Practical applications of whole genome sequencing for reservoir epidemiology, molecular surveillance, and antimicrobial susceptibility testing of extended-spectrum beta-lactamase-producing *Escherichia coli*

Tess D Verschuuren

Practical applications of whole genome sequencing for reservoir epidemiology, molecular surveillance, and antimicrobial susceptibility testing of extended-spectrum ß-lactamase-producing *Escherichia coli*

Tess D Verschuuren

Practical applications of whole genome sequencing for reservoir epidemiology, molecular surveillance, and antimicrobial susceptibility testing of extended-spectrum ß-lactamase-producing *Escherichia coli*

PhD thesis, Utrecht University, the Netherlands

ISBN:	978-94-6416-774-0
DOI:	https://doi.org/10.33540/779
Author:	Tess Verschuuren
Cover design:	takeadetour.eu
Layout:	Tess Verschuuren
Printing:	Ridderprint | www.ridderprint.nl

Part of the research described in this thesis was financially supported by the Netherlands Organisation for Health Research and Development (ZonMw, Projectnumber: 547001004). Printing of this thesis was supported by the Julius Center for Health Sciences and Primary Care.

Practical applications of whole genome sequencing for reservoir epidemiology, molecular surveillance, and antimicrobial susceptibility testing of extended-spectrum ß-lactamase-producing *Escherichia coli*

Praktische toepassingen van whole genome sequencing voor reservoir epidemiologie, moleculaire surveillance en antimicrobiële gevoeligheidsbepalingen van extended-spectrum ß-lactamase-producerende *Escherichia coli*

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 4 november 2021 des ochtends te 10.15 uur

door

Tess Daniëlle Verschuuren

geboren op 23 mei 1990 te Erp

Promotoren:
Prof. dr. J.A.J.W. Kluytmans
Prof. dr. R.J.L. Willems

Copromotor:
Dr. T. Bosch

-De natuurlijke wereld wemelt van organische krachten die onophoudelijk aan het werk zijn, die niet onafhankelijk, maar 'verwoven' zijn met elkaar.
De natuur is een 'reflectie van het geheel'-

Gedachtengoed van Alexander von Humboldt

Beoordelingscommissie:

dr. L. Mughini Gras	Utrecht University, Faculty of Veterinary Medicine
prof. dr. D.J. Mevius	Utrecht University, Faculty of Veterinary Medicine
prof. dr. J.A. Stegeman	Utrecht University, Faculty of Veterinary Medicine
prof. dr. H.F.L. Wertheim	Radboud Universitair Medisch Centrum
prof. dr. J.H.H.M. van de Wijgert	Universitair Medisch Centrum Utrecht

Paranimfen:

Amber Rongen

Pascale Venema

CONTENTS

Chapter 1 Introduction 9

PART 1 ESBL-*E. coli* in the Netherlands

Chapter 2 Extended-spectrum ß-lactamase (ESBL)-producing and non-ESBL-producing *Escherichia coli* isolates causing bacteraemia in the Netherlands (2014–16) differ in clonal distribution, antimicrobial resistance gene and virulence gene content 23

Chapter 3 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) 47

Chapter 4 External validation of WGS-based antimicrobial susceptibility prediction tools, KOVER-AMR and ResFinder 4.1 for *Escherichia coli* clinical isolates 73

PART 2 ESBL-*E. coli* and ESBL-*K. pneumoniae* in Europe

Chapter 5 Household acquisition and transmission of extended-spectrum ß-lactamase (ESBL) -producing *Enterobacteriaceae* after hospital discharge of ESBL-positive index patients 109

Chapter 6 Comparative genomics of ESBL-producing *Escherichia coli* and *Klebsiella pneumoniae* from household-, and long term care facility- cohorts in five European catchment areas assessing human carriage, the human-polluted environment, and food 136

Chapter 7 General discussion 161

Chapter 8 Nederlandse samenvatting 184

Contributing authors 191

List of publications 193

Dankwoord 194

About the author 199

1

Introduction

Tess D Verschuuren

Nature is a dynamic world, with unimaginable complex relations between organisms and their (in)animate environments.[1] All animals carry numerous and diverse microbial communities, a partnership that has been present since the start of animal evolution.[1] In humans, the largest, and arguably most important microbial community, can be found in the gut. The so-called gut microbiome facilitates a healthy immune system, several digestive processes, and protects the host against potential pathogenic microorganisms. This was demonstrated by, amongst others, comparing mice with germ-free mice. The latter were prone to immune defects, had poorer intestinal vascularisation and were more susceptible to infections.[1,2]

ESCHERICHIA COLI (COLONISATION VERSUS INFECTION)

The human innate immune system recognises microbes that colonise the gut through pattern recognition receptors normally involved in inflammation.[3] One of these microbes is *Escherichia coli*, a gram-negative bacterium that is part of the *Enterobacterales* order.[4] *E. coli* strains, constitute approximately 2% of the microbial intestinal relative abundance, and are pathobionts of the healthy human intestinal tract. Here, *E. coli* may help to deplete oxygen, creating a hospitable environment for strict anaerobes to flourish. Furthermore, *E. coli* colonisation may provide protection against colonisation of other gram-negative bacterial pathogens (a mechanism called colonisation resistance), by competition for space and nutrients.[4-10]

E. coli colonisation can progress to extra-intestinal infection by bacterial adhesion and invasion of epithelial cells through surface appendages like type 1 fimbriae.[11] *E. coli* can also cause diarrheagenic disease through other mechanisms, however, this is outside the scope of this thesis.[4] Extra-intestinal pathogenic *E. coli* (ExPEC), is the most important cause of urinary tract infections (UTI) and bloodstream infections (i.e. bacteraemia) in the Western world.[4,12] Infections with ExPEC are predominantly community-acquired and have increased over the last decades.[13] For example, in the United Kingdom *E. coli* bacteraemia rates increased from 45 to 70 per 100,000 inhabitants between 2009 and 2018.[14] This increase might be explained by (i) a changing human population, and/or (ii) a changing *E. coli* population. (i) Population ageing could lead to an overall increase of infections. Indeed, the majority of *E. coli* infections occurs in those above 75 years of age.[14] (ii) Introduction of certain *E. coli* subpopulations with improved adaptation to the human gut (i.e. a genetic repertoire that provides a fitness advantage), may lead to an increase in infections as an accidental by-product of this adaptation.[4,15-18] From an evolutionary standpoint this makes sense, as for successful spread between humans, *E. coli* depends on its ability

to maintain intestinal colonisation, not on its ability to cause infection. Interestingly, in parallel with increasing infections, a global spread of certain clonal lineages like sequence type (ST)131 has been observed. Evidence suggests that ST131 might be associated with an increased acquisition likelihood and colonisation duration.[17-21] Another aspect of gut adaptation might be the ability to easily acquire, sustain, and share resistance genes, as antibiotic resistance is a relevant quality for self-preservation in an environment where antibiotics frequently occur, such as hospital-, or primary care settings.[4,15] Indeed, some ST131 lineages have acquired antimicrobial resistance (AMR) to several classes of the most important treatment options for UTI and bacteraemia. These antimicrobial classes include: most ß-lactam antibiotics including third-generation cephalosporins (3GC) (but excluding carbapenems), fluoroquinolones, aminoglycosides, and trimethoprim-sulfamethoxazole.[19-22]

3RD GENERATION CEPHALOSPORIN RESISTANCE SURVEILLANCE

Since the late 2000s, surveillance of AMR bacteria is performed in the Netherlands by the infectious disease surveillance information system for antibiotic resistance (ISIS-AR), and Europe at large by the European antimicrobial resistance surveillance network (EARS-Net). The goal of these surveillance systems is to provide information on the prevalence and trends of antimicrobial resistance.[23,24] 3GCresistant (3GCR)-*E. coli* are of particular interest for AMR-surveillance, as with 44%, this group is the most important cause of multidrug-resistant bacterial infections in Europe.[25]

ISIS-AR and EARS-Net currently report phenotypic antimicrobial susceptibility data derived from clinical cultures. Here, connected public-health- and hospital-laboratories report test results from methodologies like broth microdilution, or automated systems such as VITEK 2.[23,24] Although very informative, this methodology does not allow for detection and tracking of (new) 3GCR-*E. coli* genetic variants.

The importance of genetic variant detection is illustrated by the COVID-19 pandemic, where new variants take over earlier variants if they possess certain advantages like increased transmissibility.[26,27] Similarly, if a novel 3GCR-*E. coli* variant would possess traits resulting in increased virulence and/or antimicrobial resistance, this could have an effect on the burden of disease of *E. coli*.[28] Genetic surveillance has recently been adopted for high-risk bacterial species such as carbapenem-resistant and colistin-resistant *Enterobacterales*. In the Netherlands, the national carbapenemase-producing pathogen surveillance program performs genetic surveillance since 2014.

Furthermore, the European antimicrobial resistance genes surveillance network (EURGen-Net) performs genetic surveillance for carbapenem- and colistin-resistant *Enterobacterales* since 2019, with potential extension of EURGen-Net to other pathogens in coming years.[29,30]

If current genetic variant surveillance systems would extend their scope to 3GCR-*E. coli*, this should ideally reflect the genetic variants that circulate in the open community, as colonisation of community-dwellers is the largest reservoir of 3GCR-*E. coli*. Several studies have assessed the genetic variants of 3GCR-*Enterobacterales* circulating in the Dutch community.[31-33] These studies yielded important insights in the molecular epidemiology of 3GCR-*Enterobacterales*, unfortunately, these studies also showed that faecal sampling of healthy community-dwellers is too labour-intensive and costly to be performed as part of routine AMR-surveillance. It is currently unknown if genetic variants of 3GCR-*E. coli* circulating in the community are similar to genetic variants found in clinical cultures.

EPIDEMIOLOGY OF EXTENDED-SPECTRUM ß-LACTAMASE (ESBL) GENES

Currently, most 3GCR in *E. coli* is caused by genetic variants that carry extended-spectrum ß-lactamase (ESBL) genes. ESBL genes encode for enzymes that hydrolyse several ß-lactam antibiotics, including small-spectrum penicillins (e.g. amoxicillin), and cephalosporins (e.g. cefotaxime), but excluding broad-spectrum penicillins (e.g. amoxicillin-clavulanic acid), or carbapenems (e.g. meropenem). Several ESBL gene classes have been described, among which $bla_{\text{TEM-}}$, $bla_{\text{SHV-}}$ and $bla_{\text{CTX-M-}}$ (of note: these gene classes also include (broad-spectrum) ß-lactamase gene types, for classification per gene type visit: http://bldb.eu/).[22,34-36] Within these classes several ESBL gene types occur, which are referred to with numbers (e.g. $bla_{\text{CTX-M-15}}$ or $bla_{\text{SHV-12}}$). The bla_{SHV} and bla_{TEM} gene classes dominated in Europe in the 1990s and were mostly associated with hospital outbreaks of *Enterobacterales*.[37-39] Currently, $bla_{\text{CTX-M}}$ is the predominant ESBL-gene class worldwide, in particular $bla_{\text{CTX-M-15}}$, which is associated with intestinal carriage of community-dwellers, and community-onset infections with *Enterobacterales*.[22,31-33,40-42]

Furthermore, ESBL-producing *E. coli* (ESBL-Ec) also occurs in numerous non-human settings, such as in the animal husbandry, wild birds, food, soil, waste- and surface-water, which all could serve as potential reservoirs of ESBL genes for *E. coli*

isolates colonising and/or causing infections in humans.[41] An increasing amount of evidence however, shows that the majority of ESBL-Ec human carriage is likely explained by human-to-human transmission, however exact ESBL transmission cycles including the role of non-human reservoirs remain to be elucidated.[40,41,43-46]

To make matters more complicated, ESBL-genes can spread in two ways; through clonal spread, and through horizontal gene transfer. ESBL-genes are concentrated in a limited number of clones that have spread globally.[47] These clones have the ability to acquire and maintain ESBL-gene-carrying plasmids.[17,48] For example, ST131 is associated with $bla_{CTX\text{-}M\text{-}15}$ and, more recently, $bla_{CTX\text{-}M\text{-}27}$, enabling these genes to 'hitchhike' on ST131s spread.[42,49] Although often maintained in certain clones, ESBL carrying plasmids have the ability to replicate and spread between different bacterial hosts. Furthermore, plasmids also may harbour mobile genetic elements, like transposons and insertion sequences which can be shared separately, or combined with the plasmid, with bacteria of the same or different species.[50]

KLEBSIELLA PNEUMONIAE

When assessing horizontal gene transfer of mobile genetic elements contained in *E. coli*, other species also need to be considered, in particular *Klebsiella pneumoniae*. Part of the *Enterobacterales* order, *K. pneumoniae* is an inhabitant of the gut in 5 to 38% of humans.[51] While *E. coli* is responsible for the majority of 3GRC-infections, *K. pneumoniae* is more likely to accumulate resistance. For example, 3GCR is present in 15% of European clinical cultures with *E. coli*, while this is 32% for *K. pneumoniae*.[14] For *K. pneumoniae*, an increase of infections has also been observed in parallel with a global spread of certain clonal lineages like ST307.[53-56] It is known that a certain genetic overlap of ESBL genes exists between both species, for example in humans $bla_{CTX\text{-}M\text{-}15}$ is the most occurring ESBL gene in both *E. coli* and *K. pneumoniae*.[31,40,45,57] Furthermore, several recent studies show that *E. coli* and *K. pneumoniae* share plasmid content, together with other members of the *Enterobacterales* family.[58-62]

WHOLE GENOME SEQUENCING

Whole genome sequencing (WGS), where the nucleotide order of the entire DNA of an organism is determined, can characterise *E. coli* in the highest possible genetic resolution. The genomic sequence of a bacterium can be used for several purposes: (i) bacterial identification, (ii) inferring phylogenetic relations, that in turn can be used for molecular surveillance, or to assess transmission, and (iii) prediction of certain

bacterial traits, amongst others antimicrobial resistance.[63,64] These applications make WGS the ideal methodology for molecular epidemiologic research, genetic surveillance, and potentially clinical antimicrobial susceptibility testing (AST). Furthermore, the digital output can be endlessly stored, re-analysed, and shared. These factors improve reproducibility and consequently the quality of research, surveillance, and clinical diagnostics.[63,64]

Advancements in technology, lower costs, and availability of free (online) analysis tools, have made WGS accessible to medical microbiology and public healthcare, in many parts of the world, with expecting increase in use in the coming decades.[63,64] WGS can be divided in (i) short-read technologies, and (ii) long-read technologies. (i) Short read sequencing technologies include (amongst others); Illumina platforms (HiSeq, NextSeq, MiSeq, and NovaSeq).[65] Advantages of short read sequencing are; relatively low costs, low error rate, and availability of (free) analysis tools.[66] However, complete assembly of the chromosome and (ESBL gene harbouring) plasmids is very challenging with this methodology. Complete assembly is inhibited by repeat sequences, often arising from ESBL harbouring plasmids, that surpass the average read length of short-read sequencing technologies (i.e. up to 600 bases).[18,64] (ii) Long read sequencing technologies include; Nanopore platforms (MinION, and PromethION) and Pacific Biosciences platforms. As the name suggests, long read technologies often produce read lengths in the range of 10-30 kb, overcoming the issues of short read sequencing. However, long read sequencing is more costly, has a higher error rate, and is less broadly adopted in medical microbiology and public healthcare.[64]

AIM AND THESIS OUTLINE

In 2019, the World Health Organisation (WHO) named antimicrobial resistance (AMR) one of the top 10 public health threats facing humanity, due to a current yearly estimated 700 000 deaths, and up to 3.4 trillion US dollars loss in gross domestic product by 2030.[67] Proposed core actions by the WHO were, amongst others: (i) increase understanding and awareness of AMR, (ii) improving global surveillance, and (iii) optimizing antimicrobial use.[67] The studies in this thesis provide information on how we can improve resilience against ESBL-Ec with practical applications of WGS, addressing the above-mentioned core actions proposed by the WHO, by (i) increasing scientific knowledge on which reservoirs contribute to human carriage

and infection of ESBL-Ec, (ii) improve detection and tracking of (new) ESBL variants that circulate in the community with genetic surveillance, and (iii) revise antimicrobial susceptibility testing (AST) of *E. coli* infections.

The first part of this thesis focuses on the use of WGS in strategies to improve surveillance and diagnosis of (ESBL-producing) *E. coli* in the Netherlands, which is the result of a collaboration between the National Institute for Public Health and the Environment (RIVM) and the UMC Utrecht. In **chapter 2** ESBL-positive and ESBL-negative *E. coli* bloodstream infections are compared, to assess differences in molecular epidemiology of *E. coli* bacteraemia in the Netherlands, and the likelihood of susceptible *E. coli* to acquire ESBL genes. **Chapter 3** proposes the use of urine and blood cultures, collected in routine clinical practice, for genetic surveillance of ESBL-Ec, as an alternative for the costly and labour intensive, faecal sampling of community-dwellers. **Chapter 4** describes an external validation of two WGS-AST phenotype prediction tools (KOVER-AMR and Resfinder 4.1) for clinical diagnostic use in *E. coli* UTI and bacteraemia.

The second part of this thesis focuses on ESBL-Ec and ESBL-producing *K. pneumoniae* (ESBL-Kp). The results are part of the MODERN-studies, a collaboration between 5 European centres with varying ESBL prevalence, with the goal to improve understanding of ESBL-Ec and ESBL-Kp occurrence in the European community. The frequency and risk factors of clonal transmission of these species in households following hospital discharge of an ESBL-positive patient are presented in **chapter 5**. In **chapter 6** the importance of ESBL-Kp from human carriage, the human-associated environment, and food, and ESBL-Ec from the human-associated environment, and food are assessed as potential reservoirs for human carriage of ESBL-Ec. Where the human-associated environment consists of frequently touched surfaces, U-bends, waste- and river-water.

Lastly, **chapter 7** contains a summary and discussion of the most important results, along with implications for clinical practice and future research.

REFERENCES

1. Gilbert SF, Bosch TCG, Ledón-Rettig C. Eco-Evo-Devo: developmental sy*MBio*sis and developmental plasticity as evolutionary agents. *Nat Rev Genet* 2015; **16**: 611-22.

2. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. *EMBO Rep* 2006; **7**: 688-93.

3. Chu H and Mazmanian SK. Innate immune recognition of the microbiota promotes hostmicrobial sy*MBio*sis. *Nat Immunol* 2013; **14**: 668–75.

4. Riley LW. Distinguishing pathovars from nonpathovars: *Escherichia coli*. *Microbiol Spectrum* 2020; **8**: 1-23.

5. Martinson JNV, Pinkham NV, Peters GW *et al*. Rethinking gut microbiome residency and the *Enterobacteriaceae* in healthy human adults. *ISME J* 2019; **13**: 2306–18.

6. Tenaillon O, Skurnik D, Picard B *et al*. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 2010; **8**: 207–17.

7. Rastegar Lari A, Gold F, Borderon JC *et al*. Implantation and in vivo antagonistic effects of antibiotic-susceptible *Escherichia coli* strains administered to premature newborns. *Biol Neonate* 1990; **2**: 73-8.

8. Vollaard EJ, Clasener HA. Colonisation resistance. *Antimicrob Agents Chemother* 1994; **38**: 409-14.

9. Hudault S, Guignot J, Servin AL. *Escherichia coli* strains colonising the gastrointestinal tract protect germfree mice against *Salmonella typhimurium* infection. *Gut* 2001; **49**: 47–55.

10. Martinson JNV, Walk ST. *Escherichia coli* residency in the gut of healthy human adults. *EcoSal Plus* 2020; **1**: 10.1128/ecosalplus.ESP-0003-2020.

11. Vizcarra IA, Hosseini V, Kollmannsberger P *et al*. How type 1 fimbriae help *Escherichia coli* to evade extracellular antibiotics. *Sci Rep* 2016; **6**: 18109.

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

13. Vihta KD, 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.

14. Public Health England 2019. Laboratory surveillance of *Escherichia coli* bacteraemia in England, Wales and Northern Ireland: 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/844788/hpr3719_ecoli18.pdf.

15. Beceiro A, Tomás M, Bou G. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin Microbiol Rev* 2013; **2**: 185-230.

16. McNally A, Kallonen T, Connor C *et al*. Diversification of colonisation factors in a multidrug-resistant *Escherichia coli* lineage evolving under negative frequency-dependent selection. *MBio* 2019; **10**: e00644-19.

17. Le Gall T, Clermont O, Gouriou S *et al*. Extraintestinal virulence is a coincidental byproduct of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol* 2007; **11**: 2373-84.

18. Dunn SJ, Connor C, McNally A. The evolution and transmission of multi-drug resistant *Escherichia coli* and *Klebsiella pneumoniae*: the complexity of clones and plasmids. *Curr Op Microb* 2019; **51**:51–56.

19. 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; **8**:1437-1449.

20. 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.

21. 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.

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

23. ECDC 2010. Surveillance of antimicrobial resistance in Europe 2009. https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/1011_SUR_annual_EARS_Net_2009.pdf.

24. ISIS-AR. https://www.rivm.nl/isis-ar.

25. Cassini A, Högberg LD, Plachouras D *et al*. Attributable deaths and disability-adjusted life-years caused by infections with antibioticresistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* 2018; **1**:56-66.

26. Davies NG, Abbott S, Barnard RC *et al*. Estimated transmissibility and impact of SARSCoV-2 lineage B.1.1.7 in England. *Science* 2021; **6538**: eabg3055.

27. Tegally H, Wilkinson E, Giovanetti M *et al*. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 2021; **7854**: 438-43.

28. Godijk NG, Bootsma MCJ, 2, van Werkhoven HC *et al*. Modelling addition and replacement mechanisms of plasmid-based beta-lactam resistant *E. coli* infections. *medRxiv* 2021: doi: https://doi.org/10.1101/2021.03.17.21253797.

29. Hendrickx APA, Landman F, de Haan A *et al*. Plasmid diversity among genetically related

*Klebsiella pneumoniae bla*KPC-2 and *bla*KPC-3 isolates collected in the Dutch national surveillance. *Sci Rep* 2020; **10**: 16778.

30. ECDC 2019. ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multicountry outbreak investigations 2019–2021. https://www.ecdc.europa.eu/sites/default/files/documents/framework-for-genomicsurveillance.pdf.

31. van den Bunt G, van Pelt W, Hidalgo L *et al*. Prevalence, risk factors and genetic characterisation of extended-spectrum betalactamase and carbapenemase-producing *Enterobacteriaceae* (ESBL-E and CPE): a community-based cross-sectional study, the Netherlands, 2014 to 2016. *Euro Surveill* 2019; **24**: 29-39.

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

33. 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.

34. Datta N, Kontomichlou P. Penicillase synthesis controlled by infectious R factors in *Enterobacteriaceae. Nature* 1965; **208**: 239-41.

35. Pitton JS. Mechanisms of bacterial resistance to antibiotics. *Ergeb Physiol* 1972; **28**: 15-93.

36. Bonnet R. Growing group of extendedspectrum beta-lactamases: the CTX-M enzymes. *Antimicrob Agents Chemother* 2004; **48**: 1-14.

37. Branger C, Bruneau B, Lesimple AL *et al*. Epidemiological typing of extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolates responsible for five outbreaks in a university hospital. *J Hosp Infect* 1997;**1**:23-36.

38. Decré D, Gachot B, Lucet JC *et al*. Clinical and bacteriologic epidemiology of extendedspectrum beta-lactamase-producing strains of *Klebsiella pneumoniae* in a medical intensive care unit. *Clin Infect Dis* 1998; **4**: 834-44.

39. Lucet JC, Chevret S, Decré D *et al*. Outbreak of multiply resistant *Enterobacteriaceae* in an intensive care unit: epidemiology and risk factors for acquisition. *Clin Infect Dis* 1996; **3**: 430-6.

40. Day MJ, Hopkins KL, Wareham DW *et al*. Extended-spectrum ß-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**: 1325-1335.

41. Mughini-Gras L, Dorado-García A, van Duijkeren E *et al*. Attributable sources of community-acquired carriage of *Escherichia coli* containing ß-lactam antibiotic resistance genes: a population-based modelling study. *Lancet Planet Health* 2019; **3**: e357–e369.

42. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M beta-lactamases: temporal and geographical shifts in genotype. *J Antimicrob Chemother* 2017; **72**:2145-2155.

43. de Been M, Lanza VF, de Toro M *et al*. Dissemination of cephalosporin resistance genes between *Escherichia coli* strains from farm animals and humans by specific plasmid lineages. *PLoS Genet* 2014; **10**: e1004776.

44. Dorado-García A, Smid JH, van Pelt W *et al*. Molecular relatedness of ESBL / AmpCproducing *Escherichia coli* from humans, animals, food and the environment: a pooled analysis. *J Antimicrob Chemother* 2018; **73**: 339–347.

45. Ludden C, Raven KE, Jamrozy D *et al*. One health genomic surveillance of *Escherichia coli* demonstrates distinct lineages and mobile genetic elements in isolates from humans versus livestock. *MBio* 2019; **1**: e02693-18.

46. Mevius D, Heederik D, van Duijkeren E *et al*. Rapport ESBL-Attributieanalyse (ES*Bla*T): Op zoek naar de bronnen van antibioticaresistentie bij de mens, 2018. https://www.1health4food.nl/upload_mm/f/5/6/ef82476f-f31d-4ceb-a0dd-563799f398e4_Rapport%20Es*bla*t.pdf.

47. Manges AR, Geum HM, Guo A *et al*. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin Microbiol Rev* 2019; **3**: e00135-18.

48. Mathers AJ, Peirano, Pitout JDD. The role of epidemic resistance plasmids and international high- risk clones in the spread of multidrug-resistant *Enterobacteriaceae. Clin Microb Rev* 2015; **3**: 565-591.

49. Peirano G, Lynch T, Matsumara Y *et al*. Trends in population dynamics of *Escherichia coli* sequence type 131, Calgary, Alberta, Canada, 2006-2016. *Emerg Infect Dis* 2020; **12**: 2907-2915.

50. Carattoli A. Plasmids and the spread of resistance. *Int J Med Microbiol* 2013; **6-7**: 298-304.

51. Podschun R, Ullmann U. *Klebsiella spp.* as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* 1998; **11**: 589-603.

53. Navon-Venezia S, Kondratyeva K, Carattoli A. *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiol Rev* 2017; **41**: 252-75.

54. Long SW, Olsen RJ, Eagar TN *et al*. Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolates, Houston, Texas: unexpected abundance of clonal group 307. *MBio* 2017; **8**: e00489-17.

55. Peirano G, Chen L, Kreiswirth BN, Pitout JDD. Emerging antimicrobial-resistant high-risk *Klebsiella pneumoniae* clones ST307 and ST147. *Antimicrob Agents Chemother* 2020; **10**:e01148-20.

56. Wyres KL, Jane Hawkey J, Hetland MAK *et al*. Emergence and rapid global dissemination of CTX-M-15-associated *Klebsiella pneumoniae* strain ST307. *J Antimicrob Chemother* 2019; **74**: 577-81.

57. Ludden C, Moradigaravand D, Jamrozy D *et al*. A One health study of the genetic relatedness of

Klebsiella pneumoniae and their mobile elements in the East of England. *Clin Infect Dis* 2020; **2**: 219–226.

58. Acman M, Dorp van L, Santini JM, Francois Balloux F. Large-scale network analysis captures biological features of bacterial plasmids. *Nat Commun* 2020; **1**: 2452.

59. Hendrickx APA, Landman F, de Haan A *et al*. *bla*OXA-48-like genome architecture among carbapenemase-producing *Escherichia coli* and *Klebsiella pneumoniae* in the Netherlands. *bioRxiv* 2020.12.18.423568.

60. Matlock W, Chau KK, AbuOun M *et al*. Genomic network analysis of environmental and livestock F-type plasmid populations. *ISME J* 2021; https://doi.org/10.1038/s41396-021-00926-w.

61. León-Sampedro R, DelaFuente J, Díaz-Agero C *et al*. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. *Nat Microbiol* 2021; https://doi.org/10.1038/s41564-021-00879-y

62. Hidalgo L, de Been M, Rogers MRC *et al*. Sequence-based epidemiology of an OXA-48 plasmid during a hospital outbreak. *Antimicrob Agents Chemother* 2019; **12**: e01204-19.

63. Köser CU, Ellington MJ, Cartwright EJP *et al*. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog* 2012; **8**: e1002824.

64. Hendriksen RS, Bortolaia V, Tate H *et al*. Using genomics to track global antimicrobial resistance. *Front Public Health* 2019; **7**:242.

65. Amarasinghe SL, Su S, Dong X *et al*. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 2020; **1**: 30.

66. Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. *Genomics* 2016; **1**:1–8.

67. World Health Organization 2019. Thirteenth General Programme Of Work. https://apps.who.int/iris/bitstream/handle/10665/324775/WHO-PRP-18.1-eng.pdf

part 1

ESBL-*E. coli* in the Netherlands

2

Extended-spectrum ß-lactamase (ESBL)-producing and non-ESBL-producing *Escherichia coli* isolates causing bacteraemia in the Netherlands (2014–16) differ in clonal distribution, antimicrobial resistance gene and virulence gene content

Denise van Hout, Tess D Verschuuren, Patricia CJ Bruijning-Verhagen, Thijs Bosch, Anita C Schürch, Rob JL Willems, Marc JM Bonten and Jan AJW Kluytmans

Published in: *PLoS One* 2020; **1**: e0227604

ABSTRACT

Background: Knowledge on the molecular epidemiology of *Escherichia coli* causing *E. coli* bacteraemia (ECB) in the Netherlands is mostly based on extended-spectrum ß-lactamase-producing *E. coli* (ESBL-Ec). We determined differences in clonality and resistance and virulence gene (VG) content between non-ESBL-producing *E. coli* (non-ESBL-Ec) and ESBL-Ec isolates from ECB episodes with different epidemiological characteristics.

Methods: A random selection of non-ESBL-Ec isolates as well as all available ESBL-Ec blood isolates was obtained from two Dutch hospitals between 2014 and 2016. Whole genome sequencing was performed to infer sequence types (STs), serotypes, acquired antibiotic resistance genes and VG scores, based on presence of 49 predefined putative pathogenic VG.

Results: ST73 was most prevalent among the 212 non-ESBL-Ec (*n*=26, 12%) and ST131 among the 69 ESBL-Ec (*n*=30, 44%). Prevalence of ST131 among non-ESBL-Ec was 10% (*n*=22, p-value <0.001 compared to ESBL-Ec). O25:H4 was the most common serotype in both non-ESBL-Ec and ESBL-Ec. Median acquired resistance gene counts were 1 (IQR 1–6) and 7 (IQR 4–9) for non-ESBL-Ec and ESBL-Ec, respectively (p-value <0.001). Among non-ESBL-Ec, acquired resistance gene count was highest among blood isolates from a primary gastro-intestinal focus (median 4, IQR 1–8). Median VG scores were 13 (IQR 9–20) and 12 (IQR 8–14) for non-ESBL-Ec and ESBL-Ec, respectively (p-value: 0.002). VG scores among non-ESBL-Ec from a primary urinary focus (median 15, IQR 11–21) were higher compared to non-ESBL-Ec from a primary gastro-intestinal (median 10, IQR 5–13) or hepatic-biliary focus (median 11, IQR 5–18) (p-values: 0.007 and 0.04, respectively). VG content varied between different *E. coli* STs.

Conclusions: Non-ESBL-Ec and ESBL-Ec blood isolates from two Dutch hospitals differed in clonal distribution, resistance gene and VG content. Also, resistance gene and VG content differed between non-ESBL-Ec from different primary foci of ECB.

INTRODUCTION

Escherichia coli is the leading causative pathogen in Gram-negative bacteraemia and is associated with 30-day mortality up to 18%.[1–4] Antibiotic treatment options of *E. coli bacteraemia* (ECB) are getting compromised by the pandemic presence of extended-spectrum ß-lactamases (ESBLs); conferring resistance to antibiotics commonly used for ECB treatment such as third-generation cephalosporins. Worryingly, the incidence of ECB is increasing and in some European countries, the incidence of ECB with antibiotic-resistant strains seems to increase faster than ECB caused by susceptible strains.[3–6] Even though the individual patient and financial burden is increased for resistant ECB episodes, ECB due to susceptible strains is far more common and therefore determines the major part of the ECB disease burden. 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.[3,7] These clinical characteristics of ECB episodes are important because they indicate different target populations for prevention. Thorough insight in the molecular epidemiology of both ESBL-negative and ESBL-positive ECB episodes with different clinical characteristics is key in identifying targets for the development of future preventive strategies, such as *E. coli* vaccines that are currently being developed.[8] Up to now, the molecular epidemiology of ECB in the Netherlands was mainly described in single-center studies or among antimicrobial resistant isolates only.[9,10]

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

METHODS

Study design

Details of the study design is fully described elsewhere.[11] In short, unique patients with ECB were retrospectively identified in the University Medical Center Utrecht, a 1,042-bed tertiary care center and the Amphia Hospital in Breda, an 837-bed teaching hospital. In each hospital, a random sample of 40 isolates for the years 2014

, 2015 and 2016 was selected, comprising ~24% of all first bacteraemic *E. coli* isolates in a year. In addition, all ESBL-Ec blood isolates from 2014 to 2016 were selected. 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). All generated raw reads were submitted to the European Nucleotide Archive (ENA) of the European Bioinformatics Institute (EBI) under the study accession number PRJEB35000. De novo assembly was performed using SPAdes genome assembler v.3.6.2 and quality of assembles was assessed using QUAST.[12] ESBL-production was defined as confirmed phenotypic ESBL-positivity, unless described otherwise.[11] Baseline characteristics were compared between non-ESBL-Ec and ESBL-Ec ECB episodes by the Fisher's Exact or Pearson Chi^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). Individual informed consent was not obtained and all study data were analyzed and stored in a pseudonymized form.
All statistical analyses were performed with Statistical Package for Social Sciences V.25.0 (SPSS, Chicago, Illinois, USA) and R Version 3.4.1.

Multi-locus sequence types (MLST)

Multi-locus sequence types (STs) were based on the allelic profile of seven housekeeping genes and 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). Clonal (i.e. ST) distribution was presented stratified for non-ESBL-Ec and ESBL-Ec isolates and by epidemiological subgroups. Genotype (ST) diversity was analysed by Simpson's diversity index.[13]

Serotyping

Serotypes were assigned 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).[14] Simpson's index for serotype diversity was calculated for non-ESBL-Ec and ESBL-Ec isolates. Serotype distribution among non-ESBL-Ec and ESBL-Ec was compared to two current *E. coli* vaccine candidates.[8,15] Excluding those isolates in which no definitive serotype could be defined.

Antimicrobial resistance genes and virulence genes

Abricate (https://github.com/tseemann/abricate) v0.8.13 was used for (i) mass screening of contigs for (acquired) antimicrobial resistance genes using ResFinder 3.1.0 (download 24 January 2019), and (ii) to determine presence of VG by *Bla*ST against the VFDB database (http://www.mgc.ac.cn/VFs) (download 8 February 2019).[16,17] We searched for 49 putative VG that were previously described as extra-intestinal pathogenic *E. coli* (ExPEC)-associated VG.[18–22] If any of the predefined VG were not included in VFDB, *Bla*ST against the ecoli_VF_collection database was performed (date 8 February 2019).[23] Coverage length and sequence identity thresholds were 80% and 95%. Resistance gene count was defined as the total number of unique identified acquired resistance genes per isolate. Resistance gene counts were compared between non-ESBL-Ec and ESBL-Ec with the non-parametric Wilcoxon rank sum test (for this comparison only, resistance gene count of ESBL-Ec was corrected for presence of the ESBL gene). The VG score was defined as the total number of pre specified VG within an isolate, 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.[20] If a VG was detected multiple times within a single isolate (i.e. different quality measures), it was only counted once. The *kpsM, afa/dra* and *sfa/foc* operons were considered present if any of the corresponding genes or allelic variants were identified. Resistance gene counts and VG scores were further analysed for non-ESBL-Ec and ESBL-Ec separately and were compared between isolates with different epidemiological characteristics and 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 Wilcoxonrank 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-Ec and 69 ESBL-Ec blood isolates (figure 1). Distribution of age, sex, onset of infection and primary foci were comparable between non-ESBL-Ec and ESBL-Ec bacteraemia episodes (table 1). As compared to non-ESBL-Ec, ECB episodes with ESBL-Ec were less often of community onset (64% versus 81%, p-value: 0.003). Crude 30-day and 1-year mortality were higher in ECB episodes caused by ESBL-Ec (28% and 51%, respectively) compared to ECB episodes caused by non-ESBL-Ec (11%, 29%) (p-values: 0.001).

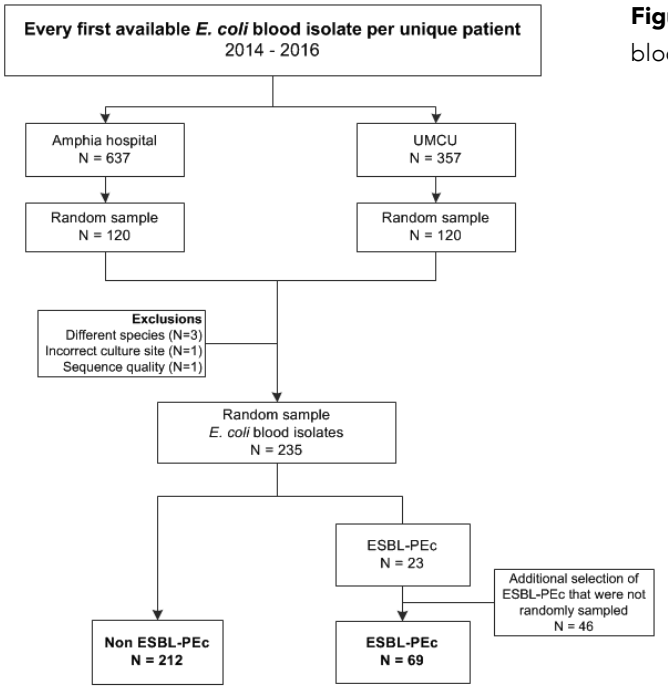

Figure 1. Selection of *E. coli* blood isolates.

Table 1. Baseline characteristics of *E. coli* bacteraemia episodes.

	Non-ESBL-Ec[a] *n*=212	**ESBL-Ec[a] *n*=69**	**p-value[b]**
Median age, years (IQR)	69 (59–77)	69 (56–76)	0.80
Female sex (%)	102 (48.1)	32 (46.4)	0.80
Community onset (%)	172 (81.1)	44 (63.8)	0.003
Primary focus of ECB (%)			
Urinary tract	103 (48.6)	30 (43.5)	0.79
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)	0.22
Ward (%)			
Non-intensive care unit (ICU)	182 (85.8)	58 (84.1)	0.71
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: *Escherichia coli* bacteraemia, ESBL: extended-spectrum ß-lactamase, ESBL-Ec: ESBL-producing *E. coli*, IQR: interquartile range.
[a]ESBL-positivity based on phenotype
[b]p-value of comparison between non-ESBL-Ec versus ESBL-Ec, calculated with Pearson's Chi^2, Fisher's exact, or Mann-Whitney U test when applicable.

Clonal distribution

Among non-ESBL-Ec, ST73 was the most frequently observed ST (*n*=26, 12%), followed by ST131 (*n*=22, 10%). Isolates of ST73, 95, 127, 141, 80 and 1193 were solely identified among non-ESBL-Ec (figure 2). ST131 was dominant among ESBL-Ec (*n*=30, 44%) and prevalence was higher than among non-ESBL-Ec (p-value <0.001). Simpson's index for clonal diversity was 96% (95% CI 94–97) and 81% (95% CI 71– 90) for non-ESBL-Ec and ESBL-Ec, respectively. The occurrence of different STs did not differ between nosocomial and community onset ECB (figure S1). ST131 was the dominant ST among ESBL-positive ECB episodes with a primary urinary (63%) and gastro-intestinal focus (57%), which was higher as compared to other primary foci of ESBL-positive ECB (i.e. 21% among primary hepatic-biliary focus, figure S1).

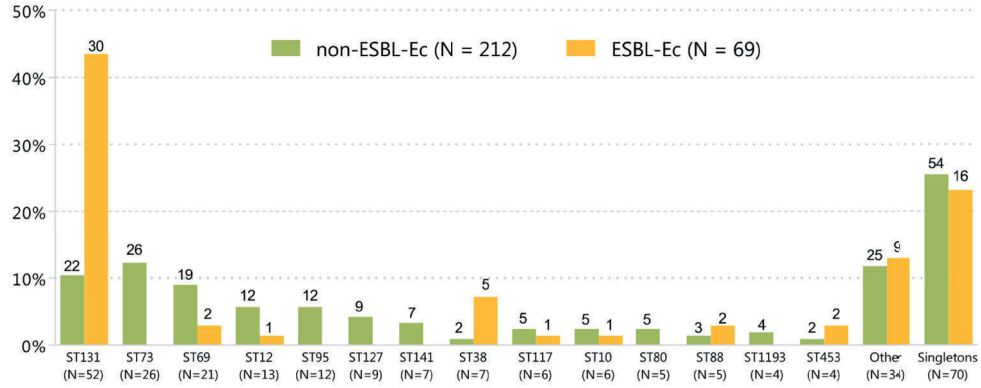

Figure 2. ST distribution among non-ESBL-Ec versus ESBL-Ec[a] in order of frequency[b].
ESBL: extended-spectrum ß-lactamase, ESBL-Ec: ESBL-producing *E. coli*, ST: sequence type.
[a] ESBL-positivity based on phenotypic ESBL production
[b] Missing STs and STs that occurred ≤3 times are grouped in "Other". STs that only occurred once are grouped in "Singletons". The height of each individual bars represents the proportion of the ST within the group of non-ESBL-Ec and ESBL-Ec, respectively.

Serotypes

The most common serotype O25:H4 was identified in 19 (9%) non-ESBL-Ec and 24 (35%) ESBL-Ec isolates, which largely reflected the prevalence of ST131 in each group (table 2). Multiple serotypes only occurred among non-ESBL-Ec, such as O6:H1 and O6:H31. ST73 was most often of serotype O6:H1 (16/26, 62%). Simpson's index for serotype diversity was 97% (95% CI 96 – 98) and 84% (95% CI 77 – 91) for non-ESBL-Ec and ESBL-Ec, respectively. Non-ESBL-Ec and ESBL-Ec isolates from ECB episodes with a primary focus in the urinary tract were most often of O-serotype O6 (15/103, 15%) and O25 (17/30, 57%), respectively (figure S2). For ECB episodes with a primary focus in the hepatic-biliary tract, O25 was the most prevalent O-serotype among

non-ESBL-Ec (7/46, 15%) and O8 (4/14, 29%) among ESBL-Ec isolates (figure S2). 53 (25%) non-ESBL-Ec and 25 (36%) ESBL-Ec 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[8,24], whereas the majority of non-ESBL-Ec (*n*=113; 53%) and ESBL-Ec isolates (*n*=35; 51%) belonged to one of the O-serotypes of the new 10-valent conjugant *E. coli* vaccine (ExPEC-10V) that is currently in development.[15]

Table 2. Serotype distribution among *E. coli* blood isolates.

Serotype	Non-ESBL-Ec *n*=212 (%)	ESBL-Ec[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 ß-lactamase, ESBL-Ec: ESBL-producing *E. coli*.
[a] ESBL-positivity based on phenotypic ESBL production.

Antimicrobial resistance genes

In total, 69 unique acquired resistance genes were identified (table S4). ESBL genes were detected in 65 (94%) of 69 *E. coli* blood isolates with phenotypic ESBL production. $bla_{CTX\text{-}M\text{-}15}$ was most prevalent (*n*=28, 43%), followed by $bla_{CTX\text{-}M\text{-}9}$ (*n*=14, 22%) and $bla_{CTX\text{-}M\text{-}27}$ (*n*=9, 14%). The median acquired resistance gene count for non-ESBL-Ec versus ESBL-Ec was 1 (IQR 1-6) versus 7 (IQR 4-9) (p-value <0.001). Among non-ESBL-Ec, acquired resistance gene count was highest among blood isolates from a primary gastro-intestinal focus (median 4, IQR 1-8). There were significant differences in resistance gene count for different primary foci of non-ESBL ECB, but absolute differences were small (figure S3, table S5). Among ESBL-Ec isolates, there were no statistical significant differences in acquired resistance gene counts between epidemiological subgroups (figure S3). We observed no significant differences among non-ESBL-Ec or ESBL-Ec isolates of different clonal backgrounds (figure 3, table S6).

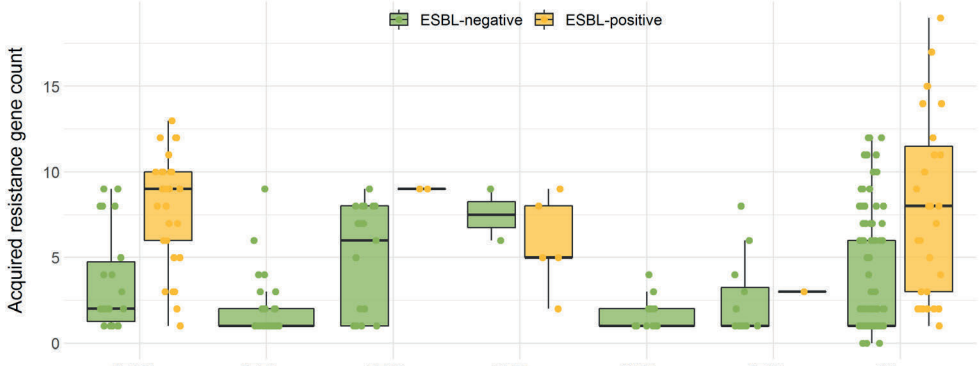

Figure 3. Acquired resistance gene count per ST, stratified for ESBL-positivity.
Boxplots display median resistance gene count and inter quartile range, each dot represents a single isolate. Only STs that occurred >5% within non-ESBL-Ec or ESBL-Ec were grouped into main groups, the rest was categorized as "Other". Results of the pairwise comparisons: table S6.

Virulence genes

Of the 49 predefined ExPEC-associated VG, 44 (90%) were detected in <1 *E. coli* blood isolate and VG scores ranged from 0 (*n*=1 non-ESBL-Ec) to 25 (*n*=2 ESBL-Ec) (figure S4). The median VG score was 13 (IQR 9-20) in non-ESBL-Ec and 12 (IQR 8-14) in ESBL-Ec blood isolates (p-value: 0.002). There were no significant differences in VG scores of epidemiological subgroups, except that the average VG score of non-ESBL-Ec isolates with a primary urinary focus (median 15, IQR 11-21) were higher compared to non-ESBL-Ec isolates with a primary focus in the gastro-intestinal (median 10, IQR 5-13) or hepatic-biliary tract (median 11, IQR 5-18) (p-values: 0.007, 0.04, respectively) (figure S4 and table S8).

There was heterogeneity in VG scores between non-ESBL-Ec of different STs, this was less pronounced for ESBL-Ec isolates (figure 4, table S9). 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).

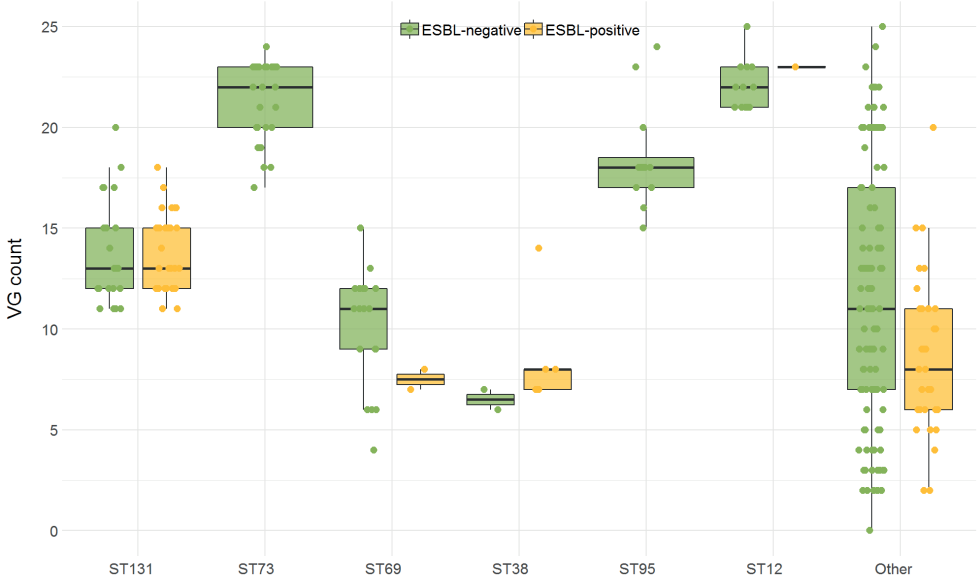

Figure 4. ExPEC-associated VG score in different STs, stratified for ESBL-positivity.
ExPEC: extra-intestinal pathogenic *E. coli*, VG: virulence gene. Boxplots display median VG score and inter quartile range; every dot represents a single isolate. Only STs that occurred >5% within non-ESBL-Ec or ESBL-Ec were grouped into main groups, the rest was categorized as "Other". Results of pairwise comparisons between STs: table S9.

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.

In line with previous research, the clonal distribution among ESBL-Ec blood isolates was less diverse compared to non-ESBL-Ec.[25–27] This was mainly caused by the predominance of ST131 within ESBL-Ec, as has been described before.[28,29] In contrast, ST73, a ST that so far is known for its susceptibility to antibiotics[28], was only identified among non-ESBL-Ec blood isolates. Previous studies have shown very different phylogeny of ST73 and ST131, with the first being characterised by a higher level of diversification in to divergent clades.[28,30] The association between

ESBL phenotype and STs in *E. coli*, which is repeatedly found, implies that the genetic make-up of strains contributes to the ability to acquire and subsequently maintain plasmids carrying ESBL genes. Indeed, 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] It has been hypothesised that this is also true for other *E. coli* lineages that are associated with multidrug resistance (MDR). Reducing the fitness costs of replicating plasmids encoding MDR will result in having competitive advantage over other intestinal strains.[33]

We hypothesised that the clonal distribution and resistance gene and VG content would differ between ECB episodes of community and hospital onset and between different primary foci, as a result of adaptive evolution of intestinal *E. coli*. We observed some statistical significant differences in resistance gene count and VG scores among non-ESBL-Ec from different primary foci of ECB, such as higher VG scores of blood isolates from a primary urinary focus as compared to isolates from a primary focus in the gastro-intestinal or hepatic-biliary tract. However, absolute differences in gene counts were small and the clinical significance remains unclear. In the current study, we found that 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.[34] In that study, there were combinations of VGs 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% for ESBL-negative and 36% for ESBL-positive ECB) and in that study, analyses were not stratified for ESBL-positivity.

Interestingly, in our study, isolates that belonged to ST73 had low resistance gene content but relatively high VG scores as compared to other STs. Furthermore, the average VG score among non-ESBL-Ec was slightly higher than among ESBL-Ec blood isolates, which supports findings of other studies that described an inverse association between antimicrobial resistance and VG content in ExPEC *E. coli*.[35–40] This historical negative association has been challenged, considering the current predominance of ST131, with its relatively broad VG profile despite being

associated with MDR.[41–43] Also in our study, ESBL-positive and ESBL-negative ST131 isolates had equal average VG scores.

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-Ec was more heterogeneous compared to ESBL-Ec, similar to the differences in clonal diversity.[44] 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, such as provided in the current study.[15] We showed that the coverage of the new potential 10-valent vaccine was higher compared to the 4-valent vaccine and was actually doubled for non-ESBL-Ec bacteraemia. Findings of the current study may help further evaluation and implementation of *E. coli* vaccines.

Strengths of the current study are the multicenter design and combination of epidemiological characteristics with 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 a small fraction of the genes that are commonly part of the accessory genome, such as VGs and acquired resistance genes, we may have missed genomic differences that could have importantly contributed to ecological specialization in the different clinically relevant primary foci. Secondly, we selected *E. coli* isolates from a tertiary care center and teaching hospital 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 generalisable to other countries since there could be differences between circulating *E. coli* strains, dependent on local population characteristics and 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, associations between clinical characteristics of ECB episodes and molecular content of *E. coli* isolates were limited. However, we did identify important differences in clonality, serotypes, antimicrobial resistance genes and VG scores between non-ESBL-Ec and ESBL-Ec blood isolates that reached beyond their

phenotypic ESBL-positivity. Future studies that aim to describe the molecular epidemiology of ECB should therefore 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.

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.

Acknowledgements: 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. Preliminary results of this paper were presented at the 29th European Congress of Clinical Microbiology & Infectious Diseases, Amsterdam, the Netherlands, 13-16 April 2019 (P1450).

Conflicts of interest: None to declare.

REFERENCES

1. Abernethy JK, Johnson AP, Guy R *et al*. Thirty day all-cause mortality in patients with *Escherichia coli* bacteraemia in England. *Clin Microbiol Infect* 2015; **3**: 251.e1–8.

2. Fitzpatrick JM, Biswas JS, Edgeworth JD *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.

3. Vihta K-D, 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**: 1138–1149.

4. de Kraker MEA, Jarlier V, Monen JCM *et al*. The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. *Clin Microbiol Infect*. 2013; **19**: 860–868.

5. 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.

6. Schlackow I, Stoesser N, Walker AS *et al*. 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.

7. Abernethy J, Guy R, Sheridan EA *et al*. Epidemiology of *Escherichia coli* bacteraemia in England: results of an enhanced sentinel surveillance programme. *J Hosp Infect* 2017; **95**: 365–375.

8. Huttner A, Hatz C, van den Dobbelsteen G *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.

9. Overdevest ITMA, Bergmans AMC, Verweij JJ *et al*. Prevalence of phylogroups and O25/ST131 in susceptible and extended-spectrum ß-lactamase-producing *Escherichia coli* isolates, the Netherlands. *Clin Microbiol Infect* 2015; **21**: 570.e1-e4.

10. Van Der Bij AK, Peirano G, Goessens WHF *et al*. Clinical and molecular characteristics of extended-spectrum-ß-lactamase-producing *Escherichia coli* causing bacteremia in the Rotterdam Area, Netherlands. *Antimicrob Agents Chemother* 2011; **55**: 3576–3578.

11. 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; 2019020066. doi:10.20944/preprints201902.0066.v1.

12. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013; **29**: 1072–1075.

13. Simpson EH. Measurement of diversity. *Nature* 1949; **163**: 688.

14. Joensen KG, Tetzschner AMM, Iguchi A *et al*. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol* 2015; **53**: 2410–2426.

15. Geurtsen J, Weerdenburg E, Davies T *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.

16. Zankari E, Hasman H, Cosentino S *et al*. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012; **67**: 2640–2644.

17. Liu B, Zheng D, Jin Q *et al*. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 2019; **47**: D687–D692.

18. 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.

19. Johnson JR, Johnston BD, Porter S *et al*. Accessory traits and phylogenetic background predict *Escherichia coli* extraintestinal virulence better than does ecological source. *J Infect Dis* 2019; **219**: 121–132.

20. Johnson JR, Porter S, Johnston B *at 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.

21. Dale AP, Woodford N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. *J Infect* 2015; **71**: 615–626.

22. Dale AP, Pandey AK, Hesp RJ *at 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.

23. Leimbach A. ecoli_VF_collection: V.0.1. Zenodo. 2016. http://dx.doi.org/10.5281/zenodo.56686.

24. Frenck RW, Ervin J, Chu L *at 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.

25. Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant *Enterobacteriaceae*. *Clin Microbiol Rev* 2015; **28**: 565–591.

26. Hertz FB, Nielsen JB, Schonning K *at al*. Population structure of drug-susceptible,-resistant and ESBL-producing *Escherichia coli* from community-acquired urinary tract. *BMC Microbiol* 2016; **16**: 63.

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

28. Kallonen T, Brodrick HJ, Harris SR *at 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.

29. Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. *Escherichia coli* ST131, an intriguing clonal group. *Clin Microbiol Rev* 2014; **27**: 543–574.

30. Alhashash F, Wang X, Paszkiewicz K *at al*. Increase in bacteraemia cases in the East Midlands region of the UK due to MDR *Escherichia coli* ST73: high levels of genomic and plasmid diversity in causative isolates. *J Antimicrob Chemother* 2015; **71**: 339–343.

31. McNally A, Oren Y, Kelly D *at 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.

32. McNally A, Kallonen T, Connor C *at al*. Diversification of Colonisation Factors in a Multidrug-Resistant *Escherichia coli* Lineage Evolving under Negative Frequency-Dependent Selection. *MBio* 2019; **10**: e00644-19.

33. Dunn SJ, Connor C, McNally A. The evolution and transmission of multi-drug resistant *Escherichia coli* and *Klebsiella pneumoniae*: the complexity of clones and plasmids. *Curr Opin Microbiol* 2019; **51**: 51–56.

34. Goswami C, Fox S, Holden M *at al*. Genetic analysis of invasive *Escherichia coli* in Scotland reveals determinants of healthcare-associated versus community-acquired infections. *Microb Genomics* 2018; **4**: e000190.

35. Velasco M, Horcajada JP, Mensa J *at al*. Decreased invasive capacity of quinolone-resistant *Escherichia coli* in patients with urinary tract infections. *Clin Infect Dis* 2001; **33**: 1682–1686.

36. Vila J, Simon K, Ruiz J *at al*. Are quinolone-resistant uropathogenic *Escherichia coli* less virulent? *J Infect Dis* 2002; **186**: 1039–1042.

37. Johnson JR, van der Schee C, Kuskowski MA *at al*. Phylogenetic background and virulence profiles of fluoroquinolone-resistant clinical *Escherichia coli* isolates from the Netherlands. *J Infect Dis* 2002; **186**: 1852–1856.

38. Johnson JR, Kuskowski MA, Owens K *at al*. Phylogenetic origin and virulence genotype in relation to resistance to fluoroquinolones and/or extended-spectrum cephalosporins and cephamycins among *Escherichia coli* isolates from animals and humans. *J Infect Dis* 2003; **188**: 759–768.

39. Horcajada JP, Soto S, Gajewski A *at al*. Quinolone-resistant uropathogenic *Escherichia coli* strains from phylogenetic group B2 have fewer virulence factors than their susceptible counterparts. *J Clin Microbiol* 2005; **43**: 2962–2964.

40. Moreno E, Prats G, Sabaté M *et al*. Quinolone, fluoroquinolone and trimethoprim/sulfamethoxazole resistance in relation to virulence determinants and phylogenetic background among uropathogenic *Escherichia coli*. *J Antimicrob Chemother* 2006; **57**: 204–211.

41. Johnson JR, Porter S, Thuras P, Castanheira M. Epidemic Emergence in the United States of *Escherichia coli* Sequence Type 131-H30 (ST131-H30), 2000 to 2009. *Antimicrob Agents Chemother* 2017; **61**: e00732-17.

42. Johnson JR, Porter S, Thuras P, Castanheira M. The pandemic H30 subclone of sequence type 131 (ST131) as the leading cause of multidrug-resistant *Escherichia coli* infections in the United States (2011-2012). *Open forum Infect Dis* 2017; **4**: ofx089–ofx089.

43. Johnson JR, Russo TA. Molecular Epidemiology of Extraintestinal Pathogenic *Escherichia coli*. *EcoSal Plus* 2018; **8**: 10.1128/ecosalplus.ESP-0004–2017.

44. Ciesielczuk H, Jenkins C, Chattaway M *et al*. Trends in ExPEC serogroups in the UK and their significance. *Eur J Clin Microbiol Infect Dis* 2016; **35**: 1661–1666.

2

SUPPLEMENTARY MATERIAL

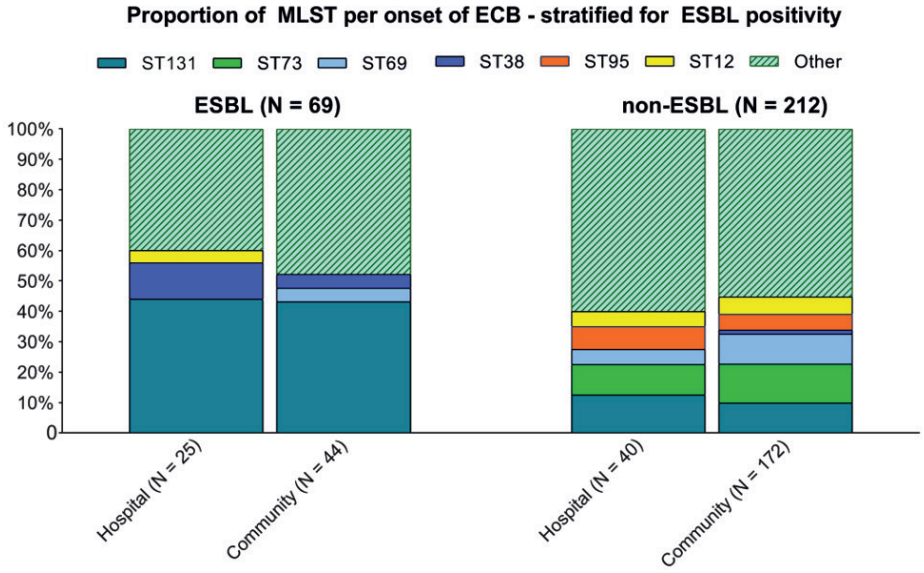

Figure S1. ST distribution among different onset of *E. coli* bacteraemia (ECB)[a]
ESBL: extended spectrum ß-lactamase, ST: sequence type. Only STs that occurred >5% within non-ESBL-Ec or ESBL-Ec were grouped into main ST groups, the rest is categorised as "Other".
[a] ESBL-positivity based on phenotypic ESBL production.

Table S1. ST distribution among different onset of *E. coli* bacteraemia

	ESBL *E. coli*		**Non-ESBL *E. coli***	
n (%)	Hospital (*n*=25)	Community (*n*=44)	Hospital (*n*=40)	Community (*n*=172)
ST131	11 (44)	19 (43)	5 (13)	17 (10)
ST73	-	-	4 (10)	22 (13)
ST69	-	2 (5)	2 (5)	17 (10)
ST38	3 (12)	2 (5)	-	2 (1)
ST95	-	-	3 (8)	9 (5)
ST12	1 (4)	-	2 (5)	10 (6)
Other	10 (40)	21 (48)	24 (60)	95 (55)

ESBL: extended spectrum ß-lactamase, ST: sequence type

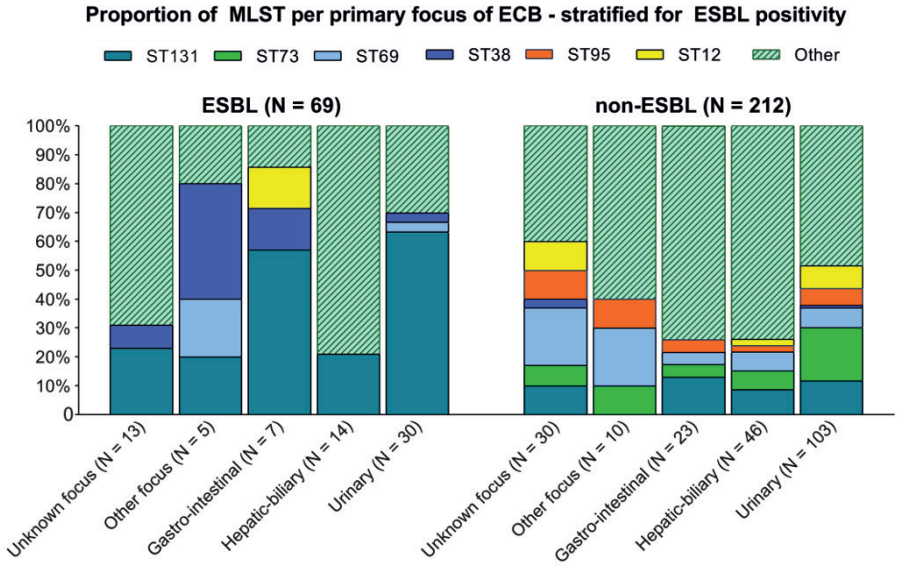

Figure S2. ST distribution among different primary foci of *E. coli* bacteraemia (ECB)
ESBL: extended spectrum ß-lactamase, ST: sequence type. Only STs that occurred >5% within non-ESBL-Ec or ESBL-Ec were grouped into main ST groups, the rest is categorised as "Other".

Table S2. ST distribution among different primary foci of *E. coli* bacteraemia

	ESBL *E. coli*					**Non-ESBL *E. coli***				
n (%)	UR[a] (*n*=30)	HB[b] (*n*=14)	GI[c] (*n*=7)	Oth[d] (*n*=5)	Unk[e] (*n*=13)	UR[a] (*n*=103)	HB[b] (*n*=46)	GI[c] (*n*=23)	Oth[d] (*n*=10)	Unk[e] (*n*=30)
ST131	19 (63)	3 (21)	4 (57)	1 (20)	3 (23)	12 (12)	4 (9)	3 (13)	-	3 (10)
ST73	-	-	-	-	-	19 (18)	3 (7)	1 (4)	1 (10)	2 (7)
ST69	1 (3)	-	-	1 (20)	-	7 (7)	3 (7)	1 (4)	2 (20)	6 (20)
ST38	1 (3)	-	1 (14)	2 (40)	1 (8)	1 (1)	-	-	-	1 (3)
ST95	-	-	-	-	-	6 (6)	1 (2)	1 (4)	1 (10)	3 (10)
ST12	-	-	1 (14)	-	-	8 (8)	1 (2)	-	-	3 (10)
Other	9 (30)	11 (79)	1 (14)	1 (20)	9 (69)	50 (49)	34 (74)	17 (74)	6 (60)	12 (40)

ESBL: extended spectrum ß-lactamase, ST: sequence type.
[a] UR: urinary
[b] HB: hepatic-biliary
[c] GI: gastro-intestinal
[d] Oth: Other primary focus than urinary, hepatic-biliary or gastro-intestinal
[e] Unk: Unknown primary focus

Table S3. Frequencies of O:serotypes per primary focus of *E. coli* bacteraemia

	ESBL *E. coli*					**Non-ESBL *E. coli***				
***n* (%)**	UR[a] (*n*=30)	HB[b] (*n*=14)	GI[c] (*n*=7)	Oth[d] (*n*=5)	Unk[e] (*n*=13)	UR[a] (*n*=103)	HB[b] (*n*=46)	GI[c] (*n*=23)	Oth[d] (*n*=10)	Unk[e] (*n*=30)
O25	17 (57)	2 (14)	3 (43)	1 (20)	1 (8)	12 (12)	7 (15)	2 (9)	-	3 (10)
O6	-	-	-	-	-	15 (15)	6 (13)	-	2 (20)	2 (7)
O4	-	-	1 (14)	-	-	8 (8)	1 (2)	-	-	3 (10)
O2/O50	-	-	-	-	-	10 (10)	6 (13)	-	1 (10)	2 (7)
O75	-	-	-	-	-	4 (4)	1 (2)	2 (9)	1 (10)	3 (10)
O1	1 (3)	-	-	-	-	2 (2)	1 (2)	-	-	1 (3)
O8	-	4 (29)	-	-	-	5 (5)	3 (7)	6 (26)	2 (20)	1 (3)
O15	1 (3)	-	-	1 (20)	-	5 (5)	-	-	-	4 (13)
O18	-	-	-	-	-	5 (5)	-	-	1 (10)	1 (3)
O16	2 (6)	1 (7)	1 (14)	-	-	2 (2)	-	2 (9)	-	-
Other	9 (30)	7 (50)	2 (29)	3 (60)	12 (92)	35 (34)	21 (46)	11 (48)	3 (30)	10 (33)

ESBL: extended spectrum ß-lactamase, ST: sequence type. Frequencies of all serotypes of the 4-valent and new potential 10-valent ExPEC vaccine are reported, the rest (including missing / unknown serotypes) is grouped as "Other serotype". Percentages are column percentages.
[a] UR: urinary
[b] HB: hepatic-biliary
[c] GI: gastro-intestinal
[d] Oth: Other primary focus than urinary, hepatic-biliary or gastro-intestinal
[e] Unk: Unknown primary focus

Table S4. Detected resistance genes with ResFinder 3.1.0 per antibiotic group

(Broad-spectrum) Beta-lactamases			ESBL and ampC			Macrolides			Fluoroquinolones		
Gene	***N***	***(%)***	***Gene***	***N***	***(%)***	***Gene***	***N***	***(%)***	***Gene***	***N***	***(%)***
blaOXA-1	23	8%	*blaCMY-146*	1	0%	*ere(A)*	1	0%	*qnrA1*	2	1%
blaTEM-1A	8	3%	*blaCMY-2*	2	1%	*mph(A)*	48	17%	*qnrS1*	5	2%
blaTEM-1B	82	29%	*blaCTX-M-1*	6	2%	*mph(B)*	3	1%			
blaTEM-1C	8	3%	*blaCTX-M-102*	8	3%						
blaTEM-1D	2	1%	*blaCTX-M-14*	9	3%						
blaTEM-30	1	0%	*blaCTX-M-15*	29	10%						
blaTEM-34	1	0%	*blaCTX-M-27*	1	0%						
blaTEM-40	1	0%	*blaCTX-M-3*	1	0%						
			blaCTX-M-55	1	0%						
			blaCTX-M-9	2	1%						
			blaSHV-102	5	2%						
			blaSHV-12	1	0%						
			blaTEM-28	1	0%						
			blaTEM-35	1	0%						
			blaTEM-52B	1	0%						

Aminoglycosides			Sulfonamides and Trimetroprim			Tetracyclines			Other		
Gene	***N***	***(%)***	***Gene***	***N***	***(%)***	***Gene***	***N***	***(%)***	***Gene***	***N***	***(%)***
aac(3)-Iia	1	0%	*dfrA1*	18	6%	*tet(A)*	72	26%	*catA1*	12	4%
aac(3)-IIa	10	4%	*dfrA12*	6	2%	*tet(B)*	27	10%	*strA*	19	7%
aac(3)-IId	4	1%	*dfrA14*	10	4%	*tet(D)*	1	0%	*strB*	10	4%
aac(3)-IId	7	2%	*dfrA17*	44	16%	*tet(J)*	1	0%	*cat*	1	0%
aac(3)-Iva	1	0%	*dfrA21*	1	0%	*tet(M)*	1	0%	*cmlA1*	6	2%
aac(3)-Via	1	0%	*dfrA5*	12	4%	*tet(X)*	1	0%	*mdf(A)*	260	93%
aac(6')-Ib-cr	12	4%	*dfrA7*	9	3%				*floR*	8	3%
aac(6')Ib-cr	8	3%	*dfrA8*	2	1%				*lnu(F)*	5	2%
aadA1	11	4%	*sul1*	64	23%						
aadA2	10	4%	*sul2*	86	31%						
aadA4	1	0%	*sul3*	6	2%						
aadA5	39	14%									
ant(2")-Ia	4	1%									
ant(3")-Ia	22	8%									
aph(3")-Ib	63	22%									
aph(3')-Ia	24	9%									
aph(3')-Ib	1	0%									
aph(4)-Ia	1	0%									
aph(6)-Id	69	25%									

In case genes were present >1 within a strain, they were only counted once in the resistance gene count.

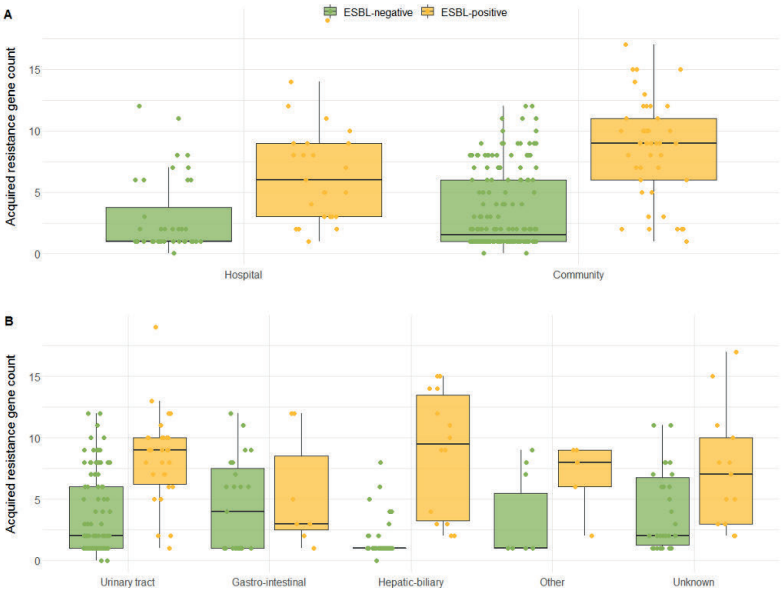

Figure S3. Acquired resistance gene count among epidemiological subgroups
ESBL: extended-spectrum ß-lactamase. Boxplots display median and inter quartile range and every dot represents a single isolate. The ResFinder 3.1.0 database was used to determine acquired resistance genes. **A:** Resistance gene count per onset of infection, stratified for non-ESBL-Ec and ESBL-Ec isolates. **B:** Resistance gene count per primary focus of ECB, stratified for non-ESBL-Ec and ESBL-Ec isolates.

Table S5. Pairwise comparisons acquired resistance gene count between epidemiological subgroups

	Median resistance gene count (IQR)		**Pairwise comparisons between groups, within ESBL- & ESBL+[f]**			
Onset of infection	**ESBL-**	**ESBL+**	**Group 1**	**Group 2**	**ESBL-**	**ESBL+**
Community (*n*=216)	2 (1–6)	9 (6–11)	Community	Hospital	NS[g]	NS
Hospital (*n*=65)	1 (1–4)	6 (3–9)				
Primary focus	**ESBL-**	**ESBL+**	**Group 1**	**Group 2**	**ESBL-**	**ESBL+**
UR[a] (*n*=133)	2 (1–6)	9 (7–10)	UR	GI	NS	NS
HB[b] (*n*=60)	1 (1–1)	10 (3–14)	UR	HB	2.8e-04	NS
GI[c] (*n*=30)	4 (1–8)	3 (3–9)	UR	Oth	NS	NS
Unk[d] (*n*=43)	2 (1–7)	7 (3–10) 8	UR	Unk	NS	NS
Oth[e] (*n*=15)	1 (1–6)	(6–9)	GI	HB	5.8e-03	NS
			GI	Oth	NS	NS
			GI	Unk	NS	NS
			HB	Oth	NS	NS
			HB	Unk	3.1e-05	NS
			Oth	Unk	NS	NS

IQR: interquartile range, NS: not significant. ResFinder 3.1.0 was used to determine presence of acquired resistances genes. Gene counts were rounded to whole numbers if applicable.[a] UR: urinary. [b] HB: hepatic-biliary. [c] GI: gastro-intestinal. [d] Oth: Other primary focus than urinary, hepatic-biliary or gastro-intestinal. [e] Unk: Unknown primary focus. [f] Groups were compared with Wilcoxon rank sum Test and p-values were adjusted with the Holm-Bonferroni correctioan to adjust for multiple testing. [g] p-value represents the adjusted p-value for the comparison of the resistance gene count of Group 1 versus Group 2, within ESBL- or ESBL+ (i.e. p-value 2.8e-04 refers to the comparison in acquired resistance gene count in urinary versus hepatic-biliary primary focus among ESBL- isolates)

Table S6. Pairwise comparisons acquired resistance gene count between dominant STs[a]

	Median resistance gene count (IQR)		Pairwise comparisons between groups, within ESBL- & ESBL+[b]			
ST	ESBL-	ESBL+	Group 1	Group 2	ESBL-	ESBL+
Other (*n*=150)	1 (1–6)	8 (3–12)	ST12	ST131	NS*	NS*
ST131 (*n*=52)	2 (1–5)	9 (6–10)	ST12	ST38		
ST73 (*n*=26)	1 (1–2)	NA	ST12	ST69		
ST69 (*n*=21)	6 (1–8)	9 (9–9)	ST12	ST73		
ST12 (*n*=13)	1 (1–3)	3 (3–3)	ST12	ST95		
ST95 (*n*=12)	1 (1–2)	NA	ST131	ST38		
ST38 (*n*=7)	8 (7–8)	5 (5–8)	ST131	ST69		
			ST131	ST73		
			ST131	ST95		
			ST38	ST69		
			ST38	ST73		
			ST38	ST95		
			ST69	ST73		
			ST69	ST95		
			ST73	ST95		

ESBL: extended-spectrum ß-lactamase, IQR: interquartile range, NA: not applicable, NS*: not significant, all comparisons. ResFinder 3.1.0 was used to determine presence of acquired resistances genes. Gene counts were rounded to whole numbers if applicable.
[a] Comparisons with category "Other" are not shown; because of heterogeneity in STs this comparison was not considered as informative.
[b] Groups were compared with Wilcoxon rank sum Test and p-values were adjusted with the Holm-Bonferroni correction to adjust for multiple testing.

Table S7. Detected extra-intestinal pathogenic *E. coli* (ExPEC) associated virulence genes (VG) per VG category

Adhesins		Siderophores		Protectins & invasins		Toxins		Other	
gene	n (%)[a]	gene	n (%)[a]	gene	n (%)[a]	gene	n (%)[a]	gene	n (%)[a]
yagZ/ecpA	271 (96)	*sitA*	233 (83)	*ompA*	235 (84)	*usp*	158 (56)	*traT*	181 (64)
FimH	266 (95)	*fyuA*	224 (80)	*ompT*	218 (78)	*vat*	101 (36)	*malX*	164 (58)
tia	124 (44)	*chuA*	158 (56)	*kpsM*[b]	78 (28)	*sat*	91 (32)	*iss*	124 (44)
iha	111 (40)	*iroN*	135 (48)	*tcpC*	53 (19)	*clbB*	80 (29)	*cvaC*	42 (15)
papC	103 (37)	*iutA*	32 (11)	*ibeA*	40 (14)	*clbN*	80 (29)	*fliC*	19 (7)
papH	100 (36)	*ireA*	39 (14)			*hlyD*	76 (27)	*rfc*	13 (5)
sfa/foc[c]	87 (40)					*hlyA*	72 (26)		
agn43	81 (29)					*cnf1*	66 (24)		
papG	57 (20)					*pic*	45 (16)		
papF	55 (20)					*astA*	29 (10)		
afa/dra[c]	43 (15)					*cdtB*	13 (45)		
nfaE	9 (3)								
gafD	8 (3)								
bmaE	7 (3)								
papE	7 (3)								
papA	5 (2)								

The following genes were not detected: *focE, hra, yfcV* and *tsh* (adhesins) and *hlyF* (toxin).
[a] *n* indicates numbers of isolates with gene, % of all isolates (*n* = 281)
[b] The *kpsM*, *afa/dra* and *sfa/foc* operons were considered present if any of the corresponding genes our allelic variants were identified.

Figure S4. Extra-intestinal pathogenic *E. coli* (ExPEC)-associated virulence gene (VG) score in different subgroups, stratified for ESBL-positivity.
Boxplots display median and inter quartile range (IQR) and every dot represents a single isolate.
A: VG count per onset of infection, stratified for non-ESBL-Ec and ESBL-Ec isolates.
B: VG count per primary focus of ECB, stratified for non-ESBL-Ec and ESBL-Ec isolates.

Table S8. Pairwise comparisons VG scores between epidemiological subgroups

	Median VG score (IQR)		Pairwise comparisons between groups, within ESBL- and ESBL+[f]			
Onset of infection	**ESBL-**	**ESBL+**	**Group 1**	**Group 2**	**ESBL-**	**ESBL+**
Community (*n*=216)	13 (9–20)	11 (7–13)	Community	Hospital	NS	NS
Hospital (*n*=65)	14 (11–20)	12 (8–15)				
Primary focus	**ESBL-**	**ESBL+**	**Group 1**	**Group 2**	**ESBL-**	**ESBL+**
UR[a] (*n*=133)	15 (11–21)	12 (11–15)	UR	GI	0.0072	NS
HB[b] (*n*=60)	11 (5–18)	7 (6–13)	UR	HB	0.036	NS
GI[c] *n*=30)	10 (5–13)	12 (8–13)	UR	Oth	NS	NS
Unk[d] (*n*=43)	16 (11–20)	12 (8–13)	UR	Unk	NS	NS
Oth[e] (*n*=15)	13 (8–20)	7 (7-14)	GI	HB	NS	NS
			GI	Oth	NS	NS
			GI	Unk	NS	NS
			HB	Oth	NS	NS
			HB	Unk	NS	NS
			Oth	Unk	NS	NS
Urinary catheter	**ESBL-**	**ESBL+**	**Group 1**	**Group 2**	**ESBL-**	**ESBL+**
No (*n*=184)	13 (9-20)	10 (7-13)	No catheter	Catheter	NS	NS
Yes (*n*=97)	13 (7-18)	13 (11-15)				
30-day mortality	**ESBL-**	**ESBL+**	**Group 1**	**Group 2**	**ESBL-**	**ESBL+**
Alive (*n*=238)	13 (9-20)	12 (8-15)	Alive	Deceased	NS	NS
Deceased (*n*=43)	12 (6-18)	11 (7-14)				
Admission ward	**ESBL-**	**ESBL+**	**Group 1**	**Group 2**	**ESBL-**	**ESBL+**
Non-ICU (*n*=240)	13 (9-20)	12 (7-15)	Non-ICU	ICU	NS	NS
ICU (*n*=41)	13 (6-18)	12 (10-14)				

IQR: interquartile range; NA, NS: not significant, gene counts were rounded to whole numbers if applicable. [a] UR: urinary. [b] HB: hepatic-biliary. [c] GI: gastro-intestinal. [d] Oth: Other primary focus than urinary, hepatic-biliary or gastro-intestinal. [e] Unk: Unknown primary focus. [f] Pairwise comparisons were made with Wilcoxon rank sum Test and p-values were adjusted with the Holm-Bonferroni correction to adjust for multiple testing.

Table S9. Pairwise comparisons virulence gene (VG) scores between dominant STs

	Median VG score (IQR)		Pairwise comparisons between groups, within ESBL- and ESBL+			
ST	**ESBL-**	**ESBL+**	**Group 1**	**Group 2**	**ESBL-**	**ESBL+**
Other (*n*=150)	11 (7–17)	8 (6 –11)	ST12	ST131	3.2e-05	NS
ST131 (*n*=52)	13 (12–15)	13 (12–15)	ST12	ST38	NS	NS
ST73 (*n*=26)	22 (20–23)	-	ST12	ST69	5.5e-05	NS
ST69 (*n*=21)	11 (9–12)	8 (7–8)	ST12	ST73	NS	-
ST12 (*n*=13)	22 (21–23)	23 (23–23)	ST12	ST95	0.032	-
ST95 (*n*=12)	18 (17–19)	-	ST131	ST38	NS	NS
ST38 (*n*=7)	7 (6–7)	8 (7–8)	ST131	ST69	4.4e-03	NS
			ST131	ST73	1.9e-07	-
			ST131	ST95	41.9e-03	-
			ST38	ST69	NS	NS
			ST38	ST73	NS	-
			ST38	ST95	NS	-
			ST69	ST73	2.0e-07	-
			ST69	ST95	5.8e-05	-
			ST73	ST95	0.032	-

NS: not significant, ST: sequence type. Comparisons with category "Other" are not shown; because of heterogeneity in STs this comparison is not considered as informative. [a] Groups were compared with Wilcoxon rank sum Test and p-values were adjusted with the Holm-Bonferroni correction to adjust for multiple testing. Gene counts were rounded to whole numbers if applicable.

3

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)

Tess D Verschuuren, Denise van Hout, Sergio Arredondo-Alonso, Ad C Fluit, Elien A Reuland, Janetta Top, Anita C Schürch, Thijs Bosch, Marc JM Bonten, Jan AJW Kluytmans and Rob JL Willems

Published in: *Journal of Antimicrobial Chemotherapy* 2021; **4**: 901-908

ABSTRACT

Background: 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 $bla_{CTX\text{-}M\text{-}1}$ 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 colonisation 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 $bla_{CTX\text{-}M}$ 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) hospitalised patients with nosocomial urinary tract infection between 2014 & 2016 [further referred to as hospital-acquired (HA) urine isolates], retrospectively collected (*n*=49); and (iii) hospitalised 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 & Table S1).

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 of the European Bioinformatics Institute under the study accession numbers PRJEB35000 and PRJEB40007. *De novo* assemblywas 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 *FimH*41 are grouped to ST131-clade A, with *FimH*22 grouped to ST131-clade B and with *FimH*30 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 $bla_{\text{CTX-M-15}}$.[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 $bla_{\text{CTX-M-27}}$ 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 visualised 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 of *k*|*p(faeces)k - p(infection)*k|

where *p* was the proportion of a observed subtype within ST- (e.g. ST131..ST*n*), SC-(e.g. SC131..SC*n*), or ESBL gene-level (e.g. $bla_{\text{CTX-M-15}}$..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%). 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 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%, $P<0.01$) (table1).

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 $bla_{\text{CTX-M}}$ family. The remaining four genes were $bla_{\text{SHV-12}}$ (1.8%, n=8), $bla_{\text{TEM-52}}$ (0.7%, n=3) and two genes belonging to the bla_{TEM} family (0.4%, n=2). The three most prevalent ESBL genes were $bla_{\text{CTX-M-15}}$ (48%), $bla_{\text{CTX-M-14}}$ (17%) and $bla_{\text{CTX-M-27}}$ (16%). Nine ESBL genes were found in both community faecal and clinical isolates, accounting for 98% of all isolates (table S4). Two urine isolates each harboured two ESBL genes ($bla_{\text{CTX-M-27}}$ & $bla_{\text{CTX-M-55}}$). Of the

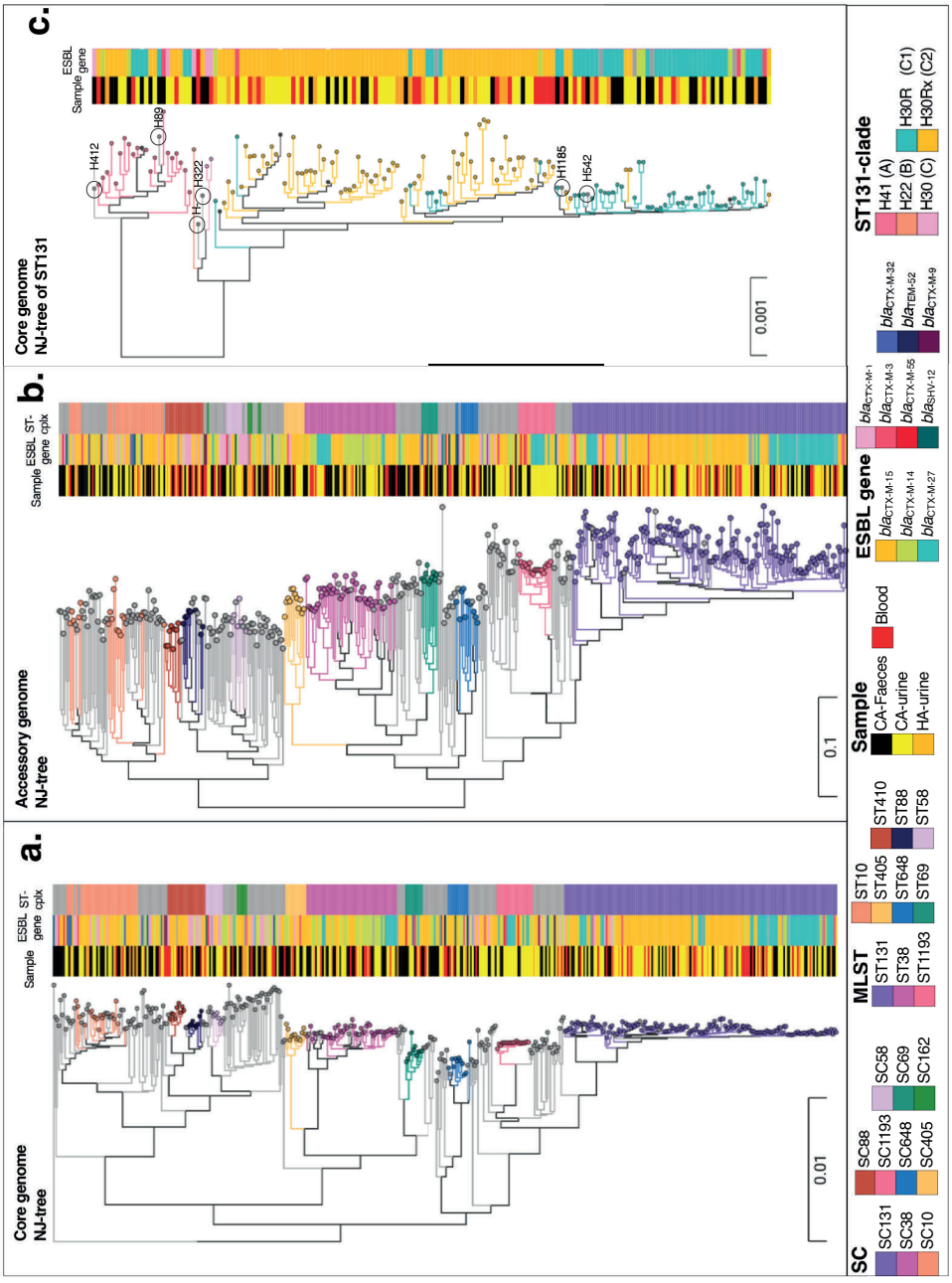

Figure 1. Neighbour-joining trees. a. Core genome, nodes coloured according to ST (10 most frequent), b. accessory genome, nodes coloured according to ST (10 most frequent), and c. core genome ST131, nodes coloured according to clade (nodes with a singleton *FimH* type indicated separately). Constructed with PopPUNK. Online view core tree (https://microreact.org/project/Vmycsy2gY/938965ce), accessory tree (https://microreact.org/project/f9lums0yo/11f42f48), core genome ST131-subtree (https://microreact.org/project/Vmycsy2gY/107f9879). Sample: community urine (CA), nosocomial urine (HA).

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 $bla_{CTX\text{-}M\text{-}9}$ (three clinical isolates) and $bla_{CTX\text{-}M\text{-}2}$ (2 clinical isolates) (figure 2c). The prevalence of $bla_{CTX\text{-}M\text{-}1}$ 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, $bla_{CTX\text{-}M\text{-}15}$ was observed throughout the trees, while $bla_{CTX\text{-}M\text{-}27}$ concentrated in ST131 (SC131), ST38 (SC38) and ST1193 (SC1193). The $bla_{CTX\text{-}M\text{-}14}$ gene concentrated mostly in ST38 (SC38), and $bla_{CTX\text{-}M\text{-}1}$ 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. 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 $bla_{CTX\text{-}M\text{-}27}$. 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 hypothesiswas 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 S2–S4). 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 S5–S8).

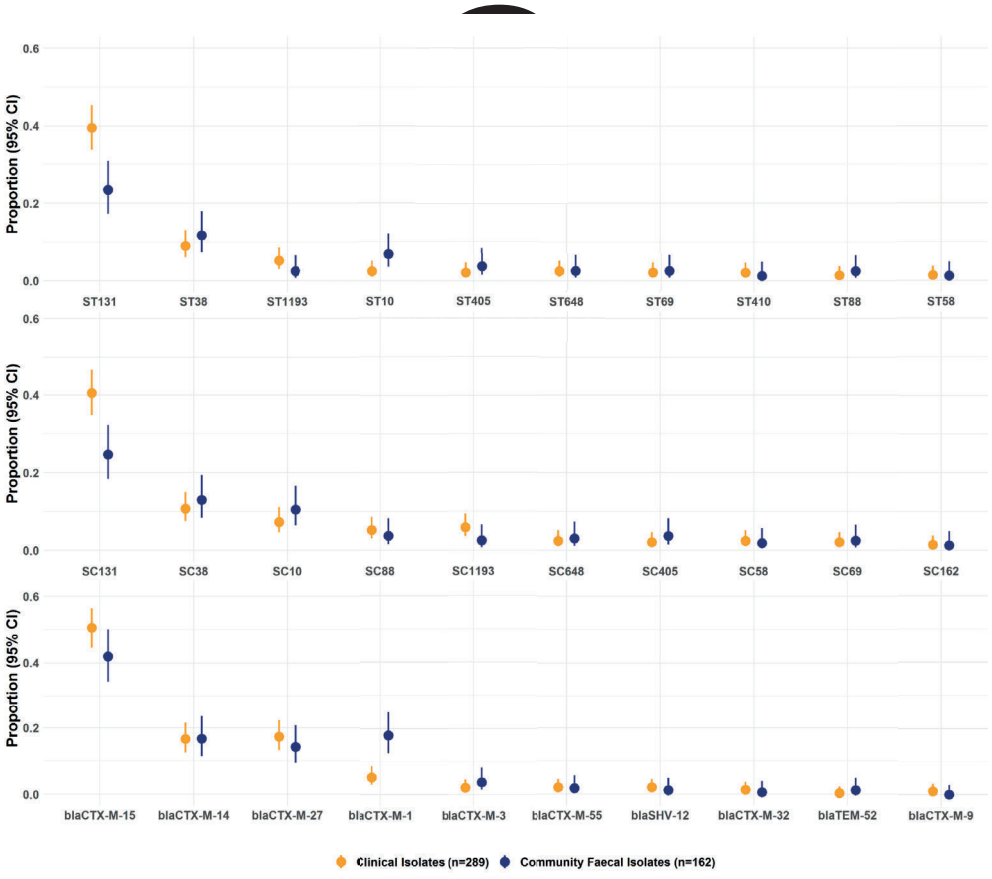

Figure 2. Proportions of the 10 most frequent genetic subtypes in clinical isolates for (upper) STs, (middle) SCs and (bottom) ESBL gene types. Proportions of ST131, SC131 and $bla_{CTX\text{-}M\text{-}1}$ differed between clinical and community faecal isolates; P<0.01. p-value derived from Chi^2 statistic. All other proportions did not differ significantly.

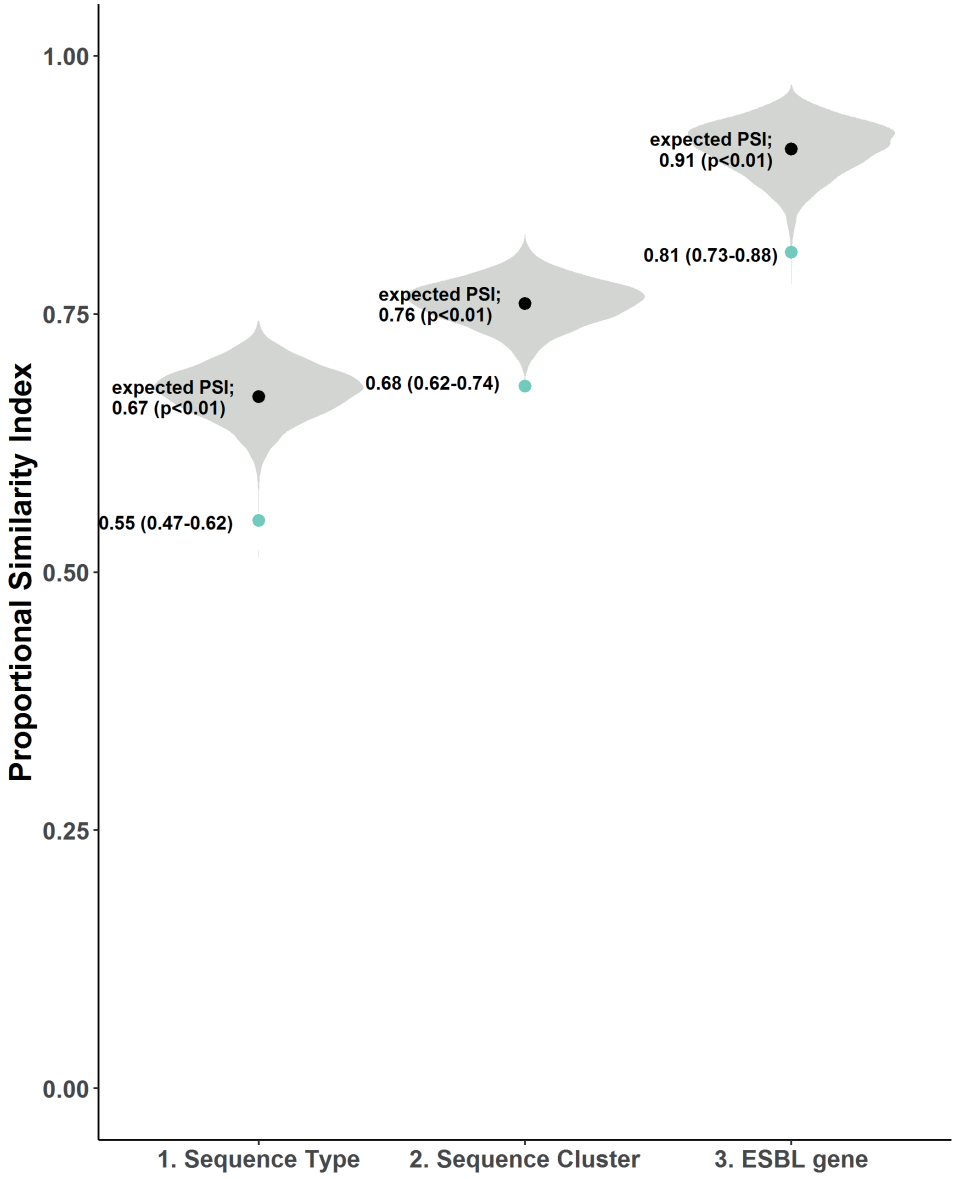

Figure 3. Mean observed PSI (PSI_{obs}), interpreted as the proportion of overlap between community faeces with clinical isolates; PSI_{obs} (green) with 95% CI calculated with 5000 bootstrap iterations. Permutation distribution with mean expected PSI (PSI_{exp}) under the null hypothesis (i.e. no difference 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).

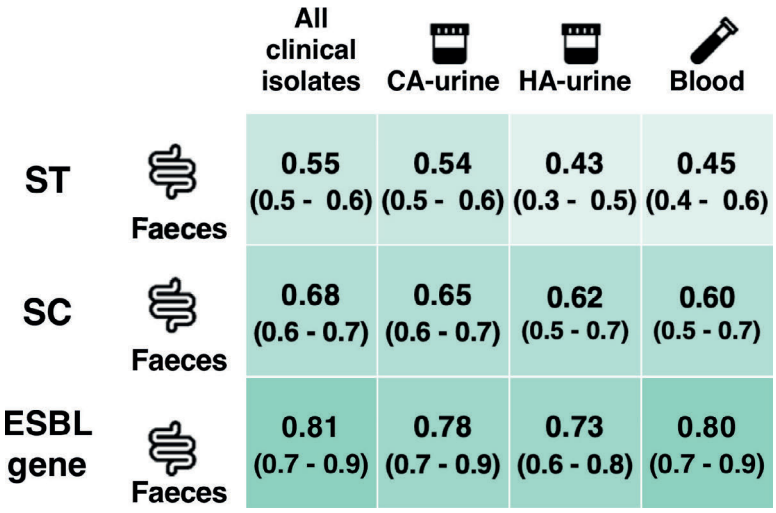

Figure 4. Mean (95% CI) PSI, interpreted as the aggregate proportion of overlap between community faeces and clinical subgroups on ST-, SC- and ESBL gene-level. Each cell is coloured according to the PSI level, with a colourgradient from 0 (light) to 1 (dark).

Table 1. Subdivision of ST131 in subclades

***n*=152**	**Clinical *n*=114**	**Faeces *n*=38**	**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 $bla_{CTX\text{-}M\text{-}27}$) (*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 Chi^2 statistic. [b] *FimH41* = clade A, [c] *FimH22* = clade B, [d] *FimH*30 = clade C. H30 subgroups: H30, no *gyrA/parC* point mutations (PM); H30R (C1), *gyrA/parC* PM and ESBL gene other than $bla_{CTX\text{-}M\text{-}15}$; H30R (C1 with $bla_{CTX\text{-}M\text{-}27}$) is described as a subgroup of H30R (C1) group; H30Rx (clade C2), *gyrA/parC* PM with $bla_{CTX\text{-}M\text{-}15}$. [e] Other *FimH* types: *FimH* type singletons are classified based on fluoroquinolone (FQ) resistance (R), ESBL gene type, and position in phylogenetic tree (fig 1): *FimH89*, FQ susceptible (S), $bla_{CTX\text{-}M\text{-}27}$, assigned to clade A; *FimH322*, FQ-S, $bla_{CTX\text{-}M\text{-}15}$, assigned to clade B/C; *FimH412*, FQ-R, $bla_{CTX\text{-}M\text{-}1}$, assigned to clade A; *FimH1185*, FQ-R, $bla_{CTX\text{-}M\text{-}27}$, assigned to clade C1; *FimH*542, FQ-R $bla_{CTX\text{-}M\text{-}27}$, assigned to clade C1; no *FimH* type, FQ-S, $bla_{CTX\text{-}M\text{-}55}$, assigned to clade B/C.[23,24]

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. 9/10 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. 8/10 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) 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 $bla_{\text{CTX-M-1}}$ in clinical isolates is also in line with earlier research. The $bla_{\text{CTX-M-1}}$ 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 hypothesise 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 $bla_{\text{CTX-M-1}}$, 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 colonisation 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 genes between community gut colonisation 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-Garcia *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 chickenmeat 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 generalisability 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 hospitalisation.

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 colonisation to assess clone/ESBL gene distribution and ESBL-Ec prevalence of community gut colonisation, 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 colonisation 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.

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). R.J.L.W. was supported by the Joint Programming Initiative in Antimicrobial Resistance (JPIAMR Third call, STARCS, JPIAMR2016-AC16/00039).

Acknowledgements: Preliminary results of this paper were presented at the Twenty-Ninth European Congress of Clinical Microbiology and Infectious Diseases, Amsterdam, the Netherlands (Poster number: P1449).

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 like to thank Lapo Mughini-Gras for his contribution in the methodological framework of the study. Lastly, we would like to thank Daniela Cianci and Peter Zuithoff for their statistical support.

Conflicts of interest: None to declare.

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. Rodriguez-Bano J, Lopez-Cerero L, Navarro MD *et al*. Faecal carriage of extended-spectrum b-lactamase-producing *Escherichia coli*: prevalence, risk factors and molecular epidemiology. *J Antimicrob Chemother* 2008; **62**: 1142–9.

6. Mughini-Gras L, Dorado-Garcıa A, van Duijkeren E *et al*. Attributable sources of community-acquired carriage of *Escherichia coli* containing b-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 b-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 b-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 b-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 b-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 SantenMG *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, GoessensWH*et al*. Clinical and molecular characteristics of extended-spectrum-b-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*. *Bla*ST!: 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 cladewith *bla*CTX-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. Argimon 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-Garcia 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, Lofmark S, Borjesson 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 colonisation 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-b 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 b-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 b-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.

3

SUPPLEMENTARY MATERIAL

Table S1. Strain collection is found in a separate csv file:
https://academic.oup.com/jac/article/76/4/901/6082778#supplementary-data

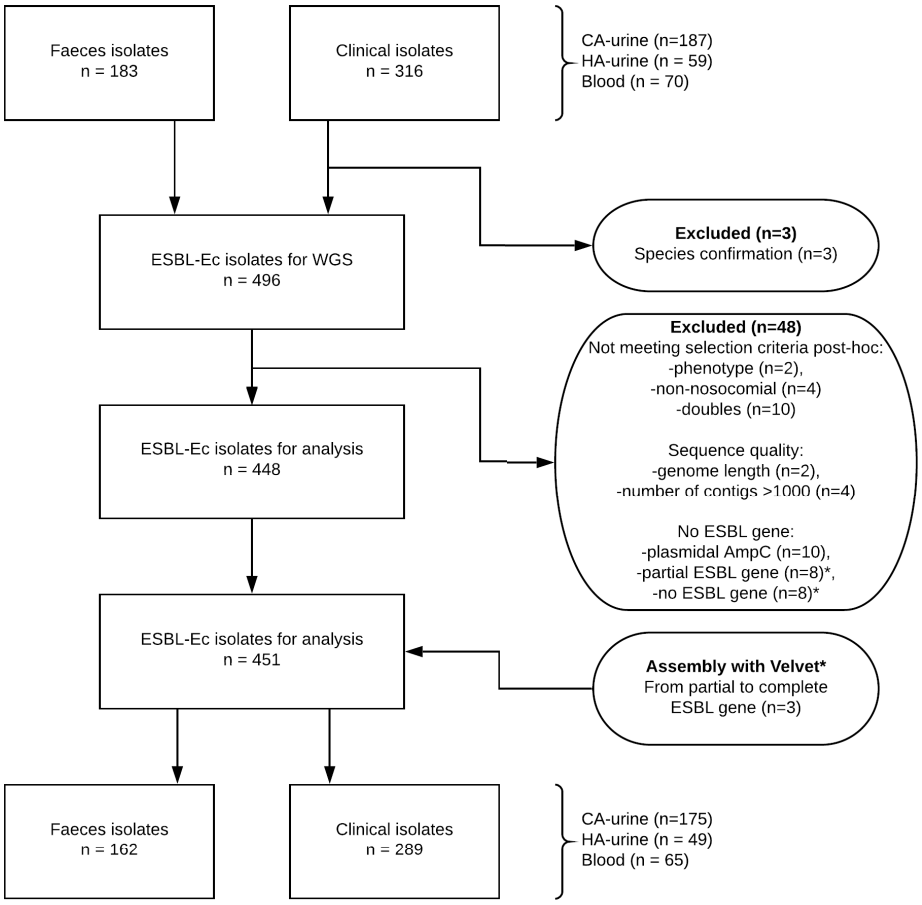

Figure S1. Flowchart of selection and exclusion of isolates.**de novo* assembly using Velvet build in, in Resfinder (v3.2) Webserver (https://cge.cbs.dtu.dk/services/ResFinder/)

Table S2. Proportions of STs. Prop: proportion.

					Clinical isolates, consisting of:							
	All		**Clinical**		*Blood*		*HA-urine*		*CA-urine*		**Faeces**	
ST	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*
Isolates	451	-	289	-	65	-	49	-	175	-	162	-
Unique ST	108	-	72	-	27	-	23	-	49	-	62	-
Unknown	9	0.02	1	0.00	0	0.00	1	0.02	0	0.00	8	0.05
131	152	0.34	114	0.39	29	0.45	19	0.39	66	0.38	38	0.23
38	45	0.10	26	0.09	5	0.08	4	0.08	17	0.10	19	0.12
1193	19	0.04	15	0.05	0	0.00	0	0.00	15	0.09	4	0.02
10	18	0.04	7	0.02	1	0.02	0	0.00	6	0.03	11	0.07
405	12	0.03	6	0.02	1	0.02	1	0.02	4	0.02	6	0.04
648	11	0.02	7	0.02	1	0.02	1	0.02	5	0.03	4	0.02
69	10	0.02	6	0.02	2	0.03	0	0.00	4	0.02	4	0.02
410	8	0.02	6	0.02	2	0.03	2	0.04	2	0.01	2	0.01
88	8	0.02	4	0.01	2	0.03	0	0.00	2	0.01	4	0.02
58	6	0.01	4	0.01	0	0.00	0	0.00	4	0.02	2	0.01
95	5	0.01	5	0.02	0	0.00	0	0.00	5	0.03	0	0.00
636	4	0.01	2	0.01	0	0.00	0	0.00	2	0.01	2	0.01
117	4	0.01	2	0.01	1	0.02	0	0.00	1	0.01	2	0.01
501	4	0.01	3	0.01	1	0.02	1	0.02	1	0.01	1	0.01
744	4	0.01	3	0.01	0	0.00	1	0.02	2	0.01	1	0.01
34	4	0.01	3	0.01	0	0.00	0	0.00	3	0.02	1	0.01
394	4	0.01	1	0.00	0	0.00	1	0.02	0	0.00	3	0.02
224	4	0.01	4	0.01	2	0.03	0	0.00	2	0.01	0	0.00
617	3	0.01	1	0.00	0	0.00	1	0.02	0	0.00	2	0.01
155	3	0.01	2	0.01	0	0.00	1	0.02	1	0.01	1	0.01
141	3	0.01	2	0.01	0	0.00	0	0.00	2	0.01	1	0.01
354	3	0.01	3	0.01	2	0.03	1	0.02	0	0.00	0	0.00
156	3	0.01	1	0.00	1	0.02	0	0.00	0	0.00	2	0.01
12	3	0.01	2	0.01	1	0.02	0	0.00	1	0.01	1	0.01
162	3	0.01	3	0.01	0	0.00	3	0.06	0	0.00	0	0.00
602	2	0.00	1	0.00	0	0.00	1	0.02	0	0.00	1	0.01
44	2	0.00	1	0.00	0	0.00	1	0.02	0	0.00	1	0.01
357	2	0.00	2	0.01	1	0.02	1	0.02	0	0.00	0	0.00
315	2	0.00	1	0.00	1	0.02	0	0.00	0	0.00	1	0.01
73	2	0.00	1	0.00	0	0.00	0	0.00	1	0.01	1	0.01
783	2	0.00	2	0.01	2	0.03	0	0.00	0	0.00	0	0.00
2197	2	0.00	2	0.01	0	0.00	1	0.02	1	0.01	0	0.00
1279	2	0.00	2	0.01	1	0.02	1	0.02	0	0.00	0	0.00
393	2	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	0.01
209	2	0.00	2	0.01	1	0.02	1	0.02	0	0.00	0	0.00
90	2	0.00	2	0.01	1	0.02	0	0.00	1	0.01	0	0.00
127	2	0.00	2	0.01	0	0.00	1	0.02	1	0.01	0	0.00
457	2	0.00	2	0.01	0	0.00	0	0.00	2	0.01	0	0.00
1304	2	0.00	1	0.00	0	0.00	1	0.02	0	0.00	1	0.01
43	2	0.00	2	0.01	0	0.00	2	0.04	0	0.00	0	0.00
517	2	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	0.01
540	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
8188	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
2562	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
1611	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
167	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
6215	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
442	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
1421	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
443	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
2086	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
450	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
3171	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01

Continuation of table S2. Proportions of STs. Prop: proportion.

					Clinical isolates, consisting of:							
	All		**Clinical**		*Blood*		*HA-urine*		*CA-urine*		**Faeces**	
ST	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*
453	1	0.00	1	0.00	1	0.02	0	0.00	0	0.00	0	0.00
7841	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
200	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
1284	1	0.00	1	0.00	0	0.00	1	0.02	0	0.00	0	0.00
500	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
1486	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
59	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
2003	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
130	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
2279	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
226	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
2952	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
295	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
5420	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
48	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
7429	1	0.00	1	0.00	1	0.02	0	0.00	0	0.00	0	0.00
609	1	0.00	1	0.00	1	0.02	0	0.00	0	0.00	0	0.00
8183	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
328	1	0.00	1	0.00	0	0.00	1	0.02	0	0.00	0	0.00
80	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
624	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
398	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
349	1	0.00	1	0.00	1	0.02	0	0.00	0	0.00	0	0.00
1431	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
641	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
1494	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
135	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
1722	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
665	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
2076	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
93	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
404	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
746	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
2301	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
773	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
2914	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
372	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
3167	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
937	1	0.00	1	0.00	1	0.02	0	0.00	0	0.00	0	0.00
3268	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
998	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
5762	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
1057	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
6355	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
1140	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
7644	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
1147	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
8181	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
1177	1	0.00	1	0.00	1	0.02	0	0.00	0	0.00	0	0.00
8187	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
23	1	0.00	1	0.00	1	0.02	0	0.00	0	0.00	0	0.00
8189	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
1196	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
101	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
569	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01

Table S3. Proportions of Sequence Clusters (SCs) by PopPUNK. Prop: proportion. Sequence Cluster number is based on the predominant ST in the cluster. Other category contains all singletons.

					Clinical isolates. consisting of:							
	All		**Clinical**		*Blood*		*HA-urine*		*CA-urine*		**Faeces**	
SC	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*
Isolates	451		289		65		49		175		162	
Unique SC	75		46		19		17		35		53	
SC131	157	0.35	117	0.40	29	0.45	19	0.39	69	0.39	40	0.25
SC38	52	0.12	31	0.11	7	0.11	5	0.10	19	0.11	21	0.13
SC10	38	0.08	21	0.07	2	0.03	7	0.14	12	0.07	17	0.10
SC88	21	0.05	15	0.05	7	0.11	3	0.06	5	0.03	6	0.04
SC1193	21	0.05	17	0.06	0	0.00	0	0.00	17	0.10	4	0.02
SC648	12	0.03	7	0.02	2	0.03	1	0.02	4	0.02	5	0.03
SC405	12	0.03	6	0.02	1	0.02	1	0.02	4	0.02	6	0.04
SC58	10	0.02	7	0.02	0	0.00	1	0.02	6	0.03	3	0.02
SC69	10	0.02	6	0.02	2	0.03	0	0.00	4	0.02	4	0.02
SC162	6	0.01	4	0.01	0	0.00	3	0.06	1	0.01	2	0.01
SC394	5	0.01	1	0.00	0	0.00	1	0.02	0	0.00	4	0.02
SC95	5	0.01	5	0.02	0	0.00	0	0.00	5	0.03	0	0.00
SC117	4	0.01	2	0.01	1	0.02	0	0.00	1	0.01	2	0.01
SC501	4	0.01	3	0.01	1	0.02	1	0.02	1	0.01	1	0.01
SC224	4	0.01	4	0.01	2	0.03	0	0.00	2	0.01	0	0.00
SC12	4	0.01	3	0.01	1	0.02	0	0.00	2	0.01	1	0.01
SC141	4	0.01	2	0.01	0	0.00	0	0.00	2	0.01	2	0.01
SC636	4	0.01	2	0.01	0	0.00	0	0.00	2	0.01	2	0.01
SC357	3	0.01	3	0.01	1	0.02	1	0.02	1	0.01	0	0.00
SC354	3	0.01	3	0.01	2	0.03	1	0.02	0	0.00	0	0.00
SC93	3	0.01	1	0.00	0	0.00	0	0.00	1	0.01	2	0.01
SC156	3	0.01	1	0.00	1	0.02	0	0.00	0	0.00	2	0.01
SC393	3	0.01	0	0.00	0	0.00	0	0.00	0	0.00	3	0.02
SC783	2	0.00	2	0.01	2	0.03	0	0.00	0	0.00	0	0.00
SC453	2	0.00	1	0.00	1	0.02	0	0.00	0	0.00	1	0.01
SC127	2	0.00	2	0.01	0	0.00	1	0.02	1	0.01	0	0.00
SC457	2	0.00	2	0.01	0	0.00	0	0.00	2	0.01	0	0.00
SC73	2	0.00	1	0.00	0	0.00	0	0.00	1	0.01	1	0.01
SC2197	2	0.00	2	0.01	0	0.00	1	0.02	1	0.01	0	0.00
SC349	2	0.00	1	0.00	1	0.02	0	0.00	0	0.00	1	0.01
SC609	2	0.00	1	0.00	1	0.02	0	0.00	0	0.00	1	0.01
SC602	2	0.00	1	0.00	0	0.00	1	0.02	0	0.00	1	0.01
SC1304	2	0.00	1	0.00	0	0.00	1	0.02	0	0.00	1	0.01
SC665	2	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2	0.01
Other	41	0.09	14	0.05	1	0.02	1	0.02	12	0.07	27	0.17

Table S4. Proportions of ESBL genes. Prop: proportion.

					Clinical isolates, consisting of:							
	All		**Clinical**		*Blood*		*HA-urine*		*CA-urine*		**Faeces**	
ESBL gene	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*	*n*	*prop*
Isolates	451		290		66		49		175		162	
Unique gene	16	0.04	15	0.06	11	0.17	9	0.18	12	0.07	10	0.07
$bla_{\text{CTX-M-15}}$	214	0.47	146	0.50	30	0.45	27	0.55	90	0.51	68	0.42
$bla_{\text{CTX-M-14}}$	75	0.17	48	0.17	14	0.21	9	0.18	25	0.14	27	0.17
$bla_{\text{CTX-M-27}}$	73	0.16	50	0.17	9	0.14	6	0.12	35	0.20	23	0.14
$bla_{\text{CTX-M-1}}$	44	0.10	15	0.05	6	0.09	2	0.04	7	0.04	29	0.18
$bla_{\text{CTX-M-3}}$	12	0.03	6	0.02	1	0.02	1	0.02	4	0.02	6	0.04
$bla_{\text{CTX-M-55}}$	9	0.02	6	0.02	1	0.02	0	0.00	3	0.02	3	0.02
$bla_{\text{SHV-12}}$	8	0.02	6	0.02	1	0.02	2	0.04	3	0.02	2	0.01
$bla_{\text{CTX-M-32}}$	5	0.01	4	0.01	0	0.00	0	0.00	4	0.02	1	0.01
$bla_{\text{TEM-52}}$	3	0.01	1	0.00	1	0.02	0	0.00	0	0.00	2	0.01
$bla_{\text{CTX-M-9}}$	3	0.01	3	0.01	2	0.03	0	0.00	1	0.01	0	0.00
$bla_{\text{CTX-M-2}}$	2	0.00	2	0.01	0	0.00	1	0.02	1	0.01	0	0.00
$bla_{\text{TEM-28}}$	1	0.00	1	0.00	1	0.02	0	0.00	0	0.00	0	0.00
$bla_{\text{CTX-M-65}}$	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
$bla_{\text{CTX-M-73}}$	1	0.00	1	0.00	0	0.00	0	0.00	1	0.01	0	0.00
$bla_{\text{CTX-M-192}}$	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01
$bla_{\text{TEM-10}}$	1	0.00	1	0.00	0	0.00	1	0.02	0	0.00	0	0.00
2 genes	2	0.00	2	0.01	0	0.00	1	0.02	1	0.01	0	0.00

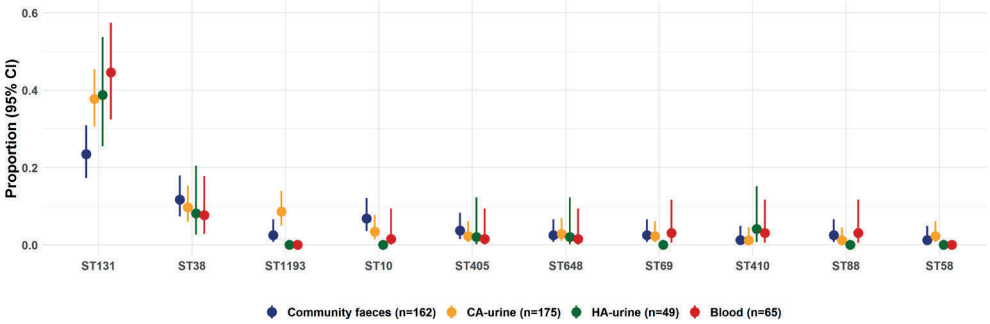

Figure S2. Proportions of the 10 most frequent sequence types (STs) in different sample groups.

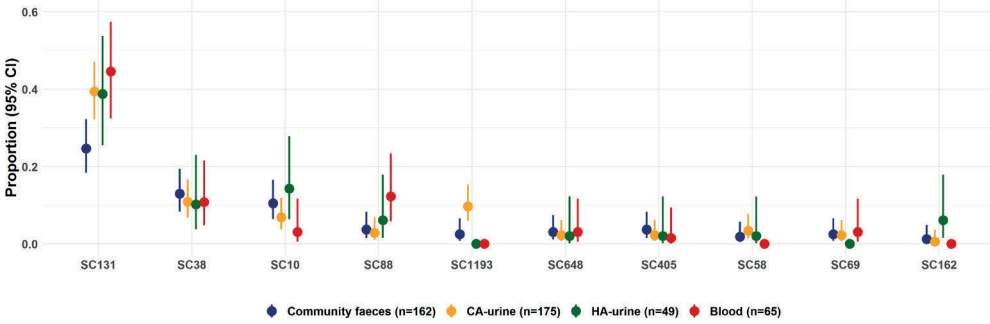

Figure S3. Proportions of the 10 most frequent sequence clusters (SCs) in different sample groups.

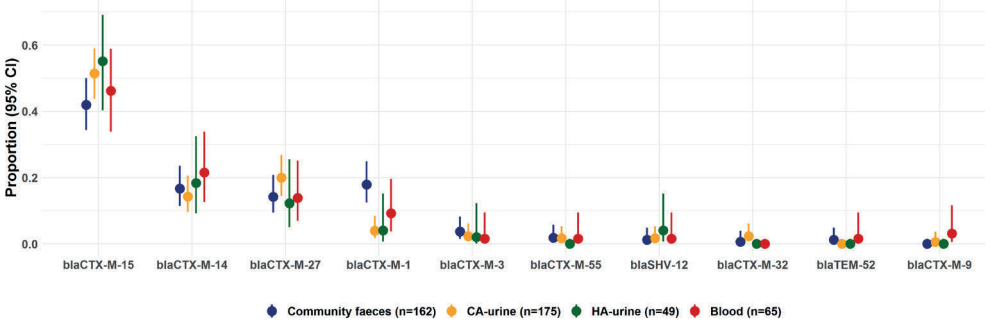

Figure S4. Proportions of the 10 most frequent ESBL gene types in different sample groups.

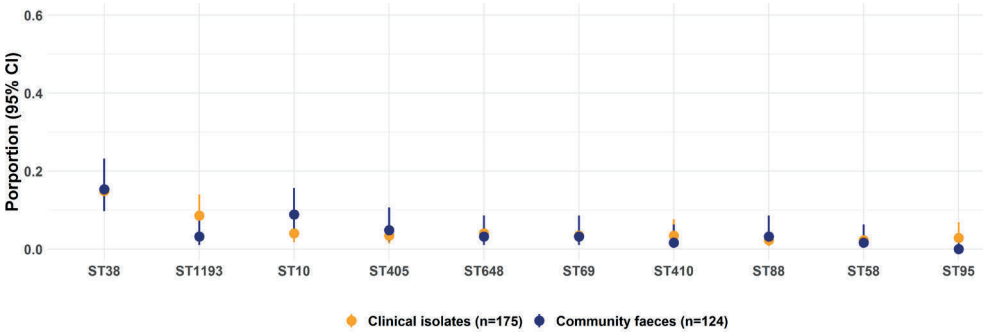

Figure S5. Sensitivity analysis excluding ST131 from dataset. Proportions of the 10 most frequent sequence types (STs) in clinical isolates and community faeces

Figure S6. Sensitivity analysis excluding ST131 from dataset. Proportions of the 10 most frequent sequence clusters (SCs) in clinical isolates and community faeces

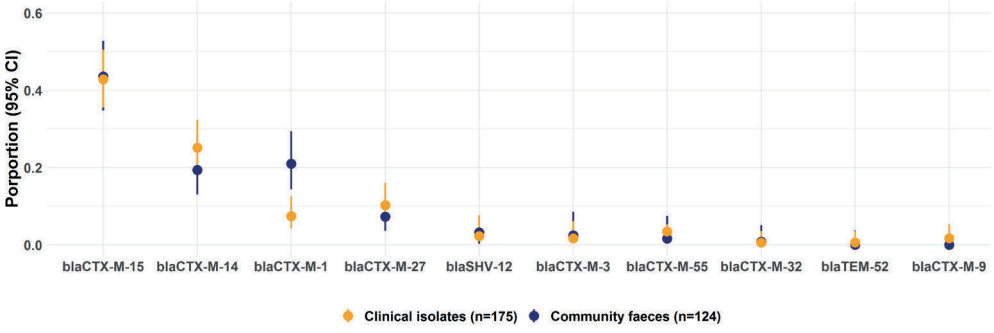

Figure S7. Sensitivity analysis excluding ST131 from dataset. Proportions of the 10 most frequent ESBL gene types in clinical isolates and community faeces

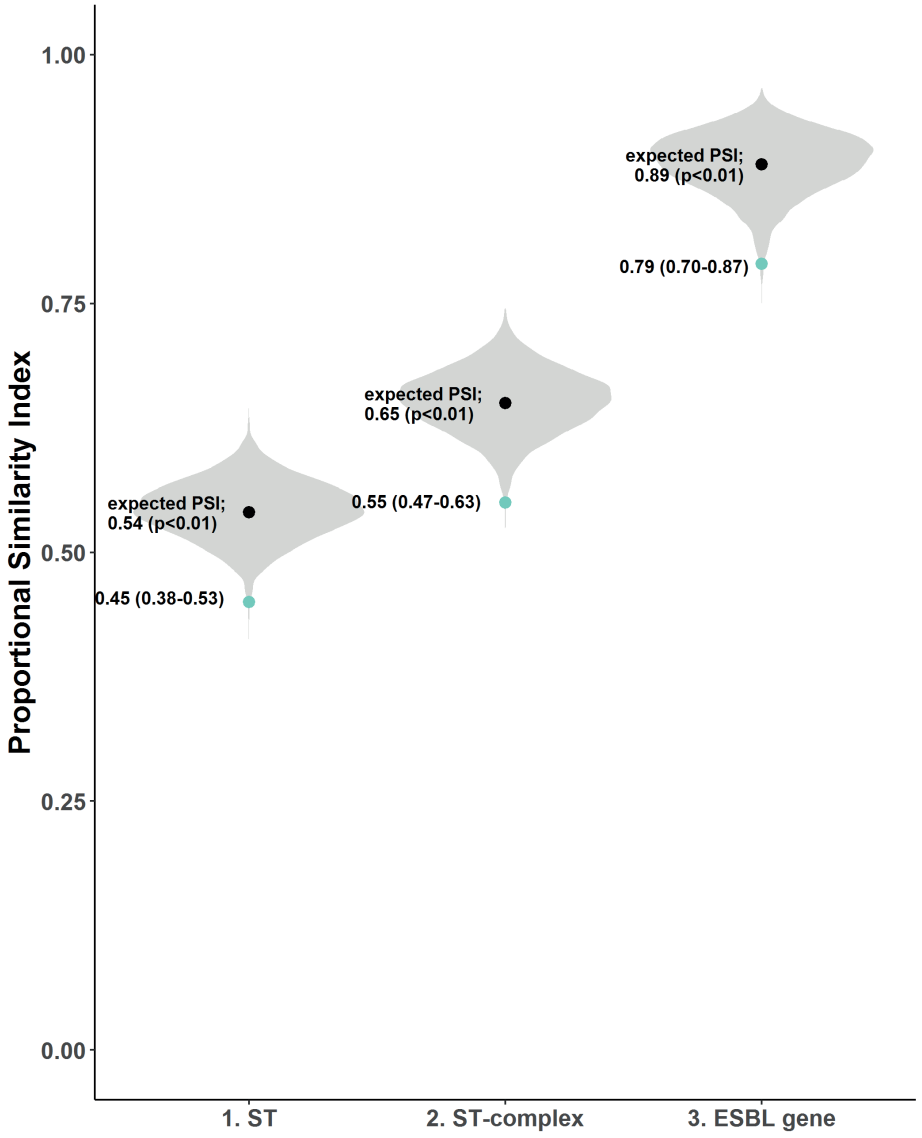

Figure S8. Sensitivity analysis excluding ST131 from dataset. Mean observed proportional similarity index (PSI), interpret as proportion of overlap between community faeces with clinical isolates (green), 95% CI (text) calculated with 5000 bootstrap iterations. Permutation distribution with mean expected PSI under the null-hypothesis (i.e. no difference between groups) (grey), calculated with 5000 permutations. p-value (text): permutation test; chance of the observed PSI under the null hypothesis (i.e. no difference between community faeces and clinical isolates).

4

External validation of WGS-based antimicrobial susceptibility prediction tools, KOVER-AMR and ResFinder 4.1 for *Escherichia coli* clinical isolates

Tess D Verschuuren, Thijs Bosch, Valentina Mascaro,
Rob JL Willems, Jan AJW Kluytmans

Manuscript in preparation

ABSTRACT

Objective: To externally validate whole genome sequence-antimicrobial susceptibility testing (WGS-AST) phenotype prediction tools KOVER-AMR and ResFinder 4.1 for *Escherichia coli* clinical isolates from Dutch routine care.

Methods: A random sample of 235 *E. coli*, and 283 3rd generation cephalosporin-resistant *E. coli* isolates from urine and blood were collected (2014-17). Culture-AST was performed using VITEK 2 and BD Phoenix. Sequences were used as input for KOVER-AMR-SCM, KOVER-AMR-CART and ResFinder 4.1. The sensitivity, specificity, positive predictive value, negative predictive value, concordance, major error rate (MER), and very major error rate (VMER) were calculated, with subsequent comparison to U.S. Food and Drug Administration (FDA) criteria (MER ≤3%, and VMER with a 95% confidence interval (CI) ≤1.5%-≤7.5%).

Results: ResFinder 4.1 performed better than KOVER-AMR-models, however, neither tools achieved (V)MERs below FDA criteria. KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1, MER (cumulative all antimicrobials) were: 5.2%, 4.7%, and 5.2%, respectively. MERs ≤3% were achieved for 6/11 tested antimicrobials for KOVER-AMR-models, and for 9/13 antimicrobials tested with ResFinder 4.1. KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1, cumulative VMERs were: 26% (24-26), 29% (27-31), and 10% (9.1-12). VMERs with a 95%CI ≤1.5-≤7.5 were only achieved only for 4/13 tested antimicrobials with ResFinder 4.1.

Conclusions: In this study, WGS-AST phenotype prediction tools, KOVER-AMR and ResFinder 4.1, did not meet the FDA criteria, (i.e. MER ≤3%, and VMER 95%CI ≤1.5%-≤7.5%), needed for clinical diagnostic use in 518 *E. coli* clinical isolates from Dutch routine care. The tested tools should be further improved before they can be used for clinical decision making.

INTRODUCTION

Escherichia coli is the most important cause of urinary tract-, bloodstream-, and antimicrobial resistant infections in Europe.[1,2] Culture-based antimicrobial susceptibility testing (AST) (culture-AST) is performed to provide adequate treatment for *E. coli* infections.[3,4] Culturing followed by AST is affordable, but takes 2-3 days, as the process depends on bacterial growth. Furthermore, results can vary between laboratories, re-analysis is labour-intensive, and storage of bacteria requires freezer capacity.[3-5] Awaiting culture-AST results, a patient is treated with empiric broad-spectrum therapy. As a consequence, patients can be temporarily over-treated (resulting in unnecessary antimicrobial use), or under-treated (resulting in potential adverse patient outcomes).

Whole genome sequencing (WGS), at least in theory, provides the possibility to replace the practice of empiric therapy with point of care AST-directed therapy, and solve issues of between-lab comparability, re-analysis, and storage.[6-8] For this, the following developments are needed in clinical microbiology: (i) high through-put sequencing and analyses workflows[5], (ii) development of methodologies that allow sequencing directly on a clinical sample, especially on material that is sterile under normal conditions (e.g. blood),[9-11] and (iii) WGS-based predictive AST (WGS-AST) tools that provide direct phenotype predictions (e.g. amoxicillin resistance yes/no), without interpretation of genomic content (e.g. presence of $bla_{\text{TEM-1}}$), which will save time and prevent interpretation errors.[12] Furthermore, WGS-AST technologies should adhere to quality criteria set by the U.S. Food and Drug Administration (FDA). These are expressed as a major error rate (MER) of ≤3% (i.e. susceptible isolates falsely predicted as resistant, resulting in over-treatment), and a very major error rate (VMER) including an upper 95% confidence interval (95%CI) limit of ≤7.5% for the true VMER, and a lower 95%CI limit of ≤1.5%, further referred to as 95%CI of ≤1.5%-≤7.5% (i.e. resistant isolates falsely predicted as susceptible, resulting in under-treatment).[13]

Proof of principle studies have shown promising results of WGS-AST for several pathogens.[14,15] Unfortunately, (V)MERs are infrequently reported, and external validation studies are scarce.[12,14,16] KOVER-AMR[17] and ResFinder 4.1[7] are two tools that provide direct phenotype predictions. Both tools observed a high concordance with culture-AST in the original studies of 95% (KOVER-AMR), and 98% (ResFinder 4.1).[7,17] Here, we present an external validation of KOVER-AMR and ResFinder 4.1 using *E. coli* isolates recovered from infections in Dutch routine clinical care.

METHODS

Study design

Sample collection of included isolates were described previously by van Hout *et al* and Verschuuren *et al*.[18,19] In short, a random sample of 235 *E. coli* isolates from blood, and 283 3rd generation cephalosporin-resistant (3rdGCR)-*E. coli* from urine and blood were included from the years 2014-2017, from two hospitals, and one primary care laboratory: the University Medical Center Utrecht (UMCU), a tertiary centre, the Amphia hospital, a large teaching centre, and Saltro, a laboratory providing services to primary care in the Utrecht region.

Ethics

The study was judged outside the scope of the Medical Research Involving Human Subjects Act by the ethics review board of the UMCU (IRB number 18/056). Informed consent was not obtained based on the 'Code of conduct for health research'.[20]

Culture-AST

Phenotypic susceptibilities were determined using VITEK 2 (Amphia hospital and Saltro), and BD Phoenix (UMCU), and extracted for 14 antimicrobials (table 1).3 Minimum Inhibitory Concentrations (MICs) were converted to susceptible or resistant based on the EUCAST clinical MIC breakpoints for *Enterobacterales* (v9.0), intermediate results were considered resistant.[21] In consultation with a medical microbiologist, the ceftriaxone phenotypes from the UMCU were relabelled as cefotaxime, based on their similar working profile. In 11 of the 518 cultures (all blood) >1 phenotypically distinct antimicrobial resistance *E. coli* isolates were retrieved. For these cultures, the isolate with the phenotype that corresponded with the WGS-AST predictions was selected for further analysis.

WGS-AST

KOVER, a freely available, supervised machine learning algorithm, was developed and used by Drouin *et al* to produce two rule-based models for WGS-AST phenotype prediction: KOVER-AMR-Set Covering Machines (SCM) and KOVER-AMR-Classification and Regression Trees (CART), using a public genotype-phenotype database, based on the presence or absence of certain k-mers (a string of DNA with the length of *k*).[17,22] Models are available for 12 bacterial species, for 56 antimicrobial treatment options.[17] Available *E. coli*-models were accessed at https://github.com/aldro61/kb_kover_amr/tree/master/data/models, and imported in R Studio.

Bortolaia *et al* published an update of ResFinder, a freely available, and centrally curated, reference database of acquired resistance genes and point mutations.[7] This update included a genotype-to-phenotype translation for each resistance predictor in the database. ResFinder 4.1 was accessed at https://cge.cbs.dtu.dk/services/ResFinder/. Assemblies were uploaded with the following default settings: (i) search for chromosomal point mutations with an %ID of at least 90%, and a minimal length of at least 60%. (ii) search for acquired antimicrobial resistance genes in all antimicrobial classes with an %ID of at least 90%, and a minimal length of at least 60%.

Statistical analyses

The resulting output of both tools was a prediction of resistance (yes/no) for a particular antimicrobial per isolate. With this, the concordance, MER, and VMER were calculated (formulas in table 1). Furthermore, the sensitivity, specificity, positive predictive value, negative predictive value were additionally calculated (results table S1-S3). 95%CIs were calculated with bootL R package (v1.0.2) or a test of given proportions. Calculations were performed with R Studio. MERs and VMERs were compared to FDA criteria (MER ≤3%, and VMER with a 95%CI of ≤1.5%-≤7.5%).

Table 1. Formulas used to calculate outcome measures

outcome measure	formula
concordance (%)	(***n*** true positives (TP) +***n*** true negatives (TN)) /(***n*** TP +***n*** TN +***n*** false positives (FP) +***n*** false negatives (FN))
sensitivity (%)	***n*** TP/(***n*** TP +***n*** FN)
specificity (%)	***n*** TN/(***n*** TN +***n*** FP)
PPV[a] (%)	***n*** TP/(***n*** TP +***n*** FP)
NPV[b] (%)	***n*** TN/(***n*** TN+***n*** FN)
ME rate[c] (%)	(***n*** FP/(***n*** TN+***n*** FP))*100
VME rate[d] (%)	(***n*** FN/(***n*** TP+***n*** FN))*100

[a] PPV: positive predictive value
[b] NPV: negative predictive value
[c] ME rate: major error rate (i.e. false resistant rate)
[d] VME rate: very major error rate (i.e. false susceptible rate)

RESULTS

Phenotypic resistance rates of the included isolates ranged from 0% for meropenem to 76% for amoxicillin (table 2). Overall, ResFinder 4.1 performed better than KOVER-AMR-models, and provided predictions for more of the 14 assessed antimicrobials (13 versus 11) (table 3). However, neither of the tools achieved overall (V)MERs below FDA thresholds, with only ResFinder 4.1 predictions for amoxicillin meeting both criteria (MER: 1.6% (0.5-5.8), VMER: 1.8% (0.9-3.6)) (table 3).

Table 2. Observed phenotypic resistance percentages of included antimicrobials.

antimicrobial class	**antimicrobial**	**dataset 1&2 (n=518) R% (n)**	**dataset 1 (n=235) R% (n)**	**dataset 2 (n=283) R% (n)**
penicillins (small-spectrum)	amoxicillin	76% (396)	48% (113)	100% (283)
penicillins (broad-spectrum)	amoxicillin-clavulanic acid	47% (244)	23% (54)	67% (190)[a]
	piperacillin-tazobactam	11% (57)	8% (18)	14% (39)
cephalosporins	cefuroxime	60% (313)	13% (30)	100% (283)
	cefotaxime	59% (307)	11% (25)	99.6% (282)
	ceftazidime	48% (249)	9% (21)	81% (228)[a]
	cefepime	31% (101)	9% (20)[b]	82% (81)[c]
carbapenems	meropenem	0%	0%	0%[a]
	imipenem	0.2% (1)	0%[d]	0.4% (1)[e]
fluorquinolones	ciprofloxacin	49% (252)	18% (43)	74% (209)[a]
aminoglycosides	gentamicin	20% (103)	9% (20)	29% (83)
	tobramycin	24% (120)	8% (18)	39% (102)[f]
miscellaneous agents	cotrimoxazol	46% (237)	28% (65)	61% (172)[a]
	colistin	0.4% (2)	0%	2% (2)[h]

Dataset 1: a random sample of available *E. coli* isolates retrieved from blood from two hospitals. Dataset 2: 3rd generation cephalosporin resistant *E. coli* from retrieved from urine and blood from two hospitals and one primary care laboratory.
[a] phenotype available for 282 isolates
[b] phenotype available for 226 isolates
[c] phenotype available for 99 isolates
[d] phenotype available for 220 isolates
[e] phenotype available for 267 isolates
[f] phenotype available for 260 isolates
[h] phenotype available for 100 isolates

Overall MERs for KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1, were : 5.2% (4.5-6.0), 4.7% (4.0-5.5), and 5.2% (4.5-5.9), respectively (i.e. >3%). However, for KOVER-AMR-models 6/11 assessed antimicrobials had MERs below the FDA threshold (table 3), and 9/13 antimicrobials of ResFinder 4.1 predictions MERs were below 3%, indicating acceptable MERs for the majority of assessed antimicrobials (table 3).

Table 3. KOVER-AMR-SCM, KOVER-AMR-CART and ResFinder 4.1 for the phenotypic antimicrobial susceptibility pattern of *E. coli* UTI & bacteraemia (*n*=518)

	concordance % (95% CI)			ME rate % (95% CI)			VME rate % (95% CI)		
A[a]	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1
AMO *(SSP)*	75 (71-79)	75 (71-79)	98 (97-99)	3.3 (1.3-8.1)	3.3 (1.3-8.1)	1.6 (0.5-5.8)	32 (27-36)	32 (27-36)	1.8 (0.9-3.6)
AMC	72 (68-76)	75 (71-79)	72 (68-76)	3.7 (2.0-6.6)	5.8 (3.6-9.2)	2.6 (1.2-5.2)	55 (49-61)	47 (40-53)	57 (50-63)
PITA	84 (80-87)	84 (80-87)	81 (78-84)	7.2 (5.1-9.9)	7.2 (5.1-9.9)	16 (13-20)	88 (77-94)	88 (77-94)	39 (27-52)
BRSP	78 (75-81)	79 (77-82)	76 (46-52)	5.9 (4.4-7.8)	6.7 (5.1-8.7)	11 (9.1-14)	61 (56-67)	55 (49-60)	54 (48-60)
CER	96 (94-98)	96 (94-98)	NA[b]	0 (0-1.8)	0 (0-1.8)	NA[b]	6.4 (4.2-9.7)	6.4 (4.2-9.7)	NA[f]
CFO	97 (95-98)	97 (95-98)	97 (95-98)	0 (0-1.8)	0 (0-1.8)	4.3 (2.3-7.9)	5.2 (3.2-8.3)	5.2 (3.2-8.3)	2.6 (1.3-5.0)
CFZ	79 (75-82)	64 (59-68)	87 (83-89)	2.2 (1.0-4.8)	6.3 (4.0-9.9)	24 (19-29)	41 (35-48)	69 (63-74)	2.0 (0.9-4.6)
CFE	72 (67-77)	78 (73-82)	92 (88-95)	22 (17-28)	11 (7.7-16)	10 (6.9-15)	40 (31-49)	48 (38-57)	3.0 (1.0-8.4)
CEPH	87 (86-89)	84 (82-86)	92 (90-93)	6.2 (4.8-7.9)	4.6 (3.4-6.2)	14 (11-16)	19 (16-21)	26 (24-29)	2.3 (1.4-3.7)
MER	88 (84-90)	89 (86-92)	100 (99-100)	12 (10-16)	11 (8.1-14)	0 (0-0.9)	-[c]	-[c]	-[c]
IMI	NA[b]	NA[b]	100 (99-100)	NA[b]	NA[b]	0 (0-0.9)	NA[b]	NA[b]	100 (5.5-100)
CARB	88 (84-90)	89 (86-92)	100 (99-100)	12 (10-16)	11 (8.1-14)	0 (0-0.6)	-[c]	-[c]	100 (5.5-100)
CIP *(FQ)*	92 (88-94)	92 (88-94)	96 (94-97)	0 (0-1.4)	0 (0-1.4)	8.3 (5.5-12)	18 (14-23)	18 (14-23)	0 (0-1.5)
GEN	96 (94-98)	96 (94-98)	98 (96-99)	0.5 (0.1-1.7)	0.5 (0.1-1.7)	0.5 (0.1-1.7)	17 (11-25)	17 (11-25)	9.7 (1.7-15)
TOB	97 (95-98)	97 (95-98)	98 (96-99)	1.3 (0.6-3.1)	1.3 (0.6-3.1)	1.9 (0.9-3.8)	8.3 (4.6-15)	8.3 (4.6-15)	4.2 (1.8-9.4)
AMIN	97 (95-98)	97 (95-98)	98 (96-98)	0.9 (0.4-1.8)	0.9 (0.4-1.8)	1.1 (0.6-2.2)	12 (8.5-17)	12 (8.5-17)	6.7 (4.1-11)
COT	NA[b]	NA[b]	97 (94-98)	NA[b]	NA[b]	2.1 (1.0-4.6)	NA[b]	NA[b]	5.1 (2.9-8.6)
COL	NA[b]	NA[b]	99 (96-99)	NA[b]	NA[b]	0.9 (0.2-2.8)	NA[b]	NA[b]	100 (20-100)
MISC	NA[b]	NA[b]	97 (96-98)	NA[b]	NA[b]	1.5 (0.8-2.8)	NA[b]	NA[b]	5.9 (3.5-9.6)
ALL	**87 (86-88)**	**86 (85-87)**	**93 (92-94)**	**5.2 (4.5-6.0)**	**4.7 (4.0-5.5)**	**5.2 (4.5-5.9)**	**26 (24-28)**	**29 (27-31)**	**10 (9.1-12)**

[a] antimicrobial: AMO: amoxicillin, SSP: small-spectrum penicillin (AMO), AMC: amoxicillin-clavulanic acid, PITA: piperacillin-tazobactam, BRSP: broad-spectrum penicillins (sum AMO, PITA), CER: cefuroxime, CFO: cefotaxime, CFZ: ceftazidime, CFE: cefepime, CEPH: cephalosporins (sum CER, CFO, CFZ, CFE), MER: meropenem, IMI: imipenem, CARB: carbapenems (sum MER, IMI), CIP: ciprofloxacin, FQ: fluorquinolones (CIP), GEN: gentamicin, TOB: tobramycin, AMIN: aminoglycosides (sum GEN, TOB) COT: cotrimoxazol, COL: colistin, MISC: miscellaneous agents (sum COT, COL). [b] antimicrobial was not available in the phenotype prediction tool. [c] output not calculated, due to absence of phenotypic resistant isolates in dataset. [d]calculation: sum of all observed true positives, true negatives, false positives and false negatives

Observed MERs were particularly high for cefepime (11-22%), and meropenem (11-12%) for KOVER-AMR-models, and for piperacillin-tazobactam 16% (13-20), and ceftazidime 24% (19-29) for ResFinder 4.1 (table 3).

KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1, VMERs for all antimicrobials were: 26% (24-26), 29% (27-31), and 10% (9.1-12), exceeding the FDA threshold. KOVER-AMR-models did not achieve VMERs below FDA thresholds for any of the assessed antimicrobials, while for ResFinder 4.1, VMERs for 4/13 antimicrobials were within the limits of the FDA threshold, namely amoxicillin, cefotaxime, ceftazidime, and ciprofloxacin. For all models, observed VMERs were particularly high for the assessed broad-spectrum penicillins: amoxicillin-clavulanic acid (47-57%), and piperacillintazobactam (39-88%) (table 3).

Lastly, comparing results from the random sample of *E. coli*, and 3rdGCR-*E. coli*, showed an overall higher concordance for the first: KOVER-AMR-SCM: 92% versus 82%, KOVER-AMR-CART: 93% versus 80%, ResFinder 4.1: 96% versus 91% (table S1-S2). Furthermore, overall MERs were lower in the random sample, while VMERs were higher, compared to the 3rdGCR-*E. coli* (table S1-S2). Indicating difference in tool performance in different datasets with amongst others different resistance prevalences.

Assessment of discrepancies

In total, 730 discrepancies in genotype-to-phenotype predictions were observed for KOVER-AMR-SCM, 774 for KOVER-AMR-CART, and 424 for ResFinder 4.1 (table 4). Grouping these into possible explanatory categories, showed that 36% of all discrepancies occurred in isolates with a MIC with a factor 2 above or below the clinical breakpoint. Presence of a genetic predictor that did not predict a resistant phenotype occurred evenly between the tools, while both KOVER-AMR-models were more likely to miss a genetic predictor for a resistant phenotype compared to ResFinder 4.1 (table 4). For amoxicillin-clavulanic acid 376 VMEs were observed, where often all tools failed to predict resistance (table 3, S4). For piperacillin-tazobactam 261 discrepancies were observed, MEs for piperacillin-tazobactam occurred more often in ResFinder 4.1 (table 3,S4), where presence of the $bla_{\text{OXA-1}}$ gene often did not translate into phenotypic resistance against this penicillin/inhibitor combination. VMEs for piperacillin-tazobactam occurred more often for both KOVER-AMR tools. Lastly, ceftazidime,

cefepime, and meropenem models of KOVER-AMR-SCM and/or KOVER-AMR-CART contained genetic predictors for resistance that did not resulted in phenotypic resistance in our strain sets (table 3,S4).

Table 3. Summarization of observed discrepancies ((V)MEs) in possible categories.

discrepancy category	KOVER-AMR-SCM % (n)	KOVER-AMR-CART % (n)	ResFinder 4.1 % (n)
MIC around clinical breakpoint[a]	36% (262)	35% (269)	36% (154)
likely incorrect genetic predictor[b]	20% (148)	16% (126)	33% (139)
possible incorrect genetic predictor or incorrect phenotype[c]	1% (9)	1% (9)	4% (18)
likely missing genetic predictor[d]	32% (237)	38% (296)	6% (26)
possible missing genetic predictor or incorrect phenotype[e]	10% (74)	10% (74)	21% (87)
total	100% (730)	100% (774)	100% (424)

[a] MIC around clinical breakpoint: minimum inhibitory concentration was ≤2x above/below the clinical breakpoint[20]
[b] likely incorrect genetic predictor (ME) if ≥1 of the assessed tools predicted the isolate as susceptible
[c] possible incorrect genetic predictor or incorrect phenotype (ME) if all of the assessed tools predicted the isolate as resistant
[d] likely missing genetic predictor (VME) if ≥1 of the assessed tools predicted the isolate as resistant
[e] possible missing genetic predictor or incorrect phenotype (VME) if all of the assessed tools predicted the isolate as susceptible

DISCUSSION

This study externally validated two WGS-AST tools: KOVER-AMR-models, and ResFinder 4.1, that provide direct phenotype predictions, in 518 *E. coli* isolates recovered from patients with infections in the Netherlands. Overall, ResFinder 4.1 performed better than KOVER-AMR-models, however, neither tools achieved cumulative (V)MERs below FDA thresholds for the total assessed panel of antimicrobials. MERs ≤3% were achieved for 6/11 tested antimicrobials for KOVER-AMR-models, and for 9/13 antimicrobials tested with ResFinder 4.1. VMERs with a 95% CI ≤1.5-≤7.5 was achieved in none of the tested antimicrobials for KOVER-AMR, and only for 4/13 tested antimicrobials with ResFinder 4.1. Only phenotype predictions for amoxicillin resistance by ResFinder 4.1 achieved both FDA criteria.

We compared the observed (V)MERs and concordances with the reported internal validations of these tools (table S5-S6).[7,17] The phenotypic resistance percentages of the *E. coli* dataset used for internal validation of KOVER-AMR-models were comparable to this study, while phenotypic resistance percentages of the datasets used for ResFinder 4.1 were slightly higher (table S5).[7,17] The reported overall MERs of KOVER-AMR-models, and ResFinder 4.1 were lower than our observations: 0.6-2.1% versus 4.7-5.2%, and 0.2% versus 5.2%, respectively.

Furthermore, the reported overall VMERs of KOVER-AMR-models, and ResFinder 4.1 were also lower than our observations: 21-25% versus 26-29%, and 3.2% versus 5.2%, respectively (table S6).[7,17] The observed differences between the reported internal validations of these tools and the results observed in this external validation can have several explanations.

Firstly, Bortolaia *et al*[7] did not include amoxicillin-clavulanic acid or piperacillin-tazobactam, for which we observed the highest (V)MERs within the assessed antimicrobials for ResFinder 4.1 predictions. Secondly, KOVER-AMR-models were the result of supervised machine learning algorithm KOVER, where KOVER selected *k*-mers that were able to predict resistance to a certain antimicrobial. However, by default, no underlying biological mechanism (i.e. causal relation) is needed for this prediction. The risk of this methodology is that the model might not predict resistance in a new dataset, as the selected genetic predictors were associated with resistance in the dataset used for model training. We observed this for the models predicting resistance against cefepime, and meropenem. Furthermore, if the training dataset of KOVER-AMR-models had an absence or too low prevalence of certain resistance genes/mutations, these will not be included as genetic resistance predictors.[14] We observed this in for example in the models predicting amoxicillin resistance, were $bla_{\text{CTX-M}}$ genes were missed. Both issues can be overcome when the training datasets used for KOVER are large and diverse enough, which will result in generalizability to a real-life setting (assuming that there are no large geographical differences in occurrence of resistance genes/mutations).[14] Indeed, the datasets used for training of cefepime and meropenem were among the smallest (*n*=426, and *n*=446, respectively), with low resistance rates (8%, and 6%, respectively), resulting in respective 32, and 28 isolates for KOVER to select resistance predictors. However, the great advantage of KOVER is that the machine learning tool can be used to discover new resistance traits, a feature that is absent from reference databases like ResFinder.[23]

To our knowledge this is first study that externally validated direct phenotype prediction tools: KOVER-AMR-models and ResFinder 4.1 for *E. coli* isolates recovered from clinical practice. We, additionally, compared our results to the FDA criteria needed for diagnostic use. Both are needed for progress to future implementation of WGS-AST tools in clinical practice.[12,24] Furthermore, we evaluated the tools in a random sample of *E. coli* blood isolates, representing a diverse and representative sample of Dutch clinical practice.[14] We also included 3rdGCR-*E. coli* isolates, the higher proportion of phenotypic resistance allowed us to more accurately calculate VMERs, arguably the most relevant outcome measure for patient care.[14]

This external validation study has several potential limitations. Firstly, we did not have the possibility to re-analyse discrepant phenotypes by repeating culture-AST, which could have contributed to the observed high (V)MERs. Gordon *et al* mentioned that in their study 40% of the discrepancies were solved by re-analysing discrepant phenotypes.[25] Indeed, 36% of the discrepancies in this study occurred in isolates with an MIC with a factor 2 above or below the clinical breakpoint. It is likely that a proportion of these isolates would have had a different MIC when retested. Furthermore, there was also the risk of human-induced errors that could not be excluded (e.g. mislabelling or errors during data handling). We did, however, not observe complete discordant MIC patterns, which would be a signal for mislabelling. Secondly, we did not use the gold-standard for culture-AST, namely broth microdilution.[3] Instead, we used VITEK 2, and BD Phoenix, two frequently used methodologies in clinical practice, which may have influenced the observed MICs. Thirdly, the sampling from a confined area in time and space (i.e. the Netherlands between 2014-2017) might have influenced the results. For example, the dataset could have contained a higher proportion of resistance traits that were missing from KOVER-AMR and ResFinder 4.1, which could have led to a lower performance of the tools.[19] Lastly, we were only able to include *E. coli* isolates in this study, and can therefore make no inferences on the performance of KOVER-AMR-models, and ResFinder 4.1 for other pathogens.

Although neither of the tools, at this stage, fulfil the criteria to be used for clinical decision making, the tools did provide concordant results in 86-87% and 93% when comparing culture-AST with KOVER-AMR, and ResFinder 4.1 phenotype predictions, respectively. This is in line with observed concordances in proof of principle WGS-AST studies.[14-16,26-28] These observations indicate that there are likely enough points of engagement for further improvement of these tools, or future WGS-AST technologies, to meet FDA criteria. For this the following route can be proposed. Firstly, collective effort should be put in making phenotype-genotype datasets available, as has been emphasized in literature.[14,29] These datasets should be used for a continues search for new resistance determinants with machine learning algorithms like for example KOVER, or recently developed methodologies like pyseer by Lees *et al* or INGOT-DR by Zabeti *et al*.[17,30,31] The resulting output should be used to update a centrally curated and freely available reference database like ResFinder 4.1. Secondly, the scientific community should focus on making direct sample sequencing applications feasible and affordable so that they can be implemented in clinical setting, starting with clinical samples that are normally sterile like blood.[9-11] Thirdly, more external validation studies for different pathogens

and datasets are needed, as well as clinical implementation studies in which culture-AST and WGS-AST are performed in parallel in routine practice.[14,24] Lastly, WGS-AST might not be able to capture all resistance predictors, as resistance might also be harboured in a bacterium without genetic changes.[15,32] This could for example be the case in epigenetic changes resulting in different expression of the same gene in different situations, (e.g. differential expression of efflux pumps).[32] An additional layer of interpretation of the WGS-AST results might be needed, that can overrule the prediction in certain cases.

In this study, WGS-AST phenotype prediction tools, KOVER-AMR and ResFinder 4.1, did not meet the FDA criteria needed for clinical diagnostic use in 518 *E. coli* infections from Dutch routine care. However, the tools can likely be improved by collective phenotype-genotype data-sharing, followed by continuous search for new resistance determinants with tools like KOVER, to update reference databases like ResFinder 4.1. This effort should be done as it has the potential to improve the quality and speed of obtaining information regarding the optimal treatment of infections.

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.

Acknowledgements: 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 also would sincerely like to thank Denise van Hout, Marc Bonten, and Ascelijn Reuland for their contributions in data collection. Finally we would like to thank Anita Schürch in her contribution to the assembly of the sequences.

Conflicts of interest: None to declare.

Data-availability: All generated raw reads were submitted to the European Nucleotide Archive (ENA) of the European Bioinformatics Institute (EBI) under the study accession number PRJEB35000. All phenotypes are available in the supplementary material.

REFERENCES

1. European Centre for Disease Prevention and Control. Surveillance of antimicrobial resistance in Europe 2018. Stockholm: ECDC; 2019. 110. Available from: https://www.ecdc.europa.eu/sites/default/files/documents/surveillance-antimicrobial-resistance-Europe-2018.pdf

2. Cassini, A, Högberg LD, Plachouras D *et al.* Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* 2019; **1**: 56–66.

3. Jorgensen JH, Ferraro MJ. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin Infect Dis* 2009; **49**: 1749-1755.

4. Török ME, Peacock SJ. Rapid whole-genome sequencing of bacterial pathogens in the clinical microbiology laboratory—pipe dream or reality? *J Antimicrob Chemother* 2012; **67**: 2307–8.

5. Deurenberg RH, Bathoorn E, Monika A *et al.* Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol* 2017; **243**: 16-24.

6. Andersson DI, Balaban NQ, Baquero F *et al.* Antibiotic resistance: turning evolutionary principles into clinical reality. *FEMS Microbiol Rev* 2020; **2**: 171-88.

7. Bortolaia V, Kaas RS, Ruppe E *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020; **12**: 3491-500.

8. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet* 2019; **6**: 341–355.

9. Su M, Satola SW, Read TD. Genome-based prediction of bacterial antibiotic resistance. *J Clin Microbiol* 2019; **3**: e01405-18.

10. Taxt AM, Avershina E, Frye SA, Naseer U, Ahmad R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. *Sci Rep* 2020; **1**: 7622.

11. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016; **7589**: 228–232.

12. Peiffer-Smadja N, Dellière S, Rodriguez C *et al.* Machine learning in the clinical microbiology laboratory: has the time come for routine practice? *Clin Microbiol Infect* 2020; **10**: 1300-130.

13. U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry and FDA. Class II Special Controls Guidance Document: Antimicrobial Susceptibility Test (AST) Systems. Rockville: FDA; 2009. 42. Available from: https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/antimicrobial-susceptibility-test-ast-systems-class-ii-special-controls-guidance-industry-and-fda

14. Hendriksen RS, Bortolaia V, Tate H *et al.* Using genomics to track global antimicrobial resistance. *Front Public Health* 2019; **7**: 242.

15. Moradigaravand D, Palm M, Farewell A *et al.* Prediction of antibiotic resistance in *Escherichia coli* from large-scale pan-genome data. *PLOS Comput Biol* 2018; **12**: 1–17.

16. Mahfouz N, Ferreira I, Beisken S *et al.* Large-scale assessment of antimicrobial resistance marker databases for genetic phenotype prediction: a systematic review. *J Antimicrob Chemother* 2020; **75**: 3099–108.

17. Drouin A, Letarte G, Raymond F *et al.* Interpretable genotype-to-phenotype classifiers with performance guarantees. *Sci Rep* 2019; **9**: 1-13.

18. van Hout D, Verschuuren TD, Bruijning-Verhagen PCJ *et al.* Extended-spectrum beta-lactamase (ESBL)-producing and non-ESBL-producing *Escherichia coli* isolates causing bacteremia in the Netherlands (2014 - 2016) differ in clonal distribution, antimicrobial resistance gene and virulence gene content. *PLoS One* 2020; **15**: e0227604.

19. Verschuuren TD, van Hout D, Arredondo-Alonso S *et al.* 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). *J Antimicrob Chem.* 2021; **4**: 901–908.

20. Federation of Medical Scientific Societies. FMWV Code of Conduct for Health Research. Rotterdam: Federa;2017. 14. Available from: https://www.federa.org/sites/default/files/bijlagen/coreon/code_of_conduct_for_medical_research_1.pdf.

21. The European Committee on Antimicrobial Susceptibility Testing 2019. Breakpoint tables for interpretation of MICs and zone diameters Version 9.0. Växjö: EUCAST; 2019. 100. Available from: https://eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_9.0_Breakpoint_Tables.pdf

22. Davis JJ, Wattam AR, Aziz RK *et al.*The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 2020; **D1**: D606-D612.

23. Falush D. Bacterial genomics: Microbial GWAS coming of age. *Nat Microbiol* 2016; **1**: 16059.

24. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: A review. *J Clin Epidemiol* 2008; **61**: 1085–94

25. Gordon NC, Price JR, Cole R *et al.* Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol* 2014; **52**: 1182–1191.

26. Feldgarden M, Brover V, Haft DH *et al*. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agen Chemother*2019; **11**: e00483-19.

27. Nguyen M, Brettin T, Long SW *et al*. Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Sci Rep* 2018; **8**: https://doi.org/10.1038/s41598-017-18972-w.

28. Davis JJ, Boisvert S, Brettin T *et al*. Antimicrobial resistance prediction in PATRIC and RAST. *Sci Rep* 2016; *6*:1-12.

29. Aarestup FM, Koopmans MG. Sharing data for global infectious disease surveillance and outbreak detection. *Trends Microbiol* 2016; **4**: 241-245.

30. Lees JA, Tien Mai T, Galardini M *et al*. Improved prediction of bacterial genotype-phenotype associations using interpretable pangenome-spanning regressions. *MBio* 2020; **4**: e01344-20.

31. Zabeti Z, Dexter N, Amir Safari AH *et al*. INGOT-DR: an interpretable classifier for predicting drug resistance in M. tuberculosis. *BioRxiv* 2021;doi: https://doi.org/10.1101/2020.05.31.115741

32. Pesesky MW, Hussain T, Wallace M *et al*. Evaluation of machine learning and rules-based approaches for predicting antimicrobial resistance profiles in gram-negative bacilli from whole genome sequence data. *Front Microbiol* 2016; **7**: doi: 10.3389/fmicb.2016.01887.

4

SUPPLEMENTARY MATERIAL

Table S1. The diagnostic performance of KOVER-AMR-SCM, KOVER-AMR-CART and ResFinder 4.1 for the phenotypic antimicrobial susceptibility pattern of *E. coli* UTI and bacteraemia (n=518).

	sensitivity % (95% CI)			**specificity % (95% CI)**			**PPV % (95% CI)**			**NPV % (95% CI)**		
A[a]	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1
AMO *(SSP)*	68 (64-73)	68 (64-73)	98 (96-99)	97 (92-99)	97 (92-99)	98 (94-100)	99 (96-99)	99 (96-99)	100 (98-100)	49 (42-56)	49 (42-56)	95 (89-97)
AMC	45 (39-51)	53 (47-60)	43 (37-50)	96 (93-98)	94 (91-96)	97 (95-99)	92 (85-95)	89 (83-93)	94 (88-97)	66 (62-71)	70 (65-74)	66 (61-70)
PITA	12 (6.1-23)	12 (6.1-23)	61 (48-73)	93 (90-95)	93 (90-95)	84 (80-87)	18 (8.7-32)	18 (8.7-32)	32 (24-41)	90 (87-92)	90 (87-92)	95 (92-96)
BRSP	39 (34-45)	45 (40-51)	46 (41-52)	94 (92-96)	93 (91-95)	89 (86-91)	73 (66-79)	74 (67-79)	63 (56-69)	79 (76-82)	81 (78-83)	80 (77-83)
CER	94 (90-96)	94 (90-96)	NA[b]	100 (98-100)	100 (98-100)	NA[b]	100 (99-100)	100 (99-100)	NA[b]	90 (87-94)	90 (87-94)	NA[b]
CFO	95 (92-97)	95 (92-97)	97 (95-99)	100 (95-98)	100 (95-98)	96 (92-98)	100 (99-100)	100 (99-100)	97 (95-99)	93 (89-96)	93 (89-96)	96 (93-98)
CFZ	59 (52-65)	31 (26-27)	98 (95-99)	98 (95-99)	94 (90-96)	76 (71-81)	96 (92-98)	82 (73-89)	79 (74-83)	72 (67-76)	60 (55-64)	97 (94-99)
CFE	60 (51-69)	53 (43-62)	97 (92-99)	78 (72-83)	89 (84-92)	90 (85-93)	55 (46-64)	68 (57-77)	81 (73-87)	81 (76-86)	81 (75-85)	99 (96-100)
CEPH	82 (79-84)	74 (71-76)	98 (96-99)	94 (92-95)	95 (94-97)	87 (84-89)	93 (92-95)	94 (93-96)	87 (85-89)	83 (80-85)	78 (75-80)	98 (96-99)
MER	-[c]	-[c]	-[c]	88 (84-90)	89 (86-92)	100 (99-100)	-[c]	-[c]	-[c]	-[c]	-[c]	-[c]
IMI	NA[b]	NA[b]	0 (0-95)	NA[b]	NA[b]	100 (99-100)	NA[b]	NA[b]	-[c]	NA[b]	NA[b]	100 (99-100)
CARB	-[c]	-[c]	0 (0-95)	88 (84-90)	89 (86-92)	100 (99-100)	-[c]	-[c]	-[c]	-[c]	-[c]	100 (99-100)
CIP *(FQ)*	82 (77-86)	82 (77-86)	100 (99-100)	100 (99-100)	100 (99-100)	92 (88-95)	100 (98-100)	100 (98-100)	92 (88-95)	86 (81-89)	86 (81-89)	100 (98-100)
GEN	84 (75-89)	84 (75-89)	90 (83-95)	100 (98-100)	100 (98-100)	100 (98-100)	98 (92-99)	98 (92-99)	98 (93-99)	96 (94-98)	96 (94-98)	98 (96-99)
TOB	92 (85-95)	92 (85-95)	96 (91-98)	99 (97-99)	99 (97-99)	98 (96-99)	96 (90-98)	96 (90-98)	94 (89-97)	97 (95-99)	97 (95-99)	99 (97-99)
AMIN	89 (83-92)	89 (83-92)	93 (89-96)	99 (98-100)	99 (98-100)	99 (98-99)	97 (93-98)	97 (93-98)	96 (92-98)	97 (95-98)	97 (95-98)	98 (97-99)
COT	NA[b]	NA[b]	95 (91-97)	NA[b]	NA[b]	98 (95-99)	NA[b]	NA[b]	97 (95-99)	NA[b]	NA[b]	96 (93-98)
COL	NA[b]	NA[b]	0 (0-80)	NA[b]	NA[b]	99 (97-100)	NA[b]	NA[b]	0 (0-70)	NA[b]	NA[b]	99 (98-100)
MISC	NA[b]	NA[b]	94 (90-97)	NA[b]	NA[b]	99 (97-99)	NA[b]	NA[b]	96 (93-98)	NA[b]	NA[b]	98 (96-99)
ALL[d]	**74 (72-76)**	**71 (69-73)**	**90 (88-91)**	**95 (94-96)**	**95 (95-96)**	**95 (94-96)**	**90 (84-96)**	**91 (89-92)**	**89 (88-91)**	**85 (84-86)**	**84 (83-85)**	**95 (94-96)**

The data consists of 2 datasets: a random sample of *E. coli* bacteraemia (235) (table S2) & 3rd gen cephalosporin resistant *E. coli* UTI & bacteraemia (283) (table S3). [a] antimicrobial: AMO: amoxicillin, SSP: small-spectrum penicillin (AMO), AMC: amoxicillin-clavulanic acid, PITA: piperacillin-tazobactam, BRSP: broad-spectrum penicillins (sum AMO, PITA), CER: cefuroxime, CFO: cefotaxime, CFZ: ceftazidime, CFE: cefepime, CEPH: cephalosporins (sum CER, CFO, CFZ, CFE), MER: meropenem, IMI: imipenem, CARB: carbapenems (sum MER, IMI), CIP: ciprofloxacin, FQ: fluorquinolones (CIP), GEN: gentamicin, TOB: tobramycin, AMIN: aminoglycosides (sum GEN, TOB) COT: cotrimoxazol, COL: colistin, MISC: miscellaneous agents (sum COT, COL). [b] no model available.[c] output not calculated, due to absence of phenotypic resistant isolates. [d]calculation: sum all true positives, true negatives, false positives & false negatives

Table S2. The diagnostic performance of KOVER-AMR-SCM, KOVER-AMR-CART, and ResFinder 4.1 for the phenotypic antimicrobial susceptibility pattern of a random sample of *E. coli* bacteraemia (n=235)

	concordance (95% CI)			sensitivity (95% CI)			specificity (95% CI)			PPV (95% CI)			NPV (95% CI)			ME rate (95% CI)			VME rate (95% CI)		
A[a]	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1
AMO	92% 88-95	92% 88-95	97% 94-99	88% 80-93	88% 80-93	95% 89-98	97% 92-99	97% 92-99	98% 94-100	96% 90-99	96% 90-99	98% 94-100	89% 83-94	89% 83-94	95% 90-98	3.3% 1.3-8.1	3.3% 1.3-8.1	1.6% 0.1-5.8	12% 7.5-20	12% 7.5-20	5.3% 2.5-11
AMC	84% 78-88	85% 79-89	83% 78-89	33% 22-47	50% 37-63	30% 19-43	99% 96-100	95% 91-97	99% 97-100	90% 70-97	75% 59-86	94% 73-99	83% 78-88	86% 81-91	83% 77-87	1.1% 0.3-3.9	5.0% 2.6-9.2	0.6% 0.1-3.1	67% 53-78	50% 37-63	70% 57-81
PITA	86% 81-90	86% 81-90	91% 86-94	6% 1-26	6% 1-26	33% 16-56	93% 88-95	93% 88-95	96% 92-98	6.0% 1-27	6.0% 1-27	40% 20-64	92% 88-95	92% 88-95	95% 91-97	7.4% 4.6-12	7.4% 4.6-12	4.1% 2.2-7.7	94% 74-99	94% 74-99	67% 44-84
CER	97% 93-98	97% 93-98	NAb	73% 56-86	73% 56-86	NAb	100% 98-100	100% 98-100	NAb	100 85-100	100% 85-100	NAb	96% 93-98	96% 93-98	NAb	0% 0-1.8	0% 0-1.8	NAb	27% 14-44	27% 14-44	NAb
CFO	98% 95-99	98% 95-99	95% 92-98	80% 61-91	80% 61-91	88% 70-96	100% 98-100	100% 98-100	96% 93-98	100 84-100	100 84-100	73% 56-86	98% 95-99	98% 95-99	99% 96-100	0% 0-1.8	0% 0-1.8	3.8% 1.9-7.3	20% 8.9-39	20% 8.9-39	12% 4.2-30
CFZ	95% 91-97	92% 88-96	94% 91-97	48% 28-68	19% 7.7-40	90% 71-97	100% 97-100	100% 97-100	95% 91-97	91% 62-98	80% 38-96	63% 46-78	95% 91-97	93% 89-95	99% 97-100	0.5% 0.1-2.6	0.5% 0.1-2.6	5.1% 2.9-9.0	52% 3.2-72	81% 60-92	9.5% 2.7-29
CFE	76% 69-81	85% 80-90	96% 92-98	60% 39-78	55% 34-74	90% 70-97	77% 71-82	88% 83-92	96% 93-98	20% 12-32	31% 19-48	63% 50-84	95% 91-98	95% 91-98	99% 96-100	23% 18-29	12% 8.0-17	3.9% 2.0-7.5	40% 22-61	45% 28-66	10% 2.8-30
MER	88% 83-92	89% 85-93	100 98-100	-c	-c	-c	88% 83-92	89% 85-93	100% 98-100	-c	-c	-c	-c	-c	-c	12% 8.2-17	11% 7.1-15	0% 0-2.0	-c	-c	-c
IMI	NAb	NAb	100% 98-100	NAb	NAb	-c	NAb	NAb	100% 98-100	NAb	NAb	-c	NAb	NAb	-c	NAb	NAb	0% 0-2.1	NAb	NAb	-c
CIP	96% 92-98	96% 92-98	94% 91-97	77% 62-87	77% 62-87	100% 92-100	100% 98-100	100% 98-100	93% 89-96	100% 90-100	100% 90-100	77% 64-86	95% 91-97	95% 91-97	100% 98-100	0% 0-1.8	0% 0-1.8	6.8% 4.0-11	23% 13-38	23% 13-38	0% 0-8.2
GEN	98% 95-99	98% 95-99	99% 96-100	80% 58-92	80% 58-92	85% 64-95	100% 98-100	100% 98-100	100% 98-100	100% 81-100	100% 81-100	100% 82-100	98% 95-99	98% 95-99	99% 96-100	0% 0-1.8	0% 0-1.8	0% 0-1.8	20% 8.1-42	20% 8.1-42	15% 5.2-36
TOB	99% 96-100	99% 96-100	99% 96-100	94% 74-99	94% 74-99	100% 82-100	99% 96-100	99% 96-100	98% 95-99	89% 64-95	89% 64-95	86% 62-93	100% 97-100	100% 97-100	100% 98-100	1.4% 0.5-4.0	1.4% 0.5-4.0	1.8% 0.7-4.6	5.6% 1.0-26	5.6% 1.0-26	0% 0-18
COT	NAb	NAb	97% 94-99	NAb	NAb	94% 85-98	NAb	NAb	99% 96-100	NAb	NAb	97% 89-99	NAb	NAb	98% 94-99	NAb	NAb	1.2% 0.3-4.2	NAb	NAb	6.2% 2.4-15
COL	NAb	NAb	100 97-100	NAb	NAb	-c	NAb	NAb	100 97-100	NAb	NAb	-c	NAb	NAb	-c	NAb	NAb	0.4% 0-2.7	NAb	NAb	-c
ALL[d]	92% 91-93	93% 91-93	96% 95-96	69% 64-73	69% 64-74	82% 78-86	96% 95-96	96% 96-97	98% 97-98	71% 66-76	76% 71-80	85% 81-88	95% 94-96	95% 94-96	97% 97-98	4.5% 3.7-5.5	3.7% 3.0-4.5	2.2% 1.7-2.8	32% 27-36	31% 26-36	18% 14-22

a antimicrobial: AMO: amoxicillin, AMC: amoxicillin-clavulanic acid, PITA: piperacillin-tazobactam, CER: cefuroxime, CFO: cefotaxime, CFZ: ceftazidime, CFE: cefepime, MER: meropenem, IMI: imipenem, CIP: ciprofloxacin, GEN: gentamicin, TOB: tobramycin, COT: cotrimoxazol, COL: colistin.b output not calculated, due to absence of phenotypic susceptible isolates in dataset. c antimicrobial was not available in the phenotype prediction tool. doutput not calculated, due to absence of phenotypic resistant isolates in dataset. dcalculated with the sum of all observed true positives, true negatives, false positives and false negatives

Table S3. The diagnostic performance of KOVER-AMR-SCM, KOVER-AMR-CART & ResFinder 4.1 for the phenotypic antimicrobial susceptibility pattern of 3rd generation cephalosporin resistant *E. coli* UTI and bactaeremia (n=283).

	concordance (95% CI)			sensitivity (95% CI)			specificity (95% CI)			PPV (95% CI)			NPV (95% CI)			ME rate (95% CI)			VME rate (95% CI)		
A[a]	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1	KOV SCM	KOV CAR	Res 4.1
AMO	61% 55-66	61% 55-66	100% 98-100	61% 54-66	61% 54-66	100% 98-100	_b	_b	_b	_b	_b	_b	_b	_b	_b	_b	_b	_b	39% 33-45	39% 33-45	0% 0-2.2
AMC	62% 56-68	66% 60-72	62% 56-67	48% 41-56	54% 47-61	47% 40-54	91% 84-96	92% 85-96	94% 87-97	92% 85-96	94% 85-96	94% 87-97	46% 39-53	49% 42-56	46% 39-53	8.7% 4.5-16	7.6% 3.7-15	6.4% 3.0-13	52% 45-59	46% 39-53	53% 46-60
PITA	82% 77-86	82% 77-86	73% 67-78	15% 7.2-30	15% 7.2-30	72% 56-84	93% 89-96	93% 89-96	73% 67-78	26% 13-47	26% 13-47	30% 22-40	87% 83-91	87% 83-91	94% 90-97	7.0% 4.4-11	7.0% 4.4-11	27% 22-33	85% 70-93	85% 70-93	28% 17-44
CER	96% 93-98	96% 93-98	NA[c]	96% 93-98	96% 93-98	NA[c]	_b	_b	NA[c]	_b	_b	NA[c]	_b	_b	NA[c]	_b	_b	NA[c]	4.2% 2.3-7.4	4.2% 2.3-7.4	NA[c]
CFO	96% 93-98	96% 93-98	99% 96-100	96% 93-98	96% 93-98	99% 96-100	100% 5-100	100% 5-100	100% 5-100	100% 98-100	100% 98-100	100% 98-100	8.3% 0.4-40	8.3% 0.4-40	20% 1.0-70	0% 0-95	0% 0-95	0% 0-95	3.9% 2.1-7.1	3.9% 2.1-7.1	1.8% 0.6-4.3
CFZ	66% 60-71	40% 34-46	80% 75-85	60% 53-66	33% 27-39	99% 96-100	91% 80-96	70% 57-81	1.9% 0-11	97% 92-99	82% 73-89	81% 75-85	35% 27-43	20% 15-26	20% 1-70	9.3% 4.0-20	30% 19-43	98% 89-99.9	40% 34-47	68% 61-73	1.3% 0.3-4.1
CFE	65% 54-74	60% 49-69	84% 75-90	61% 50-71	52% 41-62	99% 92-100	79% 57-92	94% 74-99	17% 4.4-42	93% 82-97	98% 88-100	83% 74-90	33% 21-46	30% 20-43	75% 22-99	21% 8.5-43	5.6% 1.0-26	83% 58-96	39% 29-50	48% 38-59	1.2% 0-7.6
MER	88% 83-91	89% 85-93	100% 98-100	_d	_d	_d	88% 83-91	89% 85-93	100% 98-100	_d	_d	_d	_d	_d	_d	12% 8.6-17	11% 7.4-15	0% 0-1.7	_d	_d	_d
IMI	NA[e]	NA[e]	100% 98-100	NA[e]	NA[e]	0% 0-95	NA[e]	NA[e]	100% 98-100	NA[e]	NA[e]	_d	NA[e]	NA[e]	100% 98-100	NA[e]	NA[e]	0% 0-1.8	NA[e]	NA[e]	100% 5.5-100
CIP	88% 83-91	88% 83-91	97% 94-98	83% 78-88	83% 78-88	100% 98-100	100% 95-100	100% 95-100	88% 78-93	100% 98-100	100% 98-100	96% 92-98	68% 58-76	68% 58-76	100% 94-100	0% 0-5.0	0% 0-5.0	12% 6.6-22	17% 12-23	17% 12-23	0% 0-1.8
GEN	95% 91-97	95% 91-97	97% 94-98	84% 75-91	84% 75-91	92% 84-96	99% 96-100	99% 96-100	99% 96-100	97% 90-99	97% 90-99	97% 91-99	94% 90-96	94% 90-96	97% 93-98	1.0% 0.3-3.6	1.0% 0.3-3.6	1.0% 0.3-3.6	1.6% 9-25	16% 9-25	8.4% 4.1-16
TOB	95% 92-97	95% 92-97	96% 93-98	91% 84-95	91% 84-95	95% 89-98	98% 95-99	98% 95-99	98% 94-99	97% 91-99	97% 91-99	96% 90-98	95% 90-97	95% 90-97	97% 93-99	1.9% 0.6-5.4	1.9% 0.6-5.4	3.8% 1.7-8.0	8.8% 4.7-16	8.8% 4.7-16	4.9% 2.1-11
COT	NA[e]	NA[e]	96% 92-98	NA[e]	NA[e]	0% 0-95	NA[e]	NA[e]	96% 91-99	NA[e]	NA[e]	98% 94-99	NA[e]	NA[e]	92% 87-96	NA[e]	NA[e]	3.6% 1.4-9.0	NA[e]	NA[e]	4.7% 2.4-8.9
COL	NA[e]	NA[e]	96% 89-99	NA[e]	NA[e]	0% 0-80	NA[e]	NA[e]	98% 92-100	NA[e]	NA[e]	0% 0-80	NA[e]	NA[e]	98% 92-100	NA[e]	NA[e]	2.0% 0.3-7.9	NA[e]	NA[e]	100% 20-100
ALL[d]	82% 81-83	80% 78-81	91% 90-92	75% 73-77	72% 70-74	91% 90-93	94% 92-95	93% 92-95	90% 88-91	95% 94-96	94% 93-96	91% 89-92	70% 68-72	67% 65-70	91% 89-92	6.5% 5.2-8.1	6.8% 5.4-8.4	10% 8.7-11	25% 23-27	28% 26-31	8.6% 7.4-10

a antimicrobial: AMO: amoxicillin, AMC: amoxicillin-clavulanic acid, PITA: piperacillin-tazobactam, CER: cefuroxime, CFO: cefotaxime, CFZ: ceftazidime, CFE: cefepime, MER: meropenem, IMI: imipenem, CIP: ciprofloxacin, GEN: gentamicin, TOB: tobramycin, COT: cotrimoxazol, COL: colistin.b output not calculated, due to absence of phenotypic susceptible isolates in dataset. c antimicrobial was not available in the phenotype prediction tool. doutput not calculated, due to absence of phenotypic resistant isolates in dataset. dcalculated with the sum of all observed true positives, true negatives, false positives and false negatives

Table S4. List of discrepancies

Antimicrobial[a]	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
amoxicillin	1	ME	S	<=2	R	R	S	Genetic predictor not conferring resistance
amoxicillin	1	ME	S	<=2	R	R	S	Genetic predictor not conferring resistance
amoxicillin	1	ME	S	<=2	R	R	R	Genetic predictor not conferring resistance /incorrect phenotype
amoxicillin	1	ME	S	4	R	R	S	MIC around breakpoint
amoxicillin	1	ME	S	<=2	S	S	R	Genetic predictor not conferring resistance
amoxicillin	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	1	VME	R	16	S	S	S	MIC around breakpoint
amoxicillin	1	VME	R	>= 32	S	S	S	Missing genetic predictor/incorrect phenotype
amoxicillin	1	VME	R	16	S	S	S	MIC around breakpoint
amoxicillin	1	VME	R	>= 32	S	S	S	Missing genetic predictor/incorrect phenotype
amoxicillin	1	VME	R	16	S	S	S	MIC around breakpoint
amoxicillin	1	VME	R	>8	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	ME	S	<=2	R	R	S	Genetic predictor not conferring resistance
amox-clav	1	ME	S	8	R	R	S	MIC around breakpoint
amox-clav	1	ME	S	4	S	S	R	MIC around breakpoint
amox-clav	1	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	1	ME	S	4	S	R	S	MIC around breakpoint
amox-clav	1	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	1	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	1	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	1	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	1	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	R	R	MIC around breakpoint
amox-clav	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amox-clav	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amox-clav	1	VME	R	>= 32	S	S	R	Missing genetic predictor
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	1	VME	R	>= 32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	>= 32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	1	VME	R	>8	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	1	VME	R	32	S	R	S	Missing genetic predictor
amox-clav	1	VME	R	32	S	R	S	Missing genetic predictor
amox-clav	1	VME	R	32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	16	R	S	S	MIC around breakpoint
amox-clav	1	VME	R	>= 32	R	S	R	Missing genetic predictor
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	>= 32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	>= 32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	>8	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	>32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	1	VME	R	32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	1	VME	R	16	S	S	S	MIC around breakpoint
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring res
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring res
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	8	R	R	S	MIC around breakpoint
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	8	R	R	S	MIC around breakpoint
piperacillin-taz	1	ME	S	8	R	R	S	MIC around breakpoint
piperacillin-taz	1	ME	S	8	R	R	S	MIC around breakpoint
piperacillin-taz	1	ME	S	8	R	R	S	MIC around breakpoint

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
piperacillin-taz	1	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	1	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	1	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	1	VME	R	>=128	S	S	R	Missing genetic predictor
piperacillin-taz	1	VME	R	>=128	S	S	R	Missing genetic predictor
piperacillin-taz	1	VME	R	>16	S	S	R	Missing genetic predictor
piperacillin-taz	1	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	1	VME	R	>16	S	S	R	Missing genetic predictor
piperacillin-taz	1	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	1	VME	R	>=128	R	R	S	Missing genetic predictor
piperacillin-taz	1	VME	R	16	S	S	S	MIC around breakpoint
piperacillin-taz	1	VME	R	16	S	S	S	MIC around breakpoint
piperacillin-taz	1	VME	R	64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	1	VME	R	32	S	S	S	MIC around breakpoint
piperacillin-taz	1	VME	R	>=128	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	1	VME	R	>=128	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	1	VME	R	16	S	S	S	MIC around breakpoint
piperacillin-taz	1	VME	R	64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	1	VME	R	>=128	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	1	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	1	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
cefuroxim	1	VME	R	16	S	S	NA	MIC around breakpoint
cefuroxim	1	VME	R	16	S	S	NA	MIC around breakpoint
cefuroxim	1	VME	R	>= 64	S	S	NA	Missing genetic predictor
cefuroxim	1	VME	R	32	S	S	NA	Missing genetic predictor
cefuroxim	1	VME	R	>8	S	S	NA	Missing genetic predictor
cefuroxim	1	VME	R	>8	S	S	NA	Missing genetic predictor
cefuroxim	1	VME	R	>8	S	S	NA	Missing genetic predictor
cefuroxim	1	VME	R	>8	S	S	NA	Missing genetic predictor
cefotaxime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	1	VME	R	4	S	S	R	MIC around breakpoint
cefotaxime	1	VME	R	>4	S	S	R	Missing genetic predictor
cefotaxime	1	VME	R	>4	S	S	R	Missing genetic predictor
cefotaxime	1	VME	R	>4	R	R	S	Missing genetic predictor
cefotaxime	1	VME	R	2	S	S	S	MIC around breakpoint
cefotaxime	1	VME	R	>4	S	S	S	Missing genetic predictor
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	R	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	1	ME	S	<=1	R	S	R	Genetic predictor not conferring resistance
ceftazidime	1	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	1	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	1	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	1	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	1	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	1	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	1	VME	R	4	R	S	R	MIC around breakpoint
ceftazidime	1	VME	R	>= 64	R	S	R	Missing genetic predictor
ceftazidime	1	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	1	VME	R	>8	S	R	R	Missing genetic predictor
ceftazidime	1	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	1	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	1	VME	R	8	S	R	R	MIC around breakpoint
ceftazidime	1	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	1	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	1	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	1	VME	R	4	S	S	S	MIC around breakpoint
ceftazidime	1	VME	R	>8	S	S	S	Missing genetic predictor
ceftazidime	1	VME	R	>8	R	S	R	Missing genetic predictor
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	R	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	R	S	Genetic predictor not conferring resistance
cefepime	1	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
gentamicin	1	VME	R	>4	S	S	R	Missing genetic predictor
gentamicin	1	VME	R	4	S	S	S	MIC around breakpoint
gentamicin	1	VME	R	4	S	S	S	MIC around breakpoint
gentamicin	1	VME	R	>4	S	S	S	Missing genetic predictor/incorrect phenotype
tobramycin	1	ME	S	2	S	S	R	MIC around breakpoint
tobramycin	1	ME	S	<=1	R	R	R	Genetic predictor not conferring resistance/incorrect phenotype
tobramycin	1	ME	S	2	R	R	R	MIC around breakpoint
tobramycin	1	VME	R	>=16	S	S	R	Missing genetic predictor
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	1	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	1	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	1	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	1	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	1	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	1	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	1	VME	R	>1	S	S	R	Missing genetic predictor
ciprofloxacin	1	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	1	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	1	VME	R	0.5	S	S	R	MIC around breakpoint
cotrimoxazol	1	ME	S	<=1	NA	NA	R	Genetic predictor not conferring resistance/ incorrect phenotype
cotrimoxazol	1	ME	S	<=1	NA	NA	R	Genetic predictor not conferring resistance/ incorrect phenotype
cotrimoxazol	1	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	1	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	1	VME	R	>4	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	1	VME	R	>4	NA	NA	S	Missing genetic predictor/incorrect phenotype

Table S4 continues on the next page

4

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
colistin	1	ME	S	<=1	NA	NA	R	Genetic predictor not conferring resistance/ incorrect phenotype
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	S	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	S	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.5	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.5	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.5	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	S	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.125	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.125	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.125	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.125	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.125	R	R	S	Genetic predictor not conferring resistance
meropenem	1	ME	S	<=0.125	R	R	S	Genetic predictor not conferring resistance

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>8	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor

Table S4 continues on the next page

4

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amoxicillin	2	VME	R	>16	S	S	R	Missing genetic predictor
amox-clav	2	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	2	ME	S	8	R	R	R	MIC around breakpoint
amox-clav	2	ME	S	4	R	S	S	MIC around breakpoint
amox-clav	2	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	2	ME	S	8	R	S	R	MIC around breakpoint
amox-clav	2	ME	S	4	S	S	R	MIC around breakpoint
amox-clav	2	ME	S	8	R	R	R	MIC around breakpoint
amox-clav	2	ME	S	8	R	S	S	MIC around breakpoint
amox-clav	2	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	2	ME	S	4	R	S	S	MIC around breakpoint
amox-clav	2	ME	S	8	R	S	S	MIC around breakpoint
amox-clav	2	ME	S	8	S	S	R	MIC around breakpoint
amox-clav	2	ME	S	8	R	R	R	MIC around breakpoint
amox-clav	2	VME	R	>8	R	S	S	Missing genetic predictor
amox-clav	2	VME	R	>8	S	R	S	Missing genetic predictor
amox-clav	2	VME	R	>32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>8	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>8	S	R	S	Missing genetic predictor
amox-clav	2	VME	R	32	S	R	S	Missing genetic predictor
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>32	R	S	R	Missing genetic predictor
amox-clav	2	VME	R	32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	ME	S	8	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	>= 32	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>= 32	R	S	R	Missing genetic predictor
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	>= 32	S	S	R	Missing genetic predictor
amox-clav	2	VME	R	16	R	R	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	R	R	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	R	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	R	S	R	Missing genetic predictor
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	R	S	S	Missing genetic predictor
amox-clav	2	VME	R	>16	R	R	S	Missing genetic predictor

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	R	Missing genetic predictor
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	R	Missing genetic predictor
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	R	Missing genetic predictor
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	R	S	Missing genetic predictor
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	R	S	Missing genetic predictor
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	R	Missing genetic predictor
amox-clav	2	VME	R	>16	S	R	R	Missing genetic predictor
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	R	S	Missing genetic predictor
amox-clav	2	VME	R	>16	S	R	S	Missing genetic predictor
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	R	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	R	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	R	Missing genetic predictor
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	R	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	R	R	S	Missing genetic predictor
amox-clav	2	VME	R	>16	S	R	S	Missing genetic predictor
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	R	R	S	Missing genetic predictor
amox-clav	2	VME	R	>16	R	S	S	Missing genetic predictor
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	16	S	S	S	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	R	S	S	Missing genetic predictor
amox-clav	2	VME	R	16	S	S	R	MIC around breakpoint
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	R	S	R	Missing genetic predictor
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
amox-clav	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	R	R	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	R	R	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	R	R	R	Genetic predictor not conferring resistance/incorrect phen
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	R	R	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	R	R	S	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	<=4	S	S	R	Genetic predictor not conferring resistance
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	ME	S	8	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	>=128	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	32	S	S	S	MIC around breakpoint
piperacillin-taz	2	VME	R	64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	>=128	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	>64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	>64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	>64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	>64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	16	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	32	S	S	R	MIC around breakpoint
piperacillin-taz	2	VME	R	>64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	>32	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	>64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	>64	S	S	S	Missing genetic predictor/incorrect phenotype
piperacillin-taz	2	VME	R	>64	S	S	R	Missing genetic predictor
piperacillin-taz	2	VME	R	64	S	S	R	Missing genetic predictor

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
cefuroxim	2	VME	R	>8	S	S	NA	Missing genetic predictor
cefuroxim	2	VME	R	>8	S	S	NA	Missing genetic predictor
cefuroxim	2	VME	R	>8	S	S	NA	Missing genetic predictor
cefuroxim	2	VME	R	>8	S	S	NA	Missing genetic predictor
cefuroxim	2	VME	R	16	S	S	NA	MIC around breakpoint
cefuroxim	2	VME	R	32	S	S	NA	Missing genetic predictor
cefuroxim	2	VME	R	>= 64	S	S	NA	Missing genetic predictor
cefuroxim	2	VME	R	>32	S	S	NA	Missing genetic predictor
cefuroxim	2	VME	R	>32	S	S	NA	Missing genetic predictor
cefuroxim	2	VME	R	16	S	S	NA	MIC around breakpoint
cefuroxim	2	VME	R	16	S	S	NA	MIC around breakpoint
cefuroxim	2	VME	R	>32	S	S	NA	Missing genetic predictor
cefotaxime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefotaxime	2	VME	R	>4	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	>4	S	S	S	Missing genetic predictor/incorrect phenotype
cefotaxime	2	VME	R	>4	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	>= 64	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	>= 64	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	>32	R	R	S	Missing genetic predictor
cefotaxime	2	VME	R	>32	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	8	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	16	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	>32	R	R	S	Missing genetic predictor
cefotaxime	2	VME	R	>32	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	>32	R	R	S	Missing genetic predictor
cefotaxime	2	VME	R	>32	S	S	R	Missing genetic predictor
cefotaxime	2	VME	R	16	S	S	S	Missing genetic predictor/incorrect phenotype
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	R	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	R	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.25	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	R	R	R	Genetic predictor not conferring resistance/incorrect phen
ceftazidime	2	ME	S	0.5	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	<=0.12	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	1	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	R	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	1	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	R	S	R	Genetic predictor not conferring resistance
ceftazidime	2	ME	S	0.5	S	R	R	Genetic predictor not conferring resistance
ceftazidime	2	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	2	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	R	R	MIC around breakpoint

Table S4 continues on the next page

4

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
ceftazidime	2	VME	R	>8	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	4	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>8	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>8	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>= 64	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	2	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	>= 64	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	16	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>= 64	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>= 64	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	16	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	R	R	Missing genetic predictor
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	R	S	Missing genetic predictor
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	R	R	Missing genetic predictor
ceftazidime	2	VME	R	32	S	R	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
ceftazidime	2	VME	R	>32	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	S	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>32	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	>32	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	>32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	S	R	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	2	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	S	R	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	R	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	S	S	S	Missing genetic predictor/incorrect phenotype
ceftazidime	2	VME	R	2	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	R	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	4	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	4	S	R	R	MIC around breakpoint
ceftazidime	2	VME	R	32	R	S	R	Missing genetic predictor
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint
ceftazidime	2	VME	R	8	S	S	R	MIC around breakpoint

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	R	S	S	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	R	R	R	Genetic predictor not conferring resistance/incorrect phen
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	R	S	R	Genetic predictor not conferring resistance
cefepime	2	ME	S	<=1	S	S	R	Genetic predictor not conferring resistance
cefepime	2	VME	R	>8	S	S	R	Missing genetic predictor
cefepime	2	VME	R	>8	S	S	R	Missing genetic predictor
cefepime	2	VME	R	>8	S	S	R	Missing genetic predictor
cefepime	2	VME	R	8	S	S	R	MIC around breakpoint
cefepime	2	VME	R	8	S	S	R	MIC around breakpoint
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	>8	R	S	R	Missing genetic predictor
cefepime	2	VME	R	8	S	S	R	MIC around breakpoint
cefepime	2	VME	R	8	S	S	R	MIC around breakpoint
cefepime	2	VME	R	8	S	S	R	MIC around breakpoint
cefepime	2	VME	R	4	S	S	R	MIC around breakpoint
cefepime	2	VME	R	>8	S	S	R	Missing genetic predictor
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	4	S	S	R	MIC around breakpoint
cefepime	2	VME	R	>8	S	S	R	Missing genetic predictor
cefepime	2	VME	R	4	S	S	R	MIC around breakpoint
cefepime	2	VME	R	8	S	S	S	MIC around breakpoint
cefepime	2	VME	R	>8	S	S	R	Missing genetic predictor
cefepime	2	VME	R	>8	R	S	R	Missing genetic predictor
cefepime	2	VME	R	4	S	S	R	MIC around breakpoint
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	2	R	S	R	MIC around breakpoint
cefepime	2	VME	R	>= 64	S	S	R	Missing genetic predictor
cefepime	2	VME	R	8	R	S	R	MIC around breakpoint
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	4	R	S	R	MIC around breakpoint
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	8	S	S	R	MIC around breakpoint
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	4	S	S	R	MIC around breakpoint
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	>= 64	S	S	R	Missing genetic predictor
cefepime	2	VME	R	8	S	S	R	MIC around breakpoint
cefepime	2	VME	R	4	S	S	R	MIC around breakpoint
cefepime	2	VME	R	>= 64	R	S	R	Missing genetic predictor
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
cefepime	2	VME	R	2	R	S	R	MIC around breakpoint
cefepime	2	VME	R	2	S	S	R	MIC around breakpoint
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	ME	S	<=0.25	S	S	R	Genetic predictor not conferring resistance
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	>1	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	1	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	0.5	S	S	R	MIC around breakpoint
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
ciprofloxacin	2	VME	R	>2	S	S	R	Missing genetic predictor
gentamicin	2	ME	S	2	R	R	R	MIC around breakpoint
gentamicin	2	ME	S	<=1	R	R	R	Genetic predictor not conferring resistance/incorrect phen
gentamicin	2	VME	R	>4	S	S	R	Missing genetic predictor
gentamicin	2	VME	R	>4	S	S	R	Missing genetic predictor
gentamicin	2	VME	R	>=16	S	S	R	Missing genetic predictor
gentamicin	2	VME	R	>=16	S	S	R	Missing genetic predictor
gentamicin	2	VME	R	>8	S	S	R	Missing genetic predictor
gentamicin	2	VME	R	>8	S	S	R	Missing genetic predictor
gentamicin	2	VME	R	>8	S	S	R	Missing genetic predictor
gentamicin	2	VME	R	4	S	S	S	MIC around breakpoint
gentamicin	2	VME	R	>4	S	S	S	Missing genetic predictor/incorrect phenotype
gentamicin	2	VME	R	4	S	S	S	MIC around breakpoint
gentamicin	2	VME	R	4	S	S	S	MIC around breakpoint
gentamicin	2	VME	R	>16	S	S	S	Missing genetic predictor/incorrect phenotype
gentamicin	2	VME	R	>8	S	S	S	Missing genetic predictor/incorrect phenotype
gentamicin	2	VME	R	>=16	R	R	S	Missing genetic predictor
tobramycin	2	ME	S	<=1	R	R	R	Genetic predictor not conferring resistance/incorrect phen
tobramycin	2	ME	S	<=1	R	R	R	Genetic predictor not conferring resistance/incorrect phen
tobramycin	2	ME	S	<=1	R	R	R	Genetic predictor not conferring resistance/incorrect phen
tobramycin	2	ME	S	2	S	S	R	MIC around breakpoint
tobramycin	2	VME	R	>4	S	S	R	Missing genetic predictor
tobramycin	2	VME	R	>=16	S	S	R	Missing genetic predictor
tobramycin	2	VME	R	8	S	S	R	MIC around breakpoint
tobramycin	2	VME	R	8	S	S	R	MIC around breakpoint
tobramycin	2	VME	R	4	S	S	S	MIC around breakpoint
tobramycin	2	VME	R	4	S	S	S	MIC around breakpoint
tobramycin	2	VME	R	4	S	S	S	MIC around breakpoint
tobramycin	2	VME	R	4	S	S	S	MIC around breakpoint
tobramycin	2	VME	R	>2	S	S	S	Missing genetic predictor/incorrect phenotype
meropenem	2	ME	S	<=0.5	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.125	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.125	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	S	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	S	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	S	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	S	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	S	S	Genetic predictor not conferring resistance
meropenem	2	ME	S	<=0.25	R	R	S	Genetic predictor not conferring resistance

Table S4 continues on the next page

Antimicrobial	Data set[a]	Error[b]	Ph[c]	MIC[d]	KOV SCM	KOV CART	Res 4.1	Category
imipenem	2	VME	R	>32	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	2	ME	S	<=1	NA	NA	R	Genetic predictor not conferring resistance/incorrect phen
cotrimoxazol	2	ME	S	<=1	NA	NA	R	Genetic predictor not conferring resistance/incorrect phen
cotrimoxazol	2	ME	S	<=1	NA	NA	R	Genetic predictor not conferring resistance/incorrect phen
cotrimoxazol	2	ME	S	<=1	NA	NA	R	Genetic predictor not conferring resistance/incorrect phen
cotrimoxazol	2	VME	R	4	NA	NA	S	MIC around breakpoint
cotrimoxazol	2	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	2	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	2	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	2	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	2	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	2	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
cotrimoxazol	2	VME	R	>=16	NA	NA	S	Missing genetic predictor/incorrect phenotype
colistin	2	ME	S	<=1	NA	NA	R	Genetic predictor not conferring resistance/incorrect phen
colistin	2	ME	S	<=0.5	NA	NA	R	Genetic predictor not conferring resistance/incorrect phen
colistin	2	VME	R	>4	NA	NA	S	Missing genetic predictor/incorrect phenotype
colistin	2	VME	R	>4	NA	NA	S	Missing genetic predictor/incorrect phenotype

MIC around breakpoint: factor 2 above or below. [a] dataset: 1: random sample (~1/3) of available *E. coli* harbouring bloodcultures from the years 2014-2016 from two hospitals consisting of 235 *E. coli* isolates from blood cultures, 2: 283 3rd generation cephalosporin resistant *E. coli* from urine and blood from two hospitals and one primary care laboratory. [b] ME: major error (false resistant result), VME: very major error (false susceptible result). [c] Phen: phenotype determined using VITEK 2 or BD Phoenix. [d] MIC: minimum inhibitory concentration.

Table S5. The reported phenotypic resistance percentages of datasets used for internal validation of KOVER-AMR-tools (Drouin *et al*)[17], and ResFinder 4.0 (Bortolaia *et al*)[7].

antimicrobial class	**antimicrobial**	**KOVER-AMR-tools R% (n)**	**ResFinder 4.1[a] R% (n)**
penicillins (small-spectrum)	amoxicillin	60% (661)	95% (378)
penicillins (broad-spectrum)	amoxicillin-clavulanic acid	30% (464)	NA[b]
	piperacillin-tazobactam	7% (99)	NA[b]
cephalosporins	cefuroxime	16% (241)	NA[c]
	cefotaxime	10% (139)	87% (488)
	ceftazidime	7% (99)	99.7% (376)
	cefepime	8% (32)	93% (215)
carbapenems	meropenem	6% (28)	1% (2)
	imipenem	NA[c]	1% (2)
fluorquinolones	ciprofloxacin	19% (289)	78% (368)a
aminoglycosides	gentamicin	8% (115)	31% (178)
	tobramycin	12% (50)	NA[b]
miscellaneous agents	cotrimoxazol	NA[c]	NA[b]
	colistin	NA[c]	6% (11)

[a] average of three different internal validation datasets
[b] not validated
[c] antimicrobial was not available in prediction tool.

Table S6. The reported diagnostic performance for the phenotypic antimicrobial susceptibility pattern of *E. coli* in the interval validation of KOVER-AMR (Drouin *et al*)[17] and ResFinder 4.0 (Bortolaia *et al*)[7]. Data of internal validations was accessed at: https://github.com/aldro61/kb_kover_amr/blob/master/data/models/scm/escherichia%20coli/(antimicrobialx)/report.txt, and https://academic.oup.com/jac/article/75/12/3491/5890997#supplementary-data, table S5,6.

	concordance (95% CI)			**ME rate (95% CI)**			**VME rate (95% CI)**		
A[a]	KOV SCM	KOV CAR	Res 4.1[b]	KOV SCM	KOV CAR	Res 4.1[b]	KOV SCM	KOV CAR	Res 4.1[b]
AMO	93% (92-95)	93% (92-95)	100 (98-100)	2.3% (1.3-4.2)	2.3% (1.3-4.2)	0% (0-18)	9.4% (7.4-12)	9.4% (7.4-12)	0.3% (0-1.5)
AMC	85% (83-86)	86% (84-88)	-c	1.0% (0.6-1.8)	7.4% (5.9-9.1)	-c	48% (44-53)	28% (24-33)	-c
PITA	94% (93-95)	94% (93-95)	-c	0.1% (0-0.5)	0.1% (0-0.5)	-c	88% (80-93)	88% (80-93)	-c
CER	93% (91-94)	93% (91-94)	-c	0.6% (0.3-1.2)	0.6% (0.3-1.2)	-c	43% (37-50)	46% (40-52)	-c
CFO	98% (97-98)	98% (97-98)	99% (97-100)	0.3% (0.1-0.8)	0.3% (0.1-0.8)	-	11% (6.7-17)	11% (6.7-17)	1.2% (0.4-2.8)
CFZ	99% (98-99)	98% (98-99)	99% (98-100)	0.1% (0-0.5)	0.3% (0.1-0.7)	-	20% (14-29)	19% (13-28)	0.5% (0.1-2.1)
CFE	98% (96-99)	99% (98-100)	89% (84-92)	1.3% (0.5-2.9)	0.3% (0-1.4)	0% (0-18)	6.2% (1.7-20)	6.2% (1.7-20)	12% (8.6-17)
MER	98% (96-99)	99% (98-100)	100% (98-100)	1.7% (0.8-3.4)	0.7% (0.2-2.1)	0% (0-1.9)	3.6% (0.6-18)	3.6% (0.6-18)	0% (0-66)
IMI	-c	-c	100% (99-100)	-c	-c	0% (0-0.1)	-c	-c	0% (0-66)
CIP	99% (98-100)	99% (98-100)	95% (92-97)	0.5% (0.2-1.1)	0.5% (0.2-1.1)	1.% (0.2-5.4)	3.8% (2.1-6.7)	3.8% (2.1-6.7)	7.2% (4.2-11)
GEN	99% (98-99)	99% (98-99)	98% (97-99)	0.3% (0.1-0.7)	0.3% (0.1-0.7)	0.5% (0.1-1.8)	10% (6.1-17)	10% (6.1-17)	4.4% (2.3-8.5)
TOB	98% (96-100)	98% (96-100)	-c	0.8% (0.3-2.3)	0.8% (0.3-2.3)	-c	7.8% (3.1-19)	7.8% (3.1-19)	-c
COT	-c	-c	-c	-c	-c	-c	-c	-c	-c
COL	-c	-c	100% (98-100)	-c	-c	0% (0-2.1)	-c	-c	0% (0-26)
ALL[d]	95% (95-96)	95% (94-95)	98% (97-98)	0.6% (0.5-0.7)	2.1% (1.9-2.4)	0.2% (0.1-0.7)	25% (23-26)	21% (19-22)	3.2% (2.5-4.1)

[a] antimicrobial: AMO: amoxicillin, AMC: amoxicillin-clavulanic acid, PITA: piperacillin-tazobactam, CER: cefuroxime, CFO: cefotaxime, CFZ: ceftazidime, CFE: cefepime, MER: meropenem, IMI: imipenem, CIP: ciprofloxacin, GEN: gentamicin, TOB: tobramycin, COT: cotrimoxazol, COL: colistin.
[b] results from up to three different datasets used for interval validation
[c] not available in original publication
[d] Calculated based on the cumulative true positives, true negatives, false positives, and false negatives.

part 2

ESBL-*E. coli* &

ESBL-*K. pneumoniae*

in Europe

5

Household acquisition and transmission of extended-spectrum ß-lactamase (ESBL) -producing *Enterobacteriaceae* after hospital discharge of ESBL-positive index patients

Maria E Riccio, Tess D Verschuuren, Nadine Conzelmann, Daniel Martak, Alexandre Meunier, Elena Salamanca, Mercedes Delgado, Julia Guther, Silke Peter, Julian Paganini, Romain Martischang, Julien Sauser, Marlieke EA de Kraker, Abdessalam Cherkaoui, Ad C Fluit, Ben S Cooper, Didier Hocquet, Jan AJW Kluytmans, Evelina Tacconelli, Jesús Rodriguez-Baño, Stephan Harbarth, * on behalf of the MODERN WP2 study group

Published in: *Clinical Microbiology and Infection* 2021; **20**: 30784-9.

ABSTRACT

Objective: This study aimed to determine rates and risk factors of extended-spectrum ß-lactamase producing *Enterobacteriaceae* (ESBL-PE) acquisition and transmission within households after hospital discharge of an ESBL-PE-positive index patient.

Methods: Two-year prospective cohort study in five European cities. Patients colonized with ESBL-producing *Escherichia coli* (ESBL-Ec) or *Klebsiella pneumoniae* (ESBL-Kp), and their household contacts were followed up for 4 months after hospital discharge of the index case. At each follow up, participants provided a faecal sample and personal information. ESBL-PE whole-genome sequences were compared using pairwise single nucleotide polymorphism-based analysis.

Results: We enrolled 71 index patients carrying ESBL-Ec (n = 45), ESBL-Kp (n = 20) or both (n = 6), and 102 household contacts. The incidence of any ESBL-PE acquisition among household members initially free of ESBL-PE was 1.9/100 participant-weeks at risk. Nineteen clonally related household transmissions occurred (case to contact: 13; contact to case: 6), with an overall rate of 1.18 transmissions/100 participant-weeks at risk. Most of the acquisition and transmission events occurred within the first 2 months after discharge. The rate of ESBL-Kp household transmission (1.16/100 participant-weeks) was higher than of ESBL-Ec (0.93/100 participant-weeks), whereas more acquisitions were noted for ESBL-Ec (1.06/100 participant-weeks) compared with ESBL-Kp (0.65/100 participant-weeks). Providing assistance for urinary and faecal excretion to the index case by household members increased the risk of ESBL-PE transmission (adjusted prevalence ratio 4.3; 95% CI 1.3-14.1).

Conclusions: ESBL-PE cases discharged from the hospital are an important source of ESBL-PE transmission within households. Most acquisition and transmission events occurred during the first 2 months after hospital discharge and were causally related to care activities at home, highlighting the importance of hygiene measures in community settings.

INTRODUCTION

Transmission of extended-spectrum b-lactamase-producing *Enterobacteriaceae* (ESBL-PE) in the clinical setting has been extensively studied[1], but little is known about the risk and pathways of transmission in the community. A recent systematic review evaluating human-to-human ESBL-PE transmission between household contacts highlighted important limitations of previous studies[2]: low discriminatory power of previously applied typing methods for identifying ESBL-PE transmission events[3]; cross-sectional study design preventing the assessment of transmission dynamics over time; and no systematic assessment of ESBL-PE transmission paths and possible epidemiological determinants. Furthermore, only two studies focused on the likelihood of household transmission of ESBL-PE after hospital discharge of an ESBL-positive patient.[4]

The aim of this study was to investigate ESBL-PE acquisition and transmission in household settings in five European cities with varying ESBL-PE baseline prevalence. Specifically, we attempted to determine the incidence and risk factors of ESBL-PE acquisition and transmission within families after hospital discharge of an ESBL-PE carrier.

METHODS

Study design

We conducted a prospective multicentre cohort study including ESBL-PE-positive patients and their household contacts from five university hospitals (Geneva, Sevilla, Tübingen, Utrecht, Besançon). The recruitment target was 20 households per centre (see Supplementary material, Appendix S1).

Population

Index cases were defined as intestinal ESBL-PE carriers discharged home into a household shared with at least one household contact. Household contacts were identified as any person sharing the same household with the index case at least three nights a week.

Inclusion and exclusion criteria

The inclusion criteria for the index cases were: to be 18 years old; to have a rectal swab or faecal sample at hospital discharge confirming intestinal colonisation with ESBL-producing *Escherichia coli* (ESBL-Ec) and/or *Klebsiella pneumoniae* (ESBL-Kp);

and to provide informed consent. Patients were excluded if they were permanently institutionalized or impossible to follow up. After inclusion, index cases were excluded if they had negative rectal samples during the first two visits. Enrolled participants who dropped out before collecting the first stool sample were also excluded.

Data collection

All participants were followed up for 4 months: at hospital discharge (baseline visit #1), 1 week (visit #2), 2 months (visit #3) and 4 months (visit #4). Questionnaires were filled out by all participants at visits #1, #2, #3, and #4. Collected variables concerned participants' health status, antibiotic intake, household conditions, dietary habits and lifestyle. All participants collected stool samples or rectal swabs themselves (or by a household contact) with Procult™ 500 kit (Ability Building Centre, Rochester, MN, USA) and faeces containers or Eswabs (Copan Diagnostics, Brescia, Italy) at visit #1, #2, #3 and #4 (±3 days). Collected information was transferred into a centralised REDCap database. The study was approved by each centre's institutional review board.

Microbiological methods

Selective culturing, enrichment broth, bacterial identificationnand antimicrobial susceptibility testing were performed for each stool sample or rectal swab at each centre's microbiology laboratory, using standardised methods (as described in the Supplementary material, Appendix S2).

Sequencing analysis

The full genome of ESBL-PE isolates was sequenced with Next-Seq sequencer (Illumina, San Diego, CA, USA). DNA extraction was performed with DNeasy UltraClean Microbial Kit (Qiagen, Hilden, Germany). The sequence type (ST) of each isolate was identified by using seven housekeeping genes, using MLST version 2.10 (https://github.com/tseemann/mlst). ESBL-encoding genes were identified by ResFinder version 2.1 of the Center for Genomic Epidemiology.[5] Neighbour-joining core genome multi-locus sequence typing (cgMLST) trees were constructed with SeqSphereþ (Ridom, Münster, Germany) using the Enterobase scheme (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6961584/) for *E. coli* (2,513 genes) and sensu lato scheme for *K. pneumoniae* (2358 genes). After removing genes not present in all strains, trees were built by comparing 1,863 and 2,088 genes, respectively. For strains presenting the same cgMLST alongside a strong epidemiological link, pairwise single nucleotide polymorphism (SNPs) distances were estimated using the CFSAN pipeline.[6]

Definitions

Genomes of ESBL-PE isolates were considered clonally related and closely related when having, respectively, a pairwise distance of 10 or 11-25 SNP differences.[7] Acquisition was defined as newly identified carriage of an ESBL-Ec or ESBL-Kp strain during follow up, not previously detected in the gut flora of the concerned participant. Transmission was defined as the newly detected intestinal carriage of ESBL-Ec and/or ESBL-Kp of a clonally related isolate previously identified in another household member. Co-carriage was defined as the simultaneous carriage by two or more household members of a clonally related isolate at the same sampling time-point.

Data analysis

Overall and species-specific incidence rates of acquisition and transmission were estimated at the genotypic level. Time at risk of ESBL-PE acquisition was estimated as the number of days between baseline and the acquisition of the corresponding pathogen in a participant previously free of it, or the drop out of the participant, or end of follow up, whichever occurred first. The time at risk of a possible ESBL-PE transmission was estimated as the time between baseline (for index cases) or the date of the first positive sample (for household contacts), and the first detection date of a clonally related isolate previously identified in another household member.

Incidence rates were calculated as the total number of acquisition or transmission events divided by the total number of participant-weeks at risk multiplied by 100.

Risk factors of acquisition and transmission were evaluated by univariable and multivariable mixed-effects Poisson regression models to compute prevalence ratios[8,9], accounting for the lack of independence between repeated samples, and multiple clustering effects. The multilevel structure of the data was composed of three levels: participant (four samples per participant), household and study site. Potential confounders were chosen on the basis of existing evidence, and were only scored if exposure preceded the event, with final model selection performed using a stepwise backward model selection based on Akaike's information criterion.[10] Analyses were performed using R (version 3.6.3.) and STATA version 15 (StataCorp., College Station, TX, USA).

RESULTS

Recruitment and household characteristics

Between November 2017 and April 2019, 71 households were included in the study,

with 71 index cases and 102 of 127 eligible household contacts (participation rate, 80%). During the 4-month follow up, 35 participants from 14 households dropped out (figure 1). Important characteristics of participating households are shown in table 1. The mean age of all participants was 53 ± 21 years; 47% were female.

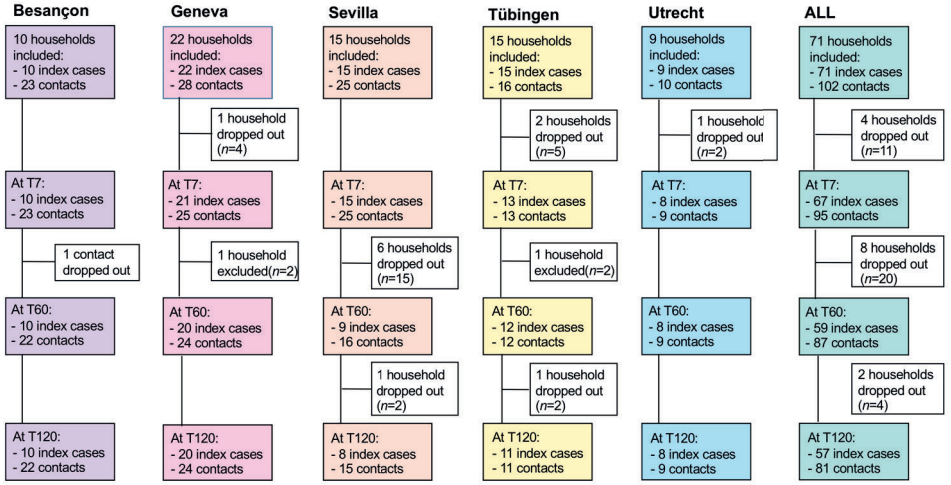

Figure 1. Study flow diagram of study participants, by centre and overall.

Table 1. Characteristics of households included in the study

		ESBL-Ec	ESBL-Kp	ESBL-Ec & Kp
Total n		45	20	6
Study site				
	Besançon	7 (15.6)	3 (15.0)	0
	Geneva	12 (26.7)	6 (30.0)	4 (66.7)
	Sevilla	9 (20.0)	6 (30.0)	0
	Tübingen	11 (24.4)	2 (10.0)	2 (33.3)
	Utrecht	6 (13.3)	3 (15.0)	0
Participating household members *n*				
	2	33 (73.3)	14 (70.0)	5 (83.3)
	3	7 (15.6)	3 (15.0)	1 (16.7)
	>4	5 (11.1)	3 (15.0)	0
Children in the household				
	<18 years	9 (20.0)	7 (35.0)	1 (16.7)
	<5 years	3 (6.7)	4 (20)	0
Household exposure to >1 antibiotics[a]				
	T60	7 (15.6)	3 (15.0)	0 (0)
	T120	7 (15.6)	1 (5.0)	0 (0)
Number of toilets in household				
	>2	17 (39.5)	8 (40.0)	3 (60.0)
Bath separated from toilet		16 (36.4)	3 (15.0)	2 (33.3)
Surface area of living space (m^2), median±SD		122.2 ± 69.7	154.2 ± 82.3	132 ± 45.7
Vegetarians in household		1 (2.3)	1 (5.0)	0

Abbreviations: ESBL, extended-spectrum ß-lactamase; ESBL-Ec, ESBL-producing *Escherichia coli*; ESBL-Kp, ESBL-producing *Klebsiella pneumoniae*; ESBL-PE, ESBL-producing *Enterobacteriaceae*. Data are reported as n (%), unless stated otherwise.[a] during follow-up

Profile of index cases and household contacts

Baseline characteristics of index cases and household contacts are presented in table 2 and table S1. During the hospital stay, 32% (*n*=23) of index cases had an ESBL-PE infection and 39% (*n*=28) received antibiotics at hospital discharge.

ESBL-PE carriage and acquisition

At baseline, index cases were carrying ESBL-Ec (*n*=45, 63%) or ESBL-Kp (*n*=20, 28%) or both (*n*=6, 8%). Among household contacts already positive at baseline (*n*=9, 31%), 79% (23/29) were carrying the same ESBL-PE as their corresponding index case.Twenty-six percent (17/65) of household contacts with complete follow up acquired ESBL-PE (ESBL-Ec,11; ESBL-Kp, 6). Most ESBL-PE acquisitions occurred during the first 2 months (1st week: 41%; 2nd to 8th week: 29%). One-third of index cases (*n*=27) were ESBL-PE negative at the end of follow up.

Genetic profiles

Overall, 38 different STs were observed for ESBL-Ec and 29 for ESBL-Kp (figure S1). Among ESBL-Ec strains, ST131 was the most frequent ST (46%). Less frequent STs were ST38 (6.9%), ST1193 (4%) and ST10 (3.6%). STs from ESBL-Kp showed a large heterogeneity (figure S2). Of 44 different ESBL-encoding genes identified, the most frequent was $bla_{\text{CTX-M-15}}$, detected in 142 ESBL-Ec and 79 ESBL-Kp isolates.

Clonally related co-carriage and transmission of related isolates

At baseline,14 out of 29 positive household contacts had isolates clonally related to the index case. The overall prevalence of cocarriage of clonally related isolates was 34% (32/94) over the entire study period. By combining epidemiological information with whole-genome sequencing data (figure 2), 19 clonally related transmission events were identified showing two possible directions: from the index case to his/her household contacts (*n*=13) and vice versa (*n*=6). Two additional closely related transmission events were identified for household BE07 from Besançon (18 to 24 SNP differences). The isolates belonged to ST80 and the intra-individual genome variability of the ESBL-Ec isolates retrieved from the index case throughout all sampling points ranged from 7 to 11 SNP differences. Most of the transmissions involved ESBL-Ec (14/21), with nine of them transmitted by the index case (table 3, table S2). Fifteen of 21 (71%) transmission events occurred during the first 2 months of follow up. The phylogenetic trees of retrieved ESBL-Ec and ESBL-Kp strains are shown in the Supplementary material (figures S3 and S4).

Table 2. Main characteristics of ESBL-PE-positive index cases included in the study

	ESBL-Ec (*n* = 45)	**ESBL-Kp** (*n* = 20)	**ESBL-Ec & Kp** (*n* = 6)
Demographic			
Age (years), median (range)	62 (21–89)	64 (28–96)	57.5 (51–83)
Female gender	16 (35.6)	9 (45.0)	2 (33.3)
Highest education			
Primary school	11 (24.4)	7 (35.0)	0
Secondary school	11 (24.4)	8 (40.0)	0
Technical school	11 (24.4)	4 (20.0)	0
University	5 (11.1)	1 (5.0)	5 (83.3)
Other/unknown	7 (15.6)	0	1 (16.6)
Antibiotic exposure in last 12 months	19 (42.2)	8 (40.0)	1 (16.7)
Travel abroad in last 12 months	23 (52.3)	5 (25.0)	4 (66.7)
Dietary habits			
Omnivore	42 (97.7)	19 (95)	5 (83.3)
Weekly meat consumption	38.5 (86.0)	20 (100)	4 (67)
Vegetarian	1 (2.3)	1 (5.0)	0
Hospital length of stay			
1–7 days	19 (42.2)	3 (15.0)	3 (50.0)
8–14 days	10 (22.2)	6 (30.0)	1 (16.7)
15–28 days	8 (17.8)	6 (30.0)	0
>28 days	8 (17.8)	5 (25.0)	2 (33.3)
Co-morbidities	40 (88.9)	18 (90.0)	5 (83.3)
Autoimmune disease	0	2 (10.0)	0
Cardiovascular disease	20 (44.4)	7 (35.0)	2 (33.3)
Chronic dermatological disease	4 (8.9)	1 (5.0)	1 (16.7)
Chronic renal failure	7 (15.6)	1 (5.0)	0
COPD[a]	3 (6.7)	2 (10)	0
Diabetes	14 (31.1)	3 (15.0)	0
Gastrointestinal disease	7 (15.6)	3 (15.0)	0
Chronic diarrhoea	1 (2.2)	0	0
Hepatic disease	4 (8.9)	2 (10.0)	0
Inflammatory bowel disease	3 (6.7)	2 (10)	0
Hemiplegia	0	1 (5.0)	0
Immunosuppression	5 (11.1)	4 (20.0)	1 (16.7)
Malignancy	14 (31.1)	9 (45.0)	1 (16.7)
Other	19 (42.2)	10 (50.0)	4 (66.7)
ESBL-PE infection during hospitalisation			
Yes	15 (33.3)	5 (25.0)	3 (50.0)
No	26 (57.8)	13 (65.0)	3 (50.0)
Unknown	4 (8.9)	2 (10.0)	0
Antibiotics at discharge			
Yes	19 (42.2)	8 (40.0)	1 (16.7)
No	26 (57.8)	11 (55.0)	4 (66.7)
Unknown	0	1 (5.0)	1 (16.7)
Incontinence	6 (13.3)	6 (30.0)	0
Urinary incontinence	3 (6.7)	4 (20.0)	0
Faecal incontinence	2 (4.4)	2 (10.0)	0
Both	1 (2.2)	0	0
Indwelling device at discharge	34 (75.6)	12 (60.0)	5 (83.3)
Intravascular	4 (8.9)	4 (20.0)	1 (16.7)
Urinary	1 (2.2)	2 (10.0)	0
Other	7 (15.6)	2 (10.0)	0
Patient autonomy			
Not completely autonomous	19 (42.2)	11 (55.0)	3 (50.0)
Needs support by family members	12 (26.7)	8 (40.0)	2 (33.3)
Help required for urinary/faecal excretion	2 (4.4)	6 (30.0)	0
Home care by healthcare personnel	12 (26.7)	5 (25.0)	1 (16.7)

Abbreviations: ESBL, extended-spectrum ß-lactamase; ESBL-Ec, ESBL-producing *Escherichia coli*; ESBL-Kp, ESBL-producing *Klebsiella pneumoniae*; ESBL-PE, ESBL-producing *Enterobacteriaceae*. Data are reported as n (%), unless stated otherwise. [a] Chronic obstructive pulmonary disease

Figure 2. Transmission events of clonally related and closely related isolates of extended-spectrum ß-lactamase (ESBL) -producing *Escherichia coli* and *Klebsiella pneumoniae*, with direction of the transmission pathways.

The figure gives the sequence type of the transmitted strains and pairwise single nucleotide polymorphism (SNP) differences between the isolates concerned. Each line of the table contains the information for a single household. Each square box represents a sample from a participant at a given sampling time-point (i.e. #1, #2, #3, #4). Red and green colours correspond to samples positive with ESBL-producing *E. coli* and *K. pneumoniae*, respectively. Grey colour corresponds to samples that were negative for ESBL-producing *Enterobacteriaceae*. Transmission events were identified in two directions: from index case (A) to household members (B to E) and from household contacts to index case. Red boxes (with *) represent clonally related ESBL-producing *E. coli* strains and green boxes (with *) represent clonally related ESBL-producing *K. pneumoniae*. MLST, multilocus sequence type.

Table 3. Crude numbers and incidence rates of acquisition and transmission events, based on core genome multi-locus sequence typing with pairwise single nucleotide polymorphism differences

		Acquisitions from any source			Transmissions in any direction			Transmissions from index case to household contacts		
	ESBL-	Ec	Kp	PE	Ec	Kp	PE	Ec	Kp	PE
Crude number		13	12	17	12	7	19	7	6	13
Incidence rate (per 100 participant-weeks at risk)		1.06	0.65	1.90	0.93	1.16	1.18	0.53	1.00	0.80

Abbreviations: ESBL, extended-spectrum ß-lactamase; ESBL-Ec, ESBL-producing *Escherichia coli*; ESBL-Kp, ESBL-producing *Klebsiella pneumoniae*; ESBL-PE, ESBL-producing *Enterobacteriaceae*

Incidence rates of household acquisition and transmission of ESBL-PE

The overall ESBL-PE acquisition rate was 1.9/100 participant-weeks at risk (table 3). ESBL-Ec had a higher rate of acquisition than ESBL-Kp (1.06 versus 0.65/100 participant-weeks at risk; relative risk (RR) 1.65; 95% CI 0.69e3.95). The rate of any clonally related ESBL-PE transmission within households was 1.18 events/100 participant-weeks of follow up, with the corresponding figure for transmissions only from the index case to household contacts of 0.8/100 participant-weeks (table 3). Although not statistically significant, a higher overall transmission rate was observed for ESBL-Kp than for ESBL-Ec (1.16 versus 0.93 per 100 participant-weeks at risk; RR 1.25; 95% CI 0.42e3.44) considering all possible transmission paths. A higher rate of ESBL-Kp transmission was also observed from index cases to household contacts (RR 1.87; 95% CI 0.52-6.49).

Risk factors for ESBL-PE acquisition and transmission

By univariable, mixed-effects Poisson regression, multiple explanatory factors were significantly associated with the risk of acquiring ESBL-PE among previously ESBL-PE-free household contacts (table S3): (a) index case determinants: hemiplegia, faecal incontinence, previous abdominal infection, proton-pump inhibitor therapy, >2 antibiotic courses after discharge, additional hospitalisations, and assistance provided by household members, in particular for urinary and faecal excretion; (b) household member determinants: age >50 years, travel abroad, assistance provided by healthcare personnel, help requested for various activities, regular contact with domestic animals, meat and seafood exposure, as well as the number of antibiotic courses. By multivariable analysis in a parsimonious model, assistance provided by family members to the index case (adjusted prevalence ratio (aPR) 2.9; 95% CI 1.1-8.0) showed the strongest association with ESBL-PE household acquisition, whereas frequency of meat consumption (aPR 1.4; 95% CI 0.4-5.3) and antibiotic exposure (aPR 1.4; 95% CI 0.4-4.2) showed only weak evidence of a positive association.

Fourteen variables were found to be significantly associated with the risk of ESBL-PE transmission from the index case to household members in the univariable analysis (table S4): (a) index case determinants: higher education (protective), full autonomy (protective), malignancy, faecal incontinence, previous abdominal infection, urinary catheter, proton-pump inhibitor therapy, three or more antibiotic courses, >1 hospitalisations, and assistance provided by family members, in particular for urinary and faecal excretion;

(b) household member determinants: spouse of index case, antibiotic intake and active helper of index case.

In the final multilevel Poisson regression model, assistance provided by household members for urinary and faecal excretion was strongly associated with increased risk of ESBL-PE transmission (aPR 4.3; 95% CI 1.3-14.1), whereas household antibiotic exposure showed weaker evidence of a positive association (aPR 2.1; 95% CI 0.7-7.0).

DISCUSSION

The principal findings of this international cohort study were: (a) clonally related ESBL-PE household transmission after hospital discharge of an ESBL-PE carrier occurred in 19 of 94 participants; (b) most acquisition and transmission events were observed during the first 2 months; (c) other household members were potential sources of cross-transmission, but to a lesser degree; (d) the ESBLPE acquisition rate was higher than the transmission rate; so, exogenous acquisition events occurred even without intra-household transmission; (e) the rate of household transmission was higher for ESBL-Kp than for ESBL-Ec; and (f) assistance provided by family members for urinary and faecal excretion of the index case was the most important risk factor for ESBL-PE transmission.

A recent meta-analysis examining clonally related ESBL-PE among household members documented co-carriage proportions of 12% (95% CI 8%-16%), and acquisition rates ranging from 0.16 to 0.20 events/100 participant-weeks of follow up.[2] In contrast, our study observed higher co-carriage proportions (34%) and 10-fold higher acquisition rates (1.9 events per 100 participant-weeks at risk). The higher proportion of co-carriage in the present study might have been influenced by sampling and detection methods, as the use of enrichment broths and selection of multiple colonies per sample might have improved the yield. Furthermore, it may reflect a higher risk of ESBL-PE transmission within enrolled households before study participation. The differences in acquisition rates depend on the length of follow up: longer follow up periods result in smaller rates. Indeed, 12-month follow-up studies found lower acquisition rates in contrast to shorter follow-up studies, which reported acquisition rates of up to 1.74 closely related ESBL-PE/100 person-weeks.[2,8,11] Furthermore, the higher proportion of infected, dependent and antibiotic-treated index cases in our study might have increased early transmission risk for household members compared with previous studies.

The incidence of ESBL-Ec acquisition was higher than the rate for ESBL-Kp.

In contrast, household transmission rates were higher for ESBL-Kp compared with ESBL-Ec. This apparent contradiction is explained by the acquisition of ESBL-Ec from a wide range of sources (e.g. food, animals, travel)[12,13], whereas transmission, as defined here, only involved human-to-human transfer. Similar observations have also been described for healthcare settings, suggesting that biological differences between bacterial species could explain higher ESBL-Kp transmission rates.[14,15] An alternative explanation might be the slightly higher intra-species diversity of ESBL-Ec within households (mean number of different STs observed per family 1.6 in ESBL-Ec versus 1.3 in ESBL-Kp). Furthermore, the frequency and intensity of human interactions may facilitate transmission of ESBL-KP, especially among elderly patients.[16] Indeed, in our study, index patients carrying ESBL-Kp were sicker and more dependent on external care, leading to increased proximity and risk of transmission.

As *Enterobacteriaceae* are coloniwers of the intestinal tract, the faecal-oral route plays an important role in the transmission chain. As in healthcare settings, where hand hygiene has been shown to be a key factor to reduce pathogen transmission[17], general hygiene measures rather than decreased intake or inappropriate handling of contaminated food may become an important preventive measure to reduce ESBL-PE transmission within households, especially if family members provide assistance to a sick relative.[18]

Hitherto, no previous study with these design characteristics and high-resolution typing methods has been conducted in high-income settings to ascertain putative transmission events within entire families, although ESBL-PE acquisition and transmission in the community or low-income settings has previously been investigated.[11,12,19-24] Therefore, the present study provides a solid methodological foundation for future studies and prioritisation of infection control interventions in the community setting.

Several limitations of this study merit consideration. First, not all members living in the same household participated in the study, omitting possible transmission events. Fortunately, the participation rate was high enough (80%) to draw meaningful conclusions. Second, by choosing not more than four colonies from a faecal sample, clonally distinct strains might have been missed, introducing a possible selection bias and underestimating the true transmission rate. As observed in a few participants (16%), each host may carry several ESBL-Ec strains simultaneously. However, we hypothesise that isolates not retrieved might present a low inoculum with lower transmission risk compared with dominating ESBL-Ec strains. Third, we did not yet

conduct plasmid typing, which is part of a complementary investigation, providing a more comprehensive picture of ESBL transmission in the community, especially for *E. coli*. Fourth, the role of intermediate vectors (i.e. animal) or environmental reservoirs (i.e. surfaces, water) in ESBL-PE transmission was not directly examined, but was assumed as part of direct human-to-human transmission. However, fomite-mediated transmission was accounted for in the estimation of exogenous risk factors by collecting relevant epidemiological information. Fifth, participants' intestinal load of ESBL-PE was not quantified, preventing the consideration of the inoculum effect as an independent risk factor. However, the bacterial load is influenced by several factors that were collected and accounted for in the analysis (e.g. antibiotic exposure, hospital length of stay).

In summary, ESBL-PE carriers discharged from the hospital were an important source of ESBL-PE transmission within households. Most acquisition and transmission events occurred during the first 2 months after hospital discharge. They were associated with care activities at home, highlighting the importance of hygiene measures to prevent community spread.

Funding: This study was part of a Joint Programming Initiative on Antimicrobial Resistance collaborative research project, under the 2016 Joint Call framework (Transnational Research Projects on the Transmission Dynamics of Antibacterial Resistance). It received funding from the following national research agencies: Instituto de Salud Carlos III (grant no. AC16/00076), Netherlands Organization for Health Research and Development (grant no. AC681055), Swiss National Science Foundation (grant no. 40AR40-173608), German Federal Ministry of Education and Research (grant no. 01KI1830) and Agence Nationale de la Recherche (grant no. ANR-16-JPEC-0007-03). As part of a separate research project, Marlieke de Kraker has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement nos 115523, 115620 and 115737 (Combatting Bacterial Resistance in Europe projects (COMBACTE)), resources of which are composed of financial contribution from the European Union's 7th Framework Programme (FP7/2007±2013) and the European Federation of Pharmaceutical Industries and associations companies' in-kind contribution. Also, Elena Salamanca, Mercedes Delgado and Jesús Rodríguez-Bano received support for research from by the Plan Nacional de IþDþi 2013-2016 and Instituto de Salud Carlos III, Subdireccion General de Redes y Centros de Investigacion Cooperativa, Ministerio de Ciencia, Innovacion y Universidades, Spanish Network for Research in Infectious Diseases (REIPI RD16/0016/ 0001), co-financed by the European Development Regional Fund 'A way to achieve Europe', Operative Programme Intelligence Growth 2014-2020.

Acknowledgements: We would like to thank all the patients and their families, as well as hospital and laboratory personnel involved in the study.

Conflicts of interest: All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. SH reports personal fees from Sandoz, outside the submitted work. SP reports advisory board and speaker honoraria from Biomerieux, Illumina and IDbyDNA outside the submitted work. All other authors have nothing to disclose.

REFERENCES

1. Tschudin-Sutter S, Lucet J-C, Mutters NT *et al*. Contact precautions for preventing nosocomial transmission of extended-spectrum ß-lactamase-producing *Escherichia coli*: a point/counterpoint review. *Clin Infect Dis* 2017; **65**: 342e7.

2. Martischang R, Riccio ME, Abbas M *et al*. Household carriage and acquisition of extended-spectrum ß-lactamase-producing *Enterobacteriaceae*: a systematic review. *Infect Control Hosp Epidemiol* 2020; **41**: 286e94.

3. Kluytmans-van den Bergh MF, Rossen JW, Bruijning-Verhagen PC *et al*. Whole-genome multilocus sequence typing of extended-spectrum-beta-lactamase-producing *Enterobacteriaceae*. *J Clin Microbiol* 2016; **54**: 2919e27.

4. Lohr IH, Rettedal S, Natas OB *et al*. Long-term faecal carriage in infants and intra-household transmission of CTX-M-15-producing *Klebsiella pneumoniae* following a nosocomial outbreak. *J Antimicrob Chemother* 2013; **68**: 1043e8.

5. Zankari E, Hasman H, Cosentino S *et al*. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012; **67**: 2640e4.

6. Davis S, Pettengill J, Luo Y *et al*. CFSAN SNP Pipeline: an automated method for constructing SNP matrices from nextgeneration sequence data. *PeerJ Comput Sci* 2015; **1**: e20.

7. Schurch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect* 2018; **24**: 350e4.

8. Stewardson AJ, Vervoort J, Adriaenssens N *et al*. Effect of outpatient antibiotics for urinary tract infections on antimicrobial resistance among commensal *Enterobacteriaceae*: a multinational prospective cohort study. *Clin Microbiol Infect* 2018; **24**: 972e9.

9. Samore MH, Tonnerre C, Hannah EL *et al*. Impact of outpatient antibiotic use on carriage of ampicillin-resistant *Escherichia coli*. *Antimicrob Agents Chemother* 2011; **55**: 1135e41.

10. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 2012; **17**: 228e43.

11. Haverkate MR, Platteel TN, Fluit AC *et al*. Quantifying within-household transmission of extended-spectrum beta-lactamase-producing bacteria. *Clin Microbiol Infect* 2017; **23**: 46. e1ee7.

12. Arcilla MS, van Hattem JM, Haverkate MR *et al*. Import and spread of extended-spectrum beta-lactamase-producing *Enterobacteriaceae* by international travellers (COMBAT study): a prospective, multicentre cohort study. *Lancet Infect Dis* 2017; **17**: 78e85.

13. Perez-Rodriguez F, Mercanoglu Taban B. A State-of-art review on multi-drug resistant pathogens in foods of animal origin: risk factors and mitigation strategies. *Front Microbiol* 2019; **10**: 2091.

14. Kluytmans-van den Bergh MFQ, Bruijning-Verhagen PCJ, Vandenbroucke-Grauls C *et al*. Contact precautions in single-bed or multiple-bed roms for patients with extended-spectrum beta-lactamase-producing *Enterobacteriaceae* in Dutch hospitals: a cluster-randomised, crossover, non-inferiority study. *Lancet Infect Dis* 2019; **19**: 1069e79.

15. Maechler F, Schwab F, Hansen S *et al*. Contact isolation versus standard precautions to decrease acquisition of extended-spectrum beta-lactamase-producing *Enterobacterales* in non-critical care wards: a cluster-randomised crossover trial. *Lancet Infect Dis* 2020; **20**: 575e84.

16. Duval A, Obadia T, Boelle PY *et al*. Close proximity interactions support transmission of ESBL-*K. pneumoniae* but not ESBL-*E. coli* in healthcare settings. *PLoS Comput Biol* 2019; **15**: e1006496.

17. Luangasanatip N, Hongsuwan M, Limmathurotsakul D *et al*. Comparative efficacy of interventions to promote hand hygiene in hospital: systematic review and network meta-analysis. *BMJ* 2015; **351**: h3728.

18. Day MJ, Hopkins KL, Wareham DW *et al*. Extended-spectrum beta-lactamase-producing *Escherichia coli* in humanderived and foodchain-derived samples from England, Wales, and Scotland: an epidemiological surveillance and typing study. *Lancet Infect Dis* 2019; **19**: 1325e35.

19. Rodriguez-Bano J, Lopez-Cerero L, Navarro MD *et al*. Faecal carriage of extended-spectrum beta-lactamase-producing *Escherichia coli*: prevalence, risk factors and molecular epidemiology. *J Antimicrob Chemother* 2008; **62**: 1142e9.

20. Valverde A, Grill F, Coque TM *et al*. High rate of intestinal colonization with extended-spectrum-beta-lactamase-producing organisms in household contacts of infected community patients. *J Clin Microbiol* 2008; **46**: 2796e9.

21. Lo WU, Ho PL, Chow KH *et al*. Fecal carriage of CTXM type extended-spectrum beta-lactamase-producing organisms by children and their household contacts. *J Infect* 2010; **60**: 286e92.

22. Mihaila L, Wyplosz B, Clermont O *et al*. Probable intra-family transmission of a highly virulent CTX-M-3-producing *Escherichia coli* belonging to the emerging phylogenetic subgroup D2 O102- ST405 clone. *J Antimicrob Chemother* 2010; **65**: 1537e9.

23. Hilty M, Betsch BY, Bogli-Stuber K *et al*. Transmission dynamics of extended-spectrum beta-lactamase-producing *Enterobacteriaceae* in the tertiary care hospital and the household setting. *Clin Infect Dis* 2012; **55**: 967e75.

24. Maataoui N, Langendorf C, Berthe F *et al*. Increased risk of acquisition and transmission of ESBL-producing *Enterobacteriaceae* in malnourished children exposed to amoxicillin. *J Antimicrob Chemother* 2020; **75**: 709e17.

5

SUPPLEMENTARY MATERIAL

Appendix 1 (sample size calculation)

The sample size was determined for the primary outcome without a pre-specified a priori hypothesis for the risk factor analysis. We assumed an ESBL-PE transmission rate of 10-20% among household members, a cluster size (i.e. number of individuals per household) of 3 and an intraclass correlation coefficient of 0.20 due to the clustering of individuals within families. With a ratio of 1:1 of ESBL-*E. coli* and ESBL-*K. pneumoniae* cases, the planned sample size of 100 index patients (with at least 1 household member) was considered sufficient for the purpose of this observational cohort study.

Appendix 2 (microbiologic methods)

Faecal samples and swabs were streaked directly on ChromID ESBL agar (bioMérieux, Marcy l'Etoile, France) plus additionally in MacConkey broth supplemented with vancomycin 64 µg/mL and 32 µg/mL cefuroxime, incubated for 24 h at 35°C. Centres using rectal swab had verified visually the presence of faecal material in sampling tubes (i.e. white swab tips having brownish stains). As stated by several expert sources, correctly performed rectal swabs remain « an acceptable and practical proxy for the collection of faecal specimens for stool microbiota analysis » (Basis CM *et al*. Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles. *BMC Microbiol*ogy. 2017. DOI: 10.1186/s12866-017-0983-9). Ten µl of the broth was then streaked on ChromID ESBL agar and further incubated for 48 h at 35°C. Each colony morphology was identified using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF). ESBL production was confirmed by double disk synergy tests (DDST20 and DDST30) and by the determination of the ß-lactamase inhibition profile (ESBL + AmpC Screen ID Kit, Rosco Diagnostica, Taastrup, Denmark). Based on distinct colony morphology, each centre stored at -80°C 1 to 4 isolates per sample in bead-containing cryotubes (Microbank, PRO-LAB Diagnostics, ON, Canada) until further analysis.

Table S1. Main characteristics of participating household contacts

	ESBL	**Ec** (*n*=63)	**Kp** (*n*=32)	**PE** (*n*=7)
Demographics				
	Age (median, range)	54 (2-79)	41 (1-92)	55 (26-84)
	Female gender	36 (57.1)	16 (50)	3 (42.9)
Highest educational level				
	Primary school	16 (25.4)	12 (37.5)	0
	Secondary school	11 (17.5)	7 (21.9)	0
	Technical school	15 (23.8)	6 (18.8)	0
	University	9 (14.3)	5 (15.6)	5 (71.4)
	Other/unknown	12 (19.1)	2 (2.3)	2 (28.6)
Healthcare & antibiotic exposure last 12 months				
	Hospitalisation	1 (1.6)	1 (3.1)	2 (28.6)
	Antibiotics last 12 months	19 (30.2)	5 (15.6)	1 (14.3)
	Antibiotics at enrolment	2 (3.2)	0	0
Travel abroad last 12 months		30 (48.4)	9 (28.1)	5 (71.4)
Dietary habits				
	Omnivore	57 (90.5)	29 (90.6)	6 (85.7)
	Vegetarian	1 (1.6)	2 (6.2)	0
Relation to the index case				
	Spouse	38 (60.3)	17 (53.1)	6 (85.7)
	Daughter/son	20 (31.8)	14 (43.8)	1 (14.3)
	Parent	1 (1.6)	0	0
	Sibling	1 (1.6)	0	0
	Grand-parent	1 (1.6)	0	0
	Parent in law	0	1 (3.1)	0
	No relationship	2 (3.2)	0	0

Abbreviations: ESBL, extended-spectrum ß-lactamase; Ec, ESBL-producing *Escherichia coli*; Kp, ESBL-producing *Klebsiella pneumoniae*; PE, ESBL-producing *Enterobacteriaceae*. Data are reported as *n* (%), unless stated otherwise.

Table S2. Clonally (*n*=19) or closely (*n*=2) related transmission events confirmed by analysis of cgMLST and SNP differences. For each centre, it shows the number of ESBL-PE transmission events identified for ESBL-*E. coli* (Ec) and ESBL-*K. pneumoniae* (Kp), at day 7 (#2), day 60 (#3) & day 120 (#4) of follow-up.

Genotypically confirmed transmission		**Sample #2**		**Sample #3**		**Sample #4**		**Total**
		Ec	**Kp**	**Ec**	**Kp**	**Ec**	**Kp**	
Besançon	TOTAL	**1**		**2**			**1**	**4**
	index case --> members	1		2			1	4
	members --> index case							0
	members --> members							0
Geneva	TOTAL	2	2	4	1	1		10
	index case --> members	1	2	2	1			6
	members --> index case	1		2		1		4
	members --> members							0
Sevilla	TOTAL	1	1	0	1	1	1	5
	index case --> members		1			1	1	3
	members --> index case	1			1			2
	members --> members							0
Tübingen	TOTAL	0	0	0	0	2		2
	index case --> members					2		2
	members --> index case							0
	members --> members							0
Utrecht	TOTAL	0	0	0	0	0		0
	index case --> members							0
	members --> index case							0
	members --> members							0
TOTAL		4	3	6	2	4	2	21

Table S3. Risk factors of acquisition of ESBL-producing *Enterobacteriaceae* by previously ESBL-free household contacts (*n*=17), stratified by index patient versus household variables, analysed by univariable mixed effects Poisson regression.

	Exposure variable	Prev[a] Ratio	Std. Err.	z	P>|z|	95% CI	
Household characteristics	Household surface >100m2	0.99	0.40	-0.03	0.98	0.45	2.18
	>1 toilet per household	1.09	0.71	0.13	0.89	0.30	3.94
	≥3 Household members	0.61	0.34	-0.88	0.38	0.21	1.81
	Presence of children ≤3 years old	0.92	0.26	-0.30	0.77	0.53	1.59
Index case: At baseline	Age of index case > 50	1.17	0.24	0.75	0.45	0.78	1.74
	Gender (male)	0.80	0.33	-0.53	0.60	0.36	1.80
	Nationality (non-Swiss)	2.69	1.40	1.90	0.06	0.97	7.46
	Higher education	0.79	0.18	-1.02	0.31	0.50	1.24
	Absence of comorbidities	0.64	0.87	-0.33	0.74	0.05	9.07
	---	---	---	---	---	---	---
	Chronic renal failure	1.15	0.43	0.38	0.70	0.56	2.38
	Cardio-vascular disease	1.36	0.84	0.50	0.62	0.41	4.59
	Diabetes	1.83	0.79	1.39	0.16	0.78	4.26
	Hemiplegia	5.35	1.14	7.87	<0.001	3.52	8.13
	Chronic dermatologic disease	0.62	0.47	-0.64	0.52	0.14	2.71
	---	---	---	---	---	---	---
	Chronic obstructive pulmonary disease	1.29	1.22	0.27	0.79	0.20	8.24
	Immunosuppression	1.03	0.50	0.05	0.96	0.39	2.67
	Gastrointestinal disease	0.36	0.38	-0.98	0.33	0.05	2.80
	Malignancy	1.45	0.82	0.67	0.50	0.48	4.37
	Inflammatory bowel disease	0.84	1.01	-0.14	0.89	0.08	8.86
	---	---	---	---	---	---	---
	Any incontinence	1.97	0.99	1.36	0.18	0.74	5.28
	Faecal incontinence	3.00	0.71	4.66	<0.001	1.89	4.76
	Urinary incontinence	1.15	0.32	0.51	0.61	0.67	1.98
	No indwelling device at hospital discharge	0.65	0.22	-1.29	0.20	0.34	1.25
	Urinary catheter at hospital discharge	1.79	0.81	1.30	0.20	0.74	4.33
	---	---	---	---	---	---	---
	Intravascular catheter at hospital discharge	0.83	0.46	-0.33	0.74	0.28	2.48
	Complete autonomy	0.92	0.27	-0.30	0.77	0.52	1.62
	Infection with ESBL-producing organisms[b]	0.72	0.30	-0.78	0.44	0.32	1.65
	Infection site: urinary tract	0.82	0.17	-0.97	0.33	0.55	1.22
	Infection site: abdominal tract	1.75	0.40	2.45	0.01	1.12	2.73
	---	---	---	---	---	---	---
	Antibiotic therapy at discharge	1.64	0.51	1.59	0.11	0.89	3.01
Index case: During follow-up	1 additional antibiotic course	1.90	0.90	1.35	0.18	0.75	4.83
	2 additional antibiotic courses	1.03	0.80	0.03	0.98	0.22	4.74
	3 additional antibiotic courses	2.43	0.99	2.19	0.03	1.10	5.38
	Proton pump inhibitors	1.90	0.45	2.70	0.01	1.19	3.02
	H2-receptor antagonists	0.71	0.95	-0.26	0.80	0.05	9.95
	---	---	---	---	---	---	---
	Oral corticosteroids/immunosuppressives	0.76	0.63	-0.33	0.74	0.15	3.92
	1 additional hospitalization	2.03	0.78	1.84	0.07	0.95	4.32
	2 additional hospitalizations	2.43	1.03	2.10	0.04	1.06	5.57
	Urinary incontinence	1.47	0.73	0.78	0.44	0.56	3.90
	Faecal incontinence	2.25	1.03	1.78	0.08	0.92	5.51
	---	---	---	---	---	---	---
	Indwelling device	2.50	1.22	1.87	0.06	0.96	6.52
	Urinary catheter	2.17	0.99	1.70	0.09	0.89	5.33
	Completely autonomous	0.29	0.25	-1.43	0.15	0.05	1.57
	Help provided by healthcare professional	3.02	2.32	1.44	0.15	0.67	13.6
	Help provided by family members	2.91	1.26	2.48	0.01	1.25	6.78
	---	---	---	---	---	---	---
	Help needed for food preparation	1.11	0.65	0.19	0.85	0.36	3.48
	Help needed for feeding	1.29	0.51	0.65	0.51	0.60	2.81
	Help needed for medication intake	1.96	1.14	1.16	0.25	0.63	6.15
	Help needed for urinary & faecal excretion	3.00	1.18	2.79	0.01	1.39	6.50
	Help needed for dressing	0.97	0.73	-0.04	0.97	0.22	4.21
	---	---	---	---	---	---	---
	Help needed for bed position shift	2.11	1.07	1.46	0.14	0.77	5.73
	Shared bath towel with other family members	1.16	0.35	0.51	0.61	0.65	2.10
	Prepared food for other household members	0.83	0.33	-0.46	0.64	0.38	1.81
	Cleaned hands before & while cooking meat	0.65	0.60	-0.46	0.64	0.11	4.00
	Stored separated raw & cooked food	0.56	0.37	-0.88	0.38	0.15	2.05
	---	---	---	---	---	---	---
	Cleaned surfaces & cooking materials between each meat preparation	0.54	0.51	-0.65	0.52	0.09	3.43
	Different cooking utensils for raw & cooked food	0.48	0.22	-1.58	0.11	0.19	1.19

	Exposure variable	Prev[a] ratio	Std. Err.	z	P>|z|	95% CI	
Household member: Baseline	Gender (male)	1.14	0.40	0.37	0.71	0.57	2.27
	Age household member > 50	1.61	0.18	4.16	<0.001	1.29	2.01
	Higher education	0.84	0.13	-1.13	0.26	0.63	1.13
	Spouse of index case	1.35	0.54	0.74	0.46	0.62	2.94
	Son/daughter of index case	0.54	0.18	-1.82	0.07	0.27	1.05
	----	---	---	---	---	---	---
	Vegetarian	1.08	0.29	0.26	0.79	0.63	1.84
	Number of travels abroad	1.16	0.09	1.98	0.05	1.00	1.34
Household member: Follow-up	Helper of the index case during follow-up	1.74	0.93	1.03	0.30	0.61	4.97
	Help provided by healthcare professional	3.71	0.76	6.37	<0.001	2.48	5.55
	Help provided by family member	1.79	1.30	0.80	0.42	0.43	7.44
	Help needed for food preparation	2.75	1.33	2.08	0.04	1.06	7.11
	Help needed for feeding	3.71	1.42	3.41	<0.001	1.75	7.87
	----	---	---	---	---	---	---
	Help needed for urinary and faecal excretion	2.75	1.76	1.58	0.11	0.78	9.64
	Help needed for dressing	2.17	1.36	1.24	0.21	0.64	7.42
	Help needed for any mobility	3.71	1.42	3.41	<0.001	1.75	7.87
	Regular contact with domestic animals	0.64	0.11	-2.55	<0.001	0.45	0.90
	Regular contact with cat	1.41	0.24	2.01	0.04	1.01	1.97
	----	---	---	---	---	---	---
	Swim in a river or lake	0.92	0.65	-0.12	0.90	0.23	3.66
	Share towel	0.97	0.16	-0.18	0.86	0.71	1.33
	Eat at least once per week: beef	1.50	0.66	0.93	0.35	0.64	3.54
	Eat at least once per week: lamb	3.14	0.44	8.24	<0.001	2.39	4.12
	Eat at least once per week: pork	1.50	0.27	2.28	0.02	1.06	2.14
	----	---	---	---	---	---	---
	Eat at least once per week: poultry	1.41	0.25	1.89	0.06	0.99	2.00
	Eat at least once per week: fish	1.91	0.68	1.81	0.07	0.95	3.85
	Eat at least once per week: other seafood	2.56	0.72	3.33	<0.001	1.47	4.46
	Spent time cooking meat products	1.12	0.81	0.16	0.88	0.27	4.61
	Prepare food for other household members	1.07	0.59	0.13	0.90	0.37	3.15
	----	---	---	---	---	---	---
	Different cooking utensils for raw & cooked food	1.02	0.51	0.05	0.96	0.39	2.71
	Number of antibiotic courses	2.18	0.38	4.44	<0.001	1.55	3.07

a prevalence ratio. b during at last hospitalisation

Table S4. Risk factors of clonally related ESBL-PE household transmission from index case to household contacts (*n*=13), analysed by univariate mixed-effects Poisson regression, stratified by index patient versus household variables, including characteristics present at baseline and during follow-up

	Exposure variable	Prev[a] Ratio	Std. Err.	z	P>|z|	95% CI	
Household characteristics	Age	0.78	0.26	-0.74	0.46	0.40	1.51
	Household surface <100m2	1.03	0.51	0.07	0.95	0.40	2.69
	Main bathroom separated from the toilet	0.52	0.29	-1.17	0.24	0.17	1.55
	Number of toilets in the household	1.43	0.37	1.39	0.17	0.86	2.36
	Number of household members	0.85	0.14	-0.94	0.35	0.62	1.19
	----	---	---	---	---	---	---
	Presence of infants ≤ 3 years old	1.35	1.14	0.35	0.73	0.26	7.09
Index case: At baseline	Absence of comorbidities	0.43	0.57	-0.63	0.53	0.03	5.72
	Chronic obstructive pulmonary disease	1.85	1.66	0.69	0.49	0.32	10.76
	Cardio-vascular disease	1.03	0.55	0.06	0.95	0.36	2.93
	Chronic dermatologic disease	0.89	0.64	-0.16	0.87	0.22	3.64
	Diabetes	2.79	1.67	1.72	0.09	0.87	8.99
	----	---	---	---	---	---	---
	Malignancy	2.63	1.29	1.97	0.05	1.01	6.90
	Inflammatory bowel disease	1.22	1.62	0.15	0.88	0.09	16.46
	Gastro-intestinal disease	0.51	0.61	-0.56	0.57	0.05	5.22
	Immunosuppression	0.69	0.38	-0.67	0.51	0.23	2.04
	Faecal incontinence	4.72	0.91	8.04	<0.001	3.23	6.89
	----	---	---	---	---	---	---
	Urinary incontinence	0.77	0.55	-0.36	0.72	0.19	3.11
	Help provided by healthcare professional	1.15	0.39	0.41	0.68	0.59	2.22
	Help provided by family member	2.35	0.76	2.65	<0.001	1.25	4.42
	Antibiotic prescribed at hospital discharge	1.20	0.66	0.33	0.75	0.41	3.51
	Higher education	0.16	0.08	-3.45	<0.001	0.05	0.45
	----	---	---	---	---	---	---
	Infection with ESBL during last hospitalisation	1.63	0.80	1.00	0.32	0.62	4.28
	Abdominal infection site	5.45	1.15	8.04	<0.001	3.61	8.25
Index case: During follow-up	1 additional antibiotic course	2.13	1.70	0.95	0.34	0.45	10.15
	2 additional antibiotic courses	2.49	1.95	1.16	0.25	0.54	11.58
	3 additional antibiotic courses	5.90	2.35	4.47	<0.001	2.71	12.86
	Oral corticosteroids/other immunosuppressives	0.52	0.48	-0.71	0.48	0.08	3.17
	Proton pump inhibitors	2.99	1.24	2.64	0.01	1.33	6.75
	----	---	---	---	---	---	---
	H2-receptor antagonists	1.04	1.57	0.02	0.98	0.05	20.1
	Faecal incontinence	3.42	1.45	2.91	<0.001	1.49	7.84
	Urinary incontinence	1.36	1.05	0.40	0.69	0.30	6.17
	Indwelling device	3.94	2.91	1.86	0.06	0.93	16.76
	Urinary catheter	3.26	1.58	2.44	0.02	1.26	8.42
	----	---	---	---	---	---	---
	Diarrhoea	1.60	0.63	1.21	0.23	0.74	3.45
	Autonomous	0.18	0.11	-2.88	<0.001	0.06	0.58
	Help provided by family member	3.97	1.83	3.00	<0.001	1.61	9.79
	Help provided by healthcare professional	3.00	1.86	1.77	0.08	0.89	10.12
	Help needed for dressing	1.44	1.14	0.46	0.65	0.30	6.82
	---	---	---	---	---	---	---
	Help needed for urinary and faecal excretion	4.73	2.04	3.60	<0.001	2.03	11.01
	Help needed for food preparation	1.70	1.11	0.81	0.42	0.47	6.10
	Help needed for personal hygiene	1.23	0.98	0.26	0.79	0.26	5.84
	Help needed for medication intake	2.98	1.46	2.23	0.03	1.14	7.78
	Help needed for mobility	1.52	1.23	0.52	0.60	0.31	7.45
	----	---	---	---	---	---	---
	Help needed for bed position shift	3.01	1.62	2.05	0.04	1.05	8.65
	Help needed for feeding	1.84	0.82	1.38	0.17	0.77	4.41
	≥ 2 hospitalisations after discharge	3.59	1.26	3.64	<0.001	1.80	7.15
	Spent time cooking meat products	0.53	0.19	-1.75	0.08	0.26	1.08
	Prepared food for other household members	0.53	0.20	-1.70	0.09	0.26	1.10
	----	---	---	---	---	---	---
	Shared bath towels with other contacts	1.13	0.39	0.35	0.72	0.57	2.24
Household member: Baseline	Age	1.06	0.28	0.23	0.82	0.64	1.76
	Current antibiotic intake	3.86	4.14	1.26	0.21	0.47	31.59
	Higher education	0.94	0.48	-0.11	0.91	0.35	2.53
	Spouse of index case	3.65	1.50	3.14	<0.001	1.63	8.19

	Exposure variable	Prev[a] ratio	Std. Err.	z	P>|z|	95% CI	
Household member: Follow-up	Antibiotic intake	2.59	0.48	5.11	<0.001	1.80	3.73
	Proton pump inhibitors	2.96	1.34	2.39	0.02	1.21	7.20
	Active helper of index case	3.84	1.75	2.95	<0.001	1.57	9.39
	Spent time cooking meat products	1.11	0.71	0.17	0.87	0.32	3.90
	Prepare food for other household members	1.78	1.08	0.94	0.35	0.54	5.87
	----	---	---	---	---	---	---
	Shared towel with index case	1.10	0.28	0.38	0.71	0.67	1.82

a prevalence ratio.

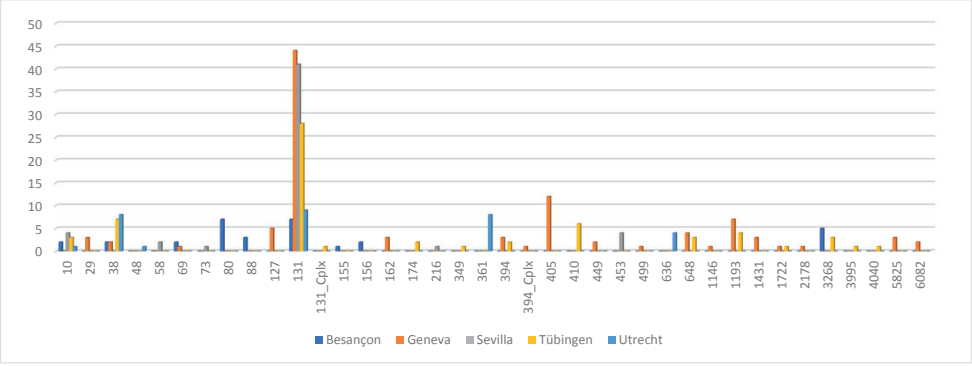

Figure S1. Sequence type distribution of ESBL-producing *E. coli* isolates per centre. Two new MLST were identified in Geneva (belonging to the clonal complex CC394) and Tübingen (belonging to CC131).

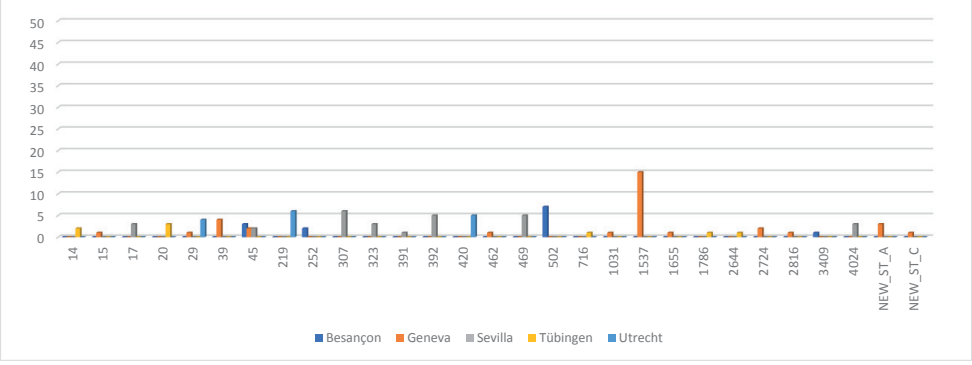

Figure S2. Sequence type distribution of ESBL-producing *K. pneumoniae* isolates per centre. Two new MLST profiles were described in Geneva, named New-ST-A and New-ST-C.

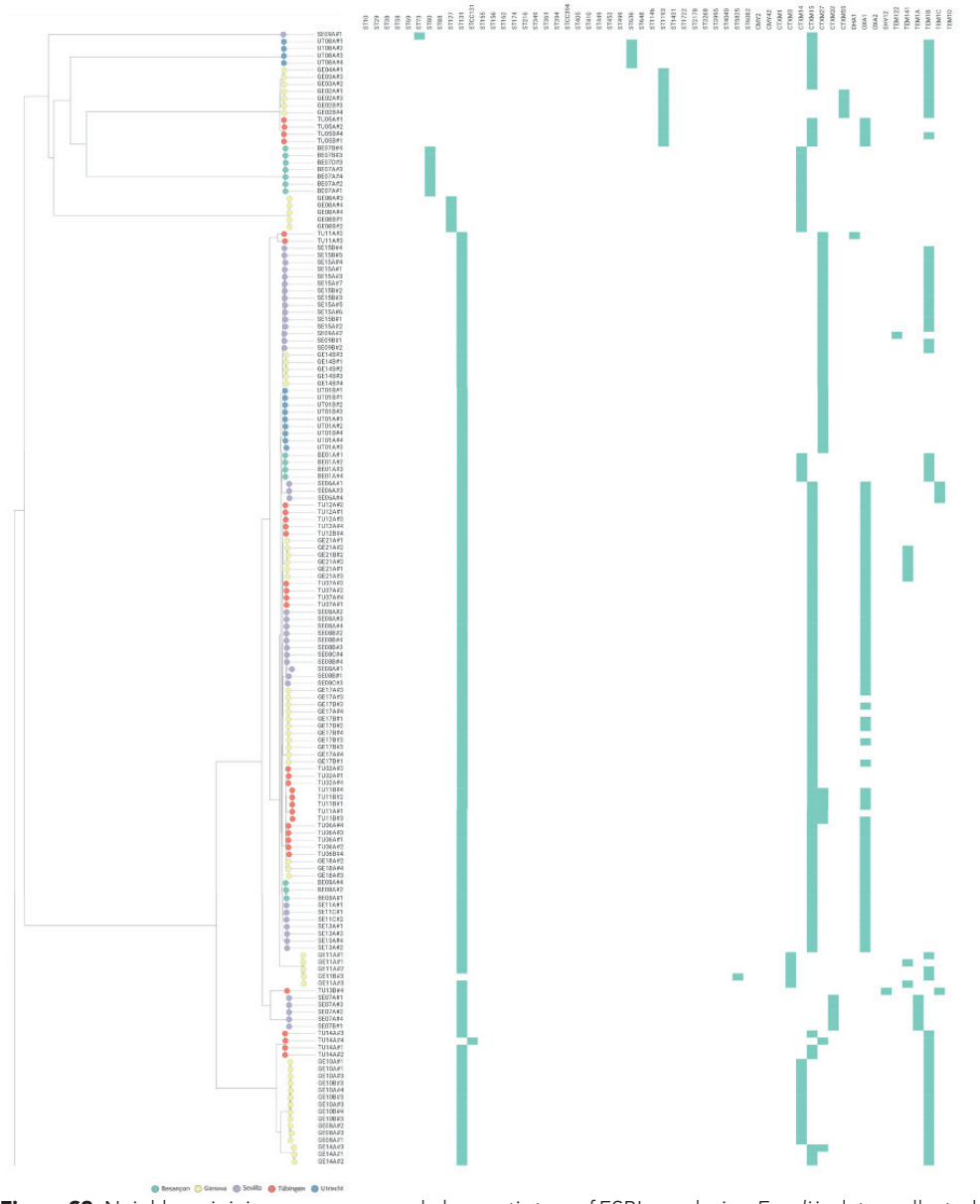

Figure S3. Neighbour joining core genome phylogenetic tree of ESBL-producing *E. coli* isolates collected during the 4-month follow-up in the 5 study centres, constructed with SeqSphere+ using the Enterobase scheme. Colour code indicates the respective MLSTs (see the legend for details).

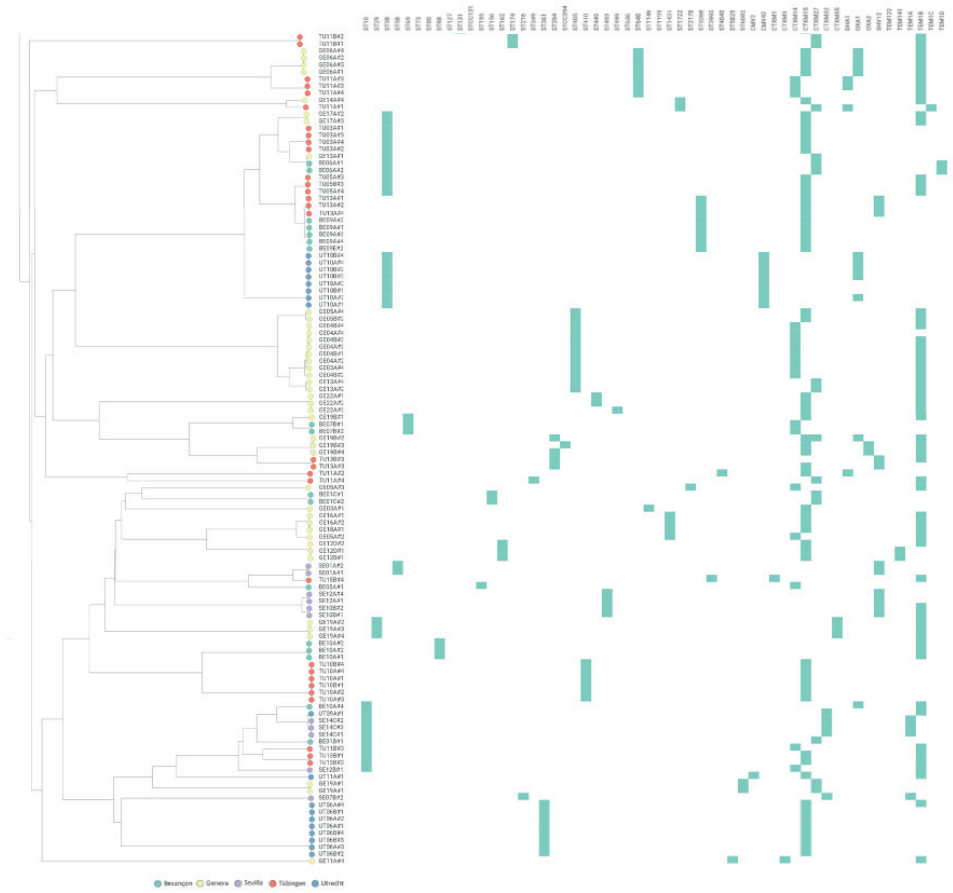

Figure S3. Continuation

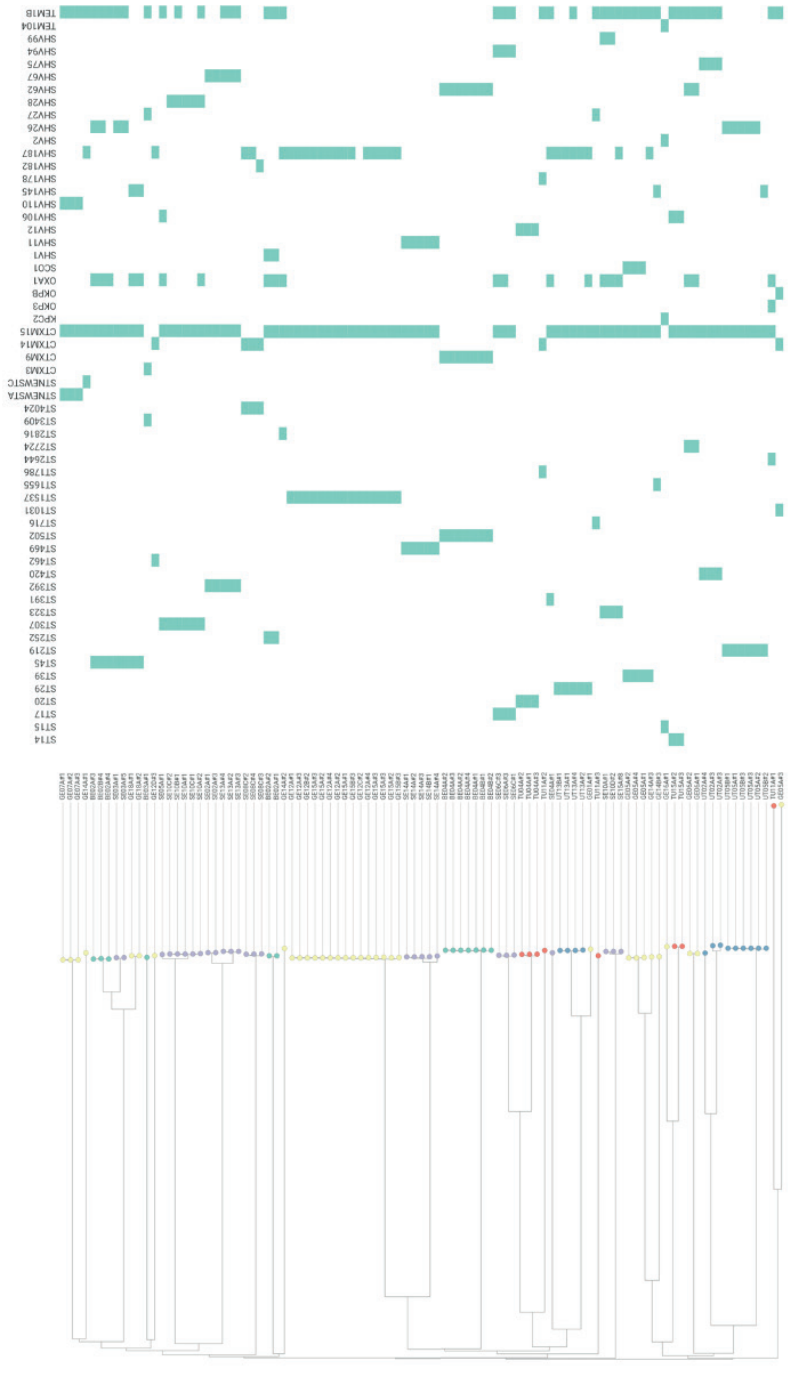

Figure S4. Neighbour joining core genome phylogenetic tree of ESBL-producing *K. pneumoniae* isolates collected during the 4-month follow-up in the 5 study centres , constructed with SeqSphere+ using the Enterobase scheme. Colour code indicates the respective MLST (see the legend for details).

6

Comparative genomics of ESBL-producing *Escherichia coli* and *Klebsiella pneumoniae* from household-, and long term care facility- cohorts in five European catchment areas assessing human carriage, the human-polluted environment, and food

Tess D Verschuuren, Julia Guther, Maria E Riccio, Daniel Martak, Elena Salamanca, Siri Göpel, Nadine Conzelmann, Jelle Scharringa, Patrick Musicha, Ingo B Autenrieth, Ben S Cooper, Didier Hocquet, Evelina Tacconelli, Jesús Rodriguez-Baño, Stephan Harbarth, Ad C Fluit, Silke Peter, and Jan AJW Kluytmans

Manuscript in preparation

ABSTRACT

Objective: To compare the antimicrobial resistance gene (ARG), and core genome (cg)-content of extended-spectrum ß-lactamase (ESBL)-producing- *Escherichia coli* (ESBL-Ec) and *Klebsiella pneumoniae* (ESBL-Kp) from human carriage, the human-polluted (hp)-environment and food from five European catchment areas.

Methods: Two cohort studies assessed ESBL carriage in: (i) index-subjects colonised and household contacts, (ii) residents of long-term care facilities (LTCFs). Additionally, hp-environment and food samples were collected. Overlap in ARG content was assessed using a proportional similarity index (PSI), and a principle component analysis. Overlap in cg-content was assessed using a cgMLST distance visualisation, and a cg-phylogeny.

Results: 482 ESBL-Ec and 171 ESBL-Kp isolates were included. During follow-up, both species were detected in 14/65 households, 3/6 LTCFs, and in 33/202 of ESBL-positive participants. The majority of ESBL gene types occurred in both species (12/20), $bla_{\text{CTX-M-15}}$ was most frequently observed, except in ESBL-Ec from food. Non-ESBL-ARGs occurrence was highest in ESBL-Kp and lowest in ESBL-Ec from food. ESBL-Kp often carried clinically relevant ARGs like *aac(3)-IIa*, *aac(6')-Ib-cr*, $bla_{\text{OXA-1}}$. Similarity of ESBL gene types was 0.59 (95% confidence interval 0.5-0.7) for ESBL-Ec and ESBL-Kp from humans, while 0.32 (95% confidence interval 0.2-0.4) for ESBL-Ec from humans and food. Additionally, the similarity between ESBL-Ec and ESBL-Kp from humans in ARG distributions encoding resistance for broad-spectrum ß-lactamases, and aminoglycosides was 0.64 and 0.68, respectively. ST131 was the most frequent ST for ESBL-Ec (29%), but absent in food. Isolates from humans and the hp-environment were frequently clonally related, limited links were observed with food.

Conclusions: ESBL-Ec and ESBL-Kp regularly co-occurred in human populations influenced by healthcare. Considerable overlap in ARG content was observed between these species. More research is needed to quantify the role of horizontal gene transfer in human-to-human transmission.

INTRODUCTION

Antimicrobial resistance was declared as one of the top 10 public health threats facing humanity by the World Health Organization in 2019.[1] In Europe, extended-spectrum ß-lactamase-producing *Escherichia coli* (ESBL-Ec) and *Klebsiella pneumoniae* (ESBL-Kp) are of particular concern, as these species are the most frequent causes of antimicrobial resistant infections.[2] Furthermore, *E. coli* infections at large are increasing, and often arise from the community.[3] Scientific effort has been put into elucidating reservoirs for human carriage of these bacteria, however yielding no clear evidence for important sources outside humans.[4-10] Overlap of antimicrobial resistance genes (ARGs) and plasmids has been described between *E. coli* and *K. pneumoniae*, which could be indicative for horizontal gene transfer (HGT).[11-16] Here, we compare the ARG-, and core genome (cg)-content of ESBL-Ec and ESBL-Kp from human carriage influenced by healthcare, the human-polluted (hp)- environment and food from five European catchment areas.

METHODS

Study design, data-collection and microbiologic methods

Two prospective cohort studies were performed (2017-19) in five European catchment areas: Besançon, Geneva, Seville, Tübingen, and Utrecht. Data-collection, sample processing, and microbiologic methods were described previously.[9,18,19]

Briefly, the household study recruited index ESBL-Ec and ESBL-Kp positive patients during hospitalisation. Participants were followed up at home for four months after discharge, together with ≥1 other household member, and provided a faecal sample at follow-up moments (*n*=4). In the long-term care facility (LTCF) study, participating residents were followed for eight months, and provided a faecal sample or perianal swab at follow-up moments (*n*=8).

Additionally, food and hp-environment samples were collected: (i) food was sampled on three time points in supermarkets were households indicated to usually shop, and eight times in the kitchen of the participating LTCFs, (ii) samples from the hp-environment were taken from: a) LTCF U-bends, b) LTCF surfaces, c) LTCF wastewater, d) wastewater treatment plant (WWTP) inflow connected to the LTCF, and e) downstream river to the WWTP. Reservoirs a) and b) were sampled twice, the other reservoirs eight times. Bacterial identification of unique morphologic isolates were performed by the local microbiology laboratory, using standardised methods.

Ethics

Approval was obtained by the ethical review board of each hospital. All enrolled participants or their representatives provided written informed consent for participation in this study.[17,18]

Sequencing and selection of unique isolates

Isolates identified as ESBL-Ec or ESBL-Kp were shipped to Tübingen or Utrecht. DNA isolation was performed (DNeasy UltraClean Microbial Kit, Qiagen). Sequencing was performed on all isolates (NextSeq and Miseq platforms, Illumina, San Diego, USA), followed by de novo assembly (Spades v3.11.1).

To reduce the influence of repeated measurements on molecular subtype distributions, a new dataset was created that contained only unique isolates. An unique isolate was defined as the first available isolate per participant or non-human sample of a species. A second isolate was included if the cgMLST pairwise genetic distance was >0.0105 for ESBL-Ec, or >0.0035 for ESBL-Kp.[19]

Distance matrices were generated (Ridom Seqsphere v5.0), using the Enterobase scheme (2,513 target genes) for ESBL-Ec, and the K. pneumonia/variicola/quasipneumoniae sensu lato scheme (2,358 target genes) for EBSL-Kp.[20,21] Pairwise distance expressed either as the proportion of allele differences (# allele differences/# good target gene shared).

In silico molecular typing

Sequence types (STs) were identified using tseemann/mlst (v2.16.2) (https://github.com/tseemann/mlst).[22] ARGs conferring resistance to ß-lactams, fluoroquinolones, aminoglycosides, fosfomycin, trimethoprim, sulfonamide, and colistin were determined with Resfinder (v3.2) using abricate (v0.8.10), with a length of ≤80% and an identity of ≤95%.[23] Detected ARGs were assigned to an antimicrobial class based on the genotype-phenotype translations, ß-lactam genes were specified as broad-spectrum ß-lactamase, ESBL, or carbapenemase.[24,25] Chromosomal combined *gyrA/parC* mutations conferring fluoroquinolone resistance (FQR) were determined with PointFinder (v4.1).[26] *FimH* types were assigned using FimTyper (v1.0).[27] Sub-clades were assigned to ST131 (ESBL-Ec) isolates: clade A (*FimH*41), clade B (*FimH*22), clade C (*FimH*30 – non-FQR), clade C1 (*FimH*30 - FQR-non $bla_{CTX\text{-}M\text{-}15}$), and lastly clade C2 (*FimH*30 - FQR-$bla_{CTX\text{-}M\text{-}15}$).[28] Observed other *FimH* types were assigned to a clade according to their position in the phylogenetic tree.[29]

Statistical and genetic analysis

The proportions of ESBL gene types, and non-ESBL-ARGs were compared using a two-proportion z-test. The mean number of ARGs were compared using a two-sample t-test.

The pairwise overlap between sampled reservoirs of acquired ARG type distributions was quantified with Czekanowski's proportional similarity index. PSI = 1 - 0.5*sum of *k|p(reservoir[n]))k - q(reservoir[nx])*k| where *p* corresponded to the relative frequency of gene type *k* in reservoir *n*, and *q* corresponded to the relative frequency of the same gene type in reservoir *nx*.[6,30] The denominator of the relative frequency was the total number genes in the corresponding ARG class. The PSI is a proportion, with 0 interpreted as no overlap, and 1 as perfect overlap in ARG type distributions between two reservoirs. Bootstrap iterations (5,000) were used to calculate 95% confidence intervals (CIs) (boot R-package, v1.3.23).[6]

The overlap in acquired ARG content of all sampled groups was analysed using a prinicipal component analysis (PCA), with 5,000 bootstrap iterations to account for uncertainty. A PCA reveals underlying structures of data by simplifying a dataset with many variables (e.g. ARG types) into principle components (PCs). PCs are expressed as a percentage of explained variance of all the observed variance in the data. ARG types with a proportion <0.01 in the total database were excluded.[6]

Clonal transmission between humans, and between non-human isolates were described previously.[9,17,18] Here, we assessed clonal relationships between humans and non-human isolates, with a visualisation where isolates were ordered based on epidemiological setting (qgraph R-package, v1.6.3). To prevent spurious relationships between the different reservoirs and epidemiologic settings, a different cgMLST threshold was chosen for ESBL-Ec of ≤0.0040 (~10 alleles) for definition of clonally related isolate pairs, based on previous work of the MODERN-studies.[9,31] The threshold for ESBL-Kp was kept at ≤0.0035 (~10 alleles).[22,31] A sensitivity analysis with increased thresholds was performed (figure S4).

Lastly, a phylogeny was created for ESBL-Ec and ESBL-Kp using PopPUNK (v1.1.3).[32] PopPUNK calculated a relative core and accessory distance for each pair in the dataset based on *k*-mer comparisons. With this, a neighbour joining cg-tree was constructed, and visualised using microreact (v98.0.0).[33]

RESULTS

In total, 202 participants carrying ESBL-Ec and/or ESBL-Kp (110 household-members, 92 LTCF residents), carrying 212 unique ESBL-Ec, and 101 unique ESBL-Kp isolates (table 1) were included. Additionally, 232 hp-environment and 108 food-isolates were included, corresponding to 270 ESBLEc, and 70 ESBL-Kp-isolates (table 1).

A list of all included strains was added to the supplement (table S1 (available upon request)). Seven ESBL-Kp isolates were retrospectively determined as *Klebsiella variicola* (all retrieved from wastewater in Utrecht).

During follow-up, 33/202 (16%) of ESBL-positive participants we colonised with both ESBL-Ec and ESBL-Kp. Furthermore, both species were detected in human samples of 14/65 (22%) households, and 3/6 LTCFs. A mean of 1.6 unique isolates (range 1-10) was detected in participants during follow-up (table 1). On average, the participants from Utrecht were colonised with 1.1 isolate while the participants from Seville were colonised with 1.7 isolates (p-value two sample t-test: 3e-06), mostly due to a higher observed mean of ESBL-Kp isolates. Five participants were colonised with two ST131 strains and 3 participants were colonised with three different ST131 strains during follow-up (cgMLST distance range: 0.017-0.29) (table S2).

Comparison of ARG content between ESBL-Ec and ESBL-Kp from humans, the hp-environment, and food

Most isolates harboured an ESBL gene from the $bla_{\text{CTX-M}}$ family: 88% (ESBL-Ec: 89%,ESBL-Kp: 86%). Followed by the bla_{SHV} family: 15% (ESBL-Ec: 12%, ESBL-Kp: 23%), and bla_{TEM}: 1% (ESBL-Ec: 2%, ESBL-Kp: 0.6%). In 26 isolates, two ESBL genes were present (ESBL-Ec: 2%, ESBL-Kp: 9%, p-value: 1e-04) (table S3).

Twelve of the 20 observed ESBL gene types occurred in both species, of which $bla_{\text{CTX-M-15}}$ was most frequent (ESBL-Ec: 44%, ESBL-Kp: 77%, p-value: 8e-14) (figure 1a). In ESBL-Ec, $bla_{\text{CTX-M-27}}$, $bla_{\text{CTX-M-1}}$, and $bla_{\text{CTX-M-55}}$ occurred significantly more, while $bla_{\text{SHV-106}}$, and $bla_{\text{SHV-129}}$ were more frequent in ESBL-Kp (figure 1a). Similar ESBL gene proportions were observed for human and hp-environment isolates (figure 1b). Compared to ESBL-Ec from human isolates (52%), $bla_{\text{CTX-M-15}}$ occurred less frequently in ESBL-Ec from food (11%, p-value: 6e-11), while $bla_{\text{SHV-12}}$, and $bla_{\text{CTX-M-1}}$ occurred more frequently (5% versus 40%, p-value: 2e-14, 3% versus 32%, p-value: 8e-12, respectively) (figure 1b).

Table 1. Characteristics of the included first unique extended-spectrum ß-lactamase-producing (ESBL) *Escherichia coli* (Ec) and *Klebsiella pneumoniae* (Kp) isolates from humans, the human-polluted environment, and food

	ESBL+	ESBL-Ec	ESBL-Kp
ESBL positive participants[a]	202	157	78
household members	110	82	41
LTCF-residents	92	75	37
Number of unique isolates per participant, mean (range)	1.6 (1-10)	1.1 (0-7)	0.48 (0-3)
household members[b]	1.4 (1-10)	1.0 (0-7)	0.43 (0-3)
LTCF residents[c]	1.7 (1-5)	1.1 (0-5)	0.53 (0-3)
Unique isolates	653	482	171
human	313	212	101
human-polluted environment	232	175	57
food	108	95	13
Composition human-polluted environment			
LTCF surface	15	7	8
LTCF U-bend	26	13	13
LTCF wastewater outflow	34	19	15[d]
wastewater treatment plant inflow	73	64	9
downstream river	84	72	12

ESBL+: extended-spectrum ß-lactamase-producing positive. Values correspond to *n* unless stated otherwise. [a] Number of participants carrying ESBL-*E. coli* and/or ESBL-*K. pneumoniae* on at least one time-point. [b] The maximum follow-up time of a household member was four months. [c] The maximum follow-up time of a LTCF resident was eight months. [d] seven isolates were retrospectively identified as *Klebsiella variicola*.

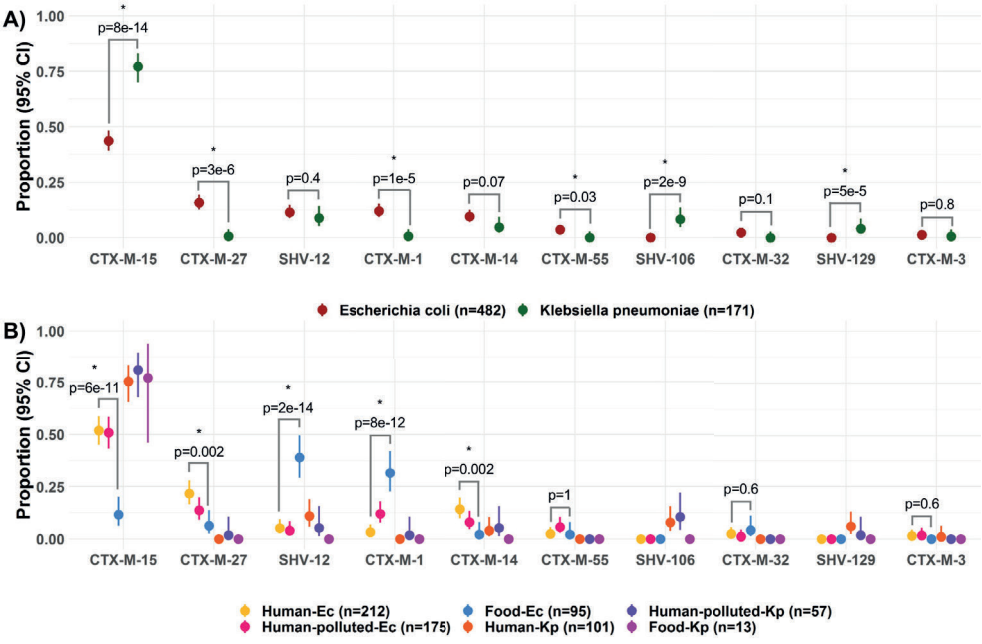

Figure 1. **A)** Proportions and 95% confidence intervals of the ten most frequent ESBL genes for *Escherichia coli* and *Klebsiella pneumoniae*. **B)** B) Proportions and 95% CIs of the ten most frequent ESBL genes for *E. coli* (Ec) and *K. pneumoniae* (Kp) per reservoir. 95% CIs of proportions of zero were removed from panel B for readability of the graph.Two-proportion z-tests were depicted (p-value), in panel B, two-proportion z-tests were depicted for *E. coli* from human- versus food -originating isolates (p-value). *statistically significant (p-value <0.05)

On average, 4.6 non-ESBL-ARGs were detected in ESBL-Ec, and 8.7 in ESBL-Kp (p-value: 2e-16) (table 2). Furthermore, the ARG prevalence was higher in ESBL-Kp for several antimicrobial classes: broadspectrum ß-lactamases (80% versus 49%), fluoroquinolones (98% versus 66%), aminoglycosides (80% versus 70%), fosfomycin (39% versus 1.5%), trimethoprim (80% versus 51%), and sulfonamide (73% versus 63%) (p-value range: 0.03-3e-39).

Within ESBL-Ec, the ARG prevalence was higher in human samples compared to food samples for: broad-spectrum ß-lactamases (62% versus 28%), aminoglycosides (77% versus 57%), and trimethoprim (54% versus 41%), (p-value range: 0.04-1e-07) (table 2).

Lastly, acquired ARGs codifying for carbapenem (ESBL-Ec (*n*=2): $bla_{\text{OXA-48}}$, $bla_{\text{OXA-181}}$, ESBL-Kp: $bla_{\text{KPC-2}}$ (*n*=2)), and colistin-resistance (ESBL-Kp: *mcr-1* (*n*=1)) were rare.

The similarity of ARG type distributions was greatest between human samples and the hp-environment of the same species (figure 2, S1). Nonetheless, 0.59-0.77 overlap was observed between ESBL-Ec and ESBL-Kp from humans for: ESBL, broad-spectrum ß-lactamase, aminoglycoside, and sulfonamide encoding ARGs. Overlap between species for fluoroquinolones, and trimethoprim was lower, around 30% (figure 2, S1). Interestingly, the similarity of ESBL genes was lowest for ESBL-Ec from food when compared to the other assessed reservoirs. 0.32 (0.2-0.4) of ESBL gene distributions overlapped between ESBL-Ec from humans and food, while 0.59 (0.5-0.7) between ESBL-Ec and ESBL-Kp from humans (figure 2a).

PCA confirmed that human samples and those from the hp-associated environment of the same species were most similar in ARG type content (figure 3, S2). Furthermore, PCA confirmed that the ESBL gene type distributions of ESBL-Ec from food were different from those of ESBL-Ec from humans due to a positive association of food with $bla_{\text{CTX-M-1}}$, $bla_{\text{SHV-12}}$, and a negative association with $bla_{\text{CTX-M-15}}$. Lastly, PCA analysis revealed that ESBL-Kp was associated with clinically relevant ARGs like: *aac(6')-Ib-cr* (tobramycin, amikacin, ciprofloxacin), *aac(3)-IIa* (gentamicin, tobramycin), $bla_{\text{OXA-1}}$ (amoxicillin-clavulanic acid, piperacillin-tazobactam), and *fosA* genes (fosfomycin) (figure 3, S2).

Table 2. Antimicrobial resistance gene prevalences per antimicrobial class (including chromosomal mutations for fluoroquinolones) of extended-spectrum ß-lactamase (ESBL)-producing *Escherichia coli* (ESBL-Ec) and ESBL-*Klebsiella pneumoniae* (ESBL-Kp).

Legend 0 10 20 30 40 50 60 70 80 90 100

	ESBL-Ec all isolates (*n*=482)		ESBL-Kp all isolates (*n*=171)		*p*[b]	ESBL-Ec			*p*[c]	ESBL-Kp			*p*[c]
	% (n)[1]	mean[2]	% (n)[1]	mean[2]		human (*n*=212) % (n)[1]	hp-envir[a] (*n*=175) % (n)[2]	food (*n*=95) % (n)[3]		human (*n*=101) % (n)[1]	hp-envir[a] (*n*=57) % (n)[2]	food (*n*=13) % (n)[3]	
broad-spectrum β-lactams	49 (238)	0.54	80 (137)	1.7	[1-2]5e-12 [2-3]2e-16	62 (131)	46 (80)	28 (27)	[1-2]0.002 [2-3]0.008 [1-3]1e-07	75 (76)	86 (76)	92 (12)	[1-2]0.16 [2-3]0.87 [1-3]0.30
fluoroquinolones	66 (317)	0.87	98 (168)	2.2	[1-2]2e-16 [2-3]2e-16	70 (149)	63 (110)	61 (58)	[1-2]0.15 [2-3]0.87 [1-3]0.14	97 (98)	100 (57)	100 (13)	[1-2]0.48 [2-3]- [1-3]1
aminoglycosides	70 (337)	1.7	80 (137)	2.7	[1-2]0.01 [2-3]2e-11	77 (163)	69 (120)	57 (54)	[1-2]0.09 [2-3]0.07 [1-3]6e-05	82 (83)	72 (41)	100 (13)	[1-2]0.19 [2-3]0.07 [1-3]0.21
carbapenems	0.4 (2)	0.004	1.1 (2)	0.011	[1-2]0.60 [2-3]0.39	0 (0)	1.1 (2)	0 (0)	[1-2]0.40 [2-3]0.76 [1-3]-	1.0 (1)	1.8 (1)	0 (0)	[1-2]1 [2-3]1 [1-3]1
colistin	0 (0)	0	0.6 (1)	0.006	[1-2]0.59 [2-3]0.32	0 (0)	0 (0)	0 (0)	-	0 (0)	0 (0)	7.7 (1)	[1-2]- [2-3]0.42 [1-3]0.22
fosfomycin	1.5 (7)	0.014	39 (66)	0.39	[1-2]3e-39 [2-3]2e-16	1.9 (4)	1.7 (3)	0 (0)	[1-2]1 [2-3]0.50 [1-3]0.42	30 (30)	53 (30)	46 (6)	[1-2]0.007 [2-3]0.91 [1-3]0.37
trimethoprim	51 (247)	0.55	80 (137)	0.81	[1-2]8e-11 [2-3]1e-09	54 (115)	53 (93)	41 (39)	[1-2]0.91 [2-3]0.08 [1-3]0.04	79 (80)	82 (47)	77 (10)	[1-2]0.78 [2-3]0.95 [1-3]1
sulfonamide	63 (306)	0.80	73 (125)	0.86	[1-2]0.03 [2-3]0.27	64 (135)	58 (102)	73 (69)	[1-2]0.33 [2-3]0.03 [1-3]0.16	75 (76)	67 (38)	85 (11)	[1-2]0.33 [2-3]0.35 [1-3]0.67
all	-	**4.47**	-	**8.74**	**[2]2e-16**	-	-	-	-	-	-	-	-

Colour gradient: 100% corresponds to full red. Each detected percentage lower than 100% or x is accompanied by a red colour with x% transparency, e.g. resistance genes detected for broad-spectrum ß-lactamases in 50% of isolates, corresponds to a red colour with 50% transparency.
[a] hp-envir: human-polluted environment.
[b] p-value 1 corresponds to a two-proportion z-test, p-value 2 corresponds to a two-sample t-test.
[c] p-values correspond to two-proportion z-tests.
P-values <0.05 were considered statistically significant.

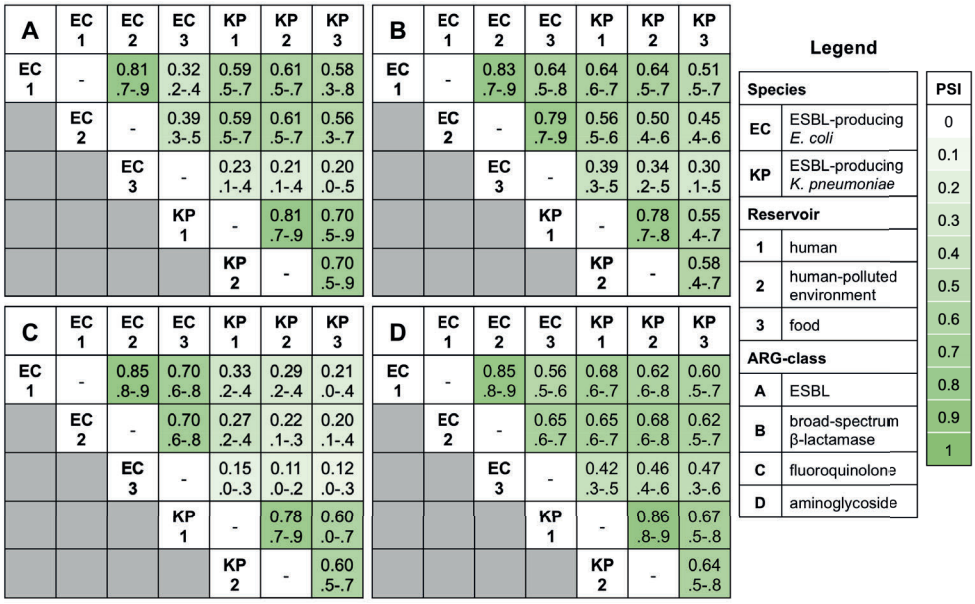

Figure 2. Czekanowski's proportional similarity index (PSI) (95% confidence interval) of acquired resistance genes of **A)** ESBL genes, **B)** broad-spectrum ß-lactamase genes, **C)** fluoroquinolone, and **D)** aminoglycosides. The PSI is calculated with the following formula: $1\text{-}0.5\sum k\ |p(\text{reservoir}[n])k\text{-}q(\text{reservoir}[nx])k|$, p: the relative frequency of gene type *k* in reservoir n, q: the relative frequency of the same gene type in reservoir nx. The denominator of the relative frequency was the total number genes of the corresponding ARG class. The PSI is a proportion, with 0 interpreted as no overlap, and 1 as perfect overlap in ARG type distributions between two reservoirs. 95% CIs were calculated with 5,000 bootstrap iterations. PSI analysis for fosfomycin, trimethoprim, and sulfonamide ARGs were described in the supplement (fig S1).

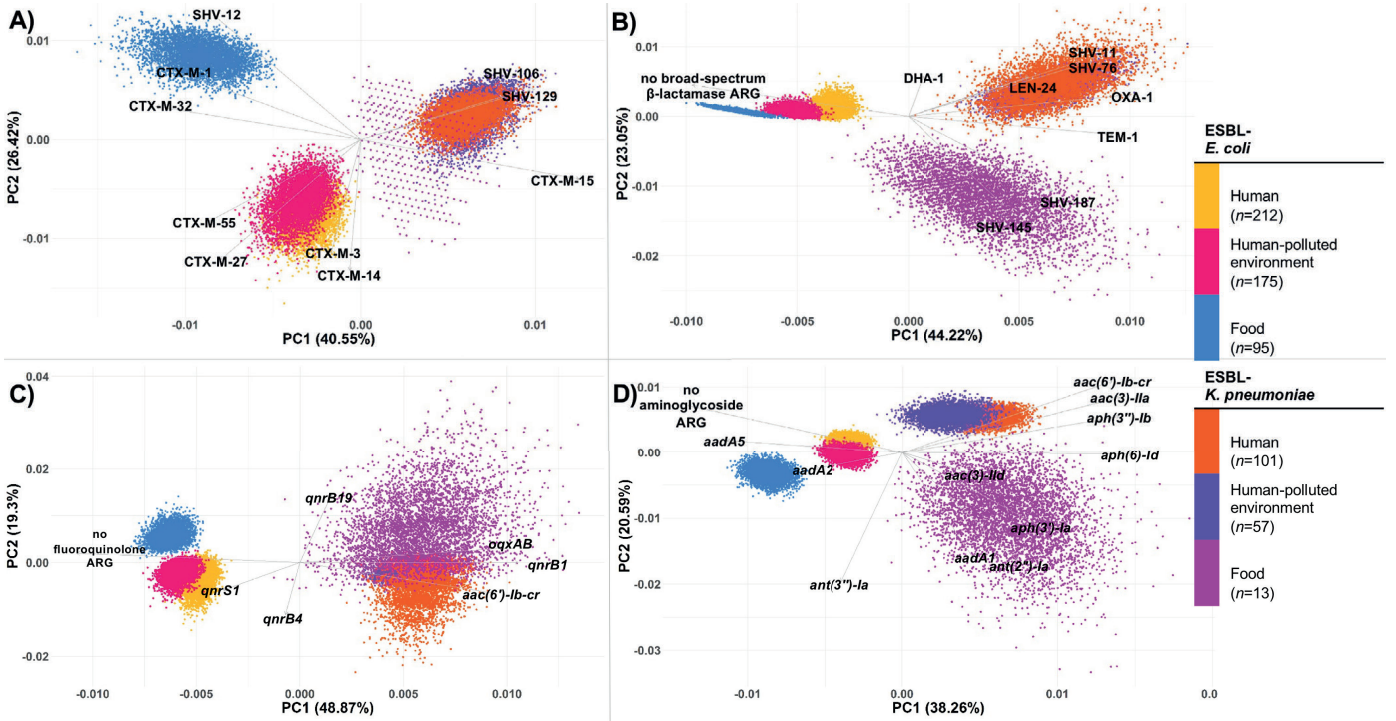

Figure 3. Principal Component Analysis (PCA) of antimicrobial resistance genes (ARGs).
The x-axis corresponds to PC1 and the corresponding explained variance, while the y-axis corresponds to PC2. **A)** Extended-spectrum ß-lactamase (ESBL) ARGs, PC1 (41% variance): positively associated with $bla_{CTX\text{-}M\text{-}15}$, and $bla_{SHV\text{-}106}$, and negatively associated with $bla_{CTX\text{-}M\text{-}1}$, and $bla_{CTX\text{-}M\text{-}32}$, PC2 (26% variance): positively associated with $bla_{SHV\text{-}12}$, and negatively associated with $bla_{CTX\text{-}M\text{-}14}$. **B)** broad-spectrum ß-lactamase ARGs, PC1 (44% variance): positively associated with $bla_{TEM\text{-}1}$, $bla_{OXA\text{-}1}$, and negatively associated with the absence of acquired resistance. **C)** fluoroquinolone ARGs (excluding chromosomal mutations), PC1 (49% variance): positively associated with *oqxAB, aac(6')-Ib-cr*, and ESBL-Kp, and negatively associated with absence of acquired resistance, and ESBL-Ec. PC2 (19% variance): positively associated with *qnrB19* and food, negatively associated with *qnrB4*. **D)** aminoglycoside ARGs, PC1 (38% variance): positively associated with ESBL-Kp, *aac(6')-Ib-cr, aac(3)-IIa, aph(6)-Id*, and *aph(3")-Ib*, and negatively associated with ESBL-Ec, *aadA5, aadA2*, and the absence of acquired resistance. Gene types with a proportion <0.01 of isolates were excluded from PCA analysis: ESBL ARGs: $bla_{CTX\text{-}M\text{-}9}$, $blaC_{TX\text{-}M\text{-}65}$, $bla_{CTX\text{-}M\text{-}8}$, $bla_{SHV\text{-}2}$, $bla_{TEM\text{-}169}$, $bla_{TEM\text{-}106}$, $bla_{CTX\text{-}M\text{-}209}$, $bla_{SHV\text{-}13}$, and $bla_{SHV\text{-}5}$. Excluded broad-spectrum ß-lactamase ARGs: $bla_{SHV\text{-}27}$, $bla_{SHV\text{-}62}$, $bla_{SHV\text{-}1}$, $bla_{TEM\text{-}122}$, $bla_{CARB\text{-}2}$, $bla_{OKP\text{-}B\text{-}1}$, $bla_{SHV\text{-}110}$, $bla_{OXA\text{-}2}$, $bla_{TEM\text{-}2}$, $bla_{SHV\text{-}75}$, $bla_{TEM\text{-}33}$, $bla_{OKP\text{-}B\text{-}3}$, $bla_{SHV\text{-}2}$, $bla_{TEM\text{-}169}$, $bla_{SCO\text{-}1}$, $bla_{TEM\text{-}210}$, $bla_{SHV\text{-}178}$, $bla_{OKP\text{-}B\text{-}9}$, $bla_{OKP\text{-}A\text{-}3}$, and $bla_{SHV\text{-}172}$. Excluded fluoroquinolone ARGs: *qnrS2, qnrA1, qepA4, qnrB52*. Excluded aminoglycoside ARGs: *aadA16, aph(4')-Ia, aac(3)-VIa, aac(3)-IVa, aph(3")-Id*. PCA analysis for fosfomycin, trimethoprim, and sulfonamide ARGs are shown in the supplement (fig S2).

Comparison of core genome content of isolates from humans, the hp-environment, and food

The three most frequent STs within ESBL-Ec were: ST131 (29%; 46% in human isolates, 23% hpenvironment, absent in food), ST10 (11%; 10% human, 10% hp-environment, 12% food), ST69 (5%; 3% human, 5% hp-environment, 9% food). For ESBL-Kp, the three most frequent STs were ST405 (22%; 27% in human isolates, 18% hp-environment, absent in food), ST307 (10%; 10% human, 12% hp-environment, absent in food), ST323 (6%; 3% human, 12% hp-environment, absent in food) (figure S3).

A pattern of clonally related isolates from humans and the hp-environment was observed (figure 4ab). This observation was most pronounced in Seville, where isolates from for residents were often clonally related to LTCF surfaces, U-bends, and LTCF wastewater outflow (figure 4c-d). These connections reflected human contamination from LTCF residents to the LTCF-associated environments, and potentially human acquisition through surface (and U-bend) contamination. Furthermore, similar patterns were observed for ESBL-Ec isolates from Besançon, Tübingen, and Geneva (figure 4a, S4a-b). Observed clonally related isolates from humans and WWTP inflow or downstream rivers, likely reflected human contamination from the total human population linked to the WWTPs. Lastly, two clonally related human-food pairs were observed for ESBL-Ec, both without a direct epidemiological link (figure 4a).

The neighbour joining cg-tree of ESBL-Ec showed no evident clustering based on sample group, with exception of the absence of food isolates within the ST131 and ST1193 phylogeny (figure 5a, https://microreact.org/project/jp35qTiHo76ntU7CYyGqKR/9b9d99f6). Both STs were observed in all cities, potentially indicating clonal human-to-human spread in a large geographic area. For ST131, clade A (24/138), B (5/138), and C (108/138) clustered separately, most of these isolates carried the $bla_{\text{CTX-M-15}}$, $bla_{\text{CTX-M-27}}$, and $bla_{\text{CTX-M-14}}$ ESBL genes (figure 5a). Clade C1 and C2 we distributed evenly with 51, and 57 isolates, respectively. Most of C1 isolates carried $bla_{\text{CTX-M-27}}$ (*n*=40).

For ESBL-Kp, most STs were observed in one catchment area, likely reflecting local clonal spread (figure 5b, https://microreact.org/project/kWAneo6PcVZKpmaH9sQvs1/e63cb4b2). However, ST405, and ST323 where observed in two cities, and ST219 (*n*=6) was observed in four cities from human, river, and chicken and turkey isolates, potentially reflecting clonal spread in a large geographic area. The few observed food isolates (*n*=13) occurred throughout the tree (figure 5b).

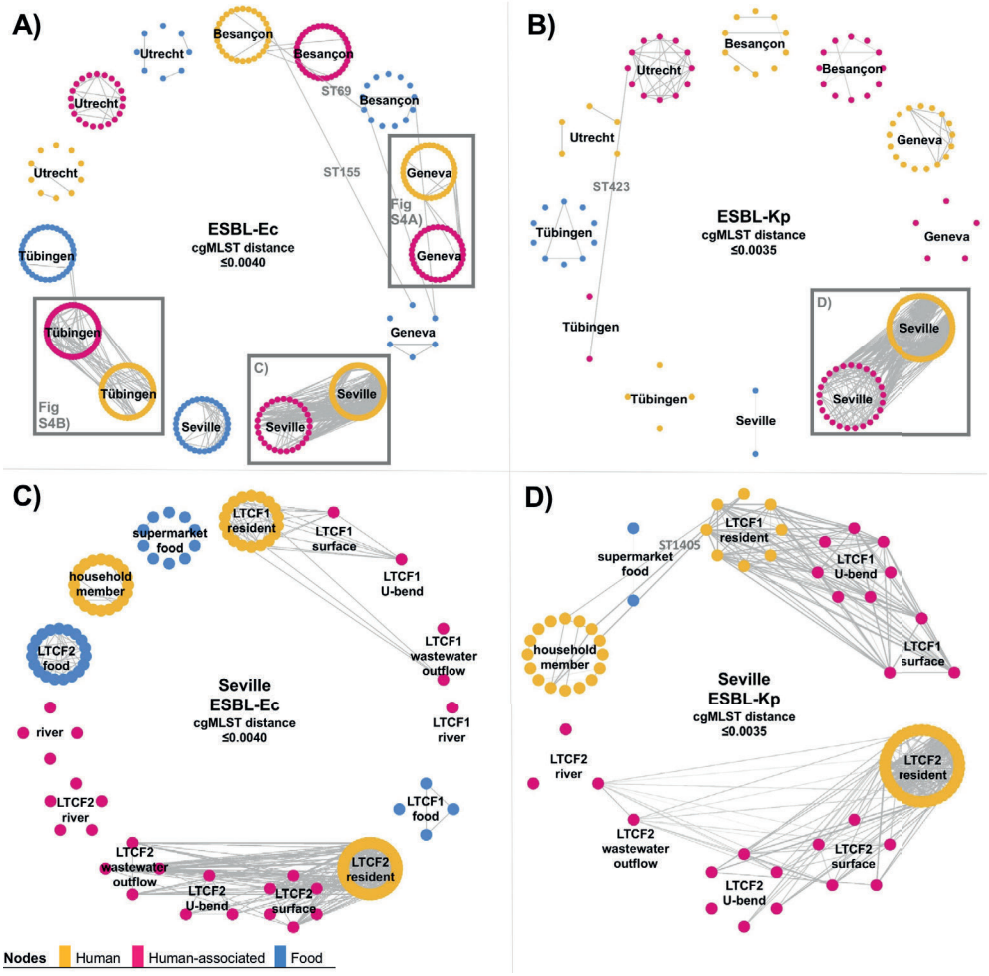

Figure 4. Visualisation of clonally related isolate pairs of extended-spectrum ß-lactamase (ESBL)-*Escherichia coli* (Ec) and ESBL-*Klebsiella pneumoniae* (Kp).
Nodes represent isolates and are grouped based on epidemiological setting, lines represent genetically similar isolate pairs. **A)** ESBL-Ec, all catchment areas, **B)** ESBL-Kp (including seven *K. variicola* isolates), **C)** ESBL-Ec from the catchment area of Seville, **D)** ESBL-Kp from the catchment area of Seville.

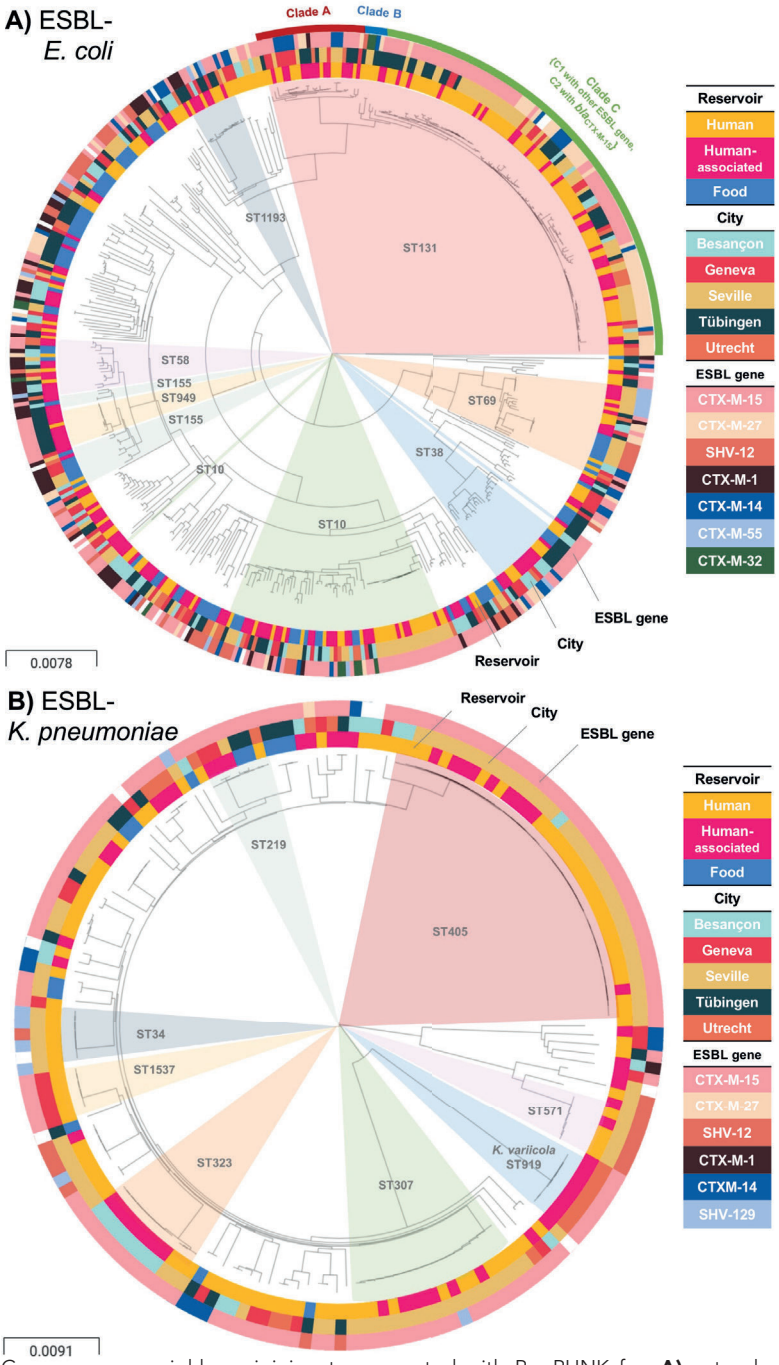

Figure 5. Core genome neighbour joining trees created with PopPUNK for: **A)** extended-spectrum ß-lactamase (ESBL)-producing *Escherichia coli* (ESBL-Ec) (*n*=482 isolates), **B)** ESBL-producing *Klebsiella pneumoniae* (ESBL-Kp) (*n*=171) isolates, including seven *Klebsiella variicola* isolates). Online access interactive trees: ESBL-Ec: https://microreact.org/project/jp35qTiHo76ntU7CYyGqKR/9b9d99f6, ESBL-Kp: https://microreact.org/project/kWAneo6PcVZKpmaH9sQvs1/e63cb4b2

DISCUSSION

In this European prospective study ESBL-Ec and ESBL-Kp regularly co-occurred in human populations influenced by healthcare. The majority of ESBL gene types occurred in both species. ESBL gene distributions of ESBL-Ec and ESBL-Kp sampled from humans were more similar than ESBL-Ec sampled from humans and food. Non-ESBL-ARGs occurrence was highest in ESBL-Kp and lowest in ESBL-Ec from food. While ESBL-Kp more often carried clinically relevant ARGs, substantial overlap was observed for broad-spectrum ß-lactamase, aminoglycoside encoding ARG types between ESBL-Ec and ESBL-Kp from humans.

Additional to the lowest ARG frequencies in ESBL-Ec from food, a low similarity in ESBL gene type distributions was observed between ESBL-Ec from food and humans. This second observation was mainly explained by a lower occurrence of $bla_{\text{CTX-M-15}}$, and higher occurrence of $bla_{\text{CTX-M-1}}$, and $bla_{\text{SHV-12}}$. Lastly, ST131, the most prevalent ST in the dataset was not found in food. Early research showed potential links between ESBL-Ec from food and humans.[34,35] However, recent evidence, demonstrated that, at least in high in-come countries, ESBL-Ec and ESBL-Kp mainly spreads through human-to-human transmission, while transmission of ESBL-Ec from food likely only occur as spill-over events (with the potential of introducing new resistance traits into humans).[4-10]

The $bla_{\text{CTX-M-15}}$ gene has been widely observed as the most frequently occurring ESBL-encoding gene in both *E. coli* and *K. pneumoniae* from human carriage and infection.[5,7,8,12,29] In this study, the majority of the detected ESBL gene types occurred in both species, with $bla_{\text{CTX-M-15}}$ detected in more than half of the included isolates. We, furthermore, observed considerable similarity in ARG content codifying resistance for broad-spectrum ß-lactamases, aminoglycosides, and sulfonamide. These observations are supported by studies assessing plasmids, demonstrating that genes harboured on *E. coli* and *K. pneumoniae* plasmids were shared amongst each other, and other species (mostly *Enterobacterales*).[11,13,15] Our findings were also supported by studies that described plasmid-mediated hospital outbreaks[13,36], amongst which a study by Hidalgo *et al.* that demonstrated the same $bla_{\text{OXA-48}}$ harbouring plasmid occurred in 14 different species during a period of 1.5 years, providing evidence for interspecies plasmid transfer.[16]

It is currently unknown how frequently HGT of ARGs occurs between *E. coli* and *K. pneumoniae*. However, based on these findings and literature, it is reasonable to assume that a certain flow of ARGs does occur in certain settings, like carriage in human populations influenced by healthcare. One hypothesis is that *K. pneumoniae* may be more likely to donate plasmids to *E. coli* than vice versa. bla_{SHV-1}, $bla_{CTX-M-1}$, bla_{KPC-1}, and bla_{NDM-1} were first described in *K. pneumoniae*, and ARG proportions are higher than in *E. coli*.[37-41] Furthermore, assessment of hospital bla_{OXA-48} outbreaks showed that the plasmid mainly originated from *K. pneumoniae*.[14,16] On the other hand, ESBL-Ec might be more likely to donate plasmids to ESBL-Kp, followed by local clonal spread of ESBL-Kp. For example, $bla_{CTX-M-15}$, bla_{OXA-1}, *aac(6')-Ib-cr* -ARG combination was first described in ESBL-Ec ST131.[42] As third possibility is that ESBL-Ec and ESBL-Kp mainly spread clonally, with only incidental exchange of ARGs through HGT. This hypothesis is supported by the observed transnational pattern of ST131, ST38, ST69, ST1193 in ESBL-Ec, and ST219, ST405, and ST323 in ESBL-Kp.[5,29,43,44] Additionally, the observed high similarity in ARG content and frequent clonal links between humans and the hp-environment, indicated that clonal spread likely was a frequent event in our study.

To our knowledge this is the first study to genetically compare ESBL-Ec and ESBL-Kp that were prospectively and longitudinally sampled from humans, the hp-environment and food in five European catchment areas with differing prevalences in humans. Due to the sampling strategy we were able to give unique insight in the co-occurrence of these species. Furthermore, the availability of short read sequences of all isolates enabled us to compare the assessed reservoirs on several genetic levels.

Several limitations need to be considered. Firstly, the results of this study couldn't assess HGT of ARGs between ESBL-Ec and ESBL-Kp due to a lack of long read sequencing. Transmission analysis using this sequence technology will be addressed in future work of the MODERN-studies. Secondly, the sample size of ESBL-Kp was considerably smaller than that of ESBL-Ec due to a lower prevalence of this species, which may have hampered comparison of the two species. ESBL-Kp from food was particularly rare, prohibiting us from drawing any conclusions about this sample group. Thirdly, we used two different cgMLST thresholds, one for unique intra-individual clones selection, and one to determine clonal tranmission between humans and non-human reservoirs. Using the same threshold would have led to either a higher number of unique isolates, or to more clonal relationships between humans and non-human groups. The optimal threshold for the sampled groups is unknown, and are likely to differ per situation.

Fourthly, the total number of ARGs per antimicrobial class that was used to calculate the PSI, differed per sample group. In particular for broad-spectrum ß-lactamases, fluoroquinolones, and aminoglycosides, ESBL-Kp often carried double the number of ARGs per isolate compared to ESBL-Ec, which could have influenced PSI results. However, results for ESBLs were not influenced, as isolates mostly carried one gene only. Longread-based plasmid analysis is need to definitively assess the overlap in ARG content between these species. Fifthly, although LTCF-associated food items were sampled directly from the LCTFS, household-associated food was sampled from supermarkets, and thus, not directly related to their food intake. A higher similarity between ESBL-Ec from human carriage and food, may have been observed if the food was sampled directly from the households, and on more time-points. As a consequence, this study could not reliably determine the exact relation between contamination of specific food items and ESBL carriage in the participating individuals. Lastly, the generalisability of the results to community dwellers is likely low, as we included populations influenced by healthcare. In these settings the prevalence of ESBL-Ec, and ESBL-Kp particularly, is higher than in the general community. This inherently resulted in a higher likelihood of human-to-human clonal transmission and HGT, likely leading to overestimation of the similarity of ARGs of ESBL-Ec and ESBL-Kp for humans in general.

Future research should elucidate to which extent, and under which circumstances, members of the *Enterobacterales* family share ARGs through HGT, and how this contributes to endemicity of amongst others ESBL-Ec, and ESBL-Kp in the European setting. Future public health interventions to reduce ESBL-Ec and ESBL-Kp infections should focus on prevention human-to-human clonal and HGT transmission (i) through adequate infection prevention in healthcare centers like LTCFs, (ii) hygienic advice for households after hospital discharge of their household member, (iii) promotion of gut health through amongst others reducing unecessary antimicrobial use.

Concluding, ESBL-Ec and ESBL-Kp regularly co-occurred in human populations influenced by healthcare in five European catchment areas. Considerable overlap in ARG content was observed between these species.

Acknowledgements: We would like to thank the staff and residents from the LTCFs of Bellevaux in Besançon – France; Ferrusola and El Recreo in Sevilla – Spain; Luise-Wetzel-Stift and Samariterstift in Tübingen, Germany; and Altenahove in Almkerk – The Netherlands for their involvement in the study. We would like to thank Judith Vlooswijk, and Heike Schmitt from Utrecht – the Netherlands; Marion Broussier from Besançon – France; John Poté and Siva Lingam from Geneva, Switzerland; Michael Eib from Tübingen – Germany for their help on microbiological analyses of the samples, Elisabeth Stoll and Steffen Ganß from Tübingen – Germany for collecting the samples. We thank the Institut Pasteur teams for the curation and maintenance of BIGSdb-Pasteur databases at http://bigsdb.pasteur.fr/. We thank the Enterobase team for the curation and maintenance of Enterobase *Escherichia coli* MLST database at https://enterobase.warwick.ac.uk.

REFERENCES

1. World Health Organization 2019. Thirteenth general programme of work. https://apps.who.int/iris/bitstream/handle/10665/324775/WHO-PRP-18.1-eng.pdf

2. Cassini, A, Högberg LD, Plachouras D *et al*. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* 2019; **1**: 56–66.

3. Vihta KD, 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.

4. de Been M, Lanza VF, de Toro M *et al*. Dissemination of cephalosporin resistance genes between *Escherichia coli* strains from farm animals and humans by specific plasmid lineages. *PLoS Genet* 2014; **10**: e1004776.

5. Day MJ, Hopkins KL, Wareham DW *et al*. Extended-spectrum ß-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**: 1325-1335.

6. 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**: 339–347.

7. Ludden C, Raven KE, Jamrozy D *et al*. One health genomic surveillance of *Escherichia coli* demonstrates distinct lineages and mobile genetic elements in isolates from humans versus livestock. *MBio* 2019; **1**: e02693-18.

8. Ludden C, Danesh Moradigaravand D, Jamrozy D *et al*. A one health study of the genetic relatedness of *Klebsiella pneumoniae* and their mobile elements in the east of England. *Clin Infect Dis* 2020; **2**: 219–226.

9. Martak D, Guther J, Verschuuren TD *et al*. Populations of extended-spectrum ß-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* are different in human-polluted environment and food items: A multicentre European study. *Clin Microbiol Infect* 2021; S1198-743X(21)00414-6.

10. Mughini-Gras L, Dorado-García A, van Duijkeren E *et al*. Attributable sources of community-acquired carriage of *Escherichia coli* containing ß-lactam antibiotic resistance genes: a population-based modelling study. *Lancet Planet Health* 2019; **3**: e357–e369.

11 Acman M, Dorp van L, Santini JM, Francois Balloux F. Large-scale network analysis captures biological features of bacterial plasmids. *Nat Commun* 2020; **1**: 2452.

12. van den Bunt G, van Pelt W, Hidalgo L *et al*. Prevalence, risk factors and genetic characterisation of extended-spectrum ß-lactamase and carbapenemase-producing *Enterobacteriaceae* (ESBL-E and CPE): a community-based cross-sectional study, the Netherlands, 2014 to 2016. *Euro Surveill* 2019; **24**: 29-39.

13. Hendrickx APA, Landman F, de Haan A *et al*. *bla*OXA-48-like genome architecture among carbapenemase-producing *Escherichia coli* and *Klebsiella pneumoniae* in the Netherlands. *bioRxiv* 2020.12.18.423568, doi: https://doi.org/10.1101/2020.12.18.423568

14. León-Sampedro R, DelaFuente J, Díaz-Agero C *et al*. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. *Nat Microbiol* 2021; https://doi.org/10.1038/s41564-021-00879-y

15. Matlock W, Chau KK, AbuOun M *et al*. Genomic network analysis of environmental and livestock F-type plasmid populations. *ISME J* 2021; https://doi.org/10.1038/s41396-021-00926-w

16. Hidalgo L, de Been M, Rogers MRC *et al*. Sequence-based epidemiology of an OXA-48 plasmid during a hospital outbreak. *Antimicrob Agents Chemother* 2019; **12**: e01204-19.

17. Göpel, in preparation

18. Riccio ME, Verschuuren T, Conzelmann N *et al*. Household acquisition and transmission of extended-spectrum ß-lactamase (ESBL)

-producing *Enterobacteriaceae* after hospital discharge of ESBL-positive index patients. *CMI* 2021; **20**: 30784-9.

19. Kluytmans-van den Bergh MFQ, Rossen JWA, Bruijning-Verhagen PCJ *et al*. Whole-genome multilocus sequence typing of extended-spectrum-ß-lactamase-producing *Enterobacteriaceae*. *J Clin Microbiol* 2016; **12**: 2919-2927.

20. Miro E, Rossen JWA, Chlebowicz MA *et al*. Core/whole genome multilocus sequence typing and core genome SNP-based typing of OXA-48-producing *Klebsiella pneumoniae* clinical isolates from Spain. Front Microbiol 2020; 10:2961.

21. Zhou Z, Alikhan NF, Mohamed K *et al*. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. *Genome Res* 2020; **30**: 138-152.

22. Larsen MV, Cosentino S, Rasmussen S *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012; **50**: 1355–1361.

23. Zankari E, Hasman H, Cosentino S *et al*. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012; **67**: 2640–2644.

24. Bortolaia V, Kaas RS, Ruppe E *et al*. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chem* 2020; **12**: 3491–3500.

25. Naas T, Oueslati S, Bonnin RA *et al*. ß-lactamase database (BLDB)—structure and function. *J Enzyme Inhib Med Chem* 2017; **32**: 917–19.

26. Zankari E, Allesøe R, Joensen KG *et al*. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother* 2017; **10**: 2764-8.

27. Roer L, Tchesnokova V, Allesoe R *et al*. Development of a web tool for *Escherichia coli* subtyping based on *FimH* alleles. *J Clin Microbiol* 2017; **8**: 2538-43.

28. 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.

29. Verschuuren TD, van Hout D, Arredondo-Alonso S *et al*. 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). *J Antimicrob Chem* 2021; **4**: 901–908.

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

31. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect* 2018; **24**: 350–4.

32. Lees JA, Harris SR, Tonkin-Hill G *et al*. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 2019; **29**: 304-316.

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

34. Kluytmans JAJW, Overdevest ITMA, Willemsen I *et al*. Extended-spectrum ß-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.

35. Smet A, Martel A, Persoons D *et al*. Diversity of extended-spectrum b-lactamases and class C b-Lactamases among cloacal *Escherichia coli* isolates in Belgian broiler farms. *Antimicrob Agents Chemother* 2008; **52**: 1238–1243.

36. Martin J, Phan HTT, Findlay J *et al*. Covert dissemination of carbapenemase-producing *Klebsiella pneumoniae* (KPC) in a successfully controlled outbreak: long- and short-read whole-genome sequencing demonstrate multiple genetic modes of transmission. *J Antimicrob Chemother* 2017; **11**: 3025-34.

37. Chaves J, Ladona MG, Segura C *et al*. SHV-1 ß-lactamase is mainly a chromosomally encoded species-specific enzyme in *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 2001; **10**: 2856-61.

38. Sirot J, C Chanal C, A Petit A *et al*. *Klebsiella pneumoniae* and other *Enterobacteriaceae* producing novel plasmid-mediated ß-lactamases markedly active against third-generation cephalosporins: epidemiologic studies. *Rev Infect Dis* 1988; **4**: 850-9

39. Nordmann P, Cuzon G, Naas T. The real threat of *Klebsiella pneumoniae* carbapenemase-producing bacteria. *Lancet Infect Dis* 2009; **4**: 228-36.

40. Nordmann P, Poirel L, Walsh TR, Livermore DM. The emerging NDM carbapenemases. *Trends Microbiol* 2011; **12**: 588-95.

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

42. Pitout JDD, Campbell L, Church DL *et al*. Molecular characteristics of travel-related extended-spectrum-beta-lactamase-producing *Escherichia coli* isolates from the Calgary Health Region. *Antimicrob Agents Chemother* 2009; **6**: 2539-43.

43. Holt KE, Wertheim H, Zadoks RN *et al*. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci* 2015; **27**: e3574-81.

44. Wyres KL, Lam MMC, Holt KE. Population genomics of *Klebsiella pneumoniae*. *Nat Rev Microbiol* 2020; **18**: 344–59

6

SUPPLEMENTARY MATERIAL

Table S1. Strain file, including molecular and metadata. Separate csv file available upon request. Sequences of unknown STs were sent to Enterobase (ESBL-Ec) and Institute Pasteur (ESBL-Kp) for ST assignment (table S1).

Table S2. Description of participants carrying multiple extended-spectrum ß-lactamase (ESBL)-producing *Escherichia coli* ST131 strains over time. Definition unique clone: Isolates within the same participant with a cgMLST distance greater than 0.0105.[19]

Isolate	City	Participant	Clade	ESBL gene	cgMLST dist[a]
1	Seville	LTCF2 resident	H30R (clade C2)	$bla_{\text{CTX-M-15}}$	-
2	Seville	LTCF2 resident	H30R (clade C1)	$bla_{\text{CTX-M-27}}$	0.0173
1	Seville	LTCF2 resident	H30R (clade C1)	$bla_{\text{CTX-M-27}}$	-
2	Seville	LTCF2 resident	H30R (clade C2)	$bla_{\text{CTX-M-15}}$	0.0205
1	Seville	LTCF2 resident	H41 (clade A)	$bla_{\text{CTX-M-14}}$	-
2	Seville	LTCF2 resident	H30R (clade C1)	$bla_{\text{CTX-M-14}}$	0.2856
1	Seville	LTCF2 resident	H41 (clade A)	$bla_{\text{CTX-M-14}}$	-
2	Seville	LTCF2 resident	H30R (clade C1)	$bla_{\text{CTX-M-27}}$	0.2833 (versus 1)
3	Seville	LTCF2 resident	H30R (clade C2)	$bla_{\text{CTX-M-15}}$	0.2894 (versus 1) 0.0209 (versus 2)
1	Seville	LTCF1 resident	H30R (clade C1)	$bla_{\text{CTX-M-27}}$	-
2	Seville	LTCF1 resident	H41 (clade A)	$bla_{\text{CTX-M-15}}$	0.2729
1	Seville	LTCF1 resident	H30R (clade C1)	$bla_{\text{CTX-M-27}}$	-
2	Seville	LTCF1 resident	H41 (clade A)	$bla_{\text{CTX-M-15}}$	0.2737
1	Tübingen	Index household	H30R (clade C2)	$bla_{\text{CTX-M-15}}$ & $bla_{\text{CTX-M-27}}$	-
2	Tübingen	Index household	H54 (clade B)	$bla_{\text{CTX-M-27}}$	0.0665 (versus 1)
3	Tübingen	Index household	H54 (clade B)	$bla_{\text{CTX-M-14}}$	0.0667 (versus 1) 0.0226 (versus 2)
1	Tübingen	LTCF1 resident	H30R (clade C1)	$bla_{\text{CTX-M-14}}$	-
2	Tübingen	LTCF1 resident	H30R (clade C1)	$bla_{\text{CTX-M-14}}$	0.0228 (versus 1)
3	Tübingen	LTCF1 resident	H30R (clade C1)	$bla_{\text{CTX-M-14}}$	0.0220 (versus 1) 0.0136 (versus 1)

[a] Pairwise cgMLST distance calculate with Enterobase scheme available in Seqsphere (v5.0)

Table S3. Isolates with >1 ESBL gene (*n* = 26)

Species[a]	City	Reservoir	ST	ESBL genes
E. coli	Geneva	Human	394	$bla_{CTX\text{-}M\text{-}15}$; $bla_{CTX\text{-}M\text{-}27}$
E. coli	Seville	Human-polluted (River)	2008	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}12}$
E. coli	Seville	Food (Chicken)	354	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}12}$
E. coli	Seville	Human-polluted (Surface)	131	$bla_{CTX\text{-}M\text{-}27}$; $bla_{SHV\text{-}12}$
E. coli	Tübingen	Human	131	$bla_{CTX\text{-}M\text{-}15}$; $bla_{CTX\text{-}M\text{-}27}$
E. coli	Tübingen	Human	131	$bla_{CTX\text{-}M\text{-}15}$; $bla_{CTX\text{-}M\text{-}27}$
E. coli	Tübingen	Human	3268	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}12}$
E. coli	Tübingen	Human	394	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}12}$
E. coli	Tübingen	Human	394	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}12}$
E. coli	Tübingen	Human	162	$bla_{TEM\text{-}52}$; $bla_{TEM\text{-}52}$
K. pneumoniae	Geneva	Human-polluted (WWTP-inflow)	281	$bla_{CTX\text{-}M\text{-}15}$; $bla_{TEM\text{-}169}$
K. pneumoniae	Seville	Human	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human-polluted (Surface)	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}12}$
K. pneumoniae	Seville	Human	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human-polluted (LTCF-outflow)	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human-polluted (River)	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Seville	Human	307	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Tübingen	Human	14	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
*K. variicola**	Utrecht	Human-polluted (WWTP-inflow)	14	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Utrecht	Human-polluted (WWTP-inflow)	14	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$
K. pneumoniae	Utrecht	Human-polluted (River)	14	$bla_{CTX\text{-}M\text{-}15}$; $bla_{SHV\text{-}106}$

Abbreviations, ST: sequence type.

A	EC 1	EC 2	EC 3	KP 1	KP 2	KP 3
EC 1	-	0.94 .9-1.	0.94 .8-1.	0.72 .6-.8	0.48 .3-.6	0.54 .3-.8
EC 2		-	0.94 .8-.1	0.72 .6-.8	0.50 .4-.6	0.55 .3-.8
EC 3			-	0.70 .6-.8	0.47 .3-.6	0.54 .3-.8
KP 1				-	0.72 .6.9	0.65 .4-.9
KP 2					-	0.64 .4-.8

B	EC 1	EC 2	EC 3	KP 1	KP 2	KP 3
EC 1	-	0.83 .7-.9	0.69 .6-.8	0.34 .2-.4	0.28 .1-.4	0.35 .1-.6
EC 2		-	0.74 .6-.8	0.41 .3-.5	0.34 .2-.5	0.45 .2-.7
EC 3			-	0.33 .2-.5	0.27 .1-.4	0.40 .1-.7
KP 1				-	0.81 .7-.9	0.70 .5-.9
KP 2					-	0.64 .4-.9

C	EC 1	EC 2	EC 3	KP 1	KP 2	KP 3
EC 1	-	0.91 .8-1.	0.75 .6-.8	0.77 .7-.9	0.78 .7-.9	0.73 .5-.9
EC 2		-	0.72 .6-.8	0.73 .6-.8	0.75 .6-.9	0.69 .5-.9
EC 3			-	0.73 .6-.8	0.74 .6-.9	0.67 .4-.9
KP 1				-	0.86 .7-.1	0.74 .5-.9
KP 2					-	0.68 .5-.9

Legend

Species	
EC	ESBL-producing *E. coli*
KP	ESBL-producing *K. pneumoniae*
Reservoir	
1	human
2	human-polluted environment
3	food
ARG-class	
A	fosfomycin
B	trimethoprim
C	sulfonamide

PSI
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure S1. Czekanowski's proportional similarity index (PSI) of acquired resistance genes of **A)** fosfomycin, **B)** trimethoprim, and **C)** sulfonamide. The PSI is calculated with the following formula: $1\text{-}0.5\sum k|p(\text{reservoir}[n])\ k\text{-}q(\text{reservoir}[nx])k|$, p: the relative frequency of gene type *k* in reservoir n, q: the relative frequency of the same gene type in reservoir nx. The denominator of the relative frequency was the total number genes of the corresponding ARG class. The PSI is a proportion, with 0 interpreted as no overlap, and 1 as perfect overlap in ARG type distributions between two reservoirs. 95% confidence intervals were calculated with 5,000 bootstrap iterations.

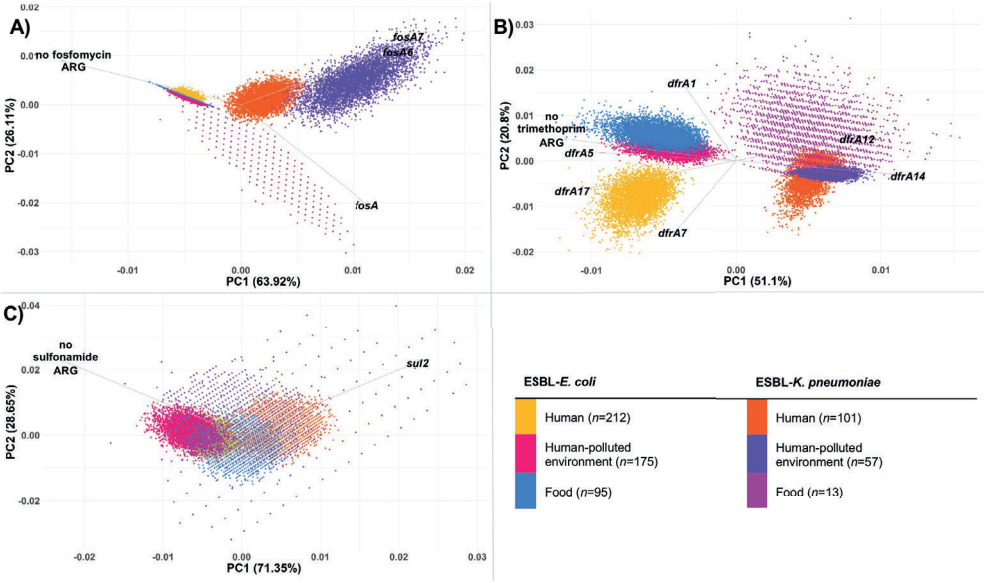

Figure S2. Principal component analysis (PCA) of antimicrobial resistance genes (ARGs). **A) fosfomycin**, PC1 (64% variance): positively associated with *fosA, fosA6, fosA7*, and extended-spectrum ß-lactamase (ESBL)-producing *Klebsiella pneumoniae* (ESBL-Kp), and negatively associated with absence of acquired resistance and ESBL-producing *Escherichia coli* (ESBL-Ec), PC2 (26% variance): positively associated with human(-associated) isolates and negatively associated with ESBL-Kp from food. **B) trimethoprim**, PC1 (51% variance): positively associated with *dfrA12, dfrA14*, and ESBL-Kp, and negatively associated with *dfrA5, dfrA17* and ESBL-Ec, PC2 (21% variance): positively associated with *dfrA1* and food isolates, and negatively associated with *dfrA7* and human isolates. **C) sulfonamide**, PC1 (71% variance): was positively associated with *sul2*, and negatively with the absence of a sulfonamide gene, however all reservoirs overlapped. Gene types with a proportion <0.01 of isolates were excluded from PCA analysis: fosfomycin ARGs: *fosA5*, trimethoprim ARGs: *dfrA27, dfrA8, dfrA15*.

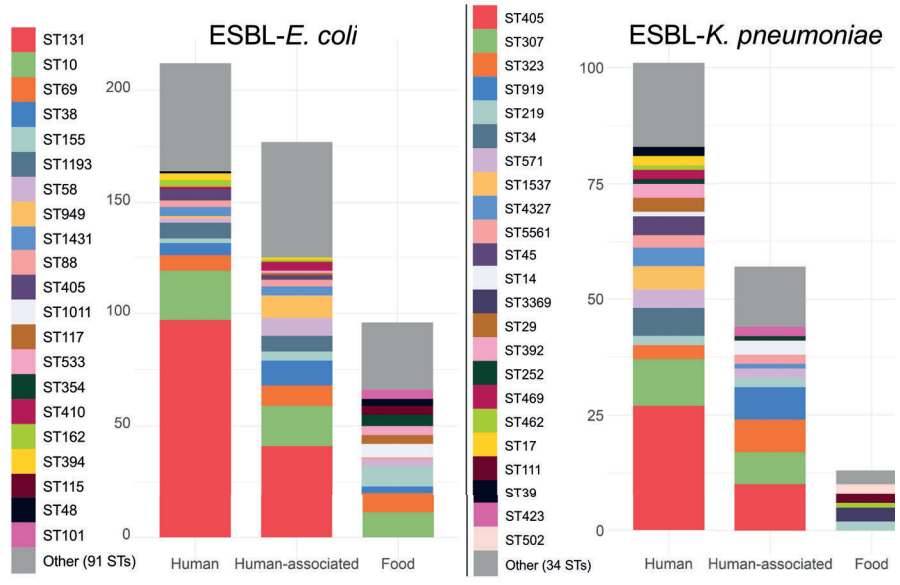

Figure S3. Observed ST counts per reservoir of extended-spectrum ß-lactamase (ESBL)-producing *Escherichia coli* (*n*=482) and ESBL-producing *Klebsiella pneumoniae* (*n*=171, including seven retrospectively identified *Klebsiella variicola* isolates).

Figure S4. Visualisation of clonally related isolate pairs of extended-spectrum ß-lactamase-producing *Escherichia coli* (ESBL-Ec) and extended-spectrum ß-lactamase-producing *Klebsiella pneumoniae* (ESBL-Kp).
A) ESBL-Ec from the catchment area of Geneva,
B) ESBL-Ec from the catchment area of Tübingen,
C) Sensitivity analysis with increased cgMLST distance results in connections between most reservoirs and cities for ESBL-Ec,
D) Sensitivity analysis with increased cgMLST distance results in additional connections between most some reservoirs and cities for ESBL-Kp.
Nodes represent isolates and are grouped based on epidemiological setting, lines represe)nt genetically similar isolate pairs.

٧

General discussion

Tess D Verschuuren

In 2019, the World Health Organisation (WHO) declared antimicrobial resistance (AMR) one of the top 10 public health threats facing humanity, due to a current yearly estimated 700 000 deaths, and up to 3.4 trillion US dollars loss in gross domestic product by 2030.[1] Proposed core actions by the WHO were, amongst others: (i) increase understanding and awareness of AMR at all levels of society, (ii) improve global surveillance, (iii) improve infection prevention and control, and (iv) optimise antimicrobial use.[1] In Europe specifically, ESBL-producing *Escherichia coli* is responsible for the majority of AMR bacterial infections.[2]

The studies in this thesis provided information on how we can improve resilience against ESBL-producing *E. coli* with practical applications of whole genome sequencing (WGS), addressing several of the before mentioned core actions proposed by the WHO:

- Increase scientific knowledge on which sources contribute to human intestinal colonisation and extra-intestinal infection with ESBL-producing *E. coli* (i)

- Improve detection and tracking of (new) ESBL variants in the community, either clones or ESBL genes, with genetic surveillance (ii)

- Revise antimicrobial susceptibility testing (AST) of *E. coli* infections for clinical diagnostic purposes (iv)

HUMAN INTESTINAL COLONISTATION AND EXTRA-INTESTINAL INFECTION WITH ESBL-PRODUCING *E. COLI*

In **chapter 2**, we showed that ESBL-positive and ESBL-negative bloodstream infections (BSIs) mainly consist of two distinct subtypes of *E. coli* lineages, by comparing sequence types (STs), acquired resistance gene (ARG) counts, virulence gene (VG) counts, and serotypes of a random sample of non-ESBL-producing *E. coli* with ESBL-producing *E. coli* derived from BSIs. The findings in this chapter suggested that the likelihood of an *E. coli* to carry antimicrobial resistance is importantly determined by its genetic 'backbone'. This notion is supported by a large body of evidence demonstrating that certain STs, like ST73, rarely carry resistance, while other STs, for example ST131 are strongly associated with ESBL-production.[3-6] Indeed, the phenomenon of AMR-associated STs has also been described for other resistance

mechanisms (e.g. carbapenemase production) and for other species (e.g. *Klebsiella pneumoniae*).[7-12] Furthermore, a handful of these AMR-associated STs are observed to be geographically widespread, subsequently translating in a disproportionately large contribution to the global AMR burden.[13]

The phenomenon of AMR-associated STs raises the question if, extra-intestinal infections with ESBL-producing *E. coli* will slowly replace their susceptible counterparts (assuming that antibiotic selection pressure stays stable over time). Alternatively, infections with AMR-associated STs may occur on top of susceptible *E. coli* infections, resulting in additional *E. coli* infections. Several reports have been published where infections with antibiotic-resistant *E. coli* seem to increase faster than those with susceptible *E. coli*.[14-18] However, longitudinal observational studies are not able to answer this question due to the large number of known and unknown time-dependent confounding factors. A mathematical modelling study by Godijk *et al* showed that only an increased infection-propensity of ß-lactam-resistant *E. coli* versus susceptible *E. coli*, would lead to an immediate, and one-time, addition of infections. In contrast, equal infection-propensity, in combination with increased transmissibility and/or intestinal colonisation duration of ß-lactam-resistant *E. coli*, would lead to a slow continuous replacement of susceptible *E. coli* infections (with no overall increase *E. coli* infections).[19] In **chapter 2** we observed no clinically relevant difference in VG counts between (non)-ESBL-producing *E. coli*, which supports the theory of replacement. However, in **chapter 3** we did observe a higher occurrence of ST131 in ESBL-producing *E. coli* from urinary tract infection (UTI) and BSI (39%) compared to ESBL-producing *E. coli* from community carriage (23%). Day *et al* also observed an increased proportion of ST131 in BSIs (64%) compared to diagnostic hospital faecal isolates (36%) (group with higher risk for infection compared to community carriage in **chapter 3**) in the U.K..[20] These observations support the hypothesis that at least ST131 could be associated with an increased likelihood of infection, which would (in part) support the theory of addition. It is, however, important to mention that neither observations in **chapter 3** or by Day *et al* were corrected for confounding factors, such as age or co-morbidities. Two other studies show that ST131 could be associated with an addition of *E. coli* occurrence: a 10-year longitudinal study of STs in BSI from the U.K., Kallonen *et al*, observed an immediate increase in occurrence of ST131 around 2002, after which the proportion of ST131 continued to be stable in the population.[4] A similar observation was made by Gladstone *et al*, where an exponential increase of ST131 occurred in Norway in the 'mid-2000s'.[21]

As stated in the introduction of this thesis, it is possible that certain *E. coli* subpopulations, like ST131, may be better adapted to the human gut (including community dwellers), due to a genetic repertoire that provides a fitness advantage in this niche. An increase in infections would then occur as an accidental by-product of this adaptation.[22-25] A limitation from the study of Godijk *et al* was that the gut abundance of susceptible and resistant *E. coli* was assumed to be equal,[19] while it is possible that this could differ, and could consequently be associated with infection-propensity, as will be discussed in the following paragraphs of this discussion section. However, all-in-all, it is likely that the observed increase in *E. coli* infections can be mainly explained by replacement and to a lesser extent due to the addition of ESBL-producing *E. coli*, however, more research is needed to understand the exact dynamics of these processes.

To improve infection prevention against ESBL-producing *E. coli*, it is important to know which factors play a role in colonisation and subsequent infection. In **chapter 3** we showed that the main reservoir for ESBL-producing *E. coli* infections is endogenous intestinal colonisation. Nine of the ten most frequent STs in extra-intestinal infections were also the most frequent in community intestinal colonisation, we furthermore did not observe distinct phylogenetic clustering of these groups. Prevention of infection with ESBL-producing *E. coli* could thus potentially be accomplished by prevention of intestinal colonisation.

In recent years it has become clear that the gut microbiome plays an important role in the health of the host, by amongst others preventing colonisation of unwanted microorganisms (a phenomenon that is known as colonisation resistance).[26] A growing body of evidence shows that antibiotic use results in dysbiosis and reduced diversity of the gut microbiome, which may take months to years to return to the state of pre-treatment.[13,27-29] A very interesting hypothesis is that as the gut microbiota are partly inherited by the mother at birth, this effect of antibiotic-induced dysbiosis could be cumulative over generations (figure 1).[28] Other factors like diet changes and urbanisation may also play a role in the loss of microbiome diversity.[29] It is thought that the transmission and infection of resistant clones, like ST131, could be facilitated by this process.[29] Under normal conditions, opportunistic pathogens represent <1% of the total intestinal microbiota. However, antibiotic-induced dysbiosis could lead to an increase in abundance of opportunistic (AMR-) pathogens and subsequent risk of translocation from the gut to extra-intestinal tissues.[29] Indeed in three studies, the relative abundance of *E. coli* was (significantly) higher (>1%) in participants that developed a bacteriuria/UTI than in comparable patients that didn't.[30-32]

One of these three studies, and a recent publication by Berkell *et al* furthermore demonstrated that ß-lactam use was associated with dysbiosis and increased relative abundance of *Enterococcus spp*.[27] Ducarmon *et al* furthermore demonstrated that long-term care facility (LTCF) residents who were colonised with AMR-bacteria compared to those who were not, had a different microbiome composition, with lower abundance of *Dorea*, *Atopobiaceae* and *Lachnospiraceae*.[33] As infections with *E. coli* occur mostly in the very young or very old,[34] it is furthermore possible that prolonged persistence of potential pathogens due to an immature microbiome could add to the risk of infection.[13] This hypothesis is supported by the observation that microbiome composition differs per age group.[35] Lastly, it is important to note that even in the non-westernised, rural communities of e.g. Nepal, intestinal colonisation with ESBL-producing *E. coli* (with $bla_{CTX\text{-}M\text{-}15}$) has become endemic.[36] This observation indicates that it is likely that other factors than loss in microbiome diversity (due to antimicrobial exposure, diet changes, or urbanisation), play a role in the spread of this bacterium.

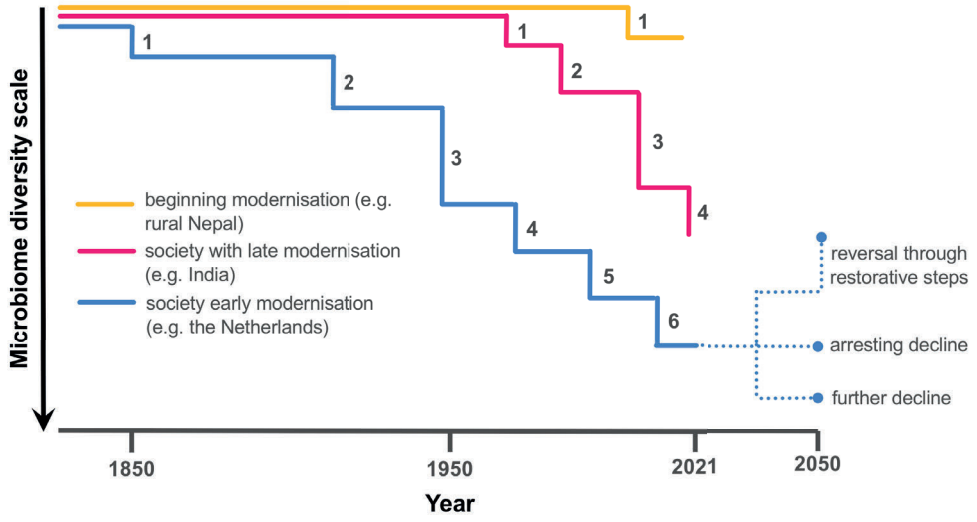

Figure 1. Cumulative microbiome diversity loss.
Derived from: Blaser MJ. Antibiotic use and its consequences for the normal microbiome. *Science* 2016; **6285**: 544-5. The early loss in microbiome diversity in the Netherlands occurred simultaneously with the introduction of sanitation (including filtered drinking water, and construction of sewage systems), followed by early antimicrobial use. Following declines could be explained by diet changes, urbanisation, and continuing antimicrobial use, and other factors. The shown scale reflects the aggregate of microbiome diversity loss, and is arbitrary. Diversity loss occurred later in societies with late modernisation, but generation times are shorter.

Based on the previous discussion of the findings in this thesis and corresponding literature, I suggest the following directions for future research, to further elucidate the role of intestinal colonisation in infection of ESBL-producing *E. coli*:

- More knowledge is needed on how ESBL-producing *E. coli* spread in areas of low antibiotic selective pressure and/or in non-modernised communities

- More knowledge is needed to understand colonisation resistance in different age groups, and under which microbial circumstances ESBL-producing *E. coli* (and specific subtypes) are able to become (i) a commensal, with (ii) increased relative abundance. Potential methodologies could be prospective cohorts or matched case-control studies, in which the microbiome composition in patients from certain care settings (e.g. NICU, ICU, LTCF[33]) with and without colonisation with an ESBL-producing *E. coli* would be compared. Participating groups should be large enough so that the outcome of relative abundance could be categorised in to groups (e.g. <0.5% 0.5-1%, 1-2%, 2-3%, >3%), and correction for potential confounding factors can be performed.

- More knowledge is needed at which level of abundance of (ESBL-producing) *E. coli* is (and specific subtypes are) most likely to cause extra intestinal infection. To date, three studies have been performed that showed a relation between bacterial pathogen relative abundance and UTI.[30-32] Suggested methodologies for future research are: prospective cohort or a matched case-control study, in which the relative abundance of (ESBL-producing) *E. coli* is quantified with e.g. qPCR targeting the 16S rRNA gene, and compared between patients with and without an extra-intestinal infection with this bacterium, with correction for (epidemiologic) confounding factors.

SOURCES CONTRIBUTING TO HUMAN INTESTINAL COLONISATION OF ESBL-PRODUCING *E. COLI*

It is important to understand what the most important sources for human intestinal colonisation of ESBL-producing *E. coli* are, as endogenous intestinal colonisation is the main reservoir for ESBL-producing *E. coli* infections. In **chapter 5 & 6** we show that, in the European setting, human carriage of ESBL-producing *E. coli* mainly is explained through human-to-human transmission. This observation is confirmed by a growing amount of literature, while transmission from non-human reservoirs (e.g. food, cattle, pets, environmental exposure) likely only occur as spill-over events.[20,37-45] The hypothesis is furthermore supported by an anecdotal drop in Dutch ESBL positive clinical samples since the start of the COVID-19 pandemic.[46] The prevalence of ESBL positivity went from 5-10% to 2.5% in some hospitals. Assuming that all exposure to non-human reservoirs stayed stable (with exception of travel abroad), this indicates that lockdown measures reduced human-to-human transmission below levels needed to maintain ESBL prevalence in the Dutch community.[41]

ROLE OF HORIZONTAL GENE TRANSFER IN INTESTINAL COLONISATION OF ESBL-PRODUCING *E. COLI*

From an evolutionary standpoint, it is known that accessory genetic content can potentially (i) expand the niche of a bacterium, (ii) give it a competitive edge against other organisms, (iii) or alter its host-interaction.[47] Considering these possible advantages it is logical to think that horizontal gene transfer (HGT) occurs very often.[48] Although HGT has quite extensively been studied in laboratory settings, due to technical difficulties, very little is known about the occurrence of this phenomenon in real-life settings.[47] Smillie *et al* demonstrated some evidence of higher rates of HGT between bacteria in the human microbiome that in between bacteria from non-human ecosystems.[49] This observation would fit with the current knowledge that human-to-human transmission is the most important source for ESBL-producing *E. coli* intestinal colonisation.

In **chapter 6** we observed that ESBL-producing *E. coli*, from healthcare-associated intestinal colonisation, was more similar in ESBL gene content to ESBL-producing *K. pneumoniae*, from healthcare-associated carriage, than to ESBL-producing *E. coli* from food. This observation, in addition to a larger proportion of ARGs in ESBL-producing *K. pneumoniae* from healthcare-associated carriage, particularly

ARGs conferring resistance to several clinically relevant antimicrobials, led us to hypothesise that ESBL-producing *K. pneumoniae* from intestinal colonisation could be an important reservoir for intestinal colonisation of ESBL-producing *E. coli* in populations that are in frequent contact with healthcare. Indeed, the ESBL gene $bla_{CTX\text{-}M\text{-}15}$ is the most occurring ESBL gene found in human carriage and infection in both *E. coli* and *K. pneumoniae*.[20,39,40,50] Furthermore, studies assessing plasmids showed that genetic plasmid content is shared between both species as well as numerous other species, mostly from the *Enterobacterales* family.[51,52] Lastly, studies assessing hospital plasmid outbreaks provided evidence for inter-species plasmid transfer, for example Hidalgo *et al* demonstrated that the same plasmid occurred in 14 different species during a period of 1.5 years.[53-55] Furthermore, various lines of evidence suggest that *K. pneumoniae* may be more likely to donate plasmids to *E. coli* than vice versa. First, *K. pneumoniae* was the species in which $bla_{SHV\text{-}1}$, $bla_{CTXM\text{-}1}$, $bla_{KPC\text{-}1}$, and $bla_{NDM\text{-}1}$ were first described, followed by identification in other Gram-negative bacteria.[48,56-58] Secondly, AMR rates of *K. pneumoniae* are higher than those of *E. coli*.[59] Lastly, assessment of two hospital $bla_{OXA\text{-}48}$ outbreaks showed that the plasmid mainly originated from *K. pneumoniae*, while *E. coli* only contributed slightly.[53,60] These studies indicate that *K. pneumoniae* is known to easily acquire, sustain and donate resistance.

However, to improve our understanding on ESBL gene dissemination and ESBL gene-carrying plasmids, there is need for studies that provide a contextual understanding of the behaviour and interactions, including HGT, between ESBL-producing bacteria in real-life settings.

Suggestions for future research of horizontal gene transfer of ESBL genes:

Longitudinal studies are needed (e.g. months), that perform intensive faecal sampling (e.g. daily/weekly), in a confined and consistent epidemiological setting (e.g. households after hospital discharge of an ESBL-positive patient, a LTCF with ≥1 ESBL-positive resident, farms, to assess HGT of ESBL genes within and between participants, using one of the following microbiologic/sequencing methods:

- A species-directed-culture-based-method: select single colonies per morphology of ESBL-producing *E. coli* and ESBL-producing *K. pneumoniae*, and subsequently perform WGS of isolated bacteria with short-read and long-read technologies to reconstruct, and assess transmission of ESBL gene carrying plasmids. This methodology is used in our ongoing work within the MODERN-studies, and is expected to give information on how

often HGT of ESBL genes occurs versus clonal transmission. Other expected results are an estimation of the likelihood of *E. coli* to donate an ESBL gene to *K. pneumoniae* and vice versa. Limitations of this study are the lack of knowledge on to the total *E. coli* and *K. pneumoniae* populations, or other species that might interact in HGT of ESBL genes.

• **A culture-based method that includes all Gram-negative bacteria (e.g. with MacConkey agar without added antibiotics).** The total content of the plate could then be used for WGS. This methodology is also referred to as 'culturomics' with the overarching goal of isolating diverse organisms from complex microbial communities.[47] This methodology would give insight into the total population of *E. coli*, *K. pneumoniae*, and other species that might harbour ESBL genes. Expected results are contextual knowledge about which genetic variants are more likely to acquire and persist in ESBL gene carriage; furthermore, it might be possible to reconstruct patterns of ESBL gene flow (e.g. from bacterium A, to bacterium B, to bacterium C). Limitations of this methodology are: (i) a risk of misidentifying HGT by a gain or loss of plasmids during culturing[47] (ii) loss of the connection between chromosomal and extra-chromosomal elements during sequencing. Advanced bioinformatics techniques are needed to solve this problem.[47] (iii) incomplete knowledge of flow of ESBL genes in the total microbiome, as not all bacteria can be cultured, which leaves the possibility of missing important steps in the flow of ESBL genes through HGT.

• **A culture-free method, i.e. metagenomic shotgun sequencing, that is performed directly on the faecal sample.** This methodology has the potential to provide the ultimate answer on (i) flow of ESBL genes, (ii) the optimal conditions under which HGT occurs, and (iii) under what circumstances ESBL genes persist within bacteria. This methodology has however major technical limitations: (i) when performed with short-read sequencing platforms, it is very difficult to determine which accessory gene content belongs to which genome, as it often contains components that are present in several genomes, (ii) a varying sequencing coverage of accessory gene content (e.g. phages, plasmids, common genes) is difficult to capture during assembly of the genomes , (iii) low abundant ESBL gene carrying species can fall below detection limits.[47] To overcome these limitations advanced bioinformatics methodologies are needed.[47]

DETECTION OF THE SPREAD OF (NEW) ESBL-PRODUCING-*E. COLI* CLONES AND ESBL GENE VARIANTS THROUGH GENETIC SURVEILLANCE

In **chapter 3** we demonstrate that routinely collected clinical isolates give a representative view of the genetic variants of ESBL-producing *E. coli* circulating in the Dutch community. Subsequently, this method is suitable to detect and track (new) genetic variants of this bacterium in at least the Dutch setting. The importance of the detection of new variants through genetic surveillance is illustrated by the COVID-19 pandemic, where new variants take over earlier variants if they possess certain advantages like increased transmissibility.[61,62] As discussed in the previous section, if a new ESBL-producing *E. coli* variant possesses traits resulting in increased virulence and/or antimicrobial resistance, this can have effects on the burden of disease of this bacterium.[19]

There are a number of bacterial surveillance systems that include ESBL-producing *E. coli*, of which I will describe the main systems on a (i) global, (ii) European, and (iii) Dutch level (table 1).

Firstly, the global antimicrobial resistance surveillance system (GLASS) was initiated in 2015 by the WHO. GLASS collects locally performed clinical phenotypic antimicrobial susceptibility data of high priority bacteria, originating from separate countries, and regional surveillance initiatives (including CAESAR (central Asia and Europe), EARS-Net (Europe), and ReLAVRA (Latin America)).[63] Furthermore, a sub-branch from GLASS, GLASS-One Health collects phenotypic (local) and genetic (local or centralised) data from animals, humans, and the environment on ESBL-producing *E. coli*, to gain knowledge on the spread of AMR in different reservoirs.[64]

In Europe, EARS-Net collects locally performed phenotypic antimicrobial susceptibilities from clinical samples.[65] Additionally, EURGen-Net performs centralised genetic characterisation of carbapenem and/or colistin-resistant *Enterobacterales* from clinical samples from Europe.[66]

In the Netherlands, a similar structure exists, where ISIS-AR collects phenotypic characteristics from hospital laboratories, and the Dutch carbapenemase-producing pathogen surveillance program performs genomic characterisation of local phenotypically confirmed clinical isolates.[67,68]

Table 1. Overview of AMR bacterial surveillance initiatives

Initiative	Region	Type of surveillance	Pathogens	Workflow
GLASS	global	phenotypic antimicrobial susceptibilities	'high priority'[a] AMR bacteria	local, phenotypic characteristion of clinical samples (+ age and gender) [a]. includes data from separate countries, and regional surveillance initiatives like CAESAR (central Asia, Europe), EARS-Net (Europe), and ReLAVRA (Latin America).
GLASS-One Health	6 LMIC's + 9 other countries	phenotypic + genetic (PCR & WGS)	ESBL-producing *E. coli*	local and/or centralised standardised surveillance phenotypic + genetic characterisation of: • animals (chicken faeces, from live bird-markets) • humans (blood + community intestinal carriage) • environment (wastewater + rivers)
EARS-Net	Europe	phenotypic antimicrobial susceptibilities	seven bacteria that often cause human infection[b]	local, phenotypic characterisation of clinical samples
EURGen-Net	Europe	genetic (WGS)	carbapenem/colistin-R *Enterobacterales* planned: carbapenem-resistant *Acinetobacter baumannii*	centralised, genetic characterisation, of locally phenotypically confirmed clinical samples
ISIS-AR	the Nether lands	phenotypic antimicrobial susceptibilities	bacteria	local, phenotypic characterisation of clinical samples
Dutch CPE surveillance program	the Nether lands	genetic (WGS)	carbapenem-R-*Enterobacterales*/ *Pseudomonas aeruginosa*	centralised, genetic characteisation, of local phenotypically confirmed clinical samples

[a] blood: *Acinetobacter spp., Escherichia coli, Klebsiella pneumoniae, Salmonella spp., Staphylococcus aureus, Streptococcus pneumoniae*. urine: *E. coli, K. pneumoniae*. genital isolates: *Neisseria gonorrhoeae*, faeces: *Salmonella spp. Shigella spp.*
[b] *E. coli, K. pneumoniae, Pseudomonas aeruginosa, Acinetobacter spp.S. pneumoniae, S. aureus, Enterococcus faecalis, Enterococcus faecium*
Abbreviations. LMIC: Low- and Middle-income Countries, R: resistant, CPE: carbapenemase-producing *Enterobacterales*.

Suggestions for future public health strategies:

Current AMR-surveillance strategies would benefit from incorporating or expanding genetic characterisation of bacteria to include ESBL-producing *E. coli*.

However, an important bottleneck can be the workflow, which often includes shipment of isolates to a centralised facility, and subsequent sequencing and analysis. This methodology ensures comparability and quality of the genomic output, however, it requires a sequencing facility with a large capacity. As sequencing becomes more available, it would be worthwhile for surveillance institutions to invest in a protocol, and practical framework, for decentralising the genomic surveillance process. Here, raw reads or assemblies could be uploaded by the participating

hospital/public health laboratory. This would be followed by a quality check (curation step), and subsequent data analysis. At least in the Netherlands, such a framework has been implemented for COVID-19 genomic surveillance, which could be an opportunity for expansion to other pathogens like ESBL-producing *E. coli* when the pandemic subsides.

If resources are available, we would thus advise de-centralised, genetic surveillance of ESBL-producing *E. coli*, using clinical isolates. Subsequently, when large shifts in ESBL variants would be detected, cross-sectional faecal sampling in the open community would be warranted. Gladstone *et al*, furthermore, suggests to perform genetic surveillance of all *E. coli*, as new genetic variants can arise irrespective of AMR.[21] However, this would require a substantial sequencing capacity, which is unlikely to be feasible at this moment. Furthermore, One-Health approaches could additionally be useful to monitor non-human sources for new AMR-variants, in potential 'high-risk' bacteria like *E. coli*.[64] Lastly, it needs to be acknowledged that local or centralised genomic surveillance, due to: i) the higher costs, ii) needed technical equipment, and iii) trained staff, is unlikely to be feasible in many low resource settings.[69] Especially in these settings, initiatives like GLASS are very valuable, as they will contribute to capacity building of phenotypic and subsequently genomic surveillance of amongst others ESBL-producing *E. coli*.

WGS BASED ANTIMICROBIAL SUSCEPTIBILITY TESTING FOR CLINICAL DIAGNOSTICS

Antimicrobial susceptibility prediction tools KOVER-AMR[70] and ResFinder 4.1[71] did not meet U.S. Food and Drug Administration (FDA) criteria for clinical diagnostic use[72], in an external validation of 518 *E. coli* infections in **chapter 4**. However, we believe that there are enough points of engagement for improvement, which makes it likely that these or comparable tools will become part of future clinical diagnostics. First I will review current clinical practice of culture-based AST, secondly discuss the potential application of WGS-AST in clinical practice, thirdly the different WGS-AST methodologies, and lastly give recommendations for future research.

At this moment, medical microbiology still relies on culture-based methods to identify and perform AST.[73] These methods are affordable, but take several days to process (24-96 hours for species identification, and 24-36 hours for AST), as they depend on bacterial growth and require a pure culture.[73,74] This means that patients with a suspected infectious disease, are treated empirically with broad-spectrum

antimicrobial therapy, to ensure coverage against the unknown pathogen. Therapy can be continued, escalated, or de-escalated when microbiology results become available. The consequence of this practice is that patients can be over-treated, resulting in unnecessary antimicrobial use, or under-treated, resulting in potential adverse patient outcomes. For example, a Dutch septic patient with an ESBL-producing *E. coli* could be undertreated, as empiric therapy in case of a suspected urogenital focus, may consist of a cephalosporin.[75]

WGS, at least in theory, provides the possibility to replace current empiric practice with AST-directed therapy.[13] However, for this, first these three requirements need to be fulfilled.

Firstly, a hospital needs to establish a high throughput workflow to be able to perform point of care sequencing and analysis.[74]

Secondly, methodologies should be developed that allow for sequencing directly on a clinical sample, especially on material that is sterile under normal conditions (e.g. blood). Especially for slow-growing pathogens like *Mycobacterium tuberculosis*, this methodology could offer improvement in AST.[76] It could furthermore, theoretically even be possible to only perform AST irrespective of the species (in normally sterile materials), which could potentially save time by skipping the step of genome assembly. Both the first and second requirement could potentially be met with sequencing technologies like for example Nanopore. This sequencing technology allows for direct sample testing, and has a reduced workload. Additionally, results could potentially be available in hours, as shown in a study addressing the Ebola virus that produced sequencing results after one hour.[77] However, long-read sequencing technologies like Nanopore, currently do have a higher error rate, which could lead to higher false susceptible, and false resistant rates.[76] Furthermore, direct sequencing of clinical samples is more complex, and is more likely to miss resistance (leading to false susceptible results), due to a low abundance of the pathogen compared to the DNA of the host.[76]

The third requirement is that WGS-AST user-friendly tools need to be developed that provide direct phenotype predictions. Furthermore, these need to perform according to FDA criteria set for clinical AST technologies.[72] FDA criteria are a major error (false resistant) rate below 3%, and a very major error (false susceptible) rate with a 95% CI ≤1.5%-≤7.5%.[72] Direct phenotype predictions (e.g. amoxicillin resistance yes/no) are important because they are likely to save time, and reduce errors in the clinical diagnostic process, when compared to genomic predictions (e.g. presence bla_{TEM-1}).[78] To our knowledge, KOVER-AMR[70] and ResFinder 4.1[71] are,

at this moment, the only tools that provide direct phenotype predictions.

Methodologies used for detection of predictors of resistance in WGS-AST can be ordered in three groups: (i) reference databases, (ii) predictions based on supervised machine learning, (iii) predictions based on pan-genome machine learning. These methodologies will be briefly discussed below.

(i) reference databases. With this method, an assembled sequence or raw reads are mapped to databases with known acquired resistance genes and/or chromosomal mutations (e.g. ResFinder[71], the comprehensive antibiotic resistance database or CARD[79], and ARG-ANNOT database[80]). Labour-intensive, genome-wide association studies (GWAS) are traditionally the source of resistance predictors in these databases.[81] Several WGS-AST proof-of-principle studies (assessing different species) that used reference databases, showed good concordance (*n* correct predictions divided by *n* total predictions) between phenotype and genotype: ≥95%, proving that a phenotype can largely be predicted based on genetic content.[82-86] The advantage of reference databases is that you can be quite sure that the detected resistance gene or chromosomal mutation predicts resistance, which translates into the low ME rates observed in **chapter 4**. However, the disadvantage is that unknown resistance predictors will be missed[87], resulting in higher VME rates observed in **chapter 4**. This disadvantage could be largely overcome, when reference databases are continuously updated with output from one of the machine learning algorithms described in (ii) and (iii). However, a phenotype can still be missed, especially when outbreaks with a new resistance variant occur, which will lead to undertreatment of the corresponding patients. Lastly, phenotypes that depend on an accumulation of mutations are difficult to predict with this method.[87]

(ii) resistance predictors based on supervised machine learning. Supervised machine learning can be used to discover and/or predict resistance through association with a phenotype.[88] A supervised machine learning algorithm is given a dataset of genomes with known antimicrobial susceptibility patterns (outcomes). With this, the algorithm constructs predictive models for the phenotype of interest (e.g. amoxicillin resistance). If a certain genetic predictor is present in the genome (e.g. a certain *k*-mer: a string of DNA with the length of *k*), then the bacterium is predicted as resistant. Several proof-of-principle studies[76,89-91], and the KOVER-AMR tool[70], have shown good results for this method. The advantages of this method are: (i) the ability to find previously unknown predictors for resistance, (ii) model flexibility to contain a combination of predictors that together predict a phenotype, (iii) the models can be used to predict minimum inhibitory concentrations.[76] All of these advantages fix issues that are encountered with reference-based databases.[7]

Supervised machine learning-based tools can be compared with diagnostic models from classical epidemiology, where one is not interested in the biological mechanism underlying the predictor, but only its ability to predict the outcome. An example of this is fever: the presence of fever can be used to predict infection, although it is not the cause of infection. This is also the main downside of this methodology. Small, non-diverse, or non-representative datasets used for training of machine learning algorithms can lead to erroneous predictions when applied to another dataset.[76] In bacteriology, the presence of population structures of bacteria is of particular concern.[81,87] When for example, the training dataset consists of 50% ESBL-positive ST131 and 50% ESBL-negative ST73 (associated with non-ESBL), it is possible that the algorithm may select parts of the genome that are associated with ST131, rather than an ESBL gene, as predictors for cefotaxime resistance. When this model is then applied in a ST131 dataset, containing both ESBL-positive and ESBL-negative isolates, all isolates might be predicted cefotaxime resistant. To prevent this, it is of the utmost importance that the data used to train the supervised machine learning algorithms represent the population the tools will be applied in.

(iii) pan-genome predictive machine learning. Recently, a new methodology has been proposed by Lees *et al*[81], based on methodologies used for human GWAS studies. This methodology uses the pan-genome, which is the total genetic content of a population, in this case of all included isolates in the training dataset. This methodology has the potential to overcome the limitations of population-structure introduced errors, while retaining the advantages of supervised machine learning. Briefly, occurring genetic variants, are efficiently expressed in unitigs (i.e. an extended sequence word), and subsequently a model is fitted using elastic net penalisation (i.e. a methodology similar to multiple regression). However, although the methodology accounts for population structure in bacterial pathogens, it remains very important to use a training dataset that is a diverse representation of the true bacterial population, in which the researcher eventually wants to implement the model.[81]

WGS-AST as a clinical diagnostic for patient-directed antibiotic therapy may seem not applicable any time soon. However, we should take into account that the first bacterial genome (*Haemophilus influenza*) was only sequenced in 1995 (which took 13 months), and the tremendous advancement of the field since then.[91] Furthermore, technology is expected to increase exponentially, as the only limitation is computational power, which is expected to double each year (Moore's law).[92] Based on the findings in **chapter 4** and discussed insights from literature, I suggest the following steps to facilitate WGS-AST development for clinical diagnostic use:

• Collective scientific, public health, and political effort should be put in making genotype-phenotype data publically available. This data is essential for the continuous discovery of resistance predictors, which can in turn be used to update free, online, tools like ResFinder. For example, steps can be taken to further centralise available sequence repositories, as at this point several repositories exist (e.g. European nucleotide archive[93], NCBI Genbank[94], PATRIC[95]). Furthermore, the upload of data could be encouraged by simplifying the upload of data (often manual and time-consuming).

• Development of direct-sample (metagenomics) sequencing for clinical applications should receive more priority. Especially in material that should normally be sterile (e.g. blood).

• External validation studies of WGS-AST tools are needed including several species and using diverse datasets.

• Proof-of-principle implementation studies, where WGS-AST is performed in parallel to current clinical diagnostic culture-based AST, need to be initiated to examine how WGS-AST could influence antibiotic therapy choices.

• When the time is right, diagnostic intervention studies could be performed to determine the impact of WGS-AST on patient care and broad-spectrum antibiotic use. Here, WGS-AST is applied in a patient population with suspected infectious disease (e.g. sepsis of suspected urogenital focus on in the emergency room). Here, the intervention would be the antibiotic therapy choice, based on WGS-AST versus empiric antibiotic therapy. The outcomes should be clinically relevant outcomes for the patient (e.g. 30/90-day mortality, days till disease clearance), and broad spectrum-antibiotic use.

CONCLUSIONS

The studies in this thesis show that it is unlikely that ESBL-producing *E. coli* will disappear from the European setting, as we observed a widespread occurrence of this bacterium in several reservoirs. However, the practical applications of WGS in science, public health, and patient care will improve our ability to understand, track, and treat ESBL-producing *E. coli*. In the grander scheme of things, it is expected that this development will change the way we look at bacteria, from a taxonomic view of pathogens, to a contextual view of the microbial world, where pathobionts are mostly beneficial, and sometimes pathogenic.

Next steps for research and/or public health that I consider necessary are studies to:

- Understand how colonisation resistance influences the ecology of ESBL-producing *E. coli* to become a (long-term) sy*MBio*nt or an opportunistic pathogen, and at which levels of relative abundance ESBL-producing *E. coli* is most likely to cause infection.

- Understand how often HGT of ESBL genes occurs and persists in real-life settings. For this, intensive sampled, longitudinal studies are needed in epidemiologically confined settings (e.g. households), using (i) culture-based-species-directed, (ii) culture-based-population-directed, and (iii) culture-free-population-directed microbiological techniques.

- Increase genomic surveillance capacity by decentralising sequencing from central laboratories to local healthcare facility laboratories, and subsequently structure sequence data sharing of healthcare facilities with central surveillance systems, which will, in turn, analyse and summarise data on genetic variants.

- Prioritise development of WGS-AST technologies by encouraging public genotype-to-phenotype datasets from various geographic locations and reservoirs, and simultaneously prioritise direct sample (metagenomic) sequencing for practical applications, starting with clinical samples that should normally be sterile, like blood.

REFERENCES

1. World Health Organization 2019. Thirteenth general programme of work. https://apps.who.int/iris/bitstream/handle/10665/324775/WHO-PRP-18.1-eng.pdf

2. Cassini, A, Högberg LD, Plachouras D *et al*. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* 2019; **1**: 56–66.

3. Hertz FB, Nielsen JB, Schønning P *et al*. Population structure of drug-susceptible,—resistant and ESBL-producing *Escherichia coli* from community-acquired urinary tract infections. *BMC Microbiol* 2016; **16**: 63

4. 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; **8**:1437-1449.

5. 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.

6. 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.

7. Manges AR, Geum HM, Guo A *et al*. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) Lineages. *Clin Microbiol Rev* 2019; **3**: e00135-18.

8. Dunn SJ, Connor C, McNally A. The evolution and transmission of multi-drug resistant *Escherichia coli* and *Klebsiella pneumoniae*: the complexity of clones and plasmids. *Curr Opin Microbiol* 2019; **51**: 51-56.

9. Mathers AJ, Peirano, Pitout JDD. The role of epidemic resistance plasmids and international high- risk clones in the spread of multidrug-resistant *Enterobacteriaceae. Clin Microb Rev* 2015; **3**: 565-591.

10. Long SW, Olsen RJ, Eagar TN *et al*. Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolates, Houston, Texas: unexpected abundance of clonal group 307. *MBio* 2017; **8**: e00489-17.

11. Peirano G, Chen L, Kreiswirth BN, Pitout JDD. Emerging antimicrobial-resistant high-risk *Klebsiella pneumoniae* clones ST307 and ST147. *Antimicrob Agents Chemother* 2020; **10**: e01148-20.

12. Wyres KL, Jane Hawkey J, Hetland MAK *et al*. Emergence and rapid global dissemination of CTX-M-15-associated *Klebsiella pneumoniae* strain ST307. *J Antimicrob Chemother* 2019; **74**: 577-81.40.

13. Andersson DI, Balaban NQ, Baquero F *et al*. Antibiotic resistance: turning evolutionary principles into clinical reality. *FEMS Microbiol Rev* 2020; **2**: 171-188.

14. Vihta KD, 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**: 1138–1149.

15. de Kraker MEA, Jarlier V, Monen JCM *et al*. The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. *Clin Microbiol Infect* 2013; **19**: 860–868.

16. Van Der Steen M, Leenstra T, Kluytmans JAJW, Van Der Bij AK. Trends in expanded-spectrum cepha- losporin-resistant *Escherichia coli* and *Klebsiella pneumoniae* among Dutch clinical isolates, from 2008 to 2012. *PLoS One* 2015; **10**: e0138088.

17. Schlackow I, Stoesser N, Walker AS *et al*. 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.

18. Ammerlaan HSM, Harbarth S, Buiting AGM *et al*. Secular trends in nosocomial bloodstream infections: antibiotic-resistant bacteria increase the total burden of infection. *Clin Infect Dis* 2013; **6** :798-805.

19. Godijk NG, Bootsma MCJ, 2, van Werkhoven HC *et al*. Modelling addition and replacement mechanisms of plasmid-based beta-lactam resistant *E. coli* infections. *medRxiv* 2021: doi: https://doi.org/10.1101/2021.03.17.21253797

20. Day MJ, Hopkins KL, Wareham DW *et al*. Extended-spectrum b-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.

21. Gladstone RA, McNally A, Pöntinen AK *et al*. Emergence and dissemination of antimicrobial resistance in *Escherichia coli* causing bloodstream infections in Norway in 2002–17: a nationwide, longitudinal, microbial population genomic study. *Lancet microb* 2021; online first: DOI:https://doi.org/10.1016/S2666-5247(21)00031-8

22. Riley LW. Distinguishing pathovars from nonpathovars: *Escherichia coli*. *Microbiol Spectrum* 2020; **8**: 1-23.

23. Beceiro A, Tomás M, Bou G. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin Microbiol Rev* 2013; **2**: 185-230.

24. 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.

25. Le Gall T, Clermont O, Gouriou S *et al.* Extraintestinal virulence Is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol* 2007; **11**: 2373-84.

26. Vollaard EJ, Clasener HA. Colonization resistance. *Antimicrob Agents Chemother* 1994; **38**: 409-14.

27. Berkell M, Mysara M, Xavier BB *et al.* Microbiota-based markers predictive of development of *Clostridioides difficile* infection. *Nat Commun* 2021; **1**: 2241.

28. Blaser MJ. Antibiotic use and its consequences for the normal microbiome. *Science* 2016; **6285**: 544-5.

29. Segata N. Gut Microbiome: westernization and the disappearance of intestinal diversity. *Curr Biol* 2015; **14**: R611-3.

30. Lee JR, Muthukumar T, Dadhania D *et al.* Gut microbial community structure and complications after kidney transplantation: a pilot study. *Transplantation* 2014; **7**: 697-705.

31. Paalanne N, Husso A, Jarmo Salo S *et al.* Intestinal microbiome as a risk factor for urinary tract infections in children. *Eur J Clin Microbiol Infect Dis* 2018; **10**:1881-1891.

32. Magruder M, Sholi AN, Gong C *et al.* Gut uropathogen abundance is a risk factor for development of bacteriuria and urinary tract infection. *Nat Commun* 2019; **1**: 5521.

33. Ducarmon QR, Terveer EM, Sam Nooij S *et al.* Microbiota-associated risk factors for asymptomatic gut colonisation with multi-drug-resistant organisms in a Dutch nursing home. *Genome Med* 2021; **13**: 54.

34. Public Health England 2019. Laboratory surveillance of *Escherichia coli* bacteraemia in England, Wales and Northern Ireland: 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/844788/hpr3719_ecoli18.pdf.

35. Chaudhari DS, Dhotre DP, Agarwal DM *et al.* Gut, oral and skin microbiome of Indian patrilineal families reveal perceptible association with age. *Sci Rep* 2020; **1**: 5685.

36. Subramanya SH, Bairy I, Metok Y *et al.* Detection and characterization of ESBL-producing *Enterobacteriaceae* from the gut of subsistence farmers, their livestock, and the surrounding environment in rural Nepal. *Sci Rep* 2021; **1**: 2091.

37. de Been M, Lanza VF, de Toro M *et al.* Dissemination of cephalosporin resistance genes between *Escherichia coli* strains from farm animals and humans by specific plasmid lineages. *PLoS Genet* 2014; **10**: e1004776.

38. 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**: 339–347.

39. Ludden C, Raven KE, Jamrozy D *et al.* One health genomic surveillance of *Escherichia coli* demonstrates distinct lineages and mobile genetic elements in isolates from humans versus livestock. *MBio* 2019; **1**: e02693-18.

40. Ludden C, Danesh Moradigaravand D, Jamrozy D *et al.* A One health study of the genetic relatedness of *Klebsiella pneumoniae* and their mobile elements in the East of England. *Clin Infect Dis* 2020; **2**: 219–226.

41. Mughini-Gras L, Dorado-García A, van Duijkeren E *et al.* Attributable sources of community-acquired carriage of *Escherichia coli* containing ß-lactam antibiotic resistance genes: a population-based modelling study. *Lancet Planet Health* 2019; **3**: e357–e369.

42. Doi Y, Paterson DL, Egea P *et al.* Extended-spectrum and CMY-type b-lactamase-producing *Escherichia coli* in clinical samples and retail meat from Pittsburgh, USA and Seville, Spain. *Clin Microbiol Infect* 2010; **16**: 33–38.

43. Kluytmans JAJW, Overdevest ITMA, Willemsen I *et al.* Extended-spectrum ß-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.

44. Machado E, Coque TM, Canton R *et al.* Antibiotic resistance integrons and extended-spectrum b-lactamases among *Enterobacteriaceae* isolates recovered from chickens and swine in Portugal. *J Antimicrob Chemother* 2008; **62**: 296–302.

45. Smet A, Martel A, Persoons D *et al.* Diversity of extended-spectrum b-lactamases and class C b-Lactamases among cloacal *Escherichia coli* isolates in Belgian broiler farms. *Antimicrob Agents Chemother* 2008; **52**: 1238–1243.

46. Collignon P, Beggs JJ. CON: COVID-19 will not result in increased antimicrobial resistance prevalence. *J Antimicrob Chemother* – AMR 2020; 3: dlaa051.

47. Brito IL. Examining horizontal gene transfer in microbial communities. *Nat Rev Microbiol* 2021. Online ahead of print. doi: 10.1038/s41579-021-00534-7.

48. Nordmann P, Cuzon G, Naas T. The real threat of *Klebsiella pneumoniae* carbapenemase-producing bacteria. *Lancet Infect Dis* 2009; **4**:228-36.

49. Smillie CS, Smith MB, Friedman J *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 2011; **7376**: 241-4.

50. van den Bunt G, van Pelt W, Hidalgo L *et al.* Prevalence, risk factors and genetic characterisation of extended-spectrum ß-lactamase and carbapenemase-producing *Enterobacteriaceae* (ESBL-E and CPE): a community-based cross-sectional study, the Netherlands, 2014 to 2016. *Euro Surveill* 2019; **24**: 29-39.

51. Acman M, Dorp van L, Santini JM, Francois Balloux F. Large-scale network analysis captures biological features of bacterial plasmids. *Nat Commun* 2020; **1**: 2452.

52. Hendrickx APA, Landman F, de Haan A *et al*. blaOXA-48-like genome architecture among carbapenemase-producing *Escherichia coli* and *Klebsiella pneumoniae* in the Netherlands. *bioRxiv* 2020.12.18.423568, doi: https://doi.org/10.1101/2020.12.18.423568

53. León-Sampedro R, DelaFuente J, Díaz-Agero C *et al*. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. *Nat Microbiol* 2021; https://doi.org/10.1038/s41564-021-00879-y

54. Hidalgo L, de Been M, Rogers MRC *et al*. Sequence-based epidemiology of an OXA-48 plasmid during a hospital outbreak. *Antimicrob Agents Chemother* 2019; **12**: e01204-19.

55. Martin J, Phan HTT, Findlay J *et al*. Covert dissemination of carbapenemase-producing *Klebsiella pneumoniae* (KPC) in a successfully controlled outbreak: long- and short-read whole-genome sequencing demonstrate multiple genetic modes of transmission. *J Antimicrob Chemother* 2017; **11**: 3025-3034.

56. Chaves J, Ladona MG, Segura C *et al*. SHV-1 ß-lactamase is mainly a chromosomally encoded species-specific enzyme in *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 2001; **10**: 2856-61.

57. Sirot J, C Chanal C, A Petit A *et al*. *Klebsiella pneumoniae* and other *Enterobacteriaceae* producing novel plasmid-mediated ß-lactamases markedly active against third-generation cephalosporins: epidemiologic studies. *Rev Infect Dis* 1988; **4**: 850-9

58. Nordmann P, Poirel L, Walsh TR, Livermore DM. The emerging NDM carbapenemases. T*rends Microbiol* 2011; **12**: 588-95.

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

60. Haverkate MR, Dautzenberg MJD, Ossewaarde TJM *et al*. Within-host and population transmission of blaOXA-48 in *K. pneumoniae* and *E. coli*. *PLoS One* 2015; **10**: e0140960.

61. Davies NG, Abbott S, Barnard RC *et al*. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021; **6538**: eabg3055.

62. Tegally H, Wilkinson E, Giovanetti M *et al*. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 2021; **7854**: 438-43.

63. World Health Organization 2020. Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report. https://apps.who.int/iris/bitstream/handle/10665/332081/9789240005587-eng.pdf?ua=1.

64. World Health Organization 2021. WHO integrated global surveillance on ESBL-producing *E. coli* using a "One Health" approach: Implementation and opportunities. Report available at: https://www.who.int/publications/i/item/who-integrated-global-surveillance-on-esbl-producing-e.-coli-using-a-one-health-approach.

65. ECDC 2020. Surveillance of antimicrobial resistance in Europe 2019. https://www.ecdc.europa.eu/sites/default/files/documents/surveillance-antimicrobial-resistance-Europe-2019.pdf.

66. ECDC 2018. ECDC study protocol for genomic-based surveillance of carbapenem-resistant and/or colistin-resistant *Enterobacteriaceae* at the EU level. https://www.ecdc.europa.eu/sites/default/files/documents/Protocol-genomic-surveillance-resistant-*Enterobacteriaceae*-update.pdf

67. ISIS-AR. https://www.rivm.nl/isis-ar.

68. Hendrickx APA, Landman F, de Haan A *et al*. Plasmid diversity among genetically related *Klebsiella pneumoniae* blaKPC-2 and blaKPC-3 isolates collected in the Dutch national surveillance. *Sci Rep* 2020; **10**: 16778.

69. Iskandar K, Molinier L, Hallit S *et al*. Surveillance of antimicrobial resistance in low- and middle-income countries: a scattered picture. *Antimicrob Resist Infect Control* 2021; **1**: 63.

70. Drouin A, Letarte G, Raymond F *et al*. Interpretable genotype-to-phenotype classifiers with performance guarantees. *Sci Rep* 2019; **9**: 1-13.

71. Bortolaia V, Kaas RS, Ruppe E *et al*. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020; **12**: 3491-500.

72. U.S. Department of health and human services food and drug administration 2009. Guidance for industry and FDA. Class II special controls guidance document: antimicrobial susceptibility test (AST) systems. https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/antimicrobial-susceptibility-test-ast-systems-class-ii-special-controls-guidance-industry-and-fda.

73. Török ME, Peacock SJ. Rapid whole-genome sequencing of bacterial pathogens in the clinical microbiology laboratory—pipe dream or reality? *J Antimicrob Chemother* 2012; **67**: 2307–2308.

74. Deurenberg RH, Bathoorn E, Monika A *et al*. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol* 2017; **243**: 16-24.

75. SWAB 2020. Richtlijn Urineweginfecties 2020. https://adult.swabid.nl/nl/node/7787.

76. Su M, Satola SW, Read TD. Genome-based prediction of bacterial antibiotic resistance. *J Clin Microbiol* 2019; **3**: e01405-18.

77. Quick J, Loman NJ, Duraffour S *et al*. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016; **7589**: 228–232.

78. Peiffer-Smadja N, Dellière S, Rodriguez C *et al*. Machine learning in the clinical microbiology laboratory: has the time come for routine practice? *Clin Microbiol Infect* 2020; **10**: 1300-130.

79. McArthur AG, Waglechner N, Nizam F *et al*. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 2013; **7**: 3348–3357.

80. Gupta SK, Padmanabhan BR, Diene SM *et al*. ARG-ANNOT, a new bioinformatic tool to discover an- tibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 2013; **1**:212–220.

81. Lees JA, Mai TT, Galardini M *et al*. Improved prediction of bacterial genotype-phenotype associations using interpretable pangenome-spanning regressions. *MBio* 2020; **4**: e01344-20.

82. Zankari E, Hasman H, Kaas RS *et al*. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J Antimicrob Chemother* 2013; **68**: 771–7.

83. Stoesser N, Batty EM, Eyre DW *et al*. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J Antimicrob Chemother* 2013; **68**: 2234–44.

84. Shelburne SA, Kim J, Munita JM *et al*. Whole-genome sequencing accurately identifies resistance to extended- spectrum beta-lactams for major gram-negative bacterial pathogens. *Clin Infect Dis* 2017; **65**: 738–45

85. Hendriksen RS, Bortolaia V, Tate H *et al*. Using genomics to track global antimicrobial resistance. *Front Public Health* 2019; **7**: 242.

86. Gordon NC, Price JR, Cole K, *et al*. Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol* 2014; **52**: 1182–1191.

87. Jaillard M, Palmieri M, van Belkum A, Mahé P. Interpreting *k*-mer-based signatures for antibiotic resistance prediction. *Gigascience* 2020; **10**: giaa110.

88. Lüftinger L, Májek P, Beisken S *et al*. Learning from limited data: towards best practice techniques for antimicrobial resistance prediction from whole genome sequencing data. *Front Cell Infect Microbiol* 2021; **11**: 610348.

89. Moradigaravand D, Palm M, Farewell A *et al*. Prediction of antibiotic resistance in *Escherichia coli* from large-scale pan-genome data. *PLOS Comput Biol* 2018; **12**: 1–17.

90. Nguyen M, Brettin T, Long SW *et al*. Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Sci Rep* 2018; **8**: https://doi.org/10.1038/s41598-017-18972-w.

91. Chen ML, Doddi A, Royer J. Beyond multidrug resistance: leveraging rare variants with machine and statistical learning models in *Mycobacterium tuberculosis* resistance prediction. *EBioMedicine* 2019; **43**: 356–369.

91. Fleischmann RD, Adams MD, White O *et al*. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 1995; **269**: 496–512.

92. Moore GE. Cramming more components onto integrated circuits. *Electronics* 1965; **8**: 114-118.

93. Leinonen R, Akhtar R, Birney E *et al*. The European nucleotide archive. *Nucleic Acids Res* 2011; **39**: D28–D31.

94. Benson DA, Cavanaugh M, Clark K *et al*. GenBank. *Nucleic Acids Res* 2013; **41**: D36-42.

95. Wattam AR, Abraham D, Dalay O *et al*. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 2014; **42**: D581–D591.

8
Appendices

Nederlandse samenvatting
Contributing authors
List of publications
Dankwoord
About the author

NEDERLANDSE SAMENVATTING

In 2019 heeft de World Health Organisation (WHO) antibioticaresistentie uitgeroepen tot een van de top tien bedreigingen voor de wereldbevolking, samen met onder andere luchtvervuiling en klimaatverandering. Jaarlijks overlijden er een geschatte 700,000 mensen aan de gevolgen van antibioticaresistentie en kost dit probleem in 2030 mogelijk jaarlijks tot 3.4 triljoen dollar.

Hoofdpijlers van het voorgestelde beleid van de WHO om antibioticaresistentie te bestrijden zijn:

(1) Verbetering van kennis over antibioticaresistentie
(2) Verbetering van surveillance
(3) Verbetering van infectiepreventie
(4) Optimalisering van antibioticagebruik

In Europa is de zogenoemde ESBL[a]-producerende *Escherichia coli* bacterie (vanaf hier afgekort als ESBL-Ec) de meest voorkomende oorzaak van antibioticaresistente infecties. De studies in dit proefschrift hebben onderzocht hoe praktische toepassingen van een nieuwe techniek ESBL-Ec beter bestreden kan worden. De nieuwe techniek (*whole genome sequencing*, WGS) decodeert en digitaliseert de genetische code van bacteriën (en andere organismen). De genetische code kan vervolgens worden gebruikt worden voor verschillende toepassingen, waarvan onder andere:

- Kennis vermeerdering over welke bronnen bijdragen aan ESBL-Ec dragerschap en infecties bij mensen (*punt 1 WHO*)
- Genetische surveillance in de algemene bevolking (*punt 2 WHO*)
- Een snellere en betere methode voor de keuze van de juiste antibiotica (*punt 4 WHO*)

Deze samenvatting geeft allereerst algemene informatie over ESBL-Ec en WGS geven en vervolgens worden de bevindingen in dit proefschrift besproken samen met de belangrijkste wetenschappelijke literatuur. Als laatste zullen er toekomstperspectieven worden besproken samen met de conclusie van dit proefschrift.

[a] Extended-spectrum ß-lactamase

ESBL-EC: DRAGERSCHAP EN INFECTIES

Mensen dragen talrijke en diverse gemeenschappen van micro-organismen op en in hun lichaam, waarvan de meest belangrijke in de darmen. Het zogenoemde microbioom[b] van de darmen (vanaf hier microbioom genoemd) draagt bij aan een gezond afweersysteem, goede vertering, en beschermt ons tegen ziekteverwekkers. De *E. coli* bacterie maakt ongeveer 2% uit van een gezond microbioom. Goede eigenschappen van *E. coli* zijn: i) een zuurstofarme omgeving creëren waarop andere gunstige bacteriën (zogenoemde anaeroben[c]) gedijen, ii) bescherming tegen ziekteverwekkers die zich proberen te vestigen in het microbioom. Dit mechanisme wordt kolonisatieresistentie genoemd en vindt plaats op basis van competitie voor ruimte en voedingsstoffen.

De aanwezigheid van *E. coli* in het microbioom kan soms leiden tot een infectie. *E. coli* kan zich buiten de darmen namelijk hechten aan oppervlaktecellen en deze binnendringen, met name in de urinewegen. *E. coli* is dan ook de meest voorkomende oorzaak van urineweginfecties en bloedvergiftiging (vaak als een complicatie van een urineweginfectie) in de westerse wereld.

Infecties met *E. coli* nemen sinds een aantal jaar toe. In het Verenigd Koninkrijk stegen bloedvergiftigingen van 45 tot 70 gevallen per 100,000 inwoners per jaar tussen 2009 en 2018. Deze toename kan mogelijk verklaard worden door (i) veranderingen onder de bevolking en/of (ii) verandering van de *E. coli* bacterie: (i) bevolkingsvergrijzing zou mogelijk kunnen leiden tot een toename van infecties met *E. coli*. De meerderheid van deze infecties komt voornamelijk voor bij een leeftijd van 75 en ouder. (ii) de *E. coli* bacterie heeft mogelijk een aantal evolutionaire veranderingen doorgemaakt waardoor deze beter aangepast is aan verblijf in de menselijke darm (vanaf hier dragerschap genoemd), met meer infecties als onbedoeld bijproduct. Gelijktijdig met de toename van infecties zijn specifieke varianten van *E. coli* bij mensen gedetecteerd over de hele wereld, waaronder de ST131-variant. Er zijn aanwijzingen dat de ST131-variant besmettelijker is én geassocieerd is met een langere dragerschapsduur dan andere varianten. Daarnaast is de ST131-variant vaak ongevoelig voor de meest voorkomende behandelingsopties tegen urineweginfecties en bloedvergiftiging door de aanwezigheid van een ESBL-gen. Dit geldt ook voor sommige andere varianten van *E. coli*, al deze varianten samen worden ESBL-Ec genoemd.

[b] De verzameling van micro-organismen in een bepaalde omgeving
[c] Anaerobe bacteriën wekken energie op zonder zuurstof en gedijen het beste in een zuurstofloze omgeving

Het monitoren van ESBL-Ec varianten en opsporen van nieuwe varianten wordt genetische surveillance genoemd. Genetische surveillance wordt reeds gebruikt voor bepaalde zeer resistente bacterie-groepen[d] aangetroffen in ziekenhuizen, maar nog niet voor ESBL-Ec. Het belang van genetische surveillance wordt verder geillustreerd door COVID-19. Tijdens de pandemie is weer gezien dat nieuwe varianten de plaats van oude varianten overnemen wanneer ze bepaalde evolutionaire voordelen hebben zoals toegenomen besmettelijkheid.

ESBL-Ec komt ook in verschillende niet-menselijke reservoirs voor, zoals in de veehouderij, wilde vogels, voedsel, aarde, afval- en oppervlaktewater. Al deze reservoirs zouden mogelijke bronnen kunnen zijn voor menselijk dragerschap en infecties.

Verder is er nog een factor die de verspreiding van ESBL-Ec extra ingewikkeld maakt. ESBL-genen kunnen zich namelijk op twee manieren verspreiden: (i) klonaal en (ii) horizontaal. (i) Een beperkt aantal *E. coli* varianten, die zich over de hele wereld hebben verspreid, is geassocieerd met het dragen van een ESBL-gen. Bijvoorbeeld de eerdergenoemde ST131-variant. Deze varianten zijn in staat ESBL-genen en andere antibioticaresistentie-genen vast te houden. Wanneer een dergelijke bacterie zich vermeerderd is er sprake van klonale verspreiding. (ii) ESBL-genen zijn meestal gelegen op zogenoemde plasmiden. Plasmiden zijn losse stukken DNA die door bacteriën worden kunnen opgenomen. Na opname kan de bacterie gebruik maken van de eigenschappen geencodeerd in het DNA van de plasmide, zoals antibioticaresistentie, wat leidt tot een overlevingsvoordeel. Wanneer het overlevingsvoordeel verdwijnt, bijvoorbeeld door verdwijnen van antibiotica uit de omgeving, de plasmide weer losgelaten worden. Het opnemen van een plasmide wordt mogelijk gemaakt door een andere naburige bacterie, die deze genetische informatie deelt. Dit kan een *E. coli*-bacterie zijn, maar ook een andere bacteriesoort. *Klebsiella pneumoniae* is een bacteriesoort die vermoedelijk regelmatig plasmiden deelt met *E. coli*. Over horizontale verspreiding is echter nog weinig bekent.

WGS

Met WGS wordt de volledige nucleotidevolgorde van het DNA van een organisme vastgesteld. Deze genetische code kan vervolgens gebruikt worden voor verschillende doeleinden: (i) De bacteriesoort en/of variant vaststellen. (ii) Vaststellen of en welk type ESBL-gen en eventuele andere antibioticaresistentie-genen de bacterie draagt. (iii) Vergelijken van de genetische code van 2 of meer bacteriën

[d] Carbapenem- en colistine-resistente *Enterobacterales*. Tot de *Enterobacterales* familie behoren onder andere *E. coli* en *K. pneumoniae*

, deze informatie kan vervolgens worden gebruikt om verspreiding, klonaal, dan wel horizontaal, vast te stellen. Verder zorgt het digitale karakter van WGS ervoor dat de genetische codes oneindig opgeslagen, geanalyseerd, en gedeeld kunnen worden. Deze factoren verbeteren de reproduceerbaarheid en hierdoor de kwaliteit van onderzoek, surveillance en klinische diagnostiek. Technische vooruitgang van WGS, lagere kosten, en beschikbaarheid van gratis en online analyseprogramma's, hebben WGS toegankelijk gemaakt voor de medische microbiologie en publieke gezondheidszorg in grote delen van de wereld. Deze toegankelijkheid zal waarschijnlijk sterk toenemen in de komende jaren.

BELANGRIJKSTE BEVINDINGEN VAN DIT PROEFSCHRIFT

Genetische verschillen tussen ESBL-positieve en ESBL-negatieve *E. coli*

In **hoofdstuk 2** tonen we aan dat ESBL-positieve en ESBL-negatieve *E. coli*, geïsoleerd uit bloedkweken, genetisch sterk van elkaar verschillen. Deze bevinding suggereert dat de waarschijnlijkheid van een *E. coli* bacterie om antibioticaresistent te zijn geassocieerd is met het type *E. coli*-variant. Deze bevinding wordt ondersteunt door ander onderzoek. Daar is namelijk ook gezien dat bepaalde varianten, zoals ST73, bijna nooit antibioticaresistent zijn, terwijl andere varianten, zoals ST131, vaak wel antibioticaresistent zijn. De varianten die vaker antibioticaresistent leveren een disproportioneel grote bijdrage aan de ziektelast door antibioticaresistente infecties en zijn daardoor interessant voor gerichte infectiepreventie en vaccinontwikkeling.

Dragerschap als bron voor infectie met ESBL-Ec

Om infectiepreventie tegen ESBL-Ec te verbeteren is het belangrijk te begrijpen in hoeverre eigen dragerschap van ESBL-Ec in het microbioom een bron is voor een mogelijke infectie met ESBL-Ec. In andere woorden: of ziekte meestal veroorzaakt wordt door een ESBL-Ec die iemand al langere tijd bij zich draagt of dat iemand een meestal ziek wordt van een ESBL-Ec van buitenaf, kort na blootstelling. In **hoofdstuk 3** observeerden we dat eigen dragerschap waarschijnlijk de belangrijkste bron is voor infecties met ESBL-Ec. Preventie van infecties met ESBL-Ec zou dus bereikt kunnen worden door dragerschap van ESBL-Ec te voorkomen.

Mens als belangrijkste bron voor ESBL-Ec dragerschap

Om dragerschap te voorkomen is het belangrijk om te begrijpen welke bronnen de belangrijk zijn voor dragerschap van ESBL-Ec. Zoals eerder genoemd komt ESBL-Ec in een diverse groep van reservoirs voor, onder andere in de veehouderij, wilde vogels, voedsel, aarde, afval-, en oppervlaktewater. In **hoofdstuk 5 & 6** tonen we

aan dat in Europa, ESBL-Ec dragerschap met name wordt verkregen door besmetting door een ander mens. Deze observaties worden ondersteund door een groeiende hoeveelheid wetenschappelijk bewijs die suggereert dat besmetting door niet-menselijke reservoirs relatief zeldzaam zijn. De hypothese dat ESBL-Ec met name van mens-tot-mens verspreidt wordt verder ondersteund door de geobserveerde daling in ESBL-Ec in ziekenhuiskweken sinds de start van de COVID-19 pandemie. De prevalentie van ESBL-positieve kweken daalde van 5-10% naar 2.5% in sommige ziekenhuizen. Wanneer we aannemen dat de blootstelling aan niet-menselijke reservoirs gelijk bleef (met uitzondering van reizen naar het buitenland), zou dit betekenen dat de lockdown-maatregelen verspreiding van mens-tot-mens heeft verminderd.

Horizontale verspreiding als bron voor ESBL-Ec dragerschap

In **hoofdstuk 6** observeerden we dat de antibioticaresistentie genen aangetroffen in ESBL-Ec en ESBL-positieve *K. pneumoniae* in menselijk dragerschap grote gelijkenissen toonden. De antibioticaresistentie genen tussen deze groepen kwamen bijvoorbeeld meer overeen dan ESBL-Ec gevonden in menselijk dragerschap en voedsel. Op basis van deze observatie stellen wij de hypothese dat horizontale verspreiding een belangrijke bron is voor ESBL-Ec dragerschap. Recente studies hebben ook bewijs gevonden voor horizontale verspreiding van plasmiden tussen *E. coli*, *K. pneumoniae* en andere bacteriële soorten. Met de vooruitgang van technische mogelijkheden zal er in de komende jaren meer duidelijkheid komen over het belang van horizontale verspreiding binnen het probleem van antibiotica-resistentie.

Genetische surveillance van ESBL-Ec in de algemene bevolking

In **hoofdstuk 3** tonen we aan dat ESBL-Ec aangetroffen in urine- en bloedkweken bij de huisarts en in het ziekenhuis een representatief beeld geven van de ESBL-Ec varianten die circuleren in de Nederlandse algemene bevolking. Routinematig verzamelde urine- en bloedkweken zouden hierom een geschikte methode voor genetische surveillance kunnen zijn in Nederland.

Zoals eerder genoemd wordt er op dit moment in Nederland (en Europa) alleen genetische surveillance uitgevoerd op een tweetal zeer resistente bacteriesoorten[d]. De surveillance vindt plaats door verzending van deze bacteriën naar een centraal lab en vervolgens uitvoering van WGS en analyse. De gecentraliseerde WGS zorgt voor een garantie in vergelijkbaarheid en kwaliteit van de genetische codes. Echter is deze methode ook een knelpunt, aangezien het uitbreiden van het aantal

[d] Carbapenem- en colistine-resistente *Enterobacterales*. Tot de *Enterobacterales* familie behoren onder andere *E. coli* en *K. pneumoniae*

bacteriesoorten voor genetische surveillance een vergroting van capaciteit vereist voor het centrale lab. Nu WGS toenemend beschikbaar wordt, zou het interessant kunnen zijn om deze werkwijze te decentraliseren. WGS zou uitgevoerd kunnen worden in het lokale ziekenhuislaboratorium en de genetische codes zouden vervolgens gedeeld kunnen worden met de centrale faciliteit. De centrale faciliteit zou, na een kwaliteitscheck, de data-analyse kunnen uitvoeren. Op dit moment wordt er in Nederland een dergelijke werkwijze gehanteerd voor de genetische surveillance van COVID-19. Dit zou een kans kunnen zijn voor uitbreiding naar, gedecentraliseerde, genetische surveillance van ESBL-Ec wanneer de pandemie in hevigheid afneemt.

WGS als snellere en betere methode voor de keuze van de juiste antibiotica

In **hoofdstuk 4** tonen we aan dat twee beschikbare programma's, die mogelijk ingezet zouden kunnen worden als antibioticakeuze-test bij patiënten in het ziekenhuis, nog niet aan de kwaliteitscriteria voldoen van de U.S. Food en Drug Administration (FDA) voor het inzetten van dergelijke testen bij patiënten. Echter denken wij dat er voldoende aangrijpingspunten voor verbetering zijn, waardoor we verwachten dat deze of vergelijkbare programma's in de toekomst ingezet kunnen worden in de gezondheidszorg.

Op dit moment worden antibioticakeuze-testen bij patiënten in het ziekenhuis uitgevoerd met een kweek van de bacterie. Deze methode is relatief goedkoop, maar kost enkele dagen. Dit betekent dat een patiënt in eerste instantie empirisch[e] wordt behandeld met een breedspectrum antibioticum, om dekking tegen de onbekende ziekteverwekker te garanderen. De antibioticabehandeling kan vervolgens zo nodig worden aangepast wanneer de kweekresultaten (antibioticakeuze-test) bekend zijn. Het gevolg van deze werkwijze is dat patiënten over- en onder-behandeld kunnen worden. Overbehandeling leidt tot onnodig antibioticagebruik. Onder-behandeling kan leiden tot een ongunstig resultaat van de patiënt (bijvoorbeeld langere ziekteduur).

WGS heeft de potentie om de werkwijze van empirische antibioticabehandeling te vervangen door gerichte antibioticabehandeling. Om dit mogelijk te maken moet aan drie voorwaarden voldaan worden: 1) ziekenhuizen moet een WGS faciliteit hebben die grote volumes aankan, 2) methoden moeten worden ontwikkeld die WGS direct op menselijk weefsel (zoals bloed) mogelijk maken (waardoor de stap van kweken niet meer nodig is), 3) Analyseprogramma's die op basis van genetische codes antibioticaresistentie voorspellen moeten aan de kwaliteitscriteria van de

[e] Behandeling gebaseerd op ervaring

FDA voldoen. Daarnaast moeten deze programma's gebruiksvriendelijk en het liefst gratis zijn.

CONCLUSIES VAN DIT PROEFSCHRIFT

De studies in dit proefschrift laten zien dat de praktische toepassing van WGS in wetenschappelijk onderzoek, publieke gezondheidszorg en patiëntenzorg de mogelijkheid biedt om ESBL-Ec en hiermee antibioticaresistentie in het algemeen beter te bestrijden. Dit is van belang omdat antibioticaresistentie wordt gezien als een van de top tien bedreigingen voor de gezondheid van de wereldbevolking.

Toekomstige stappen die ik noodzakelijk acht in onderzoek en de publieke gezondheidszorg:

- Begrip over kolonisatieresistentie en hoe dit fenomeen bijdraagt aan dragerschap en infectie met ESBL-Ec

- Begrip over hoe vaak horizontale verspreiding van antibioticaresistentie voorkomt tussen bacteriën

- De capaciteit van genetische surveillance vergroten door decentralisatie van WGS en vervolgens implementatie van genetische surveillance voor ESBL-Ec in Nederland

- Prioriteren van ontwikkeling van WGS als snellere en betere methode voor de keuze van de juiste antibiotica

CONTRIBUTING AUTHORS

Sergio Arredondo-Alonso

1) Biostatistics, University of Oslo, Oslo, Norway. 2) Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, The Netherlands

Ingo Autenrieth

Institute of Medical Microbiology and Hygiene, University Hospital Tübingen, Tübingen, Germany

Marc Bonten

Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, the Netherlands

Thijs Bosch

the Netherlands Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Patricia Bruijning-Verhagen

1) Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, the Netherlands. 2) the Netherlands Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Abdessalam Cherkaoui

Bacteriology Laboratory, Geneva University Hospital, Geneva, Switzerland

Nadine Conzelmann

Department of Infectious Diseases, University Hospital Tübingen, Tübingen, Germany

Ben Cooper

Centre for Tropical Medicine & Global Health, Nuffield Department of Medicine, University of Oxford, United Kingdom

Mercedes Delgado

1) Infectious Diseases, Microbiology and Preventive Medicine Unit, Virgen Macarena University Hospital, Sevilla, Spain. 2) Department of Medicine, University of Sevilla/Biomedicine Institute of Sevilla, Sevilla, Spain

Ad Fluit

Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, the Netherlands

Siri Göpel

Department of Infectious Diseases, University Hospital Tübingen, Tübingen, Germany

Julia Guther

Institute of Medical Microbiology and Hygiene, University Hospital Tübingen, Tübingen, Germany

Stephan Harbarth

Infection Control Program, Geneva University Hospital and Faculty of Medicine, Geneva, Switzerland

Didier Hocquet

1) Infection Control Unit, University Hospital of Besançon, Besançon, France . 2) UMR 6249, Laboratoire Chrono-environnement, CNRS-Université de Bourgogne Franche-Comté, Besançon, France

Denise van Hout

1) Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, the Netherlands. 2) the Netherlands Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Jan Kluytmans

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

Marlieke de Kraker

Infection Control Programme and WHO Collaborating Centre on Patient Safety, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

Daniel Martak

1) Infection Control Unit, University Hospital of Besançon, Besançon, France. 2) UMR 6249, Laboratoire Chrono-environnement, CNRS-Université de Bourgogne Franche-Comté, Besançon, France

Romain Martischang

Infection Control Programme and WHO Collaborating Centre on Patient Safety, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

Valentina Mascaro

ASST Rhodense, G. Salvini Hospital, Garbagnate Milanese, Italy

Alexandre Meunier

Infection Control Unit, University Hospital of Besançon, Besançon, France

Patrick Musicha

Centre for Tropical Medicine & Global Health, Nuffield Department of Medicine, University of Oxford, United Kingdom

Julian Paganini

Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, The Netherlands

Silke Peter

Institute of Medical Microbiology and Hygiene, University Hospital Tübingen, Tübingen, Germany

Elien Reuland

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

Maria Eugenia Riccio

Infection Control Program, Geneva University Hospital and Faculty of Medicine, Geneva, Switzerland

Jesús Rodriguez-Baño

1) Clinical Unit of Infectious Diseases and Microbiology, University Hospital Virgen Macarena, Seville, Spain. 2) Department of Medicina, University of Seville / IBiS, Seville, Spain

Elena Salamanca

1) Clinical Unit of Infectious Diseases and Microbiology, University Hospital Virgen Macarena, Seville, Spain. 2) Department of Medicina, University of Seville / IBiS, Seville, Spain

Julien Sauser

Infection Control Programme and WHO Collaborating Centre on Patient Safety, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

Jelle Scharringa

Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, The Netherlands

Anita Schürch

Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, The Netherlands

Evelina Tacconelli

1) Infectious Diseases, Department of Internal Medicine I, University Hospital Tübingen, Tübingen, Germany. 2) Infectious Diseases Division, Department of Diagnostic and Public Health, University Hospital Verona, Verona, Italy

Janetta Top

Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, The Netherlands

Rob Willems

Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, The Netherlands

LIST OF PUBLICATIONS

Publications included in this thesis:

van Hout D, **Verschuuren TD**, Bruijning-Verhagen PCJ, Bosch T, Schürch AC, Willems RJL, Bonten MJM, Kluytmans JAJW. Extended-spectrum beta-lactamase (ESBL)-producing and non-ESBL-producing *Escherichia coli* isolates causing bacteremia in the Netherlands (2014 - 2016) differ in clonal distribution, antimicrobial resistance gene and virulence gene content. *PLoS One* 2020; **1**: e0227604.

Verschuuren TD, van Hout D, Arredondo-Alonso S, Fluit AC, Reuland EA, Top J, Schürch AC, Bosch T, Bonten MJM, Kluytmans JAJW, Willems RJL. 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). *J Antimicrob Chemother* 2021; **4**: 901-908.

Riccio ME, **Verschuuren TD**, Conzelmann N, Martak D, Meunier A, Salamanca E, Delgado M, Guther J, Peter S, Paganini J, Martischang R, Sauser J, de Kraker MEA, Cherkaoui A, Fluit AC, Cooper BS, Hocquet D, Kluytmans JAJW, Tacconelli E, Rodriguez-Baño J, Harbarth S, MODERN WP2 study group. Household acquisition and transmission of extended-spectrum ß-lactamase (ESBL) -producing *Enterobacteriaceae* after hospital discharge of ESBL-positive index patients. *Clin Microbiol Infect* 2021; S1198-743X(20)30784-9.

Publications not included in this thesis:

Arredondo-Alonso S, Rogers MRC, Braat JC, **Verschuuren TD**, Top J, Corander J, Willems RJL, Schürch AC. mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. Microb Genom 2018; 11: e000224.

Martak D, Guther J, **Verschuuren TD**, Valot B, Conzelmann N, Bunk S, Riccio ME, Salamanca E, Meunier A, Henriot CP, Pressacco Brossier C, Bertrand X, Cooper BS, Harbarth S, Tacconelli E, Fluit AC, Rodriguez-Baño J, Kluytmans JAJW, Peter S, Hocquet D, MODERN WP3 study group. Populations of extended-spectrum ß-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* are different in human-polluted environment and food items: A multicentre European study. *Clin Microbiol Infect* 2021; S1198-743X(21)00414-6.

DANKWOORD

Dit proefschrift is tot stand gekomen door deelname van vele personen en input van vele directe en indirecte collega's alsmede deelnemers en medewerkers van de deelnemende centra. Aan iedereen die samen met mij dit pad heeft bewandeld in de afgelopen 5 jaar: dankjewel voor je inzet, tijd, deelname, lessen en inspiratie. Naast het gereedkomen van dit proefschrift heb ik persoonlijk genoten van het beoefenen van wetenschap en heb ik veel geleerd over mezelf. Ik wil graag wat personen in het bijzonder bedanken:

Beste **Jan**, dankjewel dat je mij onder je hoede hebt genomen. Je hebt me de vrijheid gegeven om binnen de EPIGENEC en MODERN studie zelf onderwerpen van interesse uit te zoeken, dankjewel voor je vertrouwen en je begeleiding. Moleculaire materie binnen een Europese studie in de community was het beste wat ik had kunnen krijgen. Met je klinische blik en nieuwsgierigheid weet je precies wat er speelt. Ik heb dan ook genoten van alle momenten waarop we de literatuur en concepten in een bredere context bespraken. De borrels bij jou thuis/tijdens de ECCMID, contacten in het Amphia en onze trip naar Madrid zijn voor mij mooie herinneringen.

Beste **Rob**, vanaf moment 1 heb ik me welkom gevoeld. Elke meeting werd gestart met een kopje koffie en was er ruimte voor niet inhoudelijke zaken. Je oog voor detail en kritische wetenschappelijke houding hebben mijn werk enorm verbetert. Ook heb ik door jouw begeleiding zelf meer oog voor detail gekregen. Je moleculaire kennis en jouw ongelofelijke en vooruitstrevende team hebben me de kans gegeven om met de nieuwste methoden te werken. Ook heb ik erg genoten van het etentje bij jou thuis, helaas heeft lila er wel een ijs-verslaving aan overgehouden.

Beste **Thijs**, het begon allemaal met 'het Burrito overleg'. Dankjewel voor je begeleiding en bereikbaarheid, ongeacht hoe druk je het had was je motto 'gewoon naar mij sturen Tess'. Je praktische tips en hulp met tijdsplanning hebben me erg geholpen. Daarnaast wil ik je ook bedanken voor het mogelijk maken van het EPIGENEC project en je input in het MODERN project. De momenten die niet over de inhoud gingen waren ook heel waardevol voor me, zoals samen het proefschrift accepteren op myphd, de ECCMID borrel (met bijbehorende aangeschoten treinrit terug), V.S. vakantietips en het etentje bij Rob. Ik ben heel benieuwd wat je de komende jaren allemaal gaat doen binnen het RIVM en ik ga het volgen!

Beste **Marc, Ad, Jelle, Janetta, Anita** en **Henri**, alhoewel geen officiële begeleiders, toch heb ik veel van jullie geleerd. **Marc**, dankjewel dat je een kans zag in een 6e jaars geneeskunde student. Jouw combinatie van kennis, scherpe visie, begrijpelijk kunnen uitleggen en ja toch ook hang naar gezelligheid, hebben voor mij een fijne werkomgeving, met veel mogelijkheden en top collega's gecreëerd. **Ad**, jouw 'matter-of-fact' houding en diepe interesse in de microbiologie werkt aanstekelijk. Verwarde vragen van mijn kant op jouw kantoor werden vaak beantwoord met iets in de trant van: 'Je moet het niet moeilijker maken dan het is'. Als ik je kantoor uitliep dacht ik dan vaak, tja het is ook zo. **Jelle**, zeker aan het begin van mijn PhD betrapte ik me erop dat ik regelmatig even bij jou langsliep om te kletsen onder het mom van materiaal/resultaten ophalen. Je bent zo relaxed en hebt zo'n dosis humor dat je je meteen op je gemak voelt. Mijn partner doet nog regelmatig je skit na waarbij je VRE verspreidt als Marc in het filmpje voor Robs inauguratie. Bedankt voor al het werk dat je voor de MODERN studie hebt gedaan! **Janetta**, je hebt me vanaf het begin betrokken in de activiteiten van de moleculaire onderzoeksgroep, waardoor ik me altijd deel voelde van de groep. Ook heb je me meermaals geholpen met onder andere analyses ongeacht of je betrokken was bij het project. Het feit dat ik altijd even kon binnenlopen vond ik heel fijn. **Anita**, dankjewel dat je het voor me mogelijk hebt gemaakt om bioinformatische analyses te doen en de tijd hebt genomen om mij hierin bij de hand te nemen. Het is voor mij een hele waardevolle vaardigheid en een die ik misschien wel de leukste vind van alle vaardigheden die ik heb opgedaan. Daarnaast vind ik jou zo goed en denk ik dat je hele bijzondere dingen gaat doen in de komende jaren. **Henri**, ik denk dat ik jou wel het hart van het epi infectieteam mag noemen, slim, integer en voor iedereen beschikbaar. Dankjewel voor al je advies, van beoordelen van mijn stageverslag, tot uitleg hoe een stuk te reviewen en alles er tussenin.

Beste leden van de beoordelingscommissie, **dr. L. Mughini Gras, prof. dr. D.J. Mevius, prof. dr. J.A. Stegeman, prof. dr. H.F.L. Wertheim** en **prof. dr. J.H.H.M. van de Wijgert**, hartelijk dank dat u de tijd heeft genomen om mijn proefschrift te lezen en beoordelen.

A big thank you to all the team members that made the MODERN-project possible. Dear **Julia**, MODERN-WP4 team! Thank you for the pleasant and fruitful collaboration. I've learned a lot from you and I enjoyed our bi-weekly meetings. Dear **Daniel, Nadine, Elena, Eugenia,** and **Patrick** I had a lot of fun working with you on the MODERN project, and am proud on the output thus far. All the best in your careers. Dear **Silke, Didier, Evelina, Jesús, Stephan, Ben,** and **Ingo**, thank you

very much for the opportunity to work with you, and the input on the MODERN-WP4 work, I enjoyed it very much.

Aan alle deelnemers en familieleden van deelnemers van de MODERN-studie en het team van Altenahove, waaronder **Adrie** en **Nicole**, dank jullie wel voor jullie bijdrage en het mogelijk maken van dit project. Dank aan alle verzorgenden voor het verzamelen van de samples. Dank aan **Arjan**, het resterende team van de technische dienst van Altenahove, **Arno** en collega's van RWZI, voor het verzamelen van de afvalwater samples en jullie interesse in dit onderzoek. Beste **Heike, Merel** en **Hetty**, hartelijk dank voor het filtreren en kweken van de afvalwater samples. Beste **Hetty** en **Camiel**, hartelijk dank voor het leveren van de lijsten van ESBL-positieve patiënten. Beste **Judith** en **Robert**, hartelijk dank voor het analyseren van de menselijke en omgevingssamples van de MODERN-studie.

Betreft het EPIGENEC project wil ik **Judith, Kim** en **Carlo** bedanken voor de fijne samenwerking en het verwerken van de samples. **Ascelijn** en **Patricia** dank jullie wel voor de gedegen input in het project. **Denise**, ik heb enorm genoten om nauw met je samen te werken in het EPIGENEC project, kamergenootjes te zijn en mede XEWMM'er. Toen we nog op kantoor werkte hoopte ik altijd dat je zelfgebakken koekjes mee zou nemen. Wat vaak ook gebeurde haha! Heel veel succes met je nieuwe baan in het RIVM.

Julian and **Sergio**, aka the bioinformatics wizards, it has been fun working together. Thank you for taking your time to, very slowly, explain to me which steps to take. Wishing you the best in your further careers! **Valentina**, thank you for joining our team for 6 months, for collecting samples in the MODERN study, and analysing genotype-phenotype data in the EPIGENEC study. Wishing you all the best and hope to visit you someday!

Fien, ik vond het erg fijn om jou als maatje te hebben tijdens de laatste loodjes. In dezelfde fase van de PhD én ook moeder had ik het gevoel dat jij het beste wist hoe ik me soms voelde haha. Heel veel succes en plezier toegewenst met je opleiding tot kinderarts!

Lieve (XE)WMM'ers, lieve **Annabel, Bastiaan, Claudia, Chiara, Darren, Denise, Diana, Els, Emma, Eva, Fien, Fleur, Gerrita, Henri, Inger, Jan, Janneke Verberk, Janneke van de Wijgert, Kelly, Kirsten, Linda, Lufang, Maaike, Marc, Marjolein, Martin, Meri, Meander, Mirjam, Mui, Nienke Paternotte, Nienke Plantinga,**

Nikki, Patricia, Noor, Peter, Tessa, Thijs, Tim, Thomas, Sharon, Valentijn, Wouter, jullie knappe en gezellige koppen waren de kers op de taart in de afgelopen jaren. Ik heb opgekeken naar jullie slimheid, vastberadenheid en werkhouding. Ook heb heb ik genoten van alle vroege meetings met pakken melk op tafel, (online) alcohol-proeverijen, bingo's, borrels op de ECCMID, teamuitjes, winterberg-weekendjes, college-momentjes, verkleedfeestjes op de koepel en hilarische Julius-kamertaferelen.

Lieve YumMums, **Geertje**, **Judith, Moniek, Anne, Sanne** en **Ellen**, jullie laten me zien hoe je moederschap en carrière kan combineren. Alles is te doen, zolang je de juiste support hebt (lees: whatsappen om 03:00).

Lieve vriendinnetjes en vriendjes, **Lieke, Lieke, Sabine, Maureen, Tom**, **Eef, Kiek, Amy, PvdP'ers, Anne & Tho, Sas & Piet, clubbes en Daphne & Josine,** dank jullie wel voor jullie onvoorwaardelijke steun, gezelligheid, alle gesprekken, lieve momenten, avonden en vakanties door de veeeele jaren heen.

Lieve paranimfen, lieve **Amber** en **Pascale**, we kennen elkaar nu bijna 30 en 20 jaar, dank jullie wel dat jullie bij vrijwel alle belangrijke momenten in mijn leven er zijn geweest en dat jullie naast mij zullen staan tijdens mijn verdediging.

Lieve Prinsjes, **Peet, Jook, Marjolijn, Marit** en **Yannick**, dank jullie voor alle steun en interesse. **Peet** en **Jook** dank jullie wel dat jullie je huis hebben opgesteld voor weken aan een stuk tijdens de eerste COVID-19 golven. Zonder alle uren dat jullie voor Lila gezorgd hebben was dit proefschrift niet afgekomen!

Verschuuren: eigenzinnig en een beetje contactgestoord. **Sophie** and **Irena**, thank you for bringing some American and Macedonian flair to us, it's really been a joy! **Pim**, het was een race naar de finish wie als eerste in het gezin de PhD titel zou behalen. Het lot heeft gezegevierd en ik heb gewonnen. Daarbij moeten we ook nog in acht nemen dat natuurkunde natuurlijk een veel simpeler vak is dan geneeskunde. **Max**, informatica en Macedonisch is een mooie combi, jij kiest je eigen pad en dat inspireert me. **Juultje**, Extinction Rebellion, diepzee oceanografie, shark-tracking, de Galapagos. Wat ben jij cool en bevlogen. **Gabriëlla**, al dan niet officieel, toch wel familie. Dankjewel voor al je vertrouwen, liefde en een rolmodel zijn in mijn leven. **Mama**, onuitputtelijke drive, in je eentje een RCT opgezet met daaruit volgend een PhD-traject, dat doet niemand je na. Dankjewel voor al je interesse en de discussies (want geneeskunde en diergeneeskunde zijn immers praktisch hetzelfde). **Papa**, je weet het, jij bent mijn grote voorbeeld in alles, dat kan ik niet onder woorden

brengen. Dankjewel.

Lila, je bent mijn liefste, mijn mooiste, mijn grappigste. Dankjewel dat je in mijn leven bent gekomen. Met jou als huisgenoot voor de komende 18 jaar kan het leven niet meer stuk.

Tom, in 5 jaar kan er veel gebeuren, van gare studenten in Wittevrouwen, naar papa en mama in de phoenix-wijk. Dankjewel voor je onvermoeide interesse in mijn onderzoek. Ondanks je beperkte begrip van geneeskunde en microbiologie hoor je al mijn verhalen aan en mag je zelf graag met medische termen gooien, je favoriet; 'erysipelas?!' en 'misschien is het VRE?' Je hebt me gehouden aan onze belofte aan elkaar om na mijn PhD op avontuur te gaan in de tropen. Op naar de volgende fase, met volle overgave!

ABOUT THE AUTHOR

Tess Verschuuren was born on May 23rd 1990 in Erp. She studied Medicine at the Utrecht University. During her electives, she focussed on (tropical) infectious diseases, and microbiology. Including internships in Cape Town (South Africa), and Paramaribo (Suriname). She was also a board member of the Lamento Foundation (part of student organisation C.S. Veritas), supporting small scale humanitarian projects, and taught anatomy and clinical lessons to medicine students. In 2016, Tess obtained her degree in Medicine.

She started a PhD trajectory in 2017 at the Epidemiology of Infectious Diseases group of the Julius Centre at the University Medical Centre Utrecht. She was supervised by prof. dr. Jan Kluytmans, prof. dr. Rob Willems, and dr. Thijs Bosch. In parallel with her PhD, Tess obtained a postgraduate master in epidemiology, with a specialisation in infectious disease epidemiology at the Utrecht University. Additionally, she taught epidemiology to medical students, and supervised research internships.

Tess lives in Utrecht with Tom and her daughter Lila. They plan to move to Bangkok (Thailand) in January 2022 where she hopes to further continue her research on molecular infectious disease epidemiology.