* (ExPEC) Lineages. *Clin Microbiol Rev* 2019; **32**: e00135–18.

17 van Hout D, Verschuuren TD, Bruijning-Verhagen PCJ *et al*. Design of the EPIGENEC study: assessing the EPIdemiology and GENetics of *Escherichia coli* in the Netherlands. *Preprints* 2019; doi:10.20944/preprints201902.0066.v1.

18 Federation of Medical Scientific Societies 2017. FMWV Code of Conduct for Health Research. https://www.federa.org/sites/default/files/bijlagen/coreon/code_of_conduct_for_medical_research_1.pdf.

19 Nurk S, Bankevich A, Antipov D *et al*. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 2013; **20**: 714–37.

20 Gurevich A, Saveliev V, Vyahhi N *et al*. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013; **29**: 1072–5.

21 Larsen MV, Cosentino S, Rasmussen S *et al*. Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *J Clin Microbiol* 2012; **50**: 1355–61.

22 Zankari E, Hasman H, Cosentino S *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012; **67**: 2640–4.

23 Camacho C, Coulouris G, Avagyan V *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* 2009; **10**: 1–9.

24 Johnson JR, Tchesnokova V, Johnston B *et al.* Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli. J Infect Dis* 2013; **207**: 919–28.

25 Banerjee R, Johnson JR. A new clone sweeps clean: the enigmatic emergence of *Escherichia coli* sequence type 131. *Antimicrob Agents Chemother* 2014; **58**: 4997–5004.

26 Matsumura Y, Pitout JDD, Gomi R *et al.* Global *Escherichia coli* Sequence Type 131 clade with blaCTX-M-27 gene. *Emerg Infect Dis* 2016; **22**: 1900–7.

27 Lees JA, Harris SR, Tonkin-Hill G *et al.* Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 2019; **29**: 304–16.

28 Hubert L, Arabie P. Comparing partitions. *J Classif* 1985; **218**: 193–218.

29 Argimón S, Abudahab K, Goater RJE *et al.* Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016; **30**: 1–11.

30 Bloom SA. Similarity indices in community studies: potential pitfalls. *Mar Ecol Prog Ser* 1981; **5**: 125–8.

31 Dorado-García A, Smid JH, van Pelt W *et al.* Molecular relatedness of ESBL/AmpC-producing *Escherichia coli* from humans, animals, food and the environment: a pooled analysis. *J Antimicrob Chemother* 2018; **73**: 339–47.

32 Wolda H. Similarity indices, sample size and diversity. *Oecologia* 1981; **50**: 296–302.

33 Berger VW. Pros and cons of permutation tests in clinical trials. *Stat Med* 2000; **19**: 1319–28.

34 Proschan MA, Dodd LE. Re-randomization tests in clinical trials. *Stat Med* 2019; **38**: 2292–302.

35 RStudio Team 2020. *RStudio: Integrated Development for R.* RStudio, PBC, Boston, MA. http://www.rstudio.com/.

36 Nielsen KL, Stegger M, Kiil K *et al.* Whole-genome comparison of urinary pathogenic *Escherichia coli* and faecal isolates of UTI patients and healthy controls. *Int J Med Microbiol* 2017; **307**: 497–507.

37 Subbiah M, Caudell MA, Mair C *et al.* Antimicrobial resistant enteric bacteria are widely distributed amongst people, animals and the environment in Tanzania. *Nat Commun* 2020; **11**: 1–12.

38 Ny S, Löfmark S, Börjesson S *et al.* Community carriage of ESBL-producing *Escherichia coli* is associated with strains of low pathogenicity: A Swedish nationwide study. *J Antimicrob Chemother* 2017; **72**: 582–8.

39 McNally A, Kallonen T, Connor 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.

40 Price LB, Johnson JR, Aziz M *et al.* The epidemic of extended-spectrum-β-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio* 2013; **4**: e00377–13.

41 Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM *et al.* Sequential Acquisition of Virulence and Fluoroquinolone Resistance Has Shaped the Evolution of *Escherichia coli* ST131. *MBio* 2016; **7**: e00347–16.

42 Stoesser N, Sheppard AE, Pankhurst L *et al.* Evolutionary history of the global emergence of the *Escherichia coli* Epidemic Clone ST131. *MBio* 2016; **7**: e02162–15.

43 Johnson JR, Johnston B, Clabots C *et al. Escherichia coli* Sequence Type ST131 as the Major Cause of Serious Multidrug-Resistant *E. coli* Infections in the United States. *Clin Infect Dis* 2010; **51**: 286–94.

44 Kluytmans JAJW, Overdevest ITMA, Willemsen I *et al.* Extended-Spectrum β-Lactamase–Producing *Escherichia coli* From Retail Chicken Meat and Humans: comparison of Strains, Plasmids, Resistance Genes, and Virulence Factors. *Clin Infect Dis* 2013; **56**: 478–87.

45 Leverstein-van Hall MA, Dierikx CM, Cohen Stuart J *et al.* Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. *Clin Microbiol Infect* 2011; **17**: 873–80.

46 Overdevest I, Willemsen I, Rijnsburger M *et al.* Extended-spectrum β-lactamase genes of *Escherichia coli* in chicken meat and humans, the Netherlands. *Emerg Infect Dis* 2011; **17**: 1216–22.

47 Vossenkuhl B, Brandt J, Fetsch A *et al.* Comparison of spa Types, SCCmec types and antimicrobial resistance profiles of MRSA isolated from Turkeys at farm, slaughter and from retail meat indicates transmission along the production chain. *PLoS One* 2014; **9**: e96308–9.

48 Ramonaite S, Tamuleviciene E, Alter T *et al.* MLST genotypes of *Campylobacter jejuni* isolated from broiler products, dairy cattle and human campylobacteriosis cases in Lithuania. *BMC Infect Dis* 2017; **17**: 1–10.

49 Wieczorek K, Wołkowicz T, Osek J. MLST-based genetic relatedness of *Campylobacter jejuni* isolated from chickens and humans in Poland. *PLoS One* 2020; **15**: 1–13.