Whole-genome characterisation of *Escherichia coli* isolates from patients with bacteraemia presenting with sepsis or septic shock in Spain: a multicentre cross-sectional study

Natalia Maldonado, Inmaculada López-Hernández*, Andrea García-Montaner, Luis Eduardo López-Cortés, Pedro María Martínez Pérez-Crespo, Pilar Retamar-Gentil, Adrián Sousa-Domínguez, Josune Goikoetxea, Ángeles Pulido-Navazo, Cristina Labayru-Echeverría, Clara Natera-Kindelán, Alfredo Jover-Sáenz, Alfonso del Arco-Jiménez, Carlos Armiñanzas-Castillo, Ana Isabel Aller, Jonathan Fernández-Suárez, Teresa Marrodán-Ciordia, Lucía Boix-Palop, Alejandro Smithson-Amat, José Mª Reguera-Iglesias, Fátima Galán-Sánchez, Alberto Bahamonde, Juan Manuel Sánchez Calvo, Isabel Gea-Lázaro, Inés Pérez-Camacho, Armando Reyes-Bertos, Berta Becerril-Carral, Jesús Rodríguez-Baño†, Álvaro Pascual†, on behalf of the Grupo PROBAC REIPI/GEIH-SEIMC/SAEI‡*

Summary

Background *Escherichia coli* is the most frequent cause of bloodstream infections (BSIs). About one-third of patients with BSIs due to *E coli* develop sepsis or shock. The objective of this study is to characterise the microbiological features of *E coli* blood isolates causing sepsis or septic shock to provide exploratory information for future diagnostic, preventive, or therapeutic interventions.

Methods *E coli* blood isolates from a multicentre cross-sectional study of patients older than 14 years presenting with sepsis or septic shock (according to the Third International Consensus Definitions for Sepsis and Septic Shock criteria) from hospitals in Spain between Oct 4, 2016, and Oct 15, 2017, were studied by whole-genome sequencing. Phylogroups, sequence types (STs), serotype, FimH types, antimicrobial resistance (AMR) genes, pathogenicity islands, and virulence factors were identified. Susceptibility testing was performed by broth microdilution. The main outcome of this study was the characterisation of the *E coli* blood isolates in terms of population structure by phylogroups, groups (group 1: phylogroups B2, F, and G; group 2: A, B1, and C; group 3: D), and STs and distribution by geographical location and bloodstream infection source. Other outcomes were virulence score and prevalence of virulence-associated genes, pathogenicity islands, AMR, and AMR-associated genes. Frequencies were compared using χ^2 or Fisher's exact tests, and continuous variables using the Mann-Whitney test, with Bonferroni correction for multiple comparisons.

Findings We analysed 224 isolates: 140 isolates (63%) were included in phylogenetic group 1, 52 (23%) in group 2, and 32 (14%) in group 3. 85 STs were identified, with four comprising 44% (n=98) of the isolates: ST131 (38 [17%]), ST73 (25 [11%]), ST69 (23 [10%]), and ST95 (12 [5%]). No significant differences in phylogroup or ST distribution were found according to geographical areas or source of bloodstream infection, except for ST95, which was more frequent in urinary tract infections than in other sources (11 [9%] of 116 *vs* 1 [1%] of 108, p=0·0045). Median virulence score was higher in group 1 (median 25·0 [IQR 20·5–29·0) than in group 2 (median 14·5 [9·0–20·0]; p<0·0001) and group 3 (median 21 [16·5–23·0]; p<0·0001); prevalence of several pathogenicity islands was higher in group 1. No significant differences were found between phylogenetic groups in proportions of resistance to antibiotics. ST73 had higher median virulence score (32 [IQR 29–35]) than the other predominant clones (median range 21–28). Some virulence genes and pathogenicity islands were significantly associated with each ST. ST131 isolates had higher prevalence of AMR and a higher proportion of AMR genes, notably $bla_{CTX\text{-}M\text{-}15}$ and $bla_{OXA\text{-}1}$.

Interpretation In this exploratory study, the population structure of *E coli* causing sepsis or shock was similar to previous studies that included all bacteraemic isolates. Virulence genes, pathogenicity islands, and AMR genes were not randomly distributed among phylogroups or STs. These results provide a comprehensive characterisation of invasive *E coli* isolates causing severe response syndrome. Future studies are required to determine the contribution of these microbiological factors to severe clinical presentation and worse outcomes in patients with *E coli* bloodstream infection.

Funding Instituto de Salud Carlos III.

Introduction

Extraintestinal pathogenic *Escherichia coli* are the most frequent cause of bloodstream infections (BSIs), mostly related to urinary tract infections (UTIs) and intra-abdominal infections.[1] The burden of disease and mortality associated with BSIs due to *E coli* is considerable, in

Lancet Microbe 2024; 5: e390–99

Published Online March 26, 2024
https://doi.org/10.1016/S2666-5247(23)00369-5

For the Spanish translation of the abstract see Online for appendix 1

*Contributed equally as first authors
†Contributed equally as senior authors
‡Members listed in appendix 2 (p 1)

Unidad Clínica de Enfermedades Infecciosas y Microbiología, Hospital Universitario Virgen Macarena, Departamentos de Medicina y Microbiología, Universidad de Sevilla, and Instituto de Biomedicina de Sevilla (IBiS)/CSIC, Sevilla, Spain (N Maldonado MSc, I López-Hernández PhD, A García-Montaner PhD, L E López-Cortes PhD, P Retamar-Gentil PhD, Prof J Rodríguez-Baño PhD, Prof Á Pascual PhD); **Centro de Investigación Biomédica en Red en Enfermedades Infecciosas (CIBERINFEC), Instituto de Salud Carlos III, Madrid, Spain** (I López-Hernández, L E López-Cortes, P Retamar-Gentil, Prof J Rodríguez-Baño, Prof Á Pascual, C Natera-Kindelán PhD, C Armiñanzas-Castillo MD); **Hospital Universitario Virgen de Valme, Sevilla, Spain** (P M M Pérez-Crespo MSc, A I Aller PhD); **Complejo Hospitalario Universitario de Vigo, Vigo, Spain** (A Sousa-Domínguez MD); **Hospital Universitario de Cruces, Baracaldo, Spain** (J Goikoetxea MD); **Hospital General de Granollers, Granollers, Spain** (Á Pulido-Navazo PhD);

Hospital Universitario de Burgos, Burgos, Spain (C Labayru-Echeverría PhD); **Hospital Universitario Reina Sofía, Córdoba, Spain** (C Natera-Kindelán); **Hospital Universitario Arnau de Vilanova, Lleida, Spain** (A Jover-Sáenz PhD); **Hospital Costa del Sol, Marbella, Spain** (A del Arco-Jiménez PhD); **Hospital Universitario de Marqués de Valdecilla, Santander, Spain** (C Armiñanzas-Castillo); **Hospital Universitario Central de Asturias, Oviedo, Spain** (J Fernández-Suárez PhD); **Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Spain** (J Fernández-Suárez); **Hospital Universitario de León, León, Spain** (T Marrodán-Ciordia PhD); **Hospital Universitario MútuaTerrassa, Terrassa, Spain** (L Boix-Palop MD); **Hospital de l'Esperit Sant, Santa Coloma de Gramanet, Spain** (A Smithson-Amat PhD); **Hospital Regional Universitario de Málaga, Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina (IBIMA Plataforma BIONAND), Málaga, Spain** (J Mª Reguera-Iglesias PhD); **Hospital Universitario Puerta del Mar, Cádiz, Spain** (F Galán-Sánchez PhD); **Hospital Universitario El Bierzo, Ponferrada, Spain** (A Bahamonde MD); **Hospital Universitario de Jerez, Instituto de Investigación e Innovación Biomédica de Cádiz (INiBICA), Jerez de la Frontera, Spain** (J M Sánchez Calvo PhD); **Complejo Hospitalario Ciudad de Jaén, Jaén, Spain** (I Gea-Lázaro MD); **Hospital Universitario Poniente-El Ejido, Almería, Spain** (I Pérez-Camacho PhD); **Hospital Universitario Torrecárdenas, Almería, Spain** (A Reyes-Bertos PhD); **Hospital Universitario Punta de Europa, Algeciras, Spain** (B Becerril-Carral PhD)

Correspondence to: Prof Jesús Rodríguez-Baño, Unidad Clínica de Enfermedades Infecciosas y Microbiología, Hospital Universitario Virgen Macarena, 41009 Sevilla, Spain
jesusrb@us.es

See Online for appendix 2

Research in context

Evidence before this study

Escherichia coli is the most frequent cause of bloodstream infections (BSIs); one-third of patients with *E coli* BSIs develop sepsis or shock. We searched PubMed, SCOPUS, and Web of Science, from database inception to Oct 31, 2023, using the terms "Escherichia coli" and ("bacteraemia" or "bloodstream infection" or "sepsis" or "septic shock") and ("virulence" or "resistance" or "pathogenicity island"), with no language restrictions, and found no previous studies providing a full characterisation of *E coli* isolates causing sepsis or septic shock using whole-genome sequencing. Previous studies in bacteraemic infections included all *E coli* blood isolates without disaggregating by clinical severity and were heterogeneous in terms of typing methods and virulence genes studied.

Added value of this study

To the best of our knowledge, this is the first multicentre study in which a high number of *E coli* blood isolates causing sepsis or septic shock were characterised by whole-genome sequencing. This exploratory study provided a comprehensive insight into the relationship between phylogenetic or clonal background and virulence, pathogenicity islands, and antibiotic resistance.

Implications of all the available evidence

The analysis of the microbiological determinants of *E coli* can improve the understanding of the pathogenesis of *E coli* bloodstream infection and of the contribution of molecular factors to the severe clinical presentation and worse outcomes in patients with bacteraemia. These insights represent a first step towards the identification of potential targets for the development of preventive or therapeutic strategies.

high-income countries and in older populations.[1] About one-third of patients with *E coli*-causing BSIs develop a disregulated host response (ie, sepsis or septic shock), which is a main driver for mortality.[2] The ability of *E coli* to colonise the human bowel and eventually cause an extraintestinal infection has been associated with the expression of a large arsenal of virulence factors, which are usually encoded on pathogenicity islands, plasmids, and other mobile genetic elements.[3]

E coli isolates are classically classified in nine phylogroups (A, B1, B2, C, D, E, F, G, and H) according to their genetic background. Typically, the so-called commensal strains, belonging primarily to phylogenetic groups A and B1, possess fewer virulence determinants than the so-called extraintestinal pathogenic B2 and D phylogenetic groups.[4] Some clonal lineages that have been predominant in invasive isolates tend to show a correlation with virulence profiles.[5]

To our knowledge, there are no studies providing specific data for isolates causing a disregulated host response. The analysis of the microbiological determinants of *E coli* isolated from patients presenting with sepsis and septic shock might contribute to a better understanding of the pathogenesis of these infections and help to identify potential targets for the development of diagnostic, preventive, or therapeutic strategies. Therefore, the objective of this exploratory study was to characterise, in detail, the microbiological factors of *E coli* blood isolates causing sepsis or septic shock using whole-genome sequencing, to provide descriptive data that would eventually be useful to raise hypotheses and potential targets for interventions.

Methods

Study design, participants, and bacterial isolates

The PROBAC-Ec is a cross-sectional study nested in a prospective multicentre cohort study (PROBAC) done in 26 hospitals in Spain (appendix 2 p 2), including all episodes of microbiologically confirmed BSIs in patients older than 14 years, between Oct 4, 2016, and Oct 15, 2017. The methodology, including clinical and demographic information collected from patients, and some results of the project were previously reported.[6]

In the PROBAC study, blood cultures were used to establish the aetiology of BSIs. For this study, we planned to include the blood isolates from the first 225 patients with monomicrobial BSIs due to *E coli* from the PROBAC cohort presenting as sepsis or septic shock at the time of sampling of the blood culture. Sepsis and septic shock were defined according to the Third International Consensus Definitions for Sepsis and Septic Shock.[7] The PROBAC study was approved by the ethics committees of the participating centres, which waived the need to obtain informed consent due to the observational design.

Procedures

Patient data were collected by trained infectious diseases physicians or microbiologists at each participating site from the medical records. The following variables were included: demographics; geographical location; chronic underlying diseases; exposure to invasive devices, procedures, or antimicrobials during the previous month; type of acquisition; source and aetiology of BSIs; antibiotic empirical treatment; and 30-day mortality. Quality of data was monitored centrally for completeness and consistency of key variables.

Standard microbiological procedures were used for blood culture processing;[8] *E coli* isolates (the first per patient) were stored at –80°C in Tryptose soy broth containing 15% (v/v) glycerol and sent to the Microbiology Laboratory of University Hospital Virgen Macarena in Seville, Spain, for further characterisation. Isolate identification was confirmed by MALDI-TOF mass spectrometry (Bruker Daltonics, Billerica, MA, USA). Susceptibility testing to fosfomycin, amoxicillin–clavulanic acid, piperacillin–tazobactam, cefoxitin, cefotaxime, ceftriaxone, cefepime, cefuroxime, ertapenem,

imipenem, meropenem, nalidixic acid, ciprofloxacin, amikacin, gentamicin, and trimethoprim–sulfamethoxazole was done by EUCAST broth microdilution method using *E coli* ATCC 25922 and 35218 as quality control strains; EUCAST version 10.0 clinical breakpoints for susceptibility were used.

Fully automated DNA extraction from *E coli* cultures was done with Dneasy Blood & Tissue Kit (Qiagen, Hilden, Germany) in QIAcube instrument (Qiagen), following the manufacturer's recommendations. Pure DNA was quantified using a Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) with a minimum DNA concentration of 20 μg/μL. The sample library was prepared with the Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA). DNA paired-end sequencing was performed in the MiSeq System (2 × 300 bp) with a MiSeq Reagent Kit V3 (600 cycles; Illumina). De novo assembly was performed using the CLC Genomics Workbench version 10 (Qiagen).

Phylogenetic group was assigned using the Clermont Scheme in the Clermont Typing platform. Multilocus sequence type, serotype, FimH type, virulence factors, and antimicrobial resistance (AMR) genes were determined using MLST 2.0, SerotypeFinder 2.0, FimTyper 1.0, VirulenceFinder 2.0, and ResFinder 3.2 databases, respectively. These services are available on the Center for Genomic Epidemiology website. ST131 isolates were classified into three major clades (A, B, and C) and subclades (C0, C1, and C2).[9,10] The assembled sequences were annotated with RAST server version 2.0 to determine the presence of the alleles II and III of the *papG* gene.

The OrthoFinder software (version 2.5.2; with options -a 14 -M msa) was used to analyse the proteomes of *E coli* isolates and to obtain a concatenated alignment of 1133 single-copy pan-orthologous genes that represented the core genome of the dataset. This alignment was used to calculate a phylogenetic tree with IQ-Tree (version 2.0.3). Branch stability was assessed by 1000 ultra-fast bootstrap pseudoreplicates. The tree visualisation was generated with the online tool Iroki Phylogenetic Tree Viewer.

The genome assemblies were screened to identify pathogenicity islands. For local use, a database was built in makeblastdb (integrated in BLAST+ with the option -dbtype nucl) including the complete pathogenicity islands from the PAI-DB database (accessed on April 1, 2022), whose nucleotide sequences were obtained from the National Center for Biotechnology Information. BLASTn searches were done between the *E coli* isolates and the local pathogenicity islands database, with an e-value cutoff of 1×10^{-6}. For each isolate, a pathogenicity island was considered as present if more than 80% of the genes of the reference island were identified.

Outcomes

The main outcome of this study was the characterisation of the *E coli* blood isolates collection from patients presenting with sepsis or septic shock in terms of population structure by phylogroups, groups, and sequence types (STs) and of distribution by geographical location and BSI source. Other outcomes analysed in the overall collection and by phylogroups and STs were: virulence score; prevalence of virulence-associated genes (including subtyping of isolates by *fimH* alleles); prevalence of pathogenicity islands; prevalence of AMR determined by susceptibility testing; and prevalence of AMR-associated genes.

For **EUCAST** see https://www.eucast.org

Statistical analysis

The sample size was based on (1) the funds available; (2) the distribution of *E coli* clones in previous studies on *E coli* BSIs,[2,11] such that ten or more representatives from the four predominant STs would be included; and (3) the inclusion of a reasonably representative sample (ie, 15–20%) of the 1211 patients with monomicrobial *E coli* BSIs presenting with sepsis or shock from the PROBAC cohort.

The virulence score was calculated for each isolate as a total number of genes detected, as well as the median values and IQR for each phylogroup and ST. For statistical comparisons of genotypic results, phylogroups were merged according to their phylogenetic relatedness into three groups: group 1 (phylogroups B2, F, and G); group 2 (phylogroups A, B1, and C), and group 3 (phylogroup D). Percentages were compared using χ^2 or Fisher's exact tests, as appropriate, and continuous variables using the Mann-Whitney test. Bonferroni correction for multiple comparisons was used for assessing statistical significance. Because of the exploratory nature of the study, estimates of associations were not calculated. SPSS version 18 was used for statistical analyses.

For the **CLC Genomics Workbench** see http://www.clcbio.com/products

For the **Clermont Scheme in the Clermont Typing platform** see http://clermontyping.iame-research.center

For the **Center for Genomic Epidemiology** see https://www.genomicepidemiology.org/

For the **RAST server** see www.nmpdr.org

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

This study planned to include the blood isolates of the first 225 patients with monomicrobial BSIs due to *E coli* presenting with sepsis or septic shock recruited for the PROBAC cohort. Since one isolate could not be retrieved, the final number of included isolates was 224. These patients were recruited from 22 of the 26 participating hospitals (appendix 2 p 2).

For the **Iroki Phylogenetic Tree Viewer** see https://www.iroki.net

For the **PAI-DB database** see www.paidb.re.kr

The median age of the patients was 75 years (IQR 65–84, range 27–97) and 135 (60%) of 224 patients were men. The urinary tract was the most frequent source of bacteraemia (116 [52%]), followed by intra-abdominal sources (73 [33%]) and respiratory tract sources (15 [7%]; figure). Most BSIs were community-acquired (121 [54%]; figure). The most frequent comorbidities were neoplasia (60 [27%]), diabetes (58 [26%]), chronic pulmonary disease (32 [14%]), and liver disease (28 [13%]). 290 (93%) patients received active empirical antibiotic therapy based on local antibiotic susceptibility results. 30-day mortality was 31% (69 patients; appendix 2 p 3).

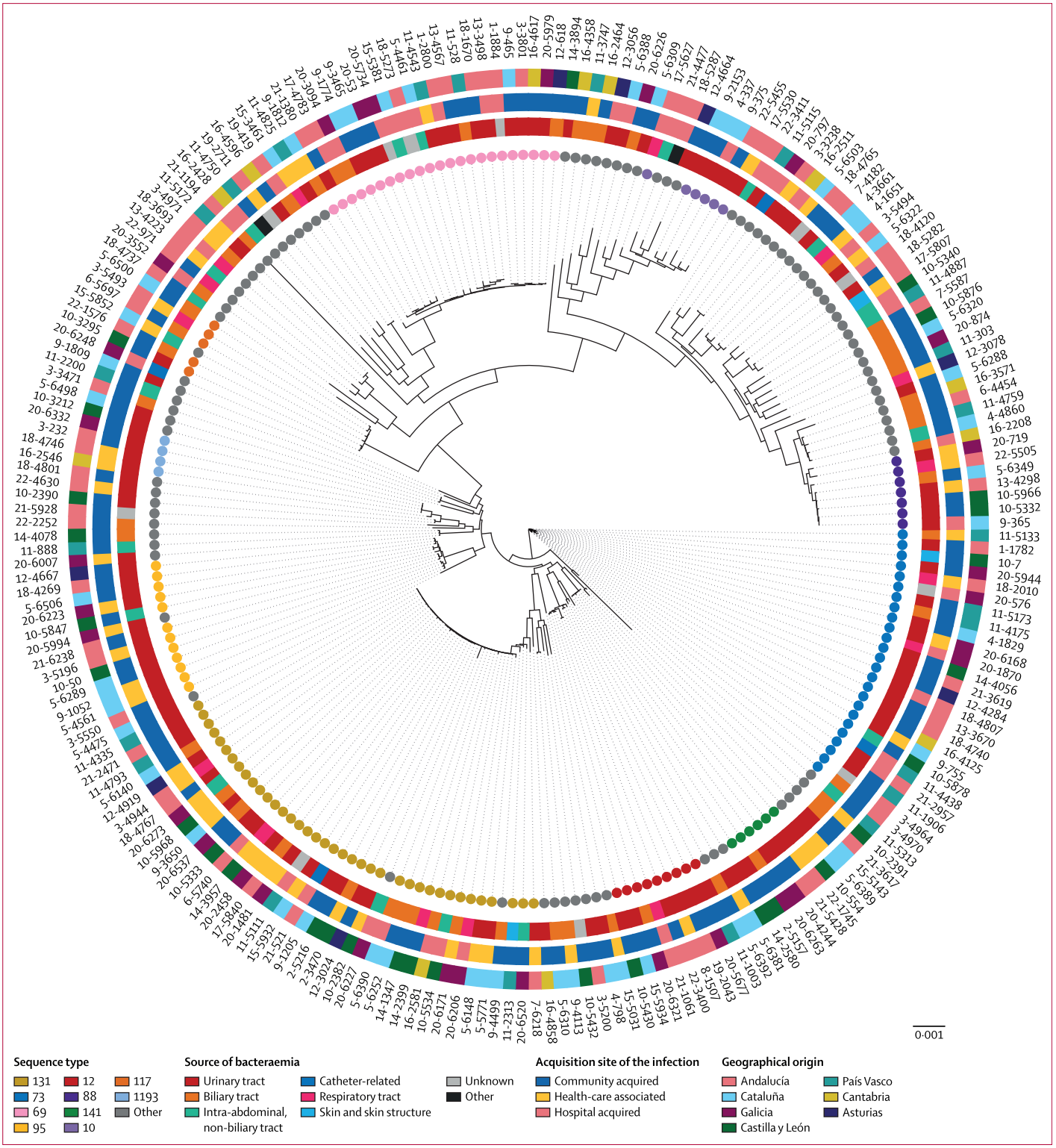

Figure: **Phylogenetic tree reconstruction**

Phylogenetic tree reconstruction based on the concatenated alignment of 1133 single-copy pan-orthologous genes that represented the core genome among the 224 *Escherichia coli* isolates from patients with bacteraemia and presentation as sepsis or septic shock. From the centre, the inner dotted ring shows sequence types, the second ring shows the source of bacteraemia, the third ring shows the acquisition site, and the outer ring shows the geographical origin.

Phylogroups B2 (127 [57%] of 224 isolates) and D (32 [14%] isolates) were predominant. 22 (10%) isolates belonged to B1, 17 (8%) isolates to A, 13 (6%) to C, 7 (3%) to F, and 6 (3%) to G (panel; figure). According to the predefined phylogenetic groups in this study, 140 isolates (63%) were included in group 1 (B2, F, and G), 52 (23%) in group 2 (A, B1, and C), and 32 (14%) in group 3 (D). Isolates were classified into 85 STs; four STs comprised 44% (n=98) of all isolates: ST131 (38 [17%]), ST73 (25 [11%]), ST69 (23 [10%]), and ST95 (12 [5%]; panel, figure). Other STs with frequencies higher than three isolates were ST12 (9 [4%]), ST88 (7 [3%]), ST141 (6 [3%]), ST10 (6 [3%]), ST117 (5 [2%]), and ST1193 (4 [2%]). No significant differences in phylogroup or ST distribution were found according to geographical areas (appendix 2 p 4) or source of BSI (appendix 2 p 5), except for ST95, which was more frequent in UTIs than in other sources (11 [9%] of 116 *vs* 1 [1%] of 108, p=0·0045).

Virulence scores varied extensively among the *E coli* isolates, ranging from 4·0 to 37·0 (median 22·0 [IQR 17·0–27·0]). Isolates from group 1 phylogroups had a higher virulence score (median 25·0 [20·5–29·0]) than isolates from group 2 (median 14·5 [9·0–20·0]; p<0·0001), and group 3 (median 21·0 [16·5–23·0]; p<0·0001). Among specific phylogroups, B2 isolates had significantly higher virulence scores (median 25·0 [20·5–29·0]) than the other major phylogroups (A, B1, C, and D), with differences in virulence profile (table 1; appendix 3). Among the main clones, virulence score was statistically higher for ST73 (median 32·0 [29·0–35·0]) than for ST131 (median 22·0 [19·0–23·0]; p<0·0001), ST69 (median 21·0 [16·5–23·5]; p<0·0001) and ST95 (median 28 [25·5–29·5]; p=0·0026; table 1).

Overall, the most frequently detected genes coding for putative virulence factors were type 1 fimbrin D-mannose specific adhesion (*fimH*; 224 [100%] isolates), tellurium ion resistance protein (*terC*; 223 [100%]), outer membrane protease (*ompT*; 193 [86%]), increased serum survival (*iss*; 192 [86%]), and iron transport protein (*sitA*; 192 [86%]). Isolates from group 1 presented a significantly higher frequency of specific virulence genes than did isolates from groups 2 and 3, in particular for fimbrial protein *yfcV*, siderophore receptor *fyuA*, iron regulatory *irp2*, and other toxin-related genes (*usp, vat, clbB, cnf1,* and *pic*; appendix 2 p 6).

Some virulence genes were significantly associated with certain STs. Specifically, genes related to adhesion or invasion, such as *iha* and *papA*-F43, were associated with ST131, *lpfA* and *air* with ST69, *focG* with ST73, and *papA*-F11 with ST95. Genes involved in the production of bacteriocins were found in most ST73 isolates (*cea*, *mchF, mcmA*, *mchB*, and *mchC*) and ST95 isolates (*mchF, cia*, and *cvaC*). Finally, genes related to K1 capsule synthesis were detected in all ST95 isolates (*neuC* and *kpsM*II-K1), and some genes encoding toxins were more frequently detected in ST73 (*cnf1*, *clbB* and *pic*), ST131 (*sat*), and ST95 (*vat* and *hlyF*; table 2; appendix 2 p 7; appendix 3).

Subtyping based on FimH revealed 55 subtypes, and two isolates were not typable. FimH30 was most frequently typed among ST131 isolates (34 [89%] of 38), FimH27 among ST69 isolates (22 [96%] of 23), FimH10 among ST73 isolates (13 [52%] of 25), and FimH27 (4 [33%] of 12) and FimH41 (3 [25%] of 12) among ST95 isolates (appendix 2 p 7).

In almost all *E coli* isolates (210 [94%] of 224) at least one pathogenicity island was detected; the most prevalent pathogenicity islands were $IV_{APEC-O1}$ (171 [76%] isolates), AGI-3 (167 [75%]), and I_{AL862} (158 [71%]). A significantly higher prevalence of pathogenicity islands was found in group 1 than in groups 2 and 3 (in particular, pathogenicity islands I_{536}, II_{536}, III_{536}, V_{536}, II_{CFT073}, and *espC*); conversely, pathogenicity islands $II_{APEC-01}$, AGI-1, and LEE-03 were significantly more frequent in group 3 than in groups 1 and 2 (appendix 2 p 8). When STs were considered, pathogenicity islands I_{536}, III_{536}, and LEE-07 were significantly more frequent in ST73 than in the other three main STs, whereas LEE-03 was found exclusively in ST69 (table 3; appendix 3).

Overall, 103 (46%) of 224 isolates were resistant to ciprofloxacin, 94 (42%) to trimethoprim–sulfamethoxazole, 80 (36%) to amoxicillin–clavulanic acid, 45 (20%) to cefuroxime, 41 (18%) to gentamicin, 38 (17%) to ceftazidime, 35 (16%) to cefotaxime, 27 (12%) to cefepime, 20 (9%) to piperacillin–tazobactam, 14 (6%) to fosfomycin, and 11 (5%) to amikacin. None of the isolates was resistant to carbapenems (appendix 2 p 9). The prevalence of AMR was similar between the groups, except for gentamicin, which was significantly higher in group 3 (phylogroup D) than in group 1 (appendix 2 p 9). Among the main clones, ST131 isolates had significantly higher prevalence of resistance to ciprofloxacin, nalidixic acid, cefuroxime, cefotaxime–ceftriaxone, and cefepime than ST73, ST69, and ST95 (appendix 2 p 10).

Panel: **Number of isolates by sequence type (ST) within each phylogroup in our collection of 224 *Escherichia coli* isolates**

- **Phylogroup A:** ST10 (n=6), ST46 (n=1), ST48 (n=1), ST167 (n=1), ST329 (n=1), ST409 (n=1), ST744 (n=1), ST1488 (n=1), ST2040 (n=1), ST2624 (n=1), ST2930 (n=1), ST3014 (n=1)
- **Phylogroup B1:** ST58 (n=2), ST224 (n=2), ST453 (n=2), ST109 (n=1), ST155 (n=1), ST162 (n=1), ST345 (n=1), ST348 (n=1), ST359 (n=1), ST380 (n=1), ST448 (n=1), ST533 (n=1), ST539 (n=1), ST641 (n=1), ST847 (n=1), ST949 (n=1), ST977 (n=1), ST1196 (n=1), ST1431 (n=1)
- **Phylogroup B2:** ST131 (n=38), ST73 (n=25), ST95 (n=12), ST12 (n=9), ST141 (n=6), ST1193 (n=4), ST127 (n=3), ST978 (n=3), ST355 (n=2), ST372 (n=2), ST537 (n=2), ST80 (n=1), ST126 (n=1), ST135 (n=1), ST144 (n=1), ST421 (n=1), ST429 (n=1), ST538 (n=1), ST567 (n=1), ST569 (n=1), ST636 (n=1), ST676 (n=1), ST681 (n=1), ST968 (n=1), ST1057 (n=1), ST1385 (n=1), ST1851 (n=1), ST2015 (n=1), ST2279 (n=1), ST2554 (n=1), ST4110 (n=1)
- **Phylogroup C:** ST88 (n=7), ST23 (n=2), ST90 (n=1), ST369 (n=1), ST410 (n=1), ST2230 (n=1)
- **Phylogroup D:** ST69 (n=23), ST38 (n=2), ST70 (n=1), ST362 (n=1), ST393 (n=1), ST1011 (n=1), ST1394 (n=1), ST1406 (n=1), ST1882 (n=1)
- **Phylogroup F:** ST624 (n=2), ST648(n=2), ST59 (n=1), ST354 (n=1), ST6469 (n=1)
- **Phylogroup G:** ST117 (n=5), ST5993 (n=1)

See Online for appendix 3

	Virulence score	Virulence profile*
A	10·0 (8·0–15·0), n=17	***fimH, terC***, *iss*, ***gad***, *traT*
B1	15·0 (10·0–19·0), n=22	***fimH, terC***, *iss*, ***gad***, *traT*, *sitA*, ***lpfA***
B2	25·0 (20·5–29·0), n=127†	***fimH, terC***, *iss*, *gad*, ***sitA***, ***ompT***, ***fyuA***, ***irp2***, ***chuA***, ***kpsE***, ***yfcV***, ***usp***
C	20·0 (15·0–23·0), n=13‡	***fimH, terC***, ***iss***, *sitA*, *ompT*, *fyuA*, *irp2*, *traT*, ***lpfA***, *iutA*, *iucC*
D, group 3	21·0 (16·5–23·0), n=32§	***fimH, terC***, ***iss***, *sitA*, *ompT*, *lpfA*, ***gad***, ***chuA***, ***kpsE***, ***eilA***, *air*
F	23·0 (20·0–27·5), n=7¶	***fimH, terC***, ***sitA***, ***lpfA***, ***gad***, ***chuA***, ***iutA***, ***iucC***, ***kpsE***, ***eilA***, *traT*, ***yfcV***, *ompT*, *air*, *fyuA*, *irp2*, *etsC*
G	23·5 (21·0–27·0), n=6‖	***fimH, terC***, ***sitA***, ***lpfA***, ***chuA***, ***iutA***, ***iucC***, ***traT***, ***ompT***, ***iss***, ***iroN***, ***hylF***, *etsC*, ***ireA***, *vat*, ***pic***, *mchF*, *fyuA*, *irp2*, *cma*
B2+F+G, group 1	25·0 (20·5–29·0), n=140**	***fimH, terC***, *iss*, *gad*, ***sitA***, ***ompT***, ***fyuA***, ***irp2***, ***chuA***, ***kpsE***, ***yfcV***, ***usp***, ***traT***
A+B1+C, group 2	14·5 (9·0–20·0), n=52	***fimH, terC***, *iss*, *gad*, *lpfA*, *traT*
ST131	22·0 (19·0–23·0), n=38	*fimH30*, ***chuA***, ***terC***, ***ompT***, ***fyuA***, ***irp2***, ***usp***, ***iss***, ***gad***, ***kpsE***, ***iucC***, ***iutA***, ***yfcV***, ***sitA***, ***iha***, *sat*, *papA_F43*, *traT*, *kpsMII_K5*
ST73	32·0 (29·0–35·0), n=25††	***chuA***, ***terC***, ***ompT***, ***fyuA***, ***irp2***, ***usp***, ***kpsE***, ***yfcV***, ***pic***, ***iss***, ***sitA***, ***clbB***, ***iroN***, ***mchF***, ***vat***, ***cnf1***, ***mchB***, ***mchC***, ***mcmA***, *cea*, *hra*, *focG*, *papC*
ST69	21·0 (16·5–23·5), n=23	***fimH27***, ***chuA***, ***terC***, ***iss***, ***gad***, ***eilA***, ***lpfA***, ***ompT***, ***kpsE***, ***sitA***, *fyuA*, *irp2*, *air*, *iucC*, *iutA*, *traT*
ST95	28·0 (25·5–29·5), n=12‡‡	***chuA***, ***terC***, ***ompT***, ***fyuA***, ***irp2***, ***usp***, ***iss***, ***sitA***, ***papGII***, ***neuC***, ***ireA***, ***kpsMII_K1***, ***papA_F11***, ***gad***, ***kpsE***, ***yfcV***, ***traT***, ***papC***, *iucC*, *iutA*, *iroN*, *mchF*, *etsC*, *hylF*, *cvaC*, *vat*, *cia*

Data are median (IQR), unless otherwise stated. For all other comparisons, p≥0·05. p values for significance are in appendix 3. *For profiles determination, the genes present in more than 66% of isolates within each phylogroup and in the four main sequence types were included; in bold, genes detected in more than 90% of the isolates within each phylogroup or sequence type. Statistically higher median virulence scores were found for: †B2 *vs* A, B2 *vs* B1, B2 *vs* C, and B2 *vs* D; ‡C *vs* A; §D *vs* A, D *vs* B1, and group 3 *vs* group 2; ¶F *vs* A and F *vs* B1; ‖G *vs* A and G *vs* B1; **Group 1 *vs* group 2 and group 1 *vs* group 3; ††ST73 *vs* ST131, ST73 *vs* ST69, and ST73 *vs* ST95; and ‡‡ST95 *vs* ST131 and ST95 *vs* ST9.

Table 1: **Virulence scores and virulence profiles of *Escherichia coli* isolates from patients with bacteraemia and presentation as sepsis or septic shock**

When screening for acquired genes mediating AMR, the bla_{OXA-48} gene was detected in one isolate and genes encoding extended-spectrum beta-lactamases (ESBLs) were detected in 27 (12%) isolates. $bla_{CTX\text{-}M\text{-}15}$ (17 [8%] isolates) was the most frequent ESBL-encoding gene, followed by $bla_{CTX\text{-}M\text{-}14}$ (4 [2%]), $bla_{CTX\text{-}M\text{-}1}$ (3 [1%]), $bla_{CTX\text{-}M\text{-}32}$ (1 [<1%]), and $bla_{SHV\text{-}12}$ (2 [1%]). 23 (10%) isolates harboured $bla_{OXA\text{-}1}$. Aminoglycoside phosphotransferase genes *aph(3″)-Ib* and *aph(6)-Id* were detected in 60 (27%) and 61 (27%) isolates, respectively. No notable differences were found in relevant acquired AMR genes between the different phylogenetic groups (appendix 2 p 11). In the analysis by STs, ST131 had a significantly higher proportion (>92% of isolates) of point mutations in DNA gyrase (*gyrA*) and topoisomerase IV (*parC* and *parE*) genes, associated with fluoroquinolone resistance, compared with the other three main clones. Also, in ST131, higher prevalences of $bla_{CTX\text{-}M\text{-}15}$ (16 [42%] of 38), $bla_{OXA\text{-}1}$ (15 [39%] of 38), and fluoroquinolone acetylating aminoglycoside-(6)-N-acetyltransferase (*aac[6′]-Ib-cr*; 15 [39%] of 38) genes were detected than in ST73, ST69, and ST95 isolates, in which these genes were not present (significant difference for ST131 *vs* ST73 and ST69 after Bonferroni correction; appendix 2 p 12).

One (3%) of the 38 ST131 isolates belonged to clade A (serotype O16:H5, *fimH*41 allele) and none to clade B (serotype O25:H4, *fimH*22 allele). Regarding the distribution by subclades, one (3%) isolate belonged to subclade C0, also known as H30 (serotype O25:H4, *fimH*30 allele, and susceptible to fluoroquinolones); 19 (50%) isolates belonged to subclade C1 (H30-R) characterised by the acquisition of fluoroquinolone resistance, and 14 (37%) isolates were classified into subclade C2 (H30-Rx), all of which harboured $bla_{CTX\text{-}M\text{-}15}$.[10]

Discussion

This study describes the microbiological features of bacteraemic *E coli* isolates causing a dysregulated response (ie, sepsis or septic shock) in a multicentre study in Spain. Overall, most of the isolates belonged to phylogroups B2 and D; isolates showed a wide clonality, but 44% of BSIs were caused by the four globally dominant clones ST131, ST73, ST95, and ST69. In general, ST131 isolates had a higher prevalence of AMR, whereas ST73 and ST95 had a higher content of virulence factors.

To the best of our knowledge, this is the first multicentre study in which a high number of *E coli* isolates causing BSIs and presenting with sepsis or septic shock were characterised by whole-genome sequencing. Previous studies included all bacteraemic isolates and not only those causing sepsis or shock[2,5,11,12–15] (approximately a third of cases of bacteraemia)[2] but they provide a useful context for our results.

The proportion of isolates belonging to phylogroups B2 and D (71%) was similar to that found in previous studies of bacteraemic isolates, in which these classic extraintestinal phylogroups ranged between 67% and 80%.[2,12–14] In addition, phylogroup B2 isolates showed more virulence genes than the other major phylogroups, also in line with previous data.[13,15] Isolates from phylogroups F and G presented virulence scores above the overall median value and closer to virulence scores of phylogroup B2 than of other phylogroups. Few studies have previously characterised virulence in these two less frequent phylogroups. Phylogroup F isolates were shown to have high prevalence of extraintestinal pathogenic *E coli* virulence genes in chicken colibacillosis;[16] and phylogroup G was described by Clermont and colleagues[17] as an intermediate lineage between phylogroups F and B2. The most common clone within phylogroup G is ST117, a poultry-associated lineage,[17] which in our collection represented five of six isolates, and had many traits typically associated with extraintestinal virulence.

Previous studies describing the population structure of *E coli* in BSI collections revealed a wide clonal distribution of isolates; however, four lineages (ST131, ST73, ST95, and ST69) account for 38–61% of isolates.[2,5,11–13,18,19] This was also the case in our study, in which 44% of isolates belong to these so-called pandemic extraintestinal pathogenic *E coli* lineages. The fact that these predominant STs were not over-represented in patients with sepsis or septic shock suggests that clonality by itself might not be determinant in the risk of inducing a severe systemic response. Regarding the source of infection, in our study, ST95 was more frequent in BSIs from UTIs than those from other sources, a finding that warrants further investigation. Previous studies comparing

	ST131 (n=38)	ST73 (n=25)	ST69 (n=23)	ST95 (n=12)
Adhesins–invasins				
fimH30	34 (89%)†,§,‖	2 (8%)	0	3 (25%)
fimH27	0	1 (4%)	22 (96%)‡,††,¶¶	4 (33%)
fimH10	0	13 (52%)*,**	0	0
yfcV	37 (97%)§	25 (100%)**	0	11 (92%)‖‖
hra	16 (42%)	20 (80%)**,‡‡	4 (17%)	0
iha	36 (95%)§,‖	15 (60%)‡‡	12 (52%)	0
lpfA	0	0	23 (100)‡,††,¶¶	0
papA-F11	1 (3%)	0	0	12 (100%)¶,§§,‖‖
papA-F14	0	7 (28%)*	1 (4%)	0
papA-F43	34 (90%)§,‖	13 (52%)	3 (13%)	0
papA-fSiA-F16	0	0	8 (35%)‡	0
papG allele II	14 (37%)	14 (56%)	10 (44%)	12 (100)¶
sfaD	0	13 (52%)*,**	0	0
air	0	0	17 (74%)‡,††,¶¶	0
focC	0	7 (28%)*	0	0
focG	0	20 (80%)*,**,‡‡	0	0
Bacteriocins				
cea	4 (11%)	22 (88%)*,**	3 (13%)	4 (33%)
cia	3 (8%)	0	2 (9%)	9 (75%)¶,§§,‖‖
mchF	3 (8%)	23 (92%)*,**	4 (17%)	10 (83%)¶,‖‖
mcmA	1 (3%)	23 (92%)*,**,‡‡	0	0
mchB	1 (3%)	23 (92%)*,**,‡‡	0	0
cvaC	2 (5%)	0	4 (17%)	10 (83%)¶,§§,‖‖
mchC	1 (3%)	23 (92%)*,**,‡‡	0	0
Iron uptake				
iucC	37 (97%)†	14 (56%)	16 (70%)	10 (83%)
iutA	37 (97%)†	14 (56%)	16 (70%)	10 (83%)
iroN	3 (8%)	23 (92%)*,**	5 (22%)	10 (83%)¶,‖‖
ireA	1 (3%)	12 (48%)*	2 (9%)	12 (100%)¶,‖‖
Protectins				
gad	37 (97%)†	9 (36%)	23 (100%)††	11 (92%)
Capsule synthesis				
kpsMII-K5	26 (68%)§,‖	8 (32%)	1 (4%)	0
neuC	1 (3%)	0	3 (13%)	12 (100%)¶,§§,‖‖
kpsMII-K1	1 (3%)	0	2 (9%)	12 (100%)¶,§§,‖‖
Toxins				
usp	38 (100%)§	25 (100%)**	0	12 (100%)‖‖
vat	1 (3%)	23 (92%)*,**	0	10 (83%)¶,‖‖
sat	34 (90%)§,‖	14 (56%)	11 (48%)	0
hylF	3 (8%)	0	4 (17%)	10 (83%)¶,§§,‖‖
cnf1	9 (24%)	23 (92%)*,**,‡‡	0	0
clbB	0	24 (96%)*,**,‡‡	0	4 (33%)
pic	0	25 (100%)*,**,‡‡	0	0
Others				
etsC	3 (8%)	0	4 (17%)	10 (83%)¶,§§,‖‖
eilA	0	0	23 (100%)‡,††,¶¶	0
tcpC	0	15 (60%)*,**	0	3 (25%)

Data are number of isolates (%). Only genes showing significant differences between at least two of the four predominant STs are presented. Data for all virulence genes detected overall and disaggregated for the main STs are shown in appendix 2 (p 7). See appendix 3 for p values calculated by Fisher's exact test or χ^2 test. All other comparisons p≥0·0005. ST=sequence type. According to the Bonferroni correction, statistically significant higher proportions were found for: *ST73 *vs* ST131; †ST131 *vs* ST73; ‡ST69 *vs* ST131; §ST131 *vs* ST69; ¶ST95 *vs* ST131; ‖ST131 *vs* ST95; **ST73 *vs* ST69; ††ST69 *vs* ST73; ‡‡ST73 *vs* ST95; §§ST95 *vs* ST73; ¶¶ST69 *vs* ST95; ‖‖ST95 *vs* ST69.

Table 2: **Virulence genes according to the four main sequence types of *Escherichia coli* in patients with bacteraemia and presentation as sepsis or septic shock**

	ST131 (n=38)	ST73 (n=25)	ST69 (n=23)	ST95 (n=12)
PAI I $_{APEC-O1}$	22 (58%)	24 (96%)*,‖	8 (35%)	12 (100%)¶¶
PAI II $_{APEC-O1}$	0	0	22 (96%)†,**	12 (100%)§,‡‡
PAI III $_{APEC-O1}$	0	0	0	3 (25%)
PAI IV $_{APEC-O1}$	38 (100%)	25 (100%)	17 (74%)	11 (92%)
AGI-3	37 (97%)	25 (100%)	21 (91%)	12 (100%)
PAI I $_{AL862}$	36 (95%)	25 (100%)	19 (83%)	11 (92%)
espC PAI	38 (100%)‡	25 (100%)‖	0	12 (100%)¶¶
OI-57	35 (92%)‡,¶	15 (60%)	6 (26%)	1 (8%)
AGI-1	0	0	22 (96%)†,**	12 (100%)§,‡‡
ETT2	18 (47%)	14 (56%)††	10 (44%)	0
HPI	19 (50%)¶	16 (64%)††	6 (26%)	0
PAI I_{536}	12 (32%)	21 (84%)*,‖,††	1 (4%)	1 (8%)
PAI II_{536}	15 (40%)	24 (96%)*,‖	3 (25%)	7 (58%)
PAI III_{536}	16 (42%)	25 (100%)*,‖,††	4 (17%)	4 (33%)
PAI V $_{536}$	28 (74%)	25 (100%)‖	9 (39%)	10 (83%)
PAI I_{CFT073}	12 (32%)	15 (60%)††	5 (22%)	0
PAI II_{CFT073}	17 (45%)	25 (100%)*,‖	3 (13%)	8 (67%)
TAI	5 (13%)	13 (52%)*,‖	1 (4%)	0
LEE-01	10 (26%)	0	8 (35%)**	0
LEE-03	0	0	23 (100%)†,**,§§	0
LEE-07	6 (16%)	14 (56%)*,‖,††	1 (4%)	0

Data are number of isolates (%). See appendix 3 for p values by Fisher's exact test or χ^2 test. All other comparisons p≥0·0018. According to the Bonferroni correction, statistically higher proportions were found for: *ST73 vs ST131; †ST69 vs ST131; ‡ST131 vs ST69; §ST95 vs ST131; ¶ST131 vs ST95; ‖ST73 vs ST69; **ST69 vs ST73; ††ST73 vs ST95; ‡‡ST95 vs ST73; §§ST69 vs ST95; ¶¶ST95 vs ST69.

Table 3: **Distribution of pathogenicity islands in to the four main sequence types of *Escherichia coli* in patients with bacteraemia and presentation with sepsis or septic shock**

urine and blood *E coli* isolates could not find any specific ST associated with bacteraemia, which suggests that some particular bacterial virulence mechanisms and host predisposing factors probably are more important than clonality in causing BSIs.[20,21]

This study showed a high prevalence of virulence genes in *E coli* isolates, with a median virulence score of 22; this was higher than the virulence score found in studies including all blood isolates (range 15–19) using whole-genome sequencing data.[20,22] The most prevalent virulence genes found in our study were *terC*, coding for one of the key proteins in tellurite resistance;[23] *ompT*, a housekeeping protease that degrades peptides released by the host's immune system, and in particular the endogenous tissue factor anticoagulant inhibitor pathway, contributing to the haemostatic imbalance associated with sepsis;[24] *iss* gene, a factor associated with complement resistance,[25] and the *sitA*, *irp2*, and *fyuA* genes, responsible for uptake of iron,[3] an essential element for supporting growth and bacterial development. Altogether, these results suggest that iron acquisition and immune evasion are important for bacteraemic *E coli* isolates.[21]

Although each isolate exhibits a particular combination of virulence factors, probably due to the plasticity of the *E coli* genome, specific genes were significantly associated with specific STs; for example, genes involved in bacteriocin production were found in most ST73 isolates (*cea, mchB, mchC, mchF* and *mcmA*) and ST95 isolates (*cia, cvaC*), genes related to K1 capsule synthesis (*neuC, kps*MII-K1) and the *ireA* siderophore receptor were detected in all ST95 isolates, and specific toxin-encoding genes were significantly more frequent in each clone (eg, *cnf1, clbB*, and *pic* in ST73, *sat* in ST131, and *vat* and *hlyF* in ST95). These observations show that virulence factors in *E coli* are not randomly distributed and reflect the underlying clonal structure of the population, in agreement with previous reports.[5,18,19]

Pathogenicity islands are a group of large (>10 kb) integrative elements of the genome encoding one or more virulence genes.[26] Daga and colleagues[15] reported a higher presence of pathogenicity island markers in blood *E coli* isolates belonging to phylogenetic group B2, in particular pathogenicity islands I_{536}, IV_{536}, I_{CFT073}, II_{CFT073}, and II_{J96}.[15] Accordingly, in our collection, several pathogenicity islands were identified in isolates from group 1 (phylogroups B2, F and G) including pathogenicity islands I–III and V of strain 536, encoding a range of virulence factors (eg, α-hemolysin, S fimbriae, adhesins, hemagglutinin-like adhesins, and iron siderophore system)[26] and pathogenicity island II_{CFT073}, encoding P fimbriae, α-hemolysin, and aerobactin.[27] In addition, in both group 1 and group 3, there was a high prevalence of pathogenicity islands $IV_{APEC-O1}$, implied in the biosynthesis of the siderophore yersiniabactin,[28] and AGI-3, related to carbohydrate assimilation and virulence.[29] We also found some pathogenicity islands to be associated with specific STs, suggesting a clonal-associated dissemination of these elements.

Regarding AMR, our isolates showed high resistance rates to ciprofloxacin, trimethoprim–sulfamethoxazole, and amoxicillin–clavulanic acid. In addition, resistance to third-generation cephalosporins was consistent with previous studies.[2,11,15] Although no significant differences were found among phylogenetic groups, ST131 isolates showed higher prevalence of AMR, acquired AMR genes, and chromosomal mutations conferring AMR, as previously reported.[11,18] ST131 has become a dominant extraintestinal pathogenic *E coli* strain worldwide, exhibiting fluoroquinolone resistance restricted almost entirely to a single, rapidly expanding ST131 subclone (H30-R). H30-Rx, a well defined multidrug-resistant clade within H30-R, comprised most of the $bla_{CTX-M-15}$-producing isolates in ST131.[10] Subtyping revealed that around one-third of the ST131 isolates in our dataset belonged to H30Rx, corresponding to 6% of the overall *E coli* collection. In another Spanish study including 425 bacteraemic UTI isolates, the prevalence of ST131-H30Rx subclone was 5%.[30]

This study has limitations. Our results might not be applicable to other geographical areas. Although the number of isolates analysed is reasonably large, the significance of results must be interpreted with caution because of the large number of comparisons made, despite the use of Bonferroni correction, and because of the small numbers in some subgroups; estimates of associations and analysis of confounders were not done due to the exploratory nature of the analyses and to avoid further complexity in interpretation. However, the strengths of this study include the use of

whole-genome sequencing, inclusion of cases and isolates from a multicentre, countrywide representation, and the use of current definitions for sepsis and septic shock.

In summary, *E coli* blood isolates causing sepsis or septic shock showed similar clonal distribution as urine and overall bloodstream isolates in previous studies; the distribution of virulence genes, pathogenicity islands, and AMR genes provided deeper insights into the epidemiology of *E coli* causing severe infections, as a basis for further studies to elucidate the contribution of those microbiological factors to severe clinical presentation and worse outcomes in patients with *E coli* BSI and, potentially, for the development of diagnostic, preventive, or therapeutic interventions.

Contributors

NM, IL-H, JR-B, and AP were responsible for conceptualisation, formulating the overall research questions, methodology, formal analysis, and writing of the original draft, and have directly accessed and verified the underlying data reported in the Article. NM, IL-H, and AG-M were responsible for data curation and bioinformatics analysis. LEL-C, PMMP-C, PR-G, JG, JMR-I, AIA, AS-D, CA-C, TM-C, JF-S, AJ-S, LB-P, CN-K, AdA-J, AB, JMSC, FG-S, AS-A, ÁP-N, AR-B, IG-L, BB-C, IP-C, and CL-E participated by reviewing the design, recruiting patients and isolates, and thoroughly reviewing the manuscript. All authors had full access to all the data in the study and accept responsibility for the decision to submit for publication. IL-H, JR-B and ÁP were responsible for funding acquisition, project administration, supervision, and coordinating the study.

Declaration of interests

LEL-C reports payments as a speaker from Angelini Pharm and Correvio Pharma Corp. PR-G reports consulting fees from and participation on a data safety monitoring board or advisory board for Advanz Pharma, support for attending meetings or travel from Gilead Sciences, and payments for presentations from Menarini Group and Shionogi & Co. LB-P reports payments for presentations in educational events from Tillotts Pharma and Menarini Group and support for attending meetings or travel from Pfizer. JMR-I reports payments as a speaker from Pfizer and Shionogi & Co. AG-M reports support for attending ECCMID 2022 from ESCMID. All other authors declare no competing interests.

Data sharing

An anonymised, de-identified version of the dataset that underlies the results reported in this Article will be available for investigators whose proposed use of the data has been approved by an independent review committee. Proposals should be directed to jesusrb@us.es; to gain access, data requestors will need to sign a data access agreement with the senior authors' institution. Genome sequence data have been deposited in the European Nucleotide Archive under the Bioproject PRJEB62601.

Acknowledgments

This study was supported by grant PI16/01432 from Instituto de Salud Carlos III (AES, European Development Regional Fund: A Way to Achieve Europe).

References

1 Bonten M, Johnson JR, van den Biggelaar AHJ, et al. Epidemiology of *Escherichia coli* bacteremia: a systematic literature review. *Clin Infect Dis* 2021; **72:** 1211–19.

2 de Lastours V, Laouénan C, Royer G, et al. Mortality in *Escherichia coli* bloodstream infections: antibiotic resistance still does not make it. *J Antimicrob Chemother* 2020; **75:** 2334–43.

3 Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, et al. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. *Gut Pathog* 2019; **11:** 10.

4 Johnson JR, Johnston BD, Porter S, Thuras P, Aziz M, Price LB. Accessory traits and phylogenetic background predict *Escherichia coli* extraintestinal virulence better than does ecological source. *J Infect Dis* 2019; **219:** 121–32.

5 Vanstokstraeten R, Crombé F, Piérard D, et al. Molecular characterization of extraintestinal and diarrheagenic *Escherichia coli* blood isolates. *Virulence* 2022; **13:** 2032–41.

6 Pérez-Crespo PMM, Lanz-García JF, Bravo-Ferrer J, et al. Revisiting the epidemiology of bloodstream infections and healthcare-associated episodes: results from a multicentre prospective cohort in Spain (PRO-BAC Study). *Int J Antimicrob Agents* 2021; **58:** 106352.

7 Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; **315:** 801–10.

8 Cercenado E, Cantón R. Procedimientos en Microbiología Clínica. Recomendaciones de la Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica editors. Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica. 2003. https://seimc.org/contenidos/documentoscientificos/procedimientosmicrobiologia/seimc-procedimientomicrobiologia1a.pdf (accessed Oct 5, 2023).

9 Pitout JDD, Finn TJ. The evolutionary puzzle of *Escherichia coli* ST131. *Infect Genet Evol* 2020; **81:** 104265.

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

11 Jauneikaite E, Honeyford K, Blandy O, et al. Bacterial genotypic and patient risk factors for adverse outcomes in *Escherichia coli* bloodstream infections: a prospective molecular epidemiological study. *J Antimicrob Chemother* 2022; **77:** 1753–61.

12 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.

13 Clermont O, Couffignal C, Blanco J, Mentré F, Picard B, Denamur E. Two levels of specialization in bacteraemic *Escherichia coli* strains revealed by their comparison with commensal strains. *Epidemiol Infect* 2017; **145:** 872–82.

14 Micenková L, Beňová A, Frankovičová L, et al. Human *Escherichia coli* isolates from hemocultures: septicemia linked to urogenital tract infections is caused by isolates harboring more virulence genes than bacteraemia linked to other conditions. *Int J Med Microbiol* 2017; **307:** 182–89.

15 Daga AP, Koga VL, Soncini JGM, et al. *Escherichia coli* bloodstream infections in patients at a university hospital: virulence factors and clinical characteristics. *Front Cell Infect Microbiol* 2019; **9:** 191.

16 Zhuge X, Zhou Z, Jiang M, et al. Chicken-source *Escherichia coli* within phylogroup F shares virulence genotypes and is closely related to extraintestinal pathogenic *E coli* causing human infections. *Transbound Emerg Dis* 2021; **68:** 880–95.

17 Clermont O, Dixit OVA, Vangchhia B, et al. Characterization and rapid identification of phylogroup G in *Escherichia coli*, a lineage with high virulence and antibiotic resistance potential. *Environ Microbiol* 2019; **21:** 3107–17.

18 Lipworth S, Vihta KD, Chau K, et al. Ten-year longitudinal molecular epidemiology study of *Escherichia coli* and *Klebsiella* species bloodstream infections in Oxfordshire, UK. *Genome Med* 2021; **13:** 144.

19 Goswami C, Fox S, Holden M, Connor M, Leanord A, Evans TJ. Genetic analysis of invasive *Escherichia coli* in Scotland reveals determinants of healthcare-associated versus community-acquired infections. *Microb Genom* 2018; **4:** e000190.

20 Kim B, Kim JH, Lee Y. Virulence factors associated with *Escherichia coli* bacteremia and urinary tract infection. *Ann Lab Med* 2022; **42:** 203–12.

21 Choi HJ, Jeong SH, Shin KS, et al. Characteristics of *Escherichia coli* urine isolates and risk factors for secondary bloodstream infections in patients with urinary tract infections. *Microbiol Spectr* 2022; **10:** e0166022.

22 Miajlovic H, Mac Aogáin M, Collins CJ, Rogers TR, Smith SGJ. Characterization of *Escherichia coli* bloodstream isolates associated with mortality. *J Med Microbiol* 2016; **65:** 71–79.

23 Turkovicova L, Smidak R, Jung G, Turna J, Lubec G, Aradska J. Proteomic analysis of the TerC interactome: novel links to tellurite resistance and pathogenicity. *J Proteomics* 2016; **136:** 167–73.

24 Yun TH, Cott JE, Tapping RI, Slauch JM, Morrissey JH. Proteolytic inactivation of tissue factor pathway inhibitor by bacterial omptins. *Blood* 2009; **113:** 1139–48.

25 Biran D, Sura T, Otto A, Yair Y, Becher D, Ron EZ. Surviving serum: the *Escherichia coli iss* gene of extraintestinal pathogenic *E coli* is required for the synthesis of group 4 capsule. *Infect Immun* 2021; **89:** e0031621.

26 Dobrindt U, Blum-Oehler G, Nagy G, et al. Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic *Escherichia coli* strain 536. *Infect Immun* 2002; **70:** 6365–72.

27 Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. *Clin Microbiol Rev* 2004; **17:** 14–56.

28 Johnson TJ, Kariyawasam S, Wannemuehler Y, et al. The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J Bacteriol* 2007; **189:** 3228–36.

29 Chouikha I, Germon P, Brée A, Gilot P, Moulin-Schouleur M, Schouler C. A selC-associated genomic island of the extraintestinal avian pathogenic *Escherichia coli* strain BEN2908 is involved in carbohydrate uptake and virulence. *J Bacteriol* 2006; **188:** 977–87.

30 Merino I, Shaw E, Horcajada JP, et al. CTX-M-15-H30Rx-ST131 subclone is one of the main causes of healthcare-associated ESBL-producing *Escherichia coli* bacteraemia of urinary origin in Spain. *J Antimicrob Chemother* 2016; **71:** 2125–30.