

From colonization to invasion: genomic and phenotypic comparison of faecal and bloodstream isolates from the same patients

Aakash Khanijau^{1,2}, Ellie Allman¹, Ralf Pulmones¹, Richard N. Goodman¹, Daire Cantillon¹, Rachel McGalliard^{2,3}, Christopher M. Parry^{1,3,4}, Enitan D. Carrol^{2,3} and Adam P. Roberts^{1,*}

Abstract

Introduction. Gram-negative bloodstream infections (GNBSIs) carry a significant global health burden. *Escherichia coli* and *Klebsiella pneumoniae* are the two most common causes of healthcare-associated GNBSI, which may arise from gastrointestinal tract (GIT) colonization.

Gap Statement. We do not fully understand how GNBSIs arise from GIT colonization.

Aim. To understand *E. coli* and *K. pneumoniae* genomic and phenotypic adaptations that underpin transition from GIT colonization to invasive bloodstream infection.

Methodology. This study identified 'linked' faecal and blood isolates from children with healthcare-associated GNBSI caused by *E. coli* and *K. pneumoniae*. Linked pairs were compared for antimicrobial resistance and biofilm formation and underwent comparative genomic analysis via whole-genome sequencing, comparative average nucleotide identity and core genome single nucleotide polymorphism (SNP) analysis.

Results. Five isolate pairs (three *E. coli*, two *K. pneumoniae*) showed high relatedness, supporting the GIT origin of bloodstream infection. Isolates within pairs had identical virulence genes, whereas phenotypic assays revealed changes in antimicrobial susceptibility, with one pair undergoing changes in resistance gene profiles and increased biofilm formation in four out of five isolates.

Conclusion. This study provides insight into within-host evolution from gastrointestinal colonization to bloodstream invasion in Gram-negative pathogens. Convergence on metabolic adaptation and biofilm formation suggests that these traits may be advantageous in healthcare-associated GNBSI. Further studies involving larger cohorts alongside functional validation of mutations are needed to better understand GNBSI pathogenesis.

DATA SUMMARY

The BioProject number for this study is PRJNA1295786, which contains the ten accession numbers for the individual genomes SRX29836134–SRX29836143.

Received 16 September 2025; Accepted 03 March 2026; Published 08 April 2026

Author affiliations: ¹Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK; ²Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK; ³Alder Hey Children's NHS Foundation Trust, Liverpool, UK; ⁴Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

*Correspondence: Adam P. Roberts, Adam.Roberts@lstm.ac.uk

Keywords: antimicrobial resistance (AMR); biofilm; bloodstream infection; *Escherichia coli*; *Klebsiella*; plasmids.

Abbreviations: AMR, antimicrobial resistance; ANI, average nucleotide identity; ARGs, antimicrobial resistance genes; AST, antimicrobial susceptibility testing; cgMLST, core genome multilocus sequence typing; CLABSIs, central line-associated bloodstream infections; COG, cluster of orthologous groups; ESBL, extended-spectrum beta-lactamase; GIT, gastrointestinal tract; GNBSIs, Gram-negative bloodstream infections; indels, insertions/deletions; iSNPs, intergenic SNPs; LB, Luria-Bertani; MHB2, Mueller–Hinton broth 2; nsSNPs, non-synonymous SNPs; SNP, single nucleotide polymorphism; sSNPs, synonymous SNPs; WGS, whole-genome sequencing.

Three supplementary figures and four supplementary tables are available with the online version of this article.

002147 © 2026 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

INTRODUCTION

Gram-negative bloodstream infections (GNBSIs) contribute significantly to the global burden of paediatric sepsis, accounting for 30–50% of all bloodstream infections worldwide [1, 2]. Healthcare-associated bloodstream infections, such as central line-associated bloodstream infections (CLABSIs), often affect neonates or other critically ill children and are associated with significant mortality [3].

The two most clinically significant organisms implicated in GNBSI are *Klebsiella pneumoniae* and *Escherichia coli* [4, 5]. The emergence of antimicrobial resistance (AMR) in these pathogens presents a significant challenge. Particularly, extended-spectrum beta-lactamase (ESBL)-producing isolates exhibit resistance to third-generation cephalosporins and beta-lactam/beta-lactamase inhibitor combinations. As such, the World Health Organization has designated ESBL-producing *K. pneumoniae* and *E. coli* as priority pathogens for which new antibiotics are required [6].

Both *K. pneumoniae* and *E. coli* are persistent colonizers of the human gastrointestinal tract (GIT), which is believed to be an important reservoir for subsequent GNBSI [7]. The transition from GIT colonization to GNBSI can occur via invasion of other body sites or medical devices, leading to secondary bloodstream infection, or direct translocation across the gut mucosa, causing primary bloodstream infection. The latter is implicated in 20–50% of cases of healthcare-associated infections [7]. The transition from GIT colonization to bloodstream invasion depends on host, clinical and environmental influences, combined with specific pathogen characteristics. Host factors that can predispose to GIT-bloodstream transition include immunosuppression, critical illness and antimicrobial exposure [7, 8]. Pathogen factors include advantageous phenotypes such as AMR, bloodstream fitness and biofilm formation, which may allow invasion, survival and dissemination in the bloodstream. Biofilm formation is of particular interest in the context of CLABSI, as growth within a biofilm provides protection against host defence and antimicrobial agents and may act as a nidus for ongoing bloodstream dissemination.

Many antimicrobial resistance genes (ARGs) and virulence genes underpinning AMR and invasive infection, respectively, are well characterized. However, predisposing genotypes and specific genetic changes undergone within-host that may underpin GIT-blood transition are seldom studied, perhaps due to the scarcity of such paired isolates from the same patient. This includes genomic adaptations like acquisition or loss of virulence genes and plasmids, and regulatory or metabolic alterations. Furthermore, the relationship between many genomic adaptations and advantageous phenotypic changes remains unclear, as well as the relationship between AMR and invasive infection. Combined phenotypic and genomic comparison of faecal and bloodstream isolates, obtained from the same patient, could improve understanding of intra-host evolution involved in the pathogenesis of GNBSI and translate to informing intervention strategies.

In this study, we aimed to identify genomically ‘linked’ pairs of faecal and blood Gram-negative isolates, which were likely to have undergone GIT-blood transition and evolution within patients to cause GNBSI. For each ‘linked’ pair, we investigated if there had been any loss or acquisition of virulence genes, ARGs and plasmids between faecal isolates and their blood counterpart. We then investigated if there were any differences in two phenotypes of interest: AMR and biofilm formation. Finally, we used a comparative genomic approach to identify further genomic changes that could explain any observed phenotypic differences or otherwise contribute to the transition from GIT colonization to bloodstream invasion.

METHODS

Clinical isolates, clinical data collection and ethics

A library of 81 stored bacterial isolates from Alder Hey Children’s Hospital, Liverpool, UK, was interrogated for pairs of ESBL-producing *E. coli* and *K. pneumoniae* isolates from the same patient, where the same pathogen had been identified on a routine faecal surveillance culture and in a subsequent blood culture. Isolates were stored in LB (Luria-Bertani) broth (Sigma-Aldrich) and 20% glycerol at –80 °C. Anonymized clinical information was obtained from patient records, including demographics, source of GNBSI and antimicrobial exposure between isolate dates. Ethical approval for linking anonymized data was granted by the University of Liverpool as part of a separate study for which the isolate library was created (REC reference 255669).

Whole-genome sequencing and comparative genomics

Short-read whole-genome sequencing (WGS) and assembly were performed at the Centre of Genomic Research (University of Liverpool) using the Illumina NovaSeq SP (2×150 bp). Raw reads were trimmed for adapter sequences using Cutadapt v1.2.1 [9]. Reads were then quality-trimmed using Sickle v1.200 (<https://github.com/najoshi/sickle>) with a minimum window quality score of Q20, and reads shorter than 20 bp after trimming were discarded. Trimmed reads were assembled using SPAdes v3.13.1 [10] with default parameters. Contigs were filtered to remove any shorter than 200 bp in length. Assembled genomes were analysed for the following characteristics, which were compared between faecal and blood isolates of each patient pair. Core genome multi-locus sequence typing (cgMLST) was performed using Kleborate [11] for *K. pneumoniae* isolates and using the Institut de Pasteur database (<https://bigsd.b.pasteur.fr/about/>) for *E. coli* isolates. Kaptive [12] was used to identify wzi-type, K-type and O-type of *K. pneumoniae* isolates, and SeroTyper 2.0 and CHType to determine the H-type, O-type and FumC-type of *E. coli* isolates.

To assess relatedness further, average nucleotide identity (ANI) was then compared between isolates using OrthoANI [13]. Following this, the number of core genome single nucleotide polymorphisms (SNPs) between faecal and blood isolates of each pair was calculated. SNPs were predicted using Snippy v.4.3.6 (<https://github.com/tseemann/snippy>) by aligning raw reads of assemblies to the reference genomes MG1655 and HS11286, for *E. coli* and *K. pneumoniae*, respectively. Core genome SNPs were extracted with Snippy-core, and intra-species SNPs were plotted and counted in R v.4.3.1 [14]. ResFinder 4.2.2 [15] was used to determine the presence of ARGs and provide a genotypic prediction of resistance for each isolate. VF analyser (<http://www.mgc.ac.cn/VFs/>) was used to identify virulence genes in each isolate. PlasmidFinder [16] was used to identify the presence of plasmids. Isolate genomes were annotated using the RAST annotation tool [17], and pairs underwent comparative genomic analysis using Breseq version 0.38.3 [18] using default parameters. Mutations between blood isolates and the putative counterpart faecal ancestor were categorized as insertions/deletions (indels), non-synonymous SNPs (nsSNPs), synonymous SNPs (sSNPs) or intergenic SNPs (iSNPs). Mutations within or <100 bp upstream of a gene were classified into Cluster of Orthologous Groups (COG) functional categories using EGGNOG 5.0. Mutations involving hypothetical proteins were classified as 'Function unknown', and mutations without genes within 100 bp were 'Unclassified.' Any mutations relating to virulence genes, ARGs or previously described bloodstream survival factors were identified. In addition, mutations in genes previously reported to affect AMR or biofilm formation were identified.

Antimicrobial susceptibility testing

Six antimicrobials from the contemporaneous local prescribing guidance for Gram-negative infections were used for antimicrobial susceptibility testing (AST): cefotaxime, gentamicin, ciprofloxacin, amoxicillin-clavulanic acid (co-amoxiclav), piperacillin-tazobactam and meropenem. Kirby–Bauer disc diffusion testing was used for initial AST following EUCAST guidelines [19]. Zone sizes were interpreted as per the EUCAST clinical breakpoints, to give a phenotype of sensitive (S), resistant (R) or in the 'Area of Technical Uncertainty' [20]. Three technical and three biological repeats were performed, and the mode was recorded as the phenotype for each isolate and compared within each pair. In addition, the mean zone size, along with SEM, was recorded for each isolate, and the Mann–Whitney U test was used to compare differences between faecal and blood isolates for each patient pair. Any identified differences between faecal and blood isolates in each patient pair were confirmed using broth microdilution to quantify the fold change in MIC. Broth microdilution was performed following EUCAST guidelines [21] and with three biological and three technical replicates.

Biofilm formation

Crystal violet assay was performed, derived from two previously published protocols [22, 23]. Isolates were incubated at 37 °C on LB agar for 18 h. A single colony of each isolate was transferred to 10 ml of Mueller–Hinton broth 2 (MHB2; Sigma-Aldrich) and incubated at 37 °C at 200 r.p.m. for a further 18 h. Cultures were diluted to an OD₆₀₀ of 0.01 in fresh MHB2. One hundred and fifty microlitres of each isolate were transferred in a U-bottom plate, with 150 µl sterile MHB2 used as a negative control. After 24 h incubation, the biomass was transferred to a flat-well plate, which was used to measure OD₆₀₀. The remaining biofilm in the U-bottom plate wells was stained with 1% crystal violet and subsequently solubilized with 30% acetic acid. The final OD₆₀₀ was measured in a flat-well plate. This value was then normalized to the initial OD₆₀₀ of the biomass to quantify biofilm formation. The assay was performed as three biological replicates and four technical replicates. Mean OD₆₀₀ of blood and faecal isolates was compared using a paired t-test, and SEM was calculated.

Statistical analysis and data visualization

All statistical analysis and data visualization were performed using ggplot2 in R version 4.3.2 [14].

RESULTS

Clinical isolates and clinical context of GNBSI

Five pairs of isolates, two *K. pneumoniae* and three *E. coli*, were identified for use in this study (Table 1). Time between faecal (first sample) and blood sample isolation ranged from 17 to 144 days. Patients from whom isolates were derived were aged 1 month to 12 years. All five cases of GNBSI were diagnosed as CLABSI, and four out of five cases occurred in the intensive care unit (ICU). Patients received between two and six different antimicrobials between faecal and blood sample isolation.

Genomic relatedness of faecal–blood isolate pairs

Genome sequencing statistics are detailed in Table S1, available in the online version of this article. To initially assess relatedness between faecal and blood isolates, cgMLST was performed. In five out of five pairs, faecal and blood isolates shared the same sequence type (Table 2). In Pair 1, *K. pneumoniae* isolates were from the sequence type ST922, and in Pair 2, *K. pneumoniae* isolates were from the sequence type ST461 (Table 2). In Pairs 3 and 4, *E. coli* isolates were from the globally dominant sequence type ST131, whereas in Pair 5, isolates were from sequence type 12 (Table 3). *K. pneumoniae* isolates were then compared for K-type, O-type and wzi-type, which were concordant between faecal and blood isolates in both pairs. *E. coli* isolates were compared for O-type, H-type and FumC-type, and again, all were concordant within each pair.

Table 1. Characteristics of isolate pairs and clinical data of patients with GNBSI

Pair	Species of isolate pair	Days from faeces – blood isolation	Age at GNBSI	CLABSI	ICU patient at time of GNBSI	Antibiotic exposure between faeces and blood isolate
1	<i>K. pneumoniae</i>	144	12 years	Y	N	Ciprofloxacin, clindamycin, ceftriaxone, piperacillin-tazobactam, co-trimoxazole, gentamicin
2	<i>K. pneumoniae</i>	87	11 months	Y	Y	Ciprofloxacin, cefalexin, piperacillin-tazobactam, ceftriaxone, vancomycin, gentamicin
3	<i>E. coli</i>	185	9 months	Y	Y	Co-amoxiclav, cefalexin
4	<i>E. coli</i>	31	12 months	Y	Y	Ciprofloxacin, meropenem
5	<i>E. coli</i>	17	1 month	Y	Y	Ciprofloxacin, vancomycin

Next, relatedness was further assessed using ANI (Fig. 1). In 4 out of 5 pairs (2, 3, 4 and 5), the faecal–blood isolates shared an ANI of 99.99%, whereas isolates in Pair 1 shared an ANI of 99.82%. Within a comparison between all isolates across all patient pairs, each faecal isolate was most closely related to its corresponding blood isolate from the same patient. Given these findings, all pairs were at this point considered to be genomically ‘linked’ and included for further study.

Isolates were next compared for core genome SNP differences (Fig. 1 and Table S2). For *K. pneumoniae* isolates, Pair 2 exhibited the lowest SNP difference (38 SNPs), whilst Pair 1 showed the highest (318 SNPs). In *E. coli* isolates, within-pair SNP distances ranged from 40 SNPs in Pair 4 to 60 SNPs in Pair 3. These findings were in alignment with the ANI comparison.

Initial genomic comparison: ARGs, virulence genes and plasmids

At this point, all pairs were considered to be ‘linked’, and an initial genomic comparison of ARGs, virulence genes and plasmids of the faecal and blood isolates within each pair was performed (Fig. 2). There were between 1 and 13 ARGs detected in each isolate, predicted to confer resistance to various drug classes (Fig. 2a). The presence of at least one ESBL gene was confirmed in each isolate. ARG profiles were identical within pairs, except for Pair 4, in which *bla*TEM-1B was detected in the blood isolate but not the faecal counterpart, suggesting acquisition between the dates of sampling, possibly during GIT–blood translocation.

There were between 51 and 86 virulence genes detected in each isolate, encoding multiple classes of virulence factors (Fig. 2b, c). Virulence gene presence was identical between faecal and blood isolates, indicating no loss or acquisition of virulence genes in GIT–blood transition in any pair.

There were between three and eight plasmids detected in each isolate (Fig. 2d). Whilst the majority of pairs demonstrated identical plasmid profiles, indicating stable plasmid carriage during the likely GIT–blood transition, Pair 1 demonstrated loss of IncFIB(K) from the faecal isolate and gain of IncN and IncFII(Yp) plasmids in the blood isolate.

Phenotypic comparison: AMR

Disc diffusion testing was used to screen for a change in AMR between faecal and blood isolates of each pair. A change in resistance phenotype was identified for three drug–bug combinations in two out of five pairs (Fig. S1). There was also a statistically significant difference in zone size between faecal and blood isolates that did not cross the EUCAST clinical breakpoint for four drug–isolate combinations in two out of five pairs (Fig. S2 and Table S3). These within-pair differences were only confirmed by broth microdilution in two pairs. Pair 2 demonstrated an increase in MIC for ciprofloxacin, piperacillin-tazobactam, cefotaxime and gentamicin in the blood isolate compared to its faecal counterpart, indicating increased resistance development to these four

Table 2. MLST, K-type, O-antigen type and wzi-type of *K. pneumoniae* isolates within pairs, determined using Kleborate and Kaptive

Species	<i>K. pneumoniae</i>			
Pair	Pair 1		Pair 2	
Isolate	Faeces	Blood	Faeces	Blood
MLST	ST922	ST922	ST461	ST461
K-type	K28	K28	K25	K25
O-type	O1	O1	O1	O1
wzi-type	wzi84	wzi84	wzi25	wzi25

Table 3. Phylotype and MLST of *E. coli* isolates within pairs using ClermonTyping and Institut de Pasteur database

Species	<i>E. coli</i>					
	Pair 3		Pair 4		Pair 5	
Pair						
Isolate	Faeces	Blood	Faeces	Blood	Faeces	Blood
MLST	ST131	ST131	ST131	ST131	ST12	ST12
FumC-type	C40	C40	C40	C40	C13	C13
O-type	H4	H4	H4	H4	H1	H1
H-type	O25	O25	O25	O25	O4	O4

antimicrobials between the two sampling timepoints (Fig. 3a). In Pair 3, MIC was decreased for piperacillin-tazobactam in the blood isolate compared to faecal, indicating increased susceptibility development in GIT–blood transition (Fig. 3b).

Phenotypic comparison: biofilm formation

Because all patients were found to be diagnosed with CLABSI, and given the importance of biofilm formation in CLABSI pathogenesis, we investigated for any difference in biofilm formation between faecal and blood isolates in each pair. In four out of five pairs, blood isolates demonstrated significantly increased biofilm formation in comparison to the faecal counterpart, as determined by a paired t-test (Fig. 4). All *P*-values are stated in Table S4. Overall, the isolates in Pair 2 demonstrated the strongest biofilm formation.

Comparative genomics: Breseq analysis

Breseq was used to identify mutations in the blood isolate when compared to the faecal isolate. Most mutations were indels, and the majority of these were ‘Function unknown’ or ‘Unclassified’ (Fig. S3 and Table S2). Mutations occurring within 100 bp of genes with a known function were assigned to COG categories (Fig. 5). There were five categories in which mutations occurred in all pairs. Of these, most occurred in ‘L: Replication, recombination and repair’, ‘E: Amino acid transport and metabolism’ and ‘G: Carbohydrate transport and metabolism’.

No mutations were identified in specific ARGs identified by ResFinder; however, four pairs demonstrated mutations in genes encoding efflux pumps and transporter proteins involved in AMR (Table 4). Four pairs demonstrated mutations in at least one known virulence factor gene. Two pairs demonstrated mutations in or related to previously reported bloodstream survival factors. All pairs demonstrated a variety of mutations in genes related to biofilm formation. In Pair 1, there was evidence of acquisition of another copy of the *hha* gene, along with modification of the Shine–Dalgarno intergenic region (Fig. 6).

DISCUSSION

In this study, we compared blood isolates with the previously acquired faecal isolates from five children with GNBSI and, in each instance, demonstrated concordance in sequence type and species-specific typing methods. This is highly suggestive that the faecal–blood isolate pairs were genomically ‘linked’ and that the invasive blood isolate was likely to have been acquired from the GIT. ANI has been used as a marker of genomic relatedness in previous studies comparing faecal–blood isolates in children [24, 25], but there is no defined consensus on a threshold that indicates isolates as similar enough to be linked. Thresholds of 99.99% have been previously used to indicate isolates being the same strain [26]. Using this threshold, four out of five pairs (Pairs 2–5) shared ANI values convincing of being the same strain. The lower ANI in Pair 1 (99.82%) could be somewhat explained by the identified plasmid changes in this pair. This is consistent with observations that MGE movement can lead to a lower ANI in isolates that are highly likely to be concordant, as noted by Vornhagen *et al.* when comparing *wzi*-type concordant rectal and blood *K. pneumoniae* isolates [27]. However, this finding could also be indicative of the blood isolate having evolved to be a different clade of the same sequence type as the faecal isolate in this pair, as indicated by the higher number of core genome SNPs calculated in further analysis.

The number of core genome SNPs between faecal and blood isolates in each pair ranged from 38 SNPs (Pair 2) to 318 SNPs (Pair 1), indicating varying degrees of evolutionary distance between faecal and blood isolates in each patient. It is possible that our BSI isolates, although related, are not immediate descendants of the cognate faecal isolate and that there could have been closer ancestors in the GIT which were not picked up during sampling. Similarly, there may have been evolution within the gut and the blood, giving rise to multiple descendants, of which we isolated one. A previous study that identified concordant faecal and blood isolates in patients with *E. coli* bloodstream infections noted that the number of core genome SNPs between faecal and blood isolates was 0 [28]. Another study comparing faecal and blood *K. pneumoniae* isolates found a core genome SNP difference of 6 SNPs between isolates considered to be concordant strains in 1 patient and 76 SNPs between concordant isolates from another patient [29]. In our study, the number of SNPs we found is generally higher and was not proportional to the number of days

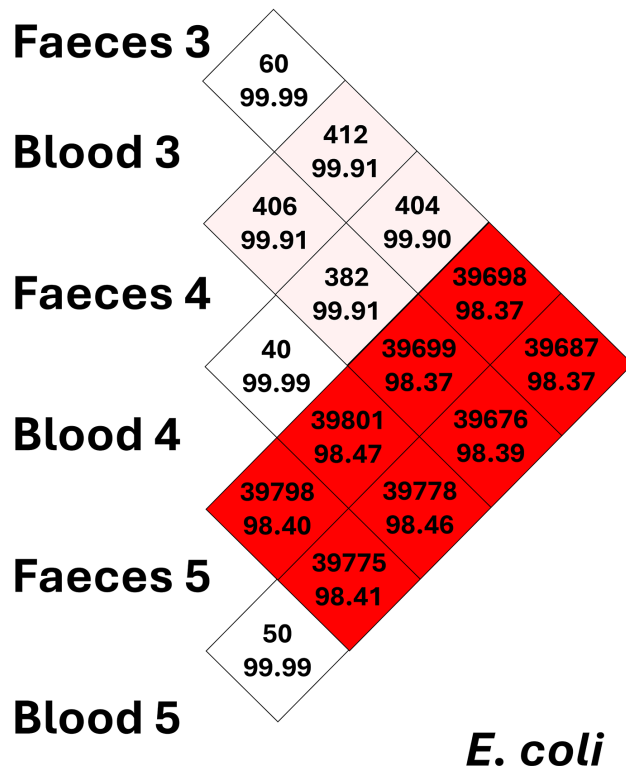
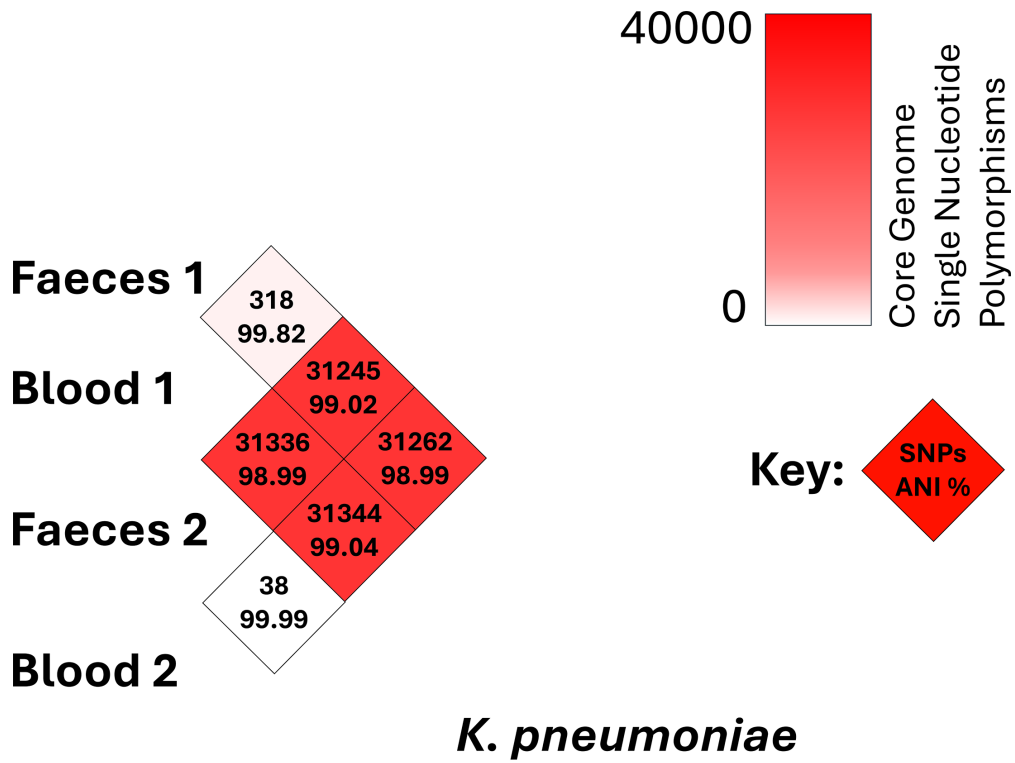


Fig. 1. Pairwise comparison of ANI and core genome SNPs between isolates.

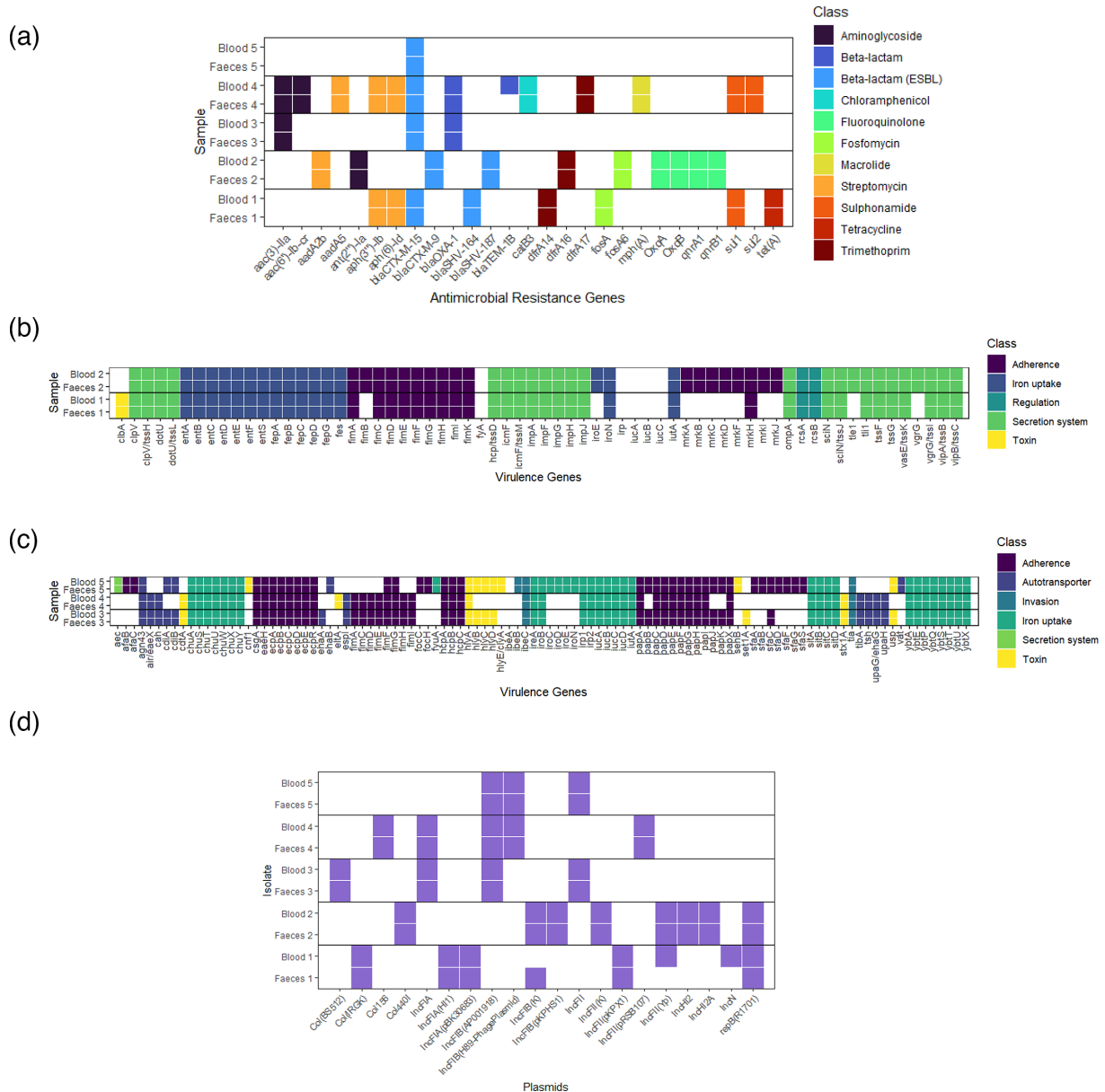


Fig. 2. (a) ARG presence in all isolates. Each tile represents the presence of an ARG as detected by ResFinder 4.2.2. using WGS data. ARGs are categorized by drug class that they are predicted to confer resistance to. (b) Virulence gene presence in *K. pneumoniae* isolates. Each tile demonstrates the presence of a virulence gene as detected by the VF analyser using WGS data. Virulence genes are categorized by class of virulence factor encoded by the gene. (c) Virulence gene presence in *E. coli* isolates. Each tile demonstrates the presence of a virulence gene as detected by the VF analyser using WGS data. Virulence genes are categorized by class of virulence factor encoded by the gene. (d) Plasmid presence in all pairs of isolates. Each tile demonstrates the presence of a plasmid replicon gene as detected by PlasmidFinder, using WGS data.

between faecal and blood isolation. This may represent varied and convoluted routes taken from the GIT to the blood, e.g. via other body tissues in each case of GNBSI, requiring adaptation to multiple niches – as opposed to direct GIT–blood translocation.

Alternatively, the higher number of core genome SNPs in Pair 1 could be considered to contradict the hypothesis that this represents a ‘linked’ pair. However, as with ANI, there is no SNP threshold that is universally agreed as defining isolates as ‘linked’. For example, whilst 1 study used a threshold of 17 core genome SNPs for this, another study looking at persistence of *E. coli* isolates between GI and urinary tracts used a threshold of ≤ 500 core genome SNPs [28, 30]. These differences could also be due to differences in sequencing and bioinformatic technologies, as well as the reference genomes used in different studies. Such differences in methodology are recognized as a key barrier to the determination of a defined threshold for strain similarity [31–33]. In addition, the expected rate of mutation of isolates moving between different body niches is currently unknown and unpredictable [34],

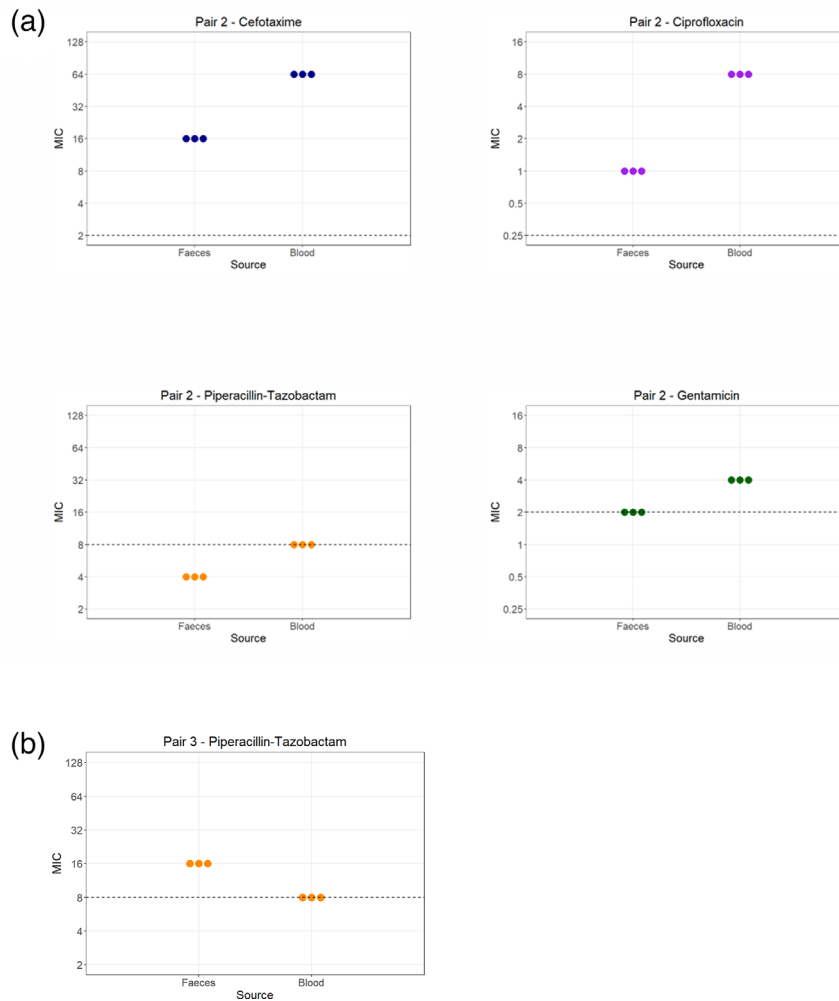


Fig. 3. (a) MIC (µg/ml) as determined by broth microdilution of ciprofloxacin, co-amoxiclav, piperacillin-tazobactam, cefotaxime and gentamicin for Pair 2. Each plotted point represents the MIC (µg/ml) of a single biological replicate in broth microdilution. Dashed line=EUCAST clinical breakpoint for MIC. (b) MIC as determined by broth microdilution of piperacillin-tazobactam for Pair 3. Each plotted point represents the MIC of a single biological replicate in broth microdilution. Dashed line=EUCAST clinical breakpoint for MIC.

adding to this challenge. When comparing isolates from the two different patients with *K. pneumoniae* infection, there are over 30,000 core genome SNPs between each. A similar number is seen when comparing the *E. coli* isolates of Patient 5 to both Patient 3 and Patient 4. However, the number of core genome SNPs between Patients 3 and 4 is more comparable to those seen within the pair of Patient 1. This raises suspicion of a common source of colonization of these patients within the intensive care unit during the time period the isolates were taken. However, granular data of overlapping hospital admission times are unavailable.

ARG profile and, subsequently, AMR phenotypes were compared within 'linked' pairs. The transition from GIT colonization to blood invasion was associated with acquisition of an ARG in one out of five pairs (Pair 4) – and in this case, the acquisition of *bla*TEM-1B did not lead to increased resistance to any of the beta-lactam/inhibitor combinations tested. Only in one pair (Pair 2) did we observe that increases in resistance to any antimicrobials occurred between the two sample timepoints. This is despite the fact that in all five cases, the patient from whom isolates were obtained received multiple antimicrobials during hospital admission. In contrast, for Pair 3, increased susceptibility to piperacillin-tazobactam was noted. These findings may reflect varying selection pressure exerted by dose and duration of antimicrobial therapy, in balance with environmental adaptation. Unfortunately, the study is limited by the lack of granular data on antimicrobial exposure. These phenotypic changes in AMR are not explained by acquisition or loss of ARGs (Fig. 2a). However, by using Breseq for whole-genome comparison, we identified deletion of drug-efflux pump repressors that could putatively explain this observation. Several drug-efflux pumps are conserved across Gram-negative species and implicated in AMR. These include MFS pumps, RND pumps and ABC transporters. The *acrAB-tolC* efflux pump is an RND pump implicated in resistance to multiple antimicrobials [35]. It has multiple regulatory pathways, including the *acrRAB* operon, in which *acrR* is a repressor, and the *marAB* operon, with *marB* as a repressor. In Pair

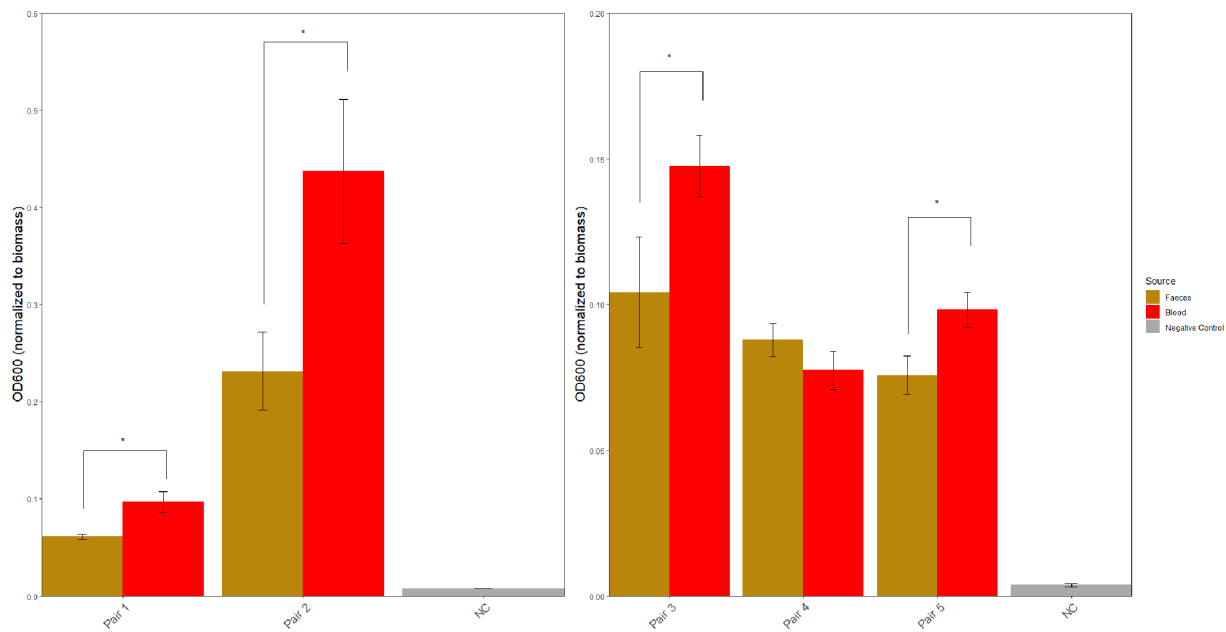


Fig. 4. Crystal violet assay for biofilm formation in a 96-well plate. Each bar represents mean OD₆₀₀ ($n=12$) for faecal and blood isolates after crystal violet staining, normalized to biomass. This is used as the measure of biofilm formation capacity. Error bars represent SEM. *=statistical significance as determined by paired t-test, $P=0.05$.

2, an nsSNP in *marB*, along with deletion of *acrR*, was identified (Table 4), which could lead to increased efflux pump activity, explaining the changes in susceptibility. The deletion of *acrR* is particularly interesting because a previous study reported that *acrR* knockout in *K. pneumoniae* isolates resulted in increased MIC to the same antimicrobial classes as seen here [36]. Previous literature also reports an association between *acrR* disruption and a multi-drug resistant phenotype [37, 38]. However, it must be noted that an *acrR* deletion was identified in Pair 5 but without a change in susceptibility, possibly explained by the variation in other efflux pump changes seen in each pair.

In five out of five cases in this study, GIT–blood transition was not associated with the acquisition of virulence genes when comparing blood isolates to the presumed faecal ancestor. Because all cases of GNBSI in this study were cases of CLABSI, biofilm formation was chosen as a virulence phenotype of interest. The finding that four out of five blood isolates exhibited an increase in biofilm formation compared to faecal counterparts suggests that in-host adaptation towards increased biofilm formation may be implicated in the pathogenesis. As noted, the observed phenotypic changes are not explained by acquisition of virulence genes involved in biofilm formation, but comparative genomic analysis using Breseq again identified a variety of putative mechanisms that could potentially underpin our phenotypic observation.

In Pair 2, the *acrR* deletion explored above could also be responsible for increased biofilm formation, as *acrAB-tolC* has a central role in biofilm formation in *E. coli* and *K. pneumoniae* [39, 40]. Studies also link biofilm formation, AMR and *acrAB-tolC* activity. In a comparison of *K. pneumoniae* urinary isolates, Vuotto *et al.* observed that those with the highest biofilm-forming capacity also demonstrated the most AMR, correlating with increased *acrAB* activity [41]. Although only a sample of three isolates, this strikingly mirrors the findings of the *K. pneumoniae* blood isolate of Pair 2, raising the possibility of a single intra-host adaptation contributing to both AMR and biofilm formation. Hennequin *et al.* demonstrated this *in vitro*, reporting that sub-MIC cephalosporin exposure in *E. coli* isolates resulted in enhanced *acrAB* expression and biofilm formation [42]. This could be an example of antimicrobial exposure selecting for a resistance-conferring mutation that also increases virulence and propensity to invasive infection. If seen consistently in a larger cohort, this could suggest a promising interventional target. For example, RND efflux pump inhibitors have been explored for use as antimicrobial-adjuvant therapy for resistant infections, and for biofilm disruption in invasive infection [43].

Furthermore, in this pair, nsSNPs in two biofilm-related virulence genes, *IcmF* and protease III precursor, were identified, possibly contributing to the phenotypic change. Similarly, in Pair 3, two nsSNPs in the biofilm-associated virulence gene ‘antigen 43’ were noted. Mutations in this gene were noted when Nielsen *et al.* compared ‘linked’ urine and faecal isolates and hypothesized this related to increased biofilm formation [44].

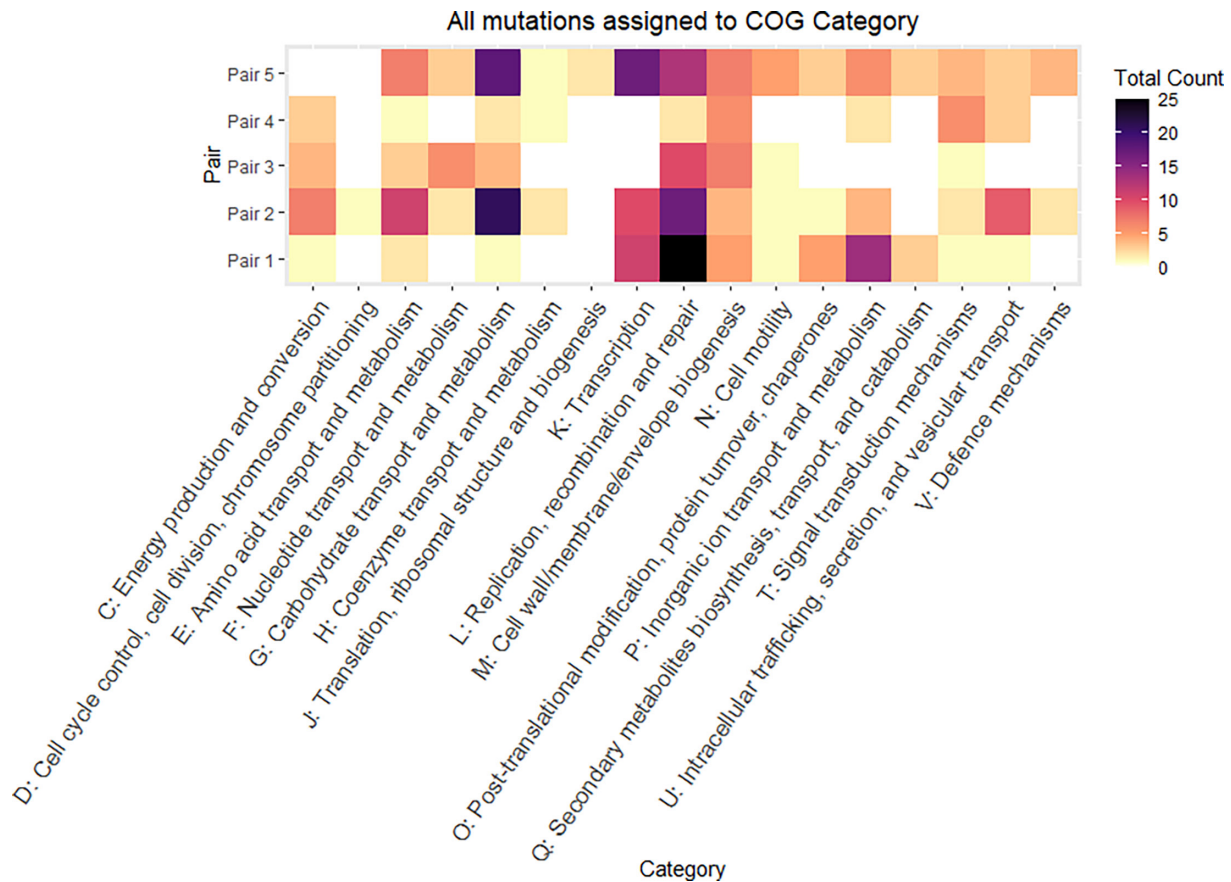


Fig. 5. Heatmap of all mutations assigned to COG categories across isolate pairs. Each tile represents the total count of mutations within each COG category for each isolate pair. White tiles represent the absence of mutations in a specific COG category. Mutations observed when comparing blood isolate to corresponding faecal isolate.

Pair 1 exhibited a potential link between biofilm formation and plasmid acquisition. This pair underwent loss of IncFIB(K) and acquisition of IncFII(Yp) between faecal and blood isolates (Fig. 2d). Breseq analysis revealed a duplication of the *hha* gene in the blood isolate compared to the faecal isolate (Table 4), with the additional copy residing on a contig containing multiple IncF-related genes, suggesting plasmid-mediated acquisition. The *hha* gene encodes a haemolysin-expression-modulating protein and is a regulator of virulence factors involved in cell adhesion and biofilm formation in Gram-negative species. Studies in *E. coli* have identified *hha* as a positive regulator of biofilm formation and of virulence [45, 46]. This is less studied in *K. pneumoniae*; however, Bandeira *et al.* described *hha* presence to be associated with *K. pneumoniae* biofilm-forming isolates, particularly on indwelling central lines [47]. Plasmid-mediated acquisition of *hha* was described by Krall *et al.* who compared 'linked' faecal and blood isolates, finding that *hha* acquisition on an IncFII plasmid increased the invasiveness of *E. coli* isolates in a gut-organoid model [48]. They concluded that *hha* was involved in GIT-blood adaptation but did not explore biofilm formation specifically. Here, one could similarly hypothesize that IncFII plasmid-mediated acquisition of *hha* could be an adaptive change leading to increased biofilm formation and procession to GNBSI. Other studies report that *hha* has a negative influence on biofilm formation [49, 50], which this hypothesis contradicts. Interestingly, in both copies of the *hha* gene in the blood isolate, a 3 bp GGT insertion was noted 1 bp downstream of *hha*. Upon gene alignment, this appeared to be an AAA duplication, disrupting the Shine–Dalgarno sequence 7 bp before the gene (Fig. 6). This could potentially reduce the efficiency of protein translation. Hence, one could equally hypothesize, assuming a negative-regulatory role of *hha*, that after plasmid acquisition during GIT colonization, the compensatory mutation in the Shine–Dalgarno region led to increased biofilm formation and aided bloodstream invasion. Further investigation is required to clarify the role of *hha* in invasion and the combined effect of the observed SNPs, Shine–Dalgarno mutation and *hha* copy number on biofilm formation.

The transition from GIT to blood and subsequent survival in the bloodstream will involve the ability to adapt to a harsh and dynamic environment, with factors such as varying nutrient abundance and host immunity to contend with. As noted, this adaptation did not appear to result from acquisition of virulence genes in these five cases but was associated with varied changes across the genome. No single specific mutation occurred in all pairs, likely reflecting the differing clinical scenarios and selection

Table 4. Specific mutations identified in relation to phenotypes of interest. Mutations observed when comparing blood isolate to corresponding faecal isolate

Pair	Mutation type	Mutation	Annotated gene	Relevance of mutation	
1	nsSNP	D47E (GAC→GAA)	<i>ompk36</i>	Bloodstream survival factor	
	nsSNP	Q62K (CAA→AAA)	<i>dam</i>	Bloodstream survival factor	
	nsSNP	T57A (ACT→GCT)	<i>dam</i>	Bloodstream survival factor	
	nsSNP	Q141L (CAG→CTG)	<i>dam</i>	Bloodstream survival factor	
	Insertion	+GGT	Haemolysin-expression-modulating protein/hypothetical protein	Regulator of virulence factors, known to regulate biofilm formation	
	iSNP	2bp→TT	Haemolysin-expression-modulating protein/hypothetical protein	Regulator of virulence factors, known to regulate biofilm formation	
	iSNP	G→T	Haemolysin-expression-modulating protein/hypothetical protein	Regulator of virulence factors, known to regulate biofilm formation	
	iSNP	A→G	Haemolysin-expression-modulating protein/hypothetical protein	Regulator of virulence factors, known to regulate biofilm formation	
	iSNP	A→T	Haemolysin-expression-modulating protein/hypothetical protein	Regulator of virulence factors, known to regulate biofilm formation	
	sSNP	A40A (GCT→GCC)	Haemolysin-expression-modulating protein	Regulator of virulence factors, known to regulate biofilm formation	
	sSNP	R11R (CGC→CGG)	Haemolysin-expression-modulating protein	Regulator of virulence factors, known to regulate biofilm formation	
	Deletion	Δ916 bp	Fec operon	Iron uptake and utilization	
	2	nsSNP	T37I (ACT→ATT)	Multiple antibiotic resistance protein MarB	Drug-efflux pump implicated in antibiotic resistance
		Deletion	Δ348 bp	YcaD – MFS transporter	Drug efflux pump implicated in antibiotic resistance
Deletion		Δ254 bp	Permease of drug/metabolite transporter	Drug efflux pump implicated in antibiotic resistance	
Deletion		Δ250 bp	Inner membrane protein/drug antiporter	Drug efflux pump implicated in antibiotic resistance	
Deletion		Δ243 bp	LysR	Transcriptional regulator implicated in antibiotic resistance	
Deletion		Δ241 bp	MFS-type transporter	Drug efflux pump implicated in antibiotic resistance	
Deletion		Δ239 bp	RND efflux transporter	Drug efflux pump implicated in antibiotic resistance	
Deletion		Δ236 bp	LysR transcriptional regulator	Drug efflux pump implicated in antibiotic resistance	
Deletion		Δ321 bp	<i>acrR</i>	Transcriptional regulator implicated in antibiotic resistance and biofilm formation	
nsSNP		A434S (CCC→TCC)	Protease III precursor (EC 3.4.24.55)	Biofilm formation	
nsSNP		V701A (GTG→GCG)	T6SS component TssM (IcmF/VasK)	Virulence factor, involved in biofilm formation	

Continued

Table 4. Continued

Pair	Mutation type	Mutation	Annotated gene	Relevance of mutation
3	Deletion	Δ235 bp	Lipopolysaccharide ABC transporter	Transporter protein implicated in antibiotic resistance
	nsSNP	T115A (ACC→GCC)	Phage endopeptidase Rz	Phage protein, adjacent to virulence factor <i>iss</i> (bloodstream survival)
	nsSNP	L237P (CTT→CCT)	Mobile element protein	MGE adjacent to Fec operon (iron uptake and utilization)
	nsSNP	A238G (GCC→GGC)	Mobile element protein	MGE adjacent to Fec operon (iron uptake and utilization)
	nsSNP	S932T (AGC→ACC)	Antigen 43	Virulence factor, involved in biofilm formation
	nsSNP	A763V (GCC→GTC)	Antigen 43	Virulence factor, involved in biofilm formation
	sSNP	G765G (GGA→GGT)	Antigen 43	Virulence factor, involved in biofilm formation
4	Deletion	Δ286 bp	MFS type transporter	Drug efflux pump implicated in antibiotic resistance
	sSNP	G977G (GGG→GGT)	Antigen 43	Virulence factor, involved in biofilm formation
	sSNP	G972G (GGT→GGA)	Antigen 43	Virulence factor, involved in biofilm formation
	nsSNP	S932T (AGC→ACC)	Antigen 43	Virulence factor, involved in biofilm formation
	nsSNP	V375A (GTC→GCC)	Antigen 43	Virulence factor, involved in biofilm formation
	Deletion	Δ265 bp	Ferrochelatase	Iron uptake and utilization
	Deletion	Δ236 bp	Ferredoxin reductase	Iron uptake and utilization
	Deletion	Δ119 bp	cog3210	Iron uptake and utilization
	Deletion	Δ95 bp	cog3210	Iron uptake and utilization
	5	sSNP	L123L (CTG→TTG)	Manganese ABC transporter, inner membrane permease protein SitC
Deletion		Δ297 bp	Transcriptional regulator, <i>acrR</i> family	Transcriptional regulator implicated in antibiotic resistance and biofilm formation
Deletion		Δ295 bp	Spermidine/putrescine ABC transporter	Transporter protein implicated in antibiotic resistance
Deletion		Δ281 bp	Fe ²⁺ ABC transporter, substrate-binding protein	Iron uptake/homeostasis
nsSNP		V32E (GTG→GAG)	Transcriptional repressor of <i>aga</i> operon	Transcriptional regulator involved in biofilm formation
nsSNP		M300L (ATG→CTG)	Lipopolysaccharide export system permease protein LptG	Biofilm formation
Deletion		Δ246 bp	TctA transporter	Drug-efflux pump implicated in antibiotic resistance
Deletion		Δ245 bp	AaeAB efflux system	Drug efflux pump implicated in antibiotic resistance

```

Faeces      tattaccgggacacctcaggatat   aaattaatgacgaaac
Blood_1     aactaccgggacacctaaggatat   aaaaaaccaatgacgaaac
Blood_2     aactaccgggacacctaaggatat   aaaaaaccaatgacgaaac

```

Fig. 6. Alignment of upstream intergenic region and first 10bp (underlined) of *hha* gene of faecal isolate and two copies in blood isolate. (Yellow=duplication of bp. Red=iSNP).

pressures driving diverse evolutionary trajectories en route to bloodstream infection. We did find a small number of common mutations within the *E. coli* strains and across two of the three *K. pneumoniae* strains (Table S2); however, in all pairs, most mutations occurred in the COG category 'Replication, recombination and repair', including a mutation in *dam*, a known bloodstream survival factor, in Pair 1 [51]. These genes are involved in environmental and stress response, and selection for mutations could be expected in adaptation to challenging environments, for example, maintaining efficient DNA repair mechanisms in the presence of reactive oxygen species, host immune defences or nutrient variability. Another commonality was mutations in 'cell wall/membrane/envelope biogenesis', 'cell motility' and 'signal transduction' categories. Such mutations might confer altered chemotaxis or adherence, allowing tissue invasion. Few mutations were observed in 'Cell cycle control and repair' and 'Translation, ribosomal structure and biogenesis'. This demonstrates the strong selective pressure to conserve key cellular processes integral to bacterial survival during the transition between GIT and blood environments. Many mutations were classified as 'Function Unknown' within COG categories but may hold significance. For example, in Pair 3, an nsSNP was identified in a phage endopeptidase gene adjacent to the gene *iss*. The *iss* gene is associated with bloodstream survival in *E. coli* through mediating serum and complement resistance, and increased mRNA copy number has been associated with increased serum survival [52, 53]. Mutations in the adjacent phage endopeptidase could influence the activation or expression of *iss*, with effects on bloodstream survival. This represents a potential avenue for further research.

When comparing blood isolates to their faecal ancestors using Breseq, we noted mutations related to carbohydrate and amino acid transport in five out of five pairs. This included an nsSNP in *ompk36* in Pair 1, a transport protein also previously identified as a bloodstream survival factor in *K. pneumoniae* [54]. Further, we identified mutations related to iron uptake utilization in four out of five pairs. Iron availability and chemical form vary in different tissues, and a 'functional hierarchy', whereby iron uptake systems are utilized variably across environments, has been discussed by Garcia *et al.* in the context of urinary tract infection [55]. The findings of *fec* operon deletion in Pair 1 and the transposase mutation near the *fec* operon in Pair 3 may suggest a genomic shift away from ferric citrate uptake, whilst the *sitC* virulence gene mutation and ferrous (Fe^{2+}) ABC transporter deletion in Pair 5 may suggest a shift away from Fe^{2+} uptake. These mutations likely reflect adaptation towards iron acquisition in the bloodstream, where ferric citrate is less abundant than the gut and Fe^{2+} is less abundant due to oxygenated conditions. Further exploration could inform therapeutic strategies. For example, *tonB*, which regulates the *fec* operon, has been explored as a potential drug target against *E. coli* [56]. However, this may not be an appropriate target in the context of bloodstream infection if functional redundancy of the *fec* operon is a common adaptation.

The strengths of this study include the combination of genomic with phenotypic comparison, using isolates from the same patient to gain insight into within-host evolution and link phenotypic differences to changes at the genomic level. To our knowledge, there are few studies in the existing literature that have investigated this. Further, we compared mutations across the whole genome, rather than focusing only on known ARGs and virulence genes, identifying potentially relevant mutations that would have been missed by a narrower approach. Limitations of our study lie in the fact that it cannot be definitively concluded that these pairs are 'linked', given the lack of an existing robustly defined threshold for strain similarity/genomic relatedness – this, in itself, highlights an area for further work. In addition, the retrospective nature and ability to only include five cases of GNBSI, without granular clinical data, mean the significance and generalizability of these findings cannot be assumed. The use of short-read rather than long-read sequencing makes it harder to discern plasmid-mediated changes that could contribute to adaptation. Further, phenotypic assays for biofilm formation demonstrated changes *in vitro* but likely do not reflect true clinical conditions. Whilst this study shows individual instances of mutations of potential clinical relevance in each pair which may represent avenues for further study, it is clear that there is a diverse range of possible evolutionary trajectories in GIT–blood transition, even in this small study, so a common pathway for clinical intervention may be difficult to elucidate. Future study with a larger sample size that used faecal isolates from proven 'linked' pairs to colonizing isolates that did not achieve bloodstream invasion could be a more promising avenue to identify pathogen factors, e.g. virulence gene SNPs that may predispose to bloodstream infection. Previous case-control studies with this approach have been used, but not including WGS data [57]. Comparative analysis, including isolates from other body sites, such as the urinary and respiratory tracts, would also be beneficial. This, along with information around antimicrobial exposure, clinical severity and GNBSI source, would give a more comprehensive insight into the relevance of adaptations to differing clinical scenarios and how findings could translate to clinical intervention.

Overall, this study used genomic and phenotypic comparison of 'linked' pairs of blood and faecal *K. pneumoniae* and *E. coli* isolates to investigate evolutionary and adaptive changes relevant to the transition from GIT colonization to bloodstream invasion and GNBSI. Phenotypic changes in AMR were variable; however, biofilm formation and metabolic flexibility were common phenotypic adaptations towards bloodstream infection in this small sample. These changes were not found to be associated with acquisition of ARGs or virulence genes, but rather, with a diverse array of genomic changes, including mutation of virulence genes, efflux pumps and acquisition of plasmid-mediated genes. A small number of common mutations were found between the *E. coli* strains and between two of the three *K. pneumoniae* strains; however, changes in functional categories of genes involved in transport and metabolism were common across all five pairs. These findings demonstrate a snapshot of the vast evolutionary landscape of intra-host adaptation that is traversed in GNBSI pathogenesis. However, significant further study is needed to assess whether these findings could feasibly translate into clinical intervention for GNBSI treatment or prevention.

Funding information

A.P.R. acknowledges funding from the Medical Research Council, Biotechnology and Biological Sciences Research Council and Natural Environmental Research Council which are all Councils of UK Research and Innovation (grant no. MR/W030578/1) under the umbrella of the JPIAMR (Joint Programming Initiative on Antimicrobial Resistance), the NIHR (grant no. NIHR200632) and UKRI through the Strength in Places Fund (grant no. S1PF 36348). E.A. and R.P. are funded by the MRC via the LSTM and Lancaster University PhD Doctoral Training Program (grant nos. MR/W007037/1 and MR/N013514/1, respectively).

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical statement

Ethical approval for linking anonymized data was granted by the University of Liverpool as part of a separate study for which the isolate library was created (REC reference 255669).

References

- Hallmaier-Wacker LK, Andrews A, Nsonwu O, Demirjian A, Hope RJ, *et al.* Incidence and aetiology of infant gram-negative bacteraemia and meningitis: systematic review and meta-analysis. *Arch Dis Child* 2022;107:988–994.
- Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the global burden of disease study. *Lancet* 2020;395:200–211.
- Wagner M, Bonhoeffer J, Erb TO, Glanzmann R, Häcker FM, *et al.* Prospective study on central venous line associated bloodstream infections. *Arch Dis Child* 2011;96:827–831.
- Kern WV, Rieg S. Burden of bacterial bloodstream infection—a brief update on epidemiology and significance of multidrug-resistant pathogens. *Clin Microbiol Infect* 2020;26:151–157.
- Lipworth S, Vihta K-D, Davies T, Wright S, Tabirao M, *et al.* Molecular epidemiology and antimicrobial resistance phenotype of paediatric bloodstream infections caused by gram-negative bacteria. *Commun Med* 2022;2:101.
- WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. Geneva: World Health Organization. 2024.
- Holmes CL, Anderson MT, Mobley HLT, Bachman MA. Pathogenesis of gram-negative bacteremia. *Clin Microbiol Rev* 2021;34:e00234–20.
- Jiménez-Rojas V, Villanueva-García D, Miranda-Vega AL, Aldana-Vergara R, Aguilar-Rodea P, *et al.* Gut colonization and subsequent infection of neonates caused by extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*. *Front Cell Infect Microbiol* 2023;13:1322874.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011;17:10.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
- Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, *et al.* A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nat Commun* 2021;12:4188.
- Lam MMC, Wick RR, Judd LM, Holt KE, Wyres KL. Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the *Klebsiella pneumoniae* species complex. *Microb Genom* 2022;8:000800.
- Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 2017;110:1281–1286.
- R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2021.
- Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom* 2022;8:000748.
- Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, *et al.* *In silico* detection and typing of plasmids using PlasmidFinder and Plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58:3895–3903.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008;9:75.
- Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* 2014;1151:165–188.
- European Committee on Antimicrobial Susceptibility Testing (EUCAST). Disk diffusion manual, version 12.0. 2024.
- European Committee on Antimicrobial Susceptibility Testing (EUCAST). Breakpoint tables for interpretation of mics and zone diameters, version 14.0. 2024.
- European Committee on Antimicrobial Susceptibility Testing (EUCAST). Broth microdilution - EUCAST reading guide, version 5.0. 2024.
- Merritt JH, Kadouri DE, O'Toole GA. Growing and analyzing static biofilms. *Curr Protoc Microbiol* 2005;Chapter 1:Unit 1B.1.
- O'Toole GA. Microtiter dish biofilm formation assay. *JoVE* 2011;2011:e2437.
- Schwartz DJ, Shalon N, Wardenburg K, DeVeaux A, Wallace MA, *et al.* Gut pathogen colonization precedes bloodstream infection in the neonatal intensive care unit. *Sci Transl Med* 2023;15:eadg5562.

25. Carl MA, Ndao IM, Springman AC, Manning SD, Johnson JR, et al. Sepsis from the gut: the enteric habitat of bacteria that cause late-onset neonatal bloodstream infections. *Clin Infect Dis* 2014;58:1211–1218.
26. Rodriguez-R LM, Conrad RE, Viver T, Feistel DJ, Lindner BG, et al. An ANI gap within bacterial species that advances the definitions of intra-species units. *mBio* 2024;15:e02696–23.
27. Vornhagen J, Roberts EK, Unverdorben L, Mason S, Patel A, et al. Combined comparative genomics and clinical modeling reveals plasmid-encoded genes are independently associated with *Klebsiella* infection. *Nat Commun* 2022;13:4459.
28. Ludden C, Coll F, Gouliouris T, Restif O, Blane B, et al. Defining nosocomial transmission of *Escherichia coli* and antimicrobial resistance genes: a genomic surveillance study. *Lancet Microbe* 2021;2:e472–e480.
29. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, et al. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clin Infect Dis* 2017;65:208–215.
30. Thänert R, Choi J, Reske KA, Hink T, Thänert A, et al. Persisting uropathogenic *Escherichia coli* lineages show signatures of niche-specific within-host adaptation mediated by mobile genetic elements. *Cell Host Microbe* 2022;30:1034–1047.
31. Rebelo AR, Bortolaia V, Leekitcharoenphon P, Hansen DS, Nielsen HL, et al. One day in Denmark: whole-genome sequence-based analysis of *Escherichia coli* isolates from clinical settings. *J Antimicrob Chemother* 2025;80:1011–1021.
32. Bush SJ, Foster D, Eyre DW, Clark EL, De Maio N, et al. Genomic diversity affects the accuracy of bacterial single-nucleotide polymorphism-calling pipelines. *Gigascience* 2020;9:giaa007.
33. Jamin C, De Koster S, van Koeveeringe S, De Coninck D, Mensaert K, et al. Harmonization of whole-genome sequencing for outbreak surveillance of *Enterobacteriaceae* and *Enterococci*. *Microb Genom* 2021;7:000567.
34. Wylie KM, Wylie TN, Minx PJ, Rosen DA. Whole-genome sequencing of *Klebsiella pneumoniae* isolates to track strain progression in a single patient with recurrent urinary tract infection. *Front Cell Infect Microbiol* 2019;9:14.
35. Okusu H, Ma D, Nikaido H. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. *J Bacteriol* 1996;178:306–308.
36. Padilla E, Llobet E, Doménech-Sánchez A, Martínez-Martínez L, Bengoechea JA, et al. *Klebsiella pneumoniae* AcrAB efflux pump contributes to antimicrobial resistance and virulence. *Antimicrob Agents Chemother* 2010;54:177–183.
37. Lv F, Cai J, He Q, Wang W, Luo Y, et al. Overexpression of efflux pumps mediate pan resistance of *Klebsiella pneumoniae* sequence type 11. *Microb Drug Resist* 2021;27:1405–1411.
38. Chowdhury N, Suhani S, Purkaystha A, Begum MK, Raihan T, et al. Identification of AcrAB-TolC efflux pump genes and detection of mutation in efflux repressor AcrR from omeprazole responsive multidrug-resistant *Escherichia coli* isolates causing urinary tract infections. *Microbiol Insights* 2019;12:1178636119889629.
39. Alav I, Sutton JM, Rahman KM. Role of bacterial efflux pumps in biofilm formation. *J Antimicrob Chemother* 2018;73:2003–2020.
40. Tang M, Wei X, Wan X, Ding Z, Ding Y, et al. The role and relationship with efflux pump of biofilm formation in *Klebsiella pneumoniae*. *Microb Pathog* 2020;147:104244.
41. Vuotto C, Longo F, Pascolini C, Donelli G, Balice MP, et al. Biofilm formation and antibiotic resistance in *Klebsiella pneumoniae* urinary strains. *J Appl Microbiol* 2017;123:1003–1018.
42. Hennequin C, Aumeran C, Robin F, Traore O, Forestier C. Antibiotic resistance and plasmid transfer capacity in biofilm formed with a CTX-M-15-producing *Klebsiella pneumoniae* isolate. *J Antimicrob Chemother* 2012;67:2123–2130.
43. Reza A, Sutton JM, Rahman KM. Effectiveness of efflux pump inhibitors as biofilm disruptors and resistance breakers in gram-negative (ESKAPEE) bacteria. *Antibiotics* 2019;8:229.
44. Nielsen KL, Stegger M, Godfrey PA, Feldgarden M, Andersen PS, et al. Adaptation of *Escherichia coli* traversing from the faecal environment to the urinary tract. *Int J Med Microbiol* 2016;306:595–603.
45. Prieto A, Urcola I, Blanco J, Dahbi G, Muniesa M, et al. Tracking bacterial virulence: global modulators as indicators. *Sci Rep* 2016;6:25973.
46. Barrios AFG, Zuo R, Ren D, Wood TK. Hha, YbaJ, and OmpA regulate *Escherichia coli* K12 biofilm formation and conjugation plasmids abolish motility. *Biotechnol Bioeng* 2006;93:188–200.
47. Bandeira M, Borges V, Gomes JP, Duarte A, Jordao L. Insights on *Klebsiella pneumoniae* biofilms assembled on different surfaces using phenotypic and genotypic approaches. *Microorganisms* 2017;5:16.
48. Krall LJ, Klein S, Boutin S, Wu CC, Sähr A, et al. Invasiveness of *Escherichia coli* is associated with an IncFII plasmid. *Pathogens* 2021;10:1645.
49. Ueda T, Takahashi H, Uyar E, Ishikawa S, Ogasawara N, et al. Functions of the Hha and YdgT proteins in transcriptional silencing by the nucleoid proteins, H-NS and StpA, in *Escherichia coli*. *DNA Research* 2013;20:263–271.
50. Sharma VK, Bearson BL. Hha controls *Escherichia coli* O157:H7 biofilm formation by differential regulation of global transcriptional regulators FlhDC and CsgD. *Appl Environ Microbiol* 2013;79:2384–2396.
51. Fang C-T, Yi W-C, Shun C-T, Tsai S-F. DNA adenine methylation modulates pathogenicity of *Klebsiella pneumoniae* genotype K1. *J Microbiol Immunol Infect* 2017;50:471–477.
52. Desvaux M, Dalmasso G, Beyrouthy R, Barnich N, Delmas J, et al. Pathogenicity factors of genomic islands in intestinal and extraintestinal *Escherichia coli*. *Front Microbiol* 2020;11:2065.
53. Xu W-Y, Li Y-J, Fan C. Different loci and mRNA copy number of the increased serum survival gene of *Escherichia coli*. *Can J Microbiol* 2018;64:147–154.
54. Chen J-H, Siu LK, Fung C-P, Lin J-C, Yeh K-M, et al. Contribution of outer membrane protein K36 to antimicrobial resistance and virulence in *Klebsiella pneumoniae*. *J Antimicrob Chemother* 2010;65:986–990.
55. Garcia EC, Brumbaugh AR, Mobley HLT. Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection. *Infect Immun* 2011;79:1225–1235.
56. Mann R, Mediati DG, Duggin IG, Harry EJ, Bottomley AL. Metabolic adaptations of uropathogenic *E. coli* in the urinary tract. *Front Cell Infect Microbiol* 2017;7:241.
57. Krawczyk B, Śledzińska A, Szemiako K, Samet A, Nowicki B, et al. Characterisation of *Escherichia coli* isolates from the blood of haematological adult patients with bacteraemia: translocation from gut to blood requires the cooperation of multiple virulence factors. *Eur J Clin Microbiol Infect Dis* 2015;34:1135–1143.