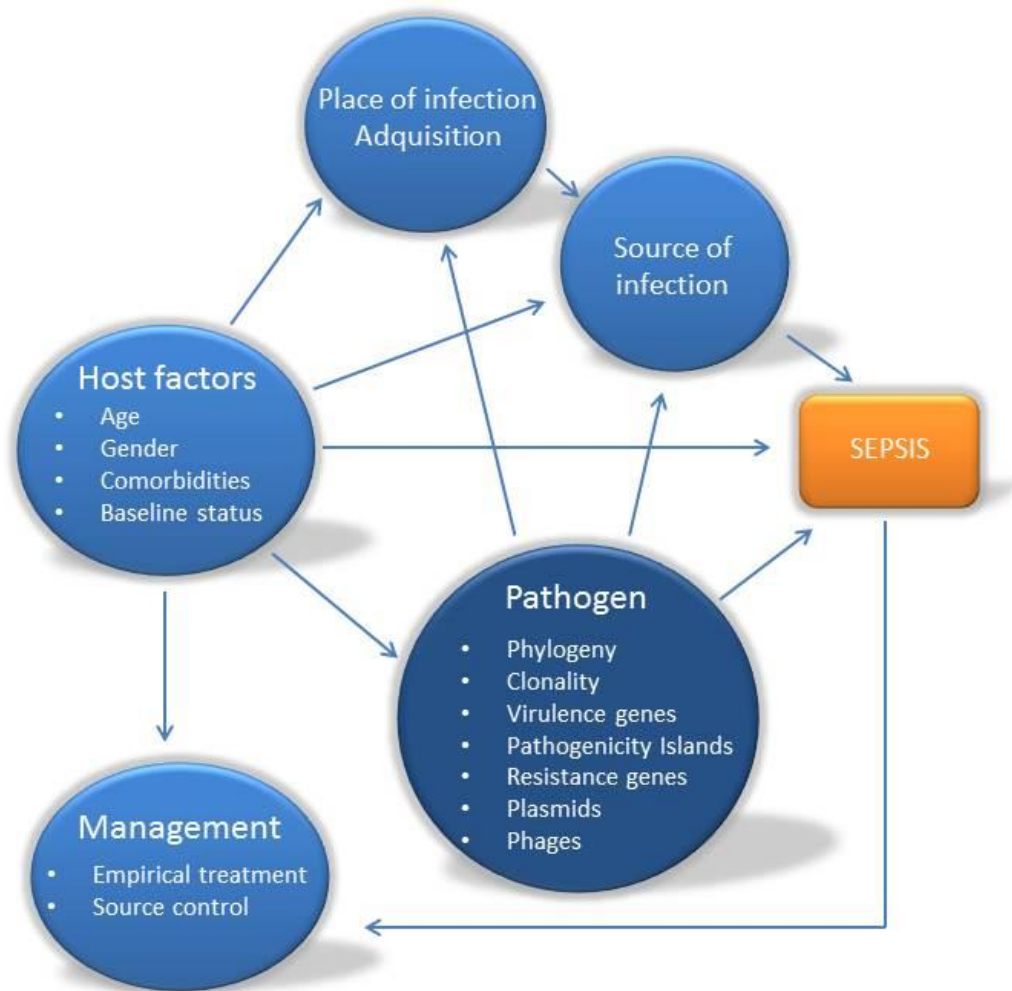


SUPPLEMENTARY APPENDIX

TABLE OF CONTENTS

List of Tables/Figures	Page
Supplementary Fig. 1. Directed Acyclic Graph (DAG) for presentation as sepsis/septic shock in patients with <i>Escherichia coli</i> bacteraemia	2
Supplementary Fig. 2. Feature selection algorithm	3
Supplementary Table 1. Random forest machine learning models to classify sepsis/septic shock in patients with <i>Escherichia coli</i> bacteremia	4
Supplementary Table 2. Crude association between comorbidities, invasive procedures and devices and presentation as sepsis/septic shock in patients with <i>Escherichia coli</i> bacteraemia.	5
Supplementary Table 3. Crude association between virulence factors and presentation as sepsis/septic shock in patients with <i>Escherichia coli</i> bacteraemia.	6-7
Supplementary Table 4. Crude association between the prevalence of pathogenicity islands (PAI) and presentation as sepsis/septic shock in patients with <i>Escherichia coli</i> bacteraemia.	8
Supplementary Table 5. Crude association between the prevalence of plasmids replicons and presentation as sepsis/septic shock in patients with <i>Escherichia coli</i> bacteraemia.	9
Supplementary Table 6. Crude association between the prevalence of prophage sequences and presentation as sepsis/septic shock in patients with <i>Escherichia coli</i> bacteraemia.	10
Supplementary Table 7. Crude association between the proportion of antibiotic resistance and presentation as sepsis/septic shock in patients with <i>Escherichia coli</i> bacteraemia.	11
Supplementary Table 8. Crude association between antibiotic resistance genes and presentation as sepsis/septic shock in patients with <i>Escherichia coli</i> bacteraemia.	12
Supplementary Table 9. Multivariate conditional logistic regression model of <i>Escherichia coli</i> microbiological factors associated with presentation as sepsis/septic shock in patients with bacteraemia, controlling for the propensity score	13
Supplementary Fig. 3. ROC curves illustrating the prediction of sepsis/septic shock presentation in patients with <i>Escherichia coli</i> bacteraemia based on A) the propensity score and B) conditional logistic regression model of microbiological factors.	14
Supplementary Table 10. Matrix of the linkage between virulence clusters and sequence types	15

Supplementary Fig. 1. Directed Acyclic Graph (DAG) for presentation as sepsis/septic shock in patients with *Escherichia coli* bacteraemia

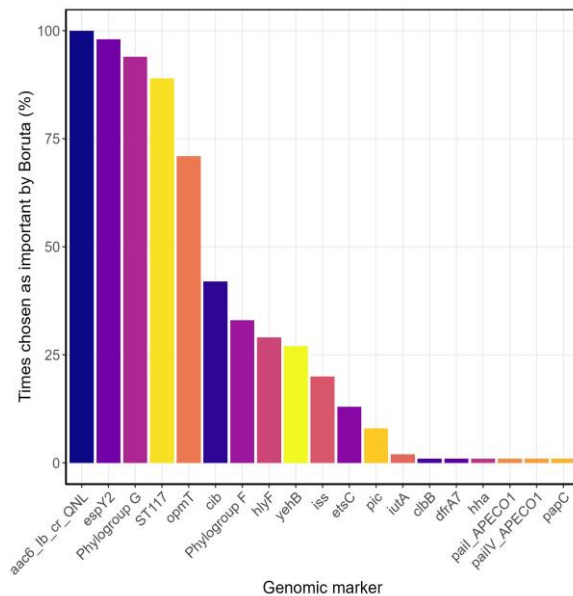


The directed acyclic graph (DAG) illustrates the hypothesised relationships between variables in the analysis of sepsis/septic shock among patients with *Escherichia coli* bacteraemia. The DAG depicts the assumed causal structure used to identify potential confounders and guide covariate selection for statistical adjustment. Arrows indicate hypothesised causal pathways.

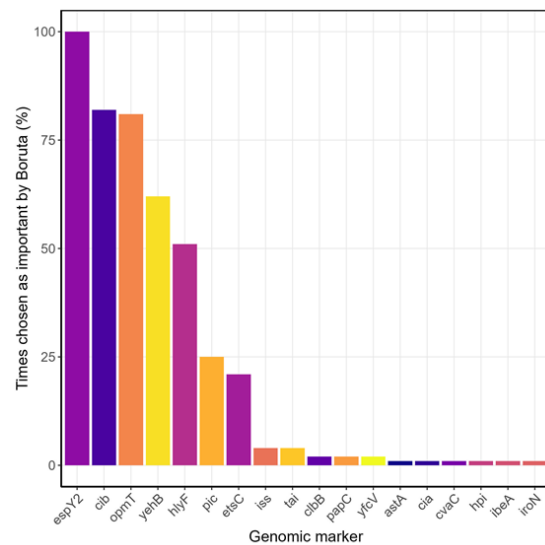
Supplementary Fig. 2. Feature selection algorithm

Feature selection to improve the machine learning model performance was conducted using the Boruta package in R¹. This method creates permuted copies for each predictor (shadow features), fits a random forest model to classify sepsis/septic shock, and then iteratively removes predictors that are statistically less important compared to the shadow feature. Two sets of predictors were used, (A) all available predictors (virulence genes, pathogenicity islands, sequence types, phylogroups, plasmids and antibiotic resistance genes), and (B) virulence genes and pathogenicity islands. The process was repeated n=100 times to account for randomness in the process. The results are presented as times a feature was deemed important by the algorithm across the repetitions. Features selected >10 times were used in subsequent analyses.

A.) All genomic predictors (n=225)



B.) Restriction to virulence genes and PAI (n=109)



Reference:

1. Kursa, M. B. & Rudnicki, W. R. 2010. Feature Selection with the Boruta Package. J. Stat. Softw. 36, 1–13.

Supplementary Table 1. Random forest machine learning models to classify sepsis/septic shock in patients with *Escherichia coli* bacteraemia

Model Name	Predictors
1) Selected bacterial factors ^a	<i>cib</i> , <i>espY2</i> , <i>hlyF</i> , <i>iss</i> , <i>opmT</i> , <i>yehB</i> , phylogroup F, phylogroup G, ST117, <i>etsC</i> , <i>aac(6')-Ib-cr</i>
2) Selected virulence genes and pathogenicity islands ^b	<i>cib</i> , <i>espY2</i> , <i>etsC</i> , <i>hlyF</i> , <i>opmT</i> , <i>pic</i> , <i>yehB</i>
3) Virulence Clusters	clusterA, clusterB, clusterC, clusterD, clusterE, clusterF, clusterG, clusterH
4) Sequence type	ST12, ST117, ST141, ST131, ST69, ST73, ST95, other_st
5) Phylogroup	A, B1, B2, C, D, F, G
6) Plasmids	IncFIA, IncFIB_AP001918, IncFIB_H89_Phage_plasmid, IncFIB_pLF82_Phage_plasmid, IncFIC_FII, IncFII, IncFII_Prsb107, IncFII_pHN7A8, IncFII_29, IncFII_pCoo, IncFII_PCRY, IncFII_pAMA1167_NDM_5, IncQ1, IncX1, IncI1_I_Gamma, IncI2, IncN2, IncB_O_K_Z, ColBS512, ColKHPS6, ColRNAI, ColPVC, Col156, Col44OII, ColMG828, IncHI1B, IncN, IncY, IncR, IncX3, IncX4, pO111, RepA_pKOX, Pkpc_cav1193, IncU, IncL
7) Resistance genes or mutations	<i>aac(6')-Ib-cr</i> , <i>aac(3)-IIa</i> , <i>aac(3)-IId</i> , <i>ant(2'')-Ia</i> , <i>aph(3'')-Ib</i> , <i>aph(6)-Id</i> , <i>aph(3')-Ia</i> , <i>aadA1</i> , <i>aadA2</i> , <i>aadA5</i> , <i>aadA12</i> , <i>bla_{TEM-1A}</i> , <i>bla_{TEM-1B}</i> , <i>bla_{TEM-1C}</i> , <i>bla_{TEM-1D}</i> , <i>bla_{TEM-30}</i> , <i>bla_{TEM-33}</i> , <i>bla_{TEM-40}</i> , <i>bla_{TEM-195}</i> , <i>bla_{CMY2}</i> , Any <i>Bla_{CTX-M}</i> , Any ESBL, <i>bla_{CTX-M1}</i> , <i>bla_{CTX-M14}</i> , <i>bla_{CTX-M15}</i> , <i>bla_{CTX-M32}</i> , <i>bla_{NDM-1}</i> , <i>bla_{CARB-2}</i> , <i>bla_{OXA-1}</i> , <i>bla_{OXA-48}</i> , <i>bla_{SVH-1}</i> , <i>bla_{SVH-12}</i> , <i>bla_{SVH-48}</i> , <i>bla_{SVH-102}</i> , <i>qnrS1</i> , <i>fosA7</i> , <i>mphA</i> , <i>mphB</i> , <i>inuF</i> , <i>catA1</i> , <i>catA2</i> , <i>catB3</i> , <i>cmlA1</i> , <i>floR</i> , <i>sul1</i> , <i>sul2</i> , <i>sul3</i> , <i>tetA</i> , <i>tetB</i> , <i>tetM</i> , <i>dfrA1</i> , <i>dfrA5</i> , <i>dfrA7</i> , <i>dfrA8</i> , <i>dfrA12</i> , <i>dfrA14</i> , <i>dfrA17</i> , Mut <i>parCS80I</i> , Mut <i>parCE84V</i> , Mut <i>gyrAS83L</i> , Mut <i>gyrAD87N</i> , Mut <i>parEI259L</i> , Mut <i>parEpS458A</i> , Mut <i>ampC</i> -promoter
8) All virulence genes and pathogenicity islands ^c	AGI-1, <i>espC</i> , ETT2, HPI, LEE-01, LEE-03, LEE-07, OI-57, PAI I ₅₃₆ , PAI-I _{AL862} , PAI I _{APEC-O1} , PAI I _{CFT073} , PAI II ₅₃₆ , PAI II _{APEC-O1} , PAI II _{CFT073} , PAI III ₅₃₆ , PAI III _{APEC-O1} , PAI IV _{APEC-O1} , PAI V ₅₃₆ , TAI, <i>aslA</i> , <i>aamR</i> , <i>afaA</i> , <i>afaB</i> , <i>afaC</i> , <i>afaD</i> , <i>air</i> , <i>anr</i> , <i>astA</i> , <i>cea</i> , <i>chuA</i> , <i>cia</i> , <i>cib</i> , <i>clbB</i> , <i>cma</i> , <i>cnf1</i> , <i>colE2_like</i> , <i>cvaC</i> , <i>dhaK</i> , <i>eilA</i> , <i>espY2</i> , <i>etsC</i> , <i>fdeC</i> , <i>focC</i> , <i>focSsfaE</i> , <i>focG</i> , <i>focI</i> , <i>fyuA</i> , <i>gad</i> , <i>hha</i> , <i>hlyA</i> , <i>hlyE</i> , <i>hlyF</i> , <i>hra</i> , <i>ibeA</i> , <i>iha</i> , <i>ireA</i> , <i>iroN</i> , <i>irp2</i> , <i>iss</i> , <i>iucC</i> , <i>iutA</i> , <i>kpsE</i> , <i>kpsMII</i> , <i>kpsMIII_K96</i> , <i>kpsMII_K1</i> , <i>kpsMII_K5</i> , <i>kpsMII_K52</i> , <i>lpfA</i> , <i>mcbA</i> , <i>mchB</i> , <i>mchC</i> , <i>mchF</i> , <i>mcmA</i> , <i>neuC</i> , <i>opmT</i> , <i>papA_F11</i> , <i>papA_F12</i> , <i>papA_F13</i> , <i>papA_f14</i> , <i>papA_F43</i> , <i>papA_F48</i> , <i>papA_F7_1</i> , <i>papA_feiA_F16</i> , <i>papC</i> , <i>papGII</i> , <i>papGIII</i> , <i>pic</i> , <i>sat</i> , <i>senB</i> , <i>sfaD</i> , <i>sfaS</i> , <i>shiA</i> , <i>shiB</i> , <i>sitA</i> , <i>tcpC</i> , <i>tia</i> , <i>traJ</i> , <i>traT</i> , <i>tsh</i> , <i>usp</i> , <i>vat</i> , <i>yehA</i> , <i>yehB</i> , <i>yfcV</i>
9) All bacterial genomic predictors	Indicated in supplementary tables 3-8

^a Features deemed important by the Bourta algorithm based on a set of all available predictors (virulence genes, pathogenicity islands, sequence types, phylogroups, plasmids and antibiotic resistance genes) (n=225).

^b Features deemed important by the Boruta algorithm based on virulence genes and pathogenicity islands (n=109).

^c Detected in more than 2%, but less than 98% of the isolates

Supplementary Table 2. Crude association between comorbidities, invasive procedures and devices and presentation as sepsis/septic shock in patients with *Escherichia coli* bacteraemia.

Variable	Cases (n=101)	Controls (n=101)	OR (95% CI)	P value
Underlying conditions				
Diabetes mellitus	30 (30)	27 (27)	1.21 (0.60-2.46)	0.591
Dementia	18 (18)	10 (10)	1.89 (0.84-4.24)	0.123
Recurrent urinary tract infection	14 (14)	6 (6)	2.60 (0.93-7.29)	0.069
Cerebrovascular disease	13 (13)	10 (10)	1.43 (0.54-3.75)	0.469
Cancer	13 (13)	23 (23)	0.38 (0.15-0.96)	0.040
Peripheral vascular disease	12 (12)	5 (5)	4.5 (0.97-20.83)	0.054
Obstructive uropathy	12 (12)	5 (5)	2.75 (0.88-8.64)	0.083
Chronic lung disease	9 (9)	14 (14)	0.62 (0.26-1.48)	0.280
Moderate or severe renal disease	8 (8)	1 (1)	65.3 (0.15-28459.8)	0.178
Congestive heart failure	8 (8)	8 (8)	1.00 (0.38-2.66)	1.000
Liver disease	7 (7)	10 (10)	0.63 (0.20-1.91)	0.410
Acute myocardial infarction	6 (6)	4 (4)	1.50 (0.42-5.32)	0.530
Peptic ulcer	6 (6)	0 (0)	65.3 (0.09-46313.7)	0.212
Connective tissue disease	4 (4)	3 (3)	1.33 (0.30-5.96)	0.706
Biliary obstruction	3 (3)	6 (6)	0.40 (0.08-2.06)	0.273
Hemiplegia / paraplegia	2 (2)	2 (2)	1.00 (0.14-7.10)	1.000
Invasive devices during the previous month				
Peripheral venous catheter	33 (33)	27 (27)	1.43 (0.72-2.83)	0.306
Urinary catheter	21 (21)	8 (8)	3.60 (1.34-9.70)	0.011
Central venous catheter	10 (10)	3 (3)	3.33 (0.92-12.11)	0.067
Joint prosthesis	7 (7)	7 (7)	1.00 (0.35-2.85)	1.000
Valvular prosthesis	4 (4)	1 (1)	4.00 (0.45-35.79)	0.215
Double J catheter	4 (4)	1 (1)	4.00 (0.45-35.79)	0.215
Peripherally inserted central catheter	3 (3)	2 (2)	2.00 (0.18-22.06)	0.571
Biliary prosthesis	3 (3)	7 (7)	0.43 (0.11-1.66)	0.220
Invasive procedures				
Surgery	3 (3)	9 (9)	0.14 (0.02-1.16)	0.069
Transurethral resection of the prostate	4 (4)	0 (0)	65.3 (0.02-202501.5)	0.308
Bronchoscopy	2 (2)	1 (1)	2.00 (0.18-22.06)	0.571
Cystoscopy	2 (2)	1 (1)	2.00 (0.18-22.06)	0.571
Gastroscopy	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
Parenteral nutrition	0 (0)	4 (4)	0.02 (0.00-47.51)	0.308

OR, odds ratio; CI, confidence interval

Supplementary Table 3. Crude association between virulence factors and presentation as sepsis/septic shock in patients with *Escherichia coli* bacteraemia. Data are number of isolates (percentage per column).

Virulence genes	Cases (n=101)	Controls (n=101)	OR (95% CI)	P value
<i>aamR</i>	15 (15)	20 (20)	0.67 (0.30-1.48)	0.321
<i>aap</i>	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
<i>afaA</i>	4 (4)	7 (7)	0.57 (0.17-1.95)	0.372
<i>afaB</i>	4 (4)	6 (6)	0.67 (0.19-2.36)	0.530
<i>afaC</i>	4 (4)	6 (6)	0.67 (0.19-2.36)	0.530
<i>afaD</i>	7 (7)	10 (10)	0.70 (0.27-1.84)	0.469
<i>afaE</i>	2 (2)	2 (2)	1.00 (0.14-7.10)	1.000
<i>air</i>	10 (10)	12 (12)	0.80 (0.32-2.03)	0.638
<i>anr</i>	44 (44)	52 (51)	0.73 (0.42-1.27)	0.269
<i>aslA</i>	84 (83)	89 (88)	0.67 (0.30-1.48)	0.321
<i>astA</i>	10 (10)	7 (7)	1.50 (0.53-4.21)	0.442
<i>cba</i>	3 (3)	1 (1)	2.00 (0.31-28.84)	0.341
<i>cdt_IVB</i>	0 (0,0)	3 (3)	0.02 (0.00-164.78)	0.378
<i>cea</i>	35 (35)	35 (35)	1.00 (0.56-1.78)	1.000
<i>chuA</i>	79 (78)	84 (83)	0.72 (0.35-1.47)	0.371
<i>cia</i>	16 (16)	25 (25)	0.64 (0.34-1.20)	0.163
<i>cib</i>	7 (7)	1 (1)	7.00 (0.86-56.89)	0.069
<i>clbB</i>	30 (30)	39 (39)	0.64 (0.34-1.20)	0.163
<i>cma</i>	7 (7)	4 (4)	2.00 (0.50-8.00)	0.327
<i>cnf1</i>	28 (28)	33 (33)	0.78 (0.42-1.45)	0.436
<i>colE2_like</i>	5 (5)	3 (3)	1.67 (0.40-6.97)	0.484
<i>colE5</i>	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
<i>csgA</i>	100 (99)	100 (99)	1.00 (0.06-15.99)	1.000
<i>cvaC</i>	20 (20)	30 (30)	0.58 (0.30-1.13)	0.109
<i>dhaK</i>	7 (7)	14 (14)	0.46 (0.18-1.21)	0.117
<i>ehxA</i>	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
<i>eilA</i>	13 (13)	14 (14)	0.92 (0.40-2.08)	0.835
<i>espI</i>	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
<i>espY2</i>	1 (1)	8 (8)	0.13 (0.02-1.00)	0.050
<i>etsC</i>	20 (20)	33 (33)	0.52 (0.27-0.99)	0.046
<i>F17A</i>	4 (4)	1 (1)	4.00 (0.45-35.79)	0.215
<i>F17C</i>	4 (4)	1 (1)	4.00 (0.45-35.79)	0.215
<i>F17D</i>	4 (4)	1 (1)	4.00 (0.45-35.79)	0.215
<i>F17G</i>	3 (3)	1 (1)	2.00 (0.31-28.84)	0.341
<i>fdeC</i>	93 (92)	99 (98)	0.25 (0.05-1.18)	0.080
<i>fimH</i>	100 (99)	100 (99)	1.00 (0.06-15.99)	1.000
<i>focC</i>	9 (9)	12 (12)	0.75 (0.32-1.78)	0.514
<i>focG</i>	15 (15)	11 (11)	1.40 (0.62-3.15)	0.416
<i>focI</i>	10 (10)	6 (6)	1.80 (0.60-5.37)	0.292
<i>focSsfaE</i>	14 (14)	10 (10)	1.44 (0.62-3.38)	0.396
<i>fyuA</i>	85 (84)	87 (86)	0.83 (0.36-1.93)	0.670
<i>gad</i>	75 (74)	79 (78)	0.82 (0.44-1.53)	0.528
<i>hha</i>	62 (61)	51 (50)	1.55 (0.88-2.72)	0.127
<i>hlyA</i>	31 (31)	36 (36)	0.77 (0.41-1.46)	0.425
<i>hlyE</i>	33 (33)	31 (31)	1.11 (0.59-2.06)	0.752
<i>hlyF</i>	21 (21)	35 (35)	0.48 (0.25-0.93)	0.030
<i>hra</i>	36 (36)	44 (44)	0.72 (0.41-1.27)	0.260
<i>ibeA</i>	7 (7)	14 (14)	0.46 (0.18-1.21)	0.117
<i>iha</i>	44 (44)	41 (41)	1.14 (0.64-2.02)	0.662
<i>ireA</i>	25 (25)	27 (27)	0.90 (0.49-1.68)	0.752
<i>iroN</i>	48 (48)	57 (56)	0.68 (0.38-1.22)	0.192
<i>irp2</i>	83 (82)	86 (85)	0.79 (0.36-1.73)	0.549
<i>iss</i>	83 (82)	93 (92)	0.33 (0.12-0.92)	0.033
<i>iucC</i>	58 (57)	69 (68)	0.59 (0.32-1.10)	0.097

<i>iutA</i>	57 (56)	69 (68)	0.57 (0.31-1.06)	0.074
<i>kpsE</i>	77 (76)	85 (84)	0.60 (0.29-1.23)	0.162
<i>kpsMII</i>	16 (16)	23 (12)	0.59 (0.27-1.28)	0.183
<i>kpsMII_K1</i>	16 (16)	26 (26)	0.57 (0.29-1.12)	0.100
<i>kpsMII_K23</i>	2 (2)	2 (2)	1.00 (0.14-7.10)	1.000
<i>kpsMII_K5</i>	23 (23)	20 (20)	1.20 (0.60-2.38)	0.602
<i>kpsMII_K52</i>	4 (4)	5 (5)	0.75 (0.17-3.35)	0.706
<i>kpsMIII_K96</i>	4 (4)	2 (2)	2.00 (0.37-10.92)	0.423
<i>lpfA</i>	32 (32)	25 (25)	1.47 (0.76-2.83)	0.253
<i>mcbA</i>	8 (8)	6 (6)	1.33 (0.46-3.84)	0.594
<i>mchB</i>	27 (27)	24 (24)	1.20 (0.60-2.38)	0.602
<i>mchC</i>	26 (26)	24 (24)	1.12 (0.57-2.21)	0.732
<i>mchF</i>	43 (43)	53 (52)	0.67 (0.38-1.17)	0.160
<i>mcmA</i>	27 (27)	28 (28)	0.95 (0.52-1.74)	0.879
<i>neuC</i>	19 (19)	28 (28)	0.60 (0.31-1.18)	0.143
<i>nfaE</i>	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
<i>nlpI</i>	100 (99)	101 (100)	0.02 (0.00-147346.84)	0.610
<i>opmT</i>	86 (85)	96 (95)	0.23 (0.07-0.81)	0.022
<i>papA_F11</i>	11 (11)	16 (16)	0.67 (0.30-1.48)	0.321
<i>papA_F12</i>	5 (5)	4 (4)	1.25 (0.34-4.65)	0.739
<i>papA_F13</i>	6 (6)	9 (9)	0.63 (0.20-1.91)	0.410
<i>papA_f14</i>	5 (5)	8 (8)	0.63 (0.20-1.91)	0.410
<i>papA_F19</i>	3 (3)	1 (1)	2.00 (0.31-28.84)	0.341
<i>papA_F20</i>	2 (2)	1 (1)	2.00 (0.18-22.06)	0.571
<i>papA_F43</i>	32 (32)	27 (27)	1.31 (0.68-2.52)	0.413
<i>papA_F48</i>	4 (4)	6 (6)	0.67 (0.19-2.36)	0.530
<i>papA_F7_1</i>	4 (4)	2 (2)	2.00 (0.37-10.92)	0.423
<i>papA_F7_2</i>	2 (2)	2 (2)	1.00 (0.14-7.10)	1.000
<i>papA_F9</i>	2 (2)	2 (2)	1.00 (0.14-7.10)	1.000
<i>papA_feiA_F16</i>	9 (9)	11 (11)	0.78 (0.29-2.09)	0.618
<i>papC</i>	57 (56)	70 (69)	0.50 (0.26-0.97)	0.041
<i>papGII</i>	43 (43)	49 (49)	0.71 (0.37-1.39)	0.320
<i>papGIII</i>	8 (8)	13 (13)	0.58 (0.23-1.48)	0.257
<i>pic</i>	27 (27)	17 (17)	1.91 (0.92-3.96)	0.082
<i>sat</i>	35 (35)	33 (33)	1.10 (0.60-2.02)	0.758
<i>senB</i>	29 (29)	26 (26)	1.19 (0.61-2.31)	0.613
<i>sfaD</i>	15 (15)	19 (19)	0.76 (0.37-1.57)	0.467
<i>sfaE</i>	1 (1,0)	3 (3)	0.33 (0.03-3.20)	0.341
<i>sfaS</i>	4 (4)	3 (3)	1.33 (0.30-5.96)	0.706
<i>shiA</i>	13 (13)	15 (15)	0.86 (0.40-1.85)	0.695
<i>shiB</i>	32 (32)	43 (43)	0.65 (0.37-1.13)	0.127
<i>sitA</i>	87 (86)	93 (92)	0.57 (0.24-1.36)	0.207
<i>tcpC</i>	25 (25)	29 (29)	0.82 (0.44-1.53)	0.528
<i>terC</i>	101 (100)	101 (100)	NC	1.000
<i>tia</i>	33 (33)	36 (36)	0.88 (0.50-1.55)	0.668
<i>traJ</i>	46 (46)	51 (50)	0.81 (0.45-1.44)	0.467
<i>traT</i>	71 (70)	69 (68)	1.10 (0.60-2.02)	0.758
<i>tsh</i>	7 (7)	10 (10)	0.67 (0.24-1.87)	0.442
<i>usp</i>	62 (61)	68 (67)	0.74 (0.39-1.38)	0.345
<i>vat</i>	50 (50)	52 (51)	0.91 (0.51-1.65)	0.763
<i>yehA</i>	97 (96)	97 (96)	1.00 (0.25-4.00)	1.000
<i>yehB</i>	96 (95)	101 (100)	0.02 (0.00-20.33)	0.255
<i>yehC</i>	100 (99)	100 (99)	1.00 (0.06-15.99)	1.000
<i>yehD</i>	99 (98)	101 (100)	0.02 (0.00-1327.44)	0.471
<i>yfcV</i>	63 (62)	75 (74)	0.56 (0.30-1.04)	0.068
Virulence Score, Me (IQR)	34 (28-40)	37 (31-42)	0.96 (0.93-1.00)	0.032

NC, not calculated; Me, median; IQR, interquartile range, OR, odds ratio; CI, confidence interval

Supplementary Table 4. Crude association between the prevalence of pathogenicity islands (PAI) and presentation as sepsis/septic shock in patients with *Escherichia coli* bacteraemia. Data are number of isolates (percentage per column).

Pathogenicity Island	Cases (n=101)	Controls (n=101)	OR (95% CI)	P value
AGI-1	79 (78)	77 (76)	1.12 (0.57-2.21)	0.732
<i>espC</i>	66 (65)	71 (71)	0.77 (0.41-1.46)	0.425
ETT2	31 (31)	30 (30)	1.06 (0.54-2.10)	0.862
HPI	28 (28)	19 (19)	1.69 (0.85-3.36)	0.133
LEE-01	14 (14)	9 (9)	1.71 (0.67-4.35)	0.257
LEE-03	21 (21)	18 (18)	1.19 (0.61-2.31)	0.613
LEE-07	15 (15)	12 (12)	1.30 (0.57-2.96)	0.533
OI-57	39 (39)	29 (29)	1.48 (0.85-2.57)	0.168
PAI I ₅₃₆	29 (29)	23 (23)	1.33 (0.72-2.46)	0.356
PAI-I _{AL862}	75 (74)	72 (71)	1.17 (0.62-2.19)	0.631
PAI I _{APEC-O1}	53 (52)	61 (60)	0.74 (0.43-1.27)	0.278
PAI I _{CFT073}	25 (25)	24 (24)	1.06 (0.54-2.10)	0.862
PAI II ₅₃₆	44 (44)	45 (45)	0.96 (0.56-1.66)	0.889
PAI II _{APEC-O1}	39 (39)	49 (49)	0.69 (0.40-1.18)	0.176
PAI II _{CFT073}	46 (46)	50 (50)	0.86 (0.50-1.48)	0.579
PAI III ₅₃₆	44 (44)	44 (44)	1.00 (0.55-1.81)	1.000
PAI III _{APEC-O1}	2 (2)	6 (6)	0.33 (0.07-1.65)	0.178
PAI IV _{APEC-O1}	81 (80)	86 (85)	0.71 (0.34-1.48)	0.356
PAI V ₅₃₆	55 (54)	57 (56)	0.92 (0.52-1.62)	0.773
TAI	12 (12)	6 (6)	2.00 (0.75-5.33)	0.166
PAI Score (Me; IQR)	9 (4-12)	9 (4-12)	1.00 (0.94-1.06)	0.988

Me, median; IQR, interquartile range, OR, odds ratio, CI, confidence interval

Supplementary Table 5. Crude association between the prevalence of plasmids replicons and presentation as sepsis/septic shock in patients with *Escherichia coli* bacteraemia. Data are number of isolates (percentage per column).

Plasmid Replicon	Accession Number	Cases (n=101)	Controls (n=101)	OR (95% CI)	P value
IncFIA	AP001918	22 (22)	23 (23)	0.94 (0.48-1.86)	0.862
IncFIB (AP001918)	AP001918	62 (61)	73 (72)	0.54 (0.28-1.06)	0.075
IncFIB (H89-Phage plasmid)	HG530657	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
IncFIB (pLF82 Phage plasmid)	CU638872	0 (0)	2 (2)	0.02 (0.00-1327.4)	0.471
IncFIC (FII)	AP001918	12 (12)	15 (15)	0.77 (0.34-1.75)	0.533
IncFII	AY458016	34 (34)	38 (38)	0.85 (0.48-1.49)	0.564
IncFII (pRSB107)	AJ851089	9 (9)	12 (12)	0.70 (0.27-1.84)	0.469
IncFII (pHN7A8)	JN232517	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
IncFII (29)	CP003035	11 (11)	10 (10)	1.11 (0.45-2.73)	0.819
IncFII (pCoo)	CR942285	4 (4)	2 (2)	2.00 (0.37-10.92)	0.423
IncFII (pAMA1167 NDM-5)	CP024805	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
IncQ1	M28829	13 (13)	12 (12)	1.09 (0.48-2.47)	0.835
IncX1	EU370913	5 (5)	5 (5)	1.00 (0.29-3.45)	1.000
Incl1	AP005147	10 (10)	5 (5)	2.00 (0.68-5.85)	0.206
Incl2	KP347127	2 (2)	1 (1)	2.00 (0.18-22.6)	0.571
IncB/O/K/Z	GU256641	15 (15)	12 (12)	1.27 (0.58-2.80)	0.549
Col(BS512)	NC010656	5 (5)	4 (4)	1.25 (0.34-4.65)	0.739
Col(KHPS6)	NC016841	2 (2)	3 (3)	0.67 (0.11-3.99)	0.657
ColpVC	JX133088	4 (4)	1 (1)	4.00 (0.45-35.79)	0.215
Col156	NC009781	30 (30)	30 (30)	1.00 (0.53-1.89)	1.000
Col440II	CP023921	3 (3)	2 (2)	1.50 (0.25-8.98)	0.657
Col(MG828)	NC008486	6 (6)	8 (8)	0.75 (0.26-2.16)	0.594
IncN	AY046276	5 (5)	2 (2)	2.50 (0.49-12.89)	0.273
IncY	K02380	6 (6)	2 (2)	3.00 (0.61-14.86)	0.178
IncX4	FN543504	2 (2)	3 (3)	0.67 (0.11-3.99)	0.657
pO111	AP010962	1 (1)	3 (3)	0.33 (0.03-3.20)	0.341
Plasmids score, Me (RIC)		3 (1-4)	3 (2-4)	1.00 (0.84-1.20)	0.964

Me, median; IQR, interquartile range

Supplementary Table 6. Crude association between the prevalence of prophage sequences and presentation as sepsis/septic shock in patients with *Escherichia coli* bacteraemia. Data are number of isolates (percentage per column).

Plasmid Replicon	Cases (n=101)	Controls (n=101)	OR (95% CI)	P value
Burkho BcepMu	0 (0)	1 (1)	0.02 (0.00-147346.84)	0.610
Edward GF 2	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
Entero BP-4795	15 (15)	17 (17)	0.86 (0.40-1.85)	0.695
Entero cdtI	1 (1)	5 (5)	0.20 (0.02-1.71)	0.142
Entero DE3	10 (10)	11 (11)	0.90 (0.37-2.21)	0.819
Entero fiAA91	11 (11)	4 (4)	3.33 (0.92-12.11)	0.067
Entero HK140	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
Entero HK544	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
Entero HK630	3 (3)	1 (1)	3.00 (0.31-28.84)	0.341
Entero I2 2	1 (1)	4 (4)	0.25 (0.03-2.24)	0.215
Entero IME10	2 (2)	0 (0)	65.29 (0.00-5658381.75)	0.471
Entero lambda	3 (3)	6 (6)	0.40 (0.08-2.06)	0.273
Entero mEp237	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
Entero mEp460	40 (40)	36 (36)	1.21 (0.66-2.22)	0.538
Entero P2	0 (0)	2 (2)	0.02 (0.00-1327.44)	0.471
Entero P4	5 (5)	6 (6)	0.80 (0.21-2.98)	0.739
Entero P88	20 (20)	16 (16)	1.31 (0.64-2.69)	0.467
Entero phiP27	0 (0)	2 (2)	0.02 (0.00-1327.44)	0.471
Entero pro483	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
Entero Sf101	14 (14)	19 (19)	0.74 (0.37-1.47)	0.386
Entero Sfi	2 (2)	1 (1)	2.00 (0.18-20.06)	0.571
Entero SfV	6 (6)	5 (5)	1.25 (0.34-4.65)	0.739
Entero Wphi	0 (0)	1 (1)	0.02 (0.00-147346.84)	0.610
Entero_Mu	1 (1)	2 (2)	0.50 (0.05-5.51)	0.571
Escher 500465_1	6 (6)	6 (6)	1.00 (0.32-3.10)	1.000
Escher D108	3 (3)	0 (0)	65.29 (0.01-702391.47)	0.378
Escher HK639	0 (0)	2 (2)	0.02 (0.00-1327.44)	0.471
Escher Lys12581	0 (0)	1 (1)	0.02 (0.00-147346.84)	0.610
Escher phiV10	7 (7)	9 (9)	0.75 (0.26-2.16)	0.594
Escher pro147	0 (0)	1 (1)	0.02 (0.00-147346.84)	0.610
Escher pro483	1 (1)	3 (3)	0.33 (0.03-3.20)	0.341
Escher SH2026SxT1	0 (0)	4 (4)	0.02 (0.00-47.51)	0.308
Escher TL 2011b	7 (7)	6 (6)	1.17 (0.39-3.47)	0.782
Escher vB EcoM ECO1230-10	2 (2)	0 (0)	65.29 (0.00-5658381.75)	0.471
Escher vB_EcoM 12474III	4 (4)	0 (0)	65.29 (0.02-202501-58)	0.308
Haemop Su-Mu	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
Klebsi_4LV2017	17 (17)	19 (19)	0.88 (0.43-1.79)	0.715
Pectob ZF40	13 (13)	12 (12)	1.11 (0.45-2.73)	0.819
Pseudo phiPSA1	6 (6)	9 (9)	0.67 (0.24-1.87)	0.442
Pseudo PPpW 3	2 (2)	1 (1)	2.00 (0.18-20.06)	0.571
Salmon 118970 sal3	9 (9)	4 (4)	2.67 (0.71-10.05)	0.147
Salmon SEN34	3 (3)	9 (9)	0.33 (0.09-1.23)	0.099
Salmon SEN4	2 (2)	0 (0)	65.29 (0.00-5658381.75)	0.471
Salmon SJ46	2 (2)	1 (1)	2.00 (0.18-20.06)	0.571
Salmon SSV5	1 (1)	3 (3)	0.33 (0.03-3.20)	0.341
Salmon SW9	0 (0)	3 (3)	0.02 (0.00-164.78)	0.378
Salmon_RE_2010	2 (2)	1 (1)	2.00 (0.18-20.06)	0.571
Shigel Sf6	2 (2)	2 (2)	1.00 (0.14-7.10)	1.000
Shigel SfiI	14 (14)	19 (19)	0.72 (0.35-1.47)	0.371
Vibrio 12B12	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
Vibrio X29	4 (4)	4 (4)	1.00 (0.20-4.95)	1.000
Yersin L-413C	10 (10)	8 (8)	1.29 (0.48-3.45)	0.618
Phage score, Me (IQR)	3 (2-3)	3 (2-4)	0.95 (0.80-1.14)	0.611

Me, median; IQR, interquartile range

Supplementary Table 7. Crude association between the proportion of antibiotic resistance and presentation as sepsis/septic shock in patients with *Escherichia coli* bacteraemia. Data are number of isolates (percentage per column).

Antibiotic	Cases (n=101)	Controls (n=101)	OR (95% CI)	P value
Amikacin	5 (5)	5 (5)	1.00 (0.25-4.00)	1.000
Amoxicillin/clavulanic acid	30 (30)	36 (36)	0.78 (0.44-1.38)	0.388
Cefepime	10 (10)	4 (4)	3.00 (0.81-11.08)	0.099
Cefotaxime/Ceftriaxone	13 (13)	5 (5)	3.67 (1.02-13.14)	0.046
Cefoxitin	8 (8)	3 (3)	3.50 (0.73-16.85)	0.118
Ceftazidime	17 (17)	4 (4)	7.50 (1.72-32.80)	0.007
Cefuroxime	19 (19)	8 (8)	3.75 (1.24-11.30)	0.019
Ciprofloxacin	43 (43)	29 (29)	2.08 (1.07-4.03)	0.030
Fosfomycin	3 (3)	6 (6)	0.50 (0.13-2.00)	0.327
Gentamicin	12 (12)	13 (13)	0.91 (0.39-2.14)	0.827
Nalidixic acid	50 (50)	38 (38)	1.80 (0.96-3.38)	0.068
Piperacillin/tazobactam	8 (8)	3 (3)	2.67 (0.71-10.05)	0.147
Trimethoprim-sulfamethoxazole	38 (38)	28 (28)	1.59 (0.87-2.91)	0.135

OR, odds ratio; CI, confidence interval

Supplementary Table 8. Crude association between antibiotic resistance genes and presentation as sepsis/septic shock in patients with *Escherichia coli* bacteraemia. Data are number of isolates (percentage per column).

Category	Cases (n=101)	Controls (n=101)	OR (95% CI)	P value
Aminoglycoside resistance				
<i>aac(3)-IIa</i>	2 (2)	2 (2)	1.00 (0.06-15.99)	1.000
<i>aac(3)-IId</i>	4 (4)	5 (5)	0.75 (0.17-3.35)	0.706
<i>aac(6')-Ib-cr</i>	6 (6)	3 (3)	2.50 (0.49-12.89)	0.273
<i>aadA1</i>	18 (18)	17 (17)	1.07 (0.52-2.22)	0.853
<i>aadA2</i>	6 (6)	4 (4)	1.67 (0.40-6.97)	0.484
<i>aadA5</i>	11 (11)	7 (7)	1.80 (0.60-5.37)	0.292
<i>aph(3')-Ia</i>	3 (3)	2 (2)	1.50 (0.25-8.98)	0.657
<i>aph(3'')-Ib</i>	27 (27)	29 (29)	0.91 (0.51-1.65)	0.763
<i>aph(6)-Id</i>	28 (28)	27 (27)	1.05 (0.58-1.90)	0.879
Betalactam resistance				
<i>bla</i> _{CTX-M-1}	1 (1)	0 (0)	65.29 (0.00-628084630.37)	0.610
<i>bla</i> _{CTX-M-14}	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
<i>bla</i> _{CTX-M-15}	7 (7)	3 (3)	3.00 (0.61-14.86)	0.178
<i>bla</i> _{CMY-2}	2 (2)	0 (0)	65.29 (0.00-5658381.75)	0.471
<i>bla</i> _{OXA-1}	7 (7)	5 (5)	1.50 (0.42-5.32)	0.530
<i>bla</i> _{OXA-48}	1 (1)	0 (0)	65.29 (0.00-628084630.37)	0.610
<i>bla</i> _{SHV-12}	1 (1)	0 (0)	65.29 (0.00-628084630.37)	0.610
<i>bla</i> _{TEM-1A}	1 (1)	4 (4)	0.25 (0.03-2.24)	0.215
<i>bla</i> _{TEM-1B}	35 (35)	31 (31)	1.17 (0.67-2.05)	0.572
<i>bla</i> _{TEM-1C}	8 (8)	14 (14)	0.57 (0.24-1.36)	0.207
<i>bla</i> _{TEM-1D}	3 (3)	1 (1)	3.00 (0.31-28.84)	0.341
Phenicol resistance				
<i>catA1</i>	6 (6)	4 (4)	1.50 (0.42-5.32)	0.530
<i>cmlA1</i>	5 (5)	1 (1)	65.29 (0.02-202501.58)	0.308
<i>catB3</i>	5 (5)	3 (3)	2.00 (0.37-10.92)	0.423
Ttrimethoprim resistance				
<i>dfrA1</i>	9 (9)	9 (9)	1.00 (0.38-2.66)	1.000
<i>dfrA12</i>	4 (4)	3 (3)	1.50 (0.25-8.98)	0.657
<i>dfrA14</i>	7 (7)	5 (5)	1.40 (0.44-4.41)	0.566
<i>dfrA17</i>	11 (11)	9 (9)	1.29 (0.48-3.45)	0.618
<i>dfrA5</i>	1 (1)	1 (1)	1.00 (0.06-15.99)	1.000
<i>dfrA7</i>	3 (3)	0 (0)	65.29 (0.01-702391.47)	0.378
Sulphonamide resistance				
<i>sul1</i>	26 (26)	19 (19)	1.50 (0.76-2.95)	0.240
<i>sul2</i>	30 (30)	33 (33)	0.88 (0.49-1.57)	0.655
<i>sul3</i>	6 (6)	2 (2)	5.00 (0.58-42.80)	0.142
Tetracycline resistance				
<i>tetA</i>	18 (18)	25 (25)	0.59 (0.27-1.28)	0.183
<i>tetB</i>	10 (10)	10 (10)	1.00 (0.40-2.52)	1.000
Nalidixic Acid, Ciprofloxacin				
<i>qnrS1</i>	3 (3)	1 (1)	3.00 (0.31-28.84)	0.341
<i>parC</i> (Ser80Ile)	30 (30)	26 (26)	1.22 (0.66-2.28)	0.528
<i>parC</i> (SGlu84Val)	15 (15)	12 (12)	1.37 (0.55-3.42)	0.493
<i>gyrA</i> (Ser83Leu)	47 (47)	36 (36)	1.79 (0.93-3.44)	0.082
<i>gyrA</i> (Asp87Asn)	32 (32)	23 (23)	1.64 (0.85-3.19)	0.143
<i>parE</i> (Ile529Leu)	15 (15)	11 (11)	1.50 (0.61-3.67)	0.374
<i>parE</i> (Ser458Ala)	7 (7)	2 (2)	3.50 (0.73-16.85)	0.118
Others				
<i>mphA</i> (Macrolide)	11 (11)	7 (7)	1.80 (0.60-5.37)	0.292
<i>mphB</i> (Macrolide)	2 (2)	2 (2)	1.00 (0.06-15.99)	1.000
Resistance genes score, Me (IQR)	4 (1-8)	3 (0-7)	1.06 (0.99-1.14)	0.114

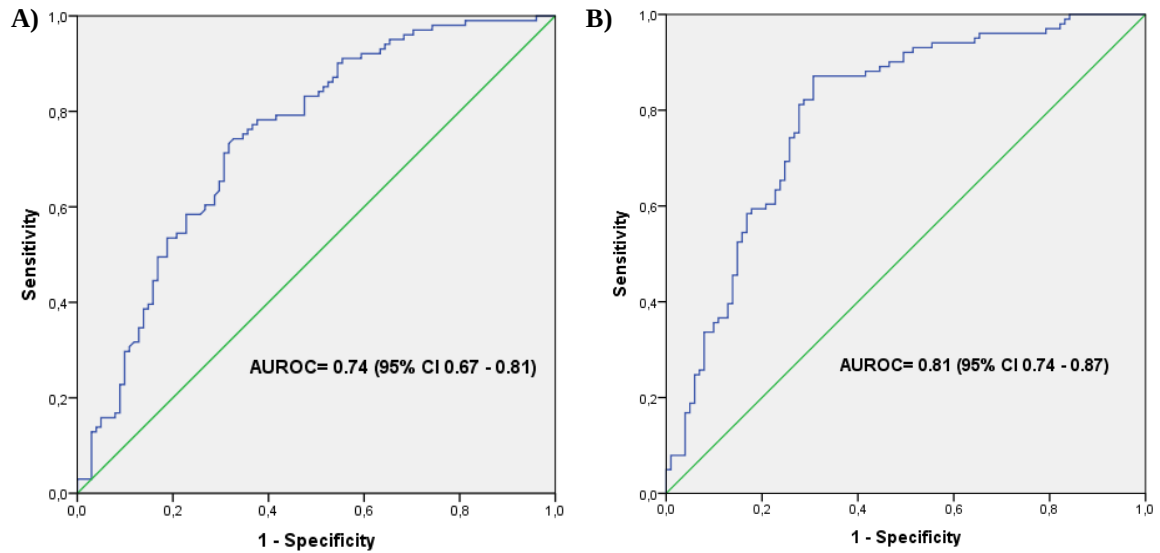
Me, median; IQR, interquartile range; OR, odds ratio, CI, confidence interval

Supplementary Table 9. Multivariate conditional logistic regression model of *Escherichia coli* microbiological factors associated with presentation as sepsis/septic shock in patients with bacteraemia, controlling for the propensity score of clinical and epidemiological variables.

Variable	Adjusted OR (95% CI)	<i>P</i> value
ST69	7.53 (1.06-35.05)	0.010
<i>pic</i>	4.38 (1.53-12.54)	0.006
<i>fdeC</i>	0.18 (0.03-1.32)	0.092
<i>papC</i>	0.30 (0.12-0.74)	0.009

OR, odds ratio; CI, confidence interval

Supplementary Fig. 3. ROC curves illustrating the prediction of sepsis/septic shock presentation in patients with *Escherichia coli* bacteraemia, based on A) the propensity score derived from clinical and epidemiological variables and B) multivariate conditional logistic regression model of microbiological factors, controlling for the propensity score.



AUROC, Area Under a Receiver Operating Characteristic Curve; CI, confidence interval

Supplementary Table 10. Matrix of the linkage between virulence clusters and sequence types

ST, n (%)*	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E	Cluster F	Cluster G	Cluster H
ST117	0 (0%)	0 (0%)	0 (0%)	3 (60%)	0 (0%)	0 (0%)	0 (0%)	2 (40%)
ST12	2 (11.8%)	0 (0%)	2 (11.8%)	0 (0%)	0 (0%)	0 (0%)	13 (76.5%)	0 (0%)
ST141	0 (0%)	0 (0%)	0 (0%)	5 (83.3%)	0 (0%)	0 (0%)	1 (16.7%)	0 (0%)
Other ST	1 (1.2%)	20 (24.4%)	2 (2.4%)	6 (7.3%)	1 (1.2%)	27 (32.9%)	4 (4.9%)	21 (25.6%)
ST131	0 (0%)	12 (42.9%)	13 (46.4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (10.7%)
ST69	0 (0%)	9 (56.2%)	1 (6.2%)	0 (0%)	0 (0%)	3 (18.8%)	0 (0%)	3 (18.8%)
ST73	21 (80.8%)	0 (0%)	4 (15.4%)	0 (0%)	0 (0%)	0 (0%)	1 (3.8%)	0 (0%)
ST95	0 (0%)	0 (0%)	0 (0%)	5 (22.7%)	17 (77.3%)	0 (0%)	0 (0%)	0 (0%)

ST, Sequence Type; n, number

*Percentages are calculated row wise