Chapter 4: Understanding the relationships between BSI and carriage *E. coli* isolates in patients attending HTD, Vietnam.

4.1. Introduction

Our results from Chapter 2 demonstrated that *E. coli* isolates from rectal swab were more likely to be members of ECOR phylogroup A or B1 than they were to belong to B2 or D. This is consistent with previous reports that have shown that phylogroups A and B1 are often identified as "commensal" or "harmless" bacteria and are less often the cause of BSIs. Although there are no single or defined sets of genes that absolutely distinguish invasive from commensal *E. coli* [10], invasive *E. coli* (ExPEC) often belonged to phylogroup B2 or D and harbour genes related to adhesion, immune evasion, iron acquisition and toxin production [109] as shown for the HTD isolates in Chapter 3.

In our 06EI study (see Section 2.2, Chapter 2), *E. coli* were cultured from 700/834 (84%) HTD ICU patients (data unpublished). Patients when staying long-term in the ICUs are highly vulnerable, and easily contract HAIs, due to their close contact with healthcare workers and other patients [150, 151]. HAI-associated outbreaks in ICU affect both adults and neonates and often involve strains that are highly resistant to therapeutic antibiotic usage in the hospital environments – these are frequently multidrug resistant (MDR) which is associated with a prolonged duration of hospital admission and requires more intense treatment and monitoring which significantly increases treatment costs as well as being associated with, and increasing the risk of a negative clinical outcome [79, 118, 152].

Chapters 2 and 3 of this study showed that there were clear phylogenetic and genetic differences between the randomly sampled rectal swab and BSI isolates from HTD. However, despite being representative of the isolates causing disease they were all derived from unlinked patients. Samples collected in study 06EI provided a unique opportunity to understand the relationship between rectal swab (carriage) and BSI (invasive) isolates taken from the same patient. The samples included in this study also included those taken from a limited number of patients who had been sampled longitudinally during the course of their hospital stay and so it was also possible to compare these *E. coli* isolates to those taken from the same patient at admission either from rectal swab and or those cultured from blood.

Aim: To understand the genetic diversity and relationships of carriage and invasive *E. coli* isolated from the same patient in HTD, Vietnam.

The specific questions to answer:

1) What is the relationship between the paired *E. coli* causing BSIs in a patient to those cultured from their rectal swabs?

2) How does *E. coli* diversity change over time during hospitalisation within a single patient? Are patients carrying the same strains for the duration of their stay or are they reinfected with new strains during their hospitalisation?

4.2. Methods

4.2.1 Study participants

Patient recruitment and sample collection was as detailed in Section 2.2. The 06EI study at the ICU in HTD was conducted from November 2014 to January 2016. Eligible patients (age ≥ 15 years old) admitted to ICU gave informed consent, and a rectal swab was taken from each patient within the first two days post-hospitalisation. This was dubbed "baseline" *E. coli* carriage. Follow-up rectal swabs were taken twice a week until the patient was discharged and were classified as "longitudinal" isolates.

4.2.2. Isolate collection

Isolates were collected and stored as described in Section 2.2. If growth on solid media revealed the presence of two or more colony morphologies (lactose/non-lactose fermented/ β-haemolytic) then multiple colonies were selected for sequencing. DNA was prepared from these isolates and sent to Sanger Institute for sequencing on Illumina HiSeq platform as described in Chapter 2.

4.2.3. Phylogenetic analysis

Phylogenetic analysis was essentially as described in Chapter 2 Section 2.2 based on whole genome SNPs built from core gene alignment. Core gene alignment was constructed by ROARY, except that we only included a subset of 94 isolates and did not include the reference genomes of *E. fergusonii* and *E. albertii*. With the smaller sample size the core genes present in at least 95% of these isolates amounted to 3,077 genes, as expected higher than the total core genes (2,796) defined for the entire sample set in Chapter 2. The pan genome of 94 isolates also comprised of fewer genes (18,886).

4.2.4. AMR genes, replicon types and virulence genes identification

These genomes of these isolates were also screened for antimicrobial resistance, plasmid replicons and virulence genes using ARIBA as describe in Chapter 3 (see Section 3.2). Phylogenetic tree and gene presence/absence as described in Section 3.2.

4.3. Results

4.3.1. Patient and sample recruitment

After QC, we successfully generated genomes of matched *E. coli* isolates taken from rectal swab (carriage) and blood collected in 2015 from a subset of twenty-eight patients (of the original thirty-one patients with matched samples. See Section 2.2) attending HTD. Of these patients twenty-five had BSI on admission or in less than 2 days after admission and so were considered to represent patients with community acquired blood stream infections. Two patients were transferred from a different ward after

8 days of treatment from the same hospital (03-0029 and 03-0514) while the third patient (03-0283) developed BSIs while being treated for tetanus; therefore, these three patients were recorded as HAIs in the 06EI study.

Of these 28 patients, 8 were male and 20 were female. The median age was 53 years of age (range 27-68 year old). Majority patients had an underlying disease that may have predisposed them to bacterial BSIs: 11 patients had liver disease as a consequence of chronic hepatitis B/C or alcoholism. In the remaining 17 patients, other comorbidities included pneumonia, AIDS, tetanus, cholecystitis, diabetes and UTIs, while three others had no known comorbidities. All 8 male patients were recorded as being moderate/heavy smokers and drinking alcoholic beverages, while none of female patients reported that they either smoked or drank alcohol.

In summary, there were 94 isolates remained after QC for this analysis. This included 28 bloodstreams and 46 rectal swab carriage isolates that were collected on day 0 or 1 of admission to HTD ICU, plus 20 longitudinal isolates taken after day 1 from 7 patients. The maximum sampling period was 20 days post admission for patient P3 (03-283).

4.3.2. Genomic diversity among rectal swab and blood derived *E. coli* isolates on admission to HTD ICU

The core genome phylogenies of all 94 isolates detailed above are shown in Figure 4.1. The phylogenetic tree shows that there are 13 BAPs phylogenetic groups (except L1 and L8) represented in the tree. BAPs groups were previously defined in Figure 2.4 Section 2.3 Chapter 2 and transposed onto this analysis. By MLST these isolates include a total of 37 ST's, of which the dominant STs were ST1193, ST131, and ST648 (Supplementary Table S4). If this is divided by sample type then the most common STs in the blood samples were ST1193 (5 samples), ST131 (4 samples) and ST69 (3 samples). For rectal isolates, the dominant STs were ST131 (9 samples), ST1193 (6 samples) and ST648 (8 samples) (Figure 4.1).

Of the 28 patients included for 16/28 (57 %) the same *E. coli* STs were cultured from the blood and rectal swab samples. For the remaining 12 patients, the *E. coli* cultured from blood and carriage samples were of different STs. Its worth noting that rectal swab isolates for 2/28 patients, P1 (03-0048) and P6 (03-0209) (Figure 4.1) with matched samples were collected after day 0. This was because the samples taken on admission failed to grow when the samples were collected for this study. Since the patients would have received treatment in this time after admission, the likelihood is that this influenced nature of the *E. coli* population in the gut. Hence, it is possible that the percentage of identical isolates on admission in the blood and rectal swab was higher 57%.

For the 28 paired samples, the genomes of the blood and carriage isolates from 16 patients were on average only differentiated by a median of 2 SNPs (IQR 1-3) (Supplementary Table S5). Clearly this located these isolates on the same phylogenetic branch of the tree. Paired rectal swab and blood isolates

from the same patient can be seen as red and orange nodes (spots) from blood and rectal swab, respectively on Figure 4.1. The arcs shown in Figure 4.1 link the respective paired sample(s) from blood and rectal swab. If multiple isolates were collected on each time point (due to multiples colony morphologies on the resultant culture plates; see methods) from the rectal swab samples then the blood samples are linked by multiple arcs to their respective paired rectal swab nodes (Figure 4.1).

To give an estimation of diversity we plotted the pairwise SNP distances between isolates from the same patient with the same MLST profile or the same ST taken from different patients (Figure 4.2). We used ST to control for the variable inter-lineage (ST) diversity. For the genomes of isolates of the same ST taken from different samples from the same patient the SNP difference between isolates was within 3 SNPs (median 1, IQR 0 – 2.5). However, if the same ST was taken from different patients, regardless of sample type then they are on average 3,670 SNPs apart at the whole genome level, (median 2,517; IQR 119 – 7,135) (Figure 4.2).

There were two exceptions, the paired *E. coli* genomes from patients P18 (03-0301) and P19 (03-0322) (Supplementary Table S5), were separated by only 40 SNPs respectively although both belonged to same ST1193 and ST131, as well as the same BAPS lineage (labelled as P18 and P19 of Figure 4.1). However, unlike the other linked samples in this study, because we know the estimated mutation rate of *E. coli* ST131 is 1 SNPs/genome/year [153] and because the same STs from 2 different patients in our study also have at least 24 SNPs difference, if 2 isolates are >40 SNPs apart it is likely that the patient has acquired the strains from a different source, including other body sites, unrelated to their own rectal carriage strains or there was unsampled diversity not captured on the rectal swab.

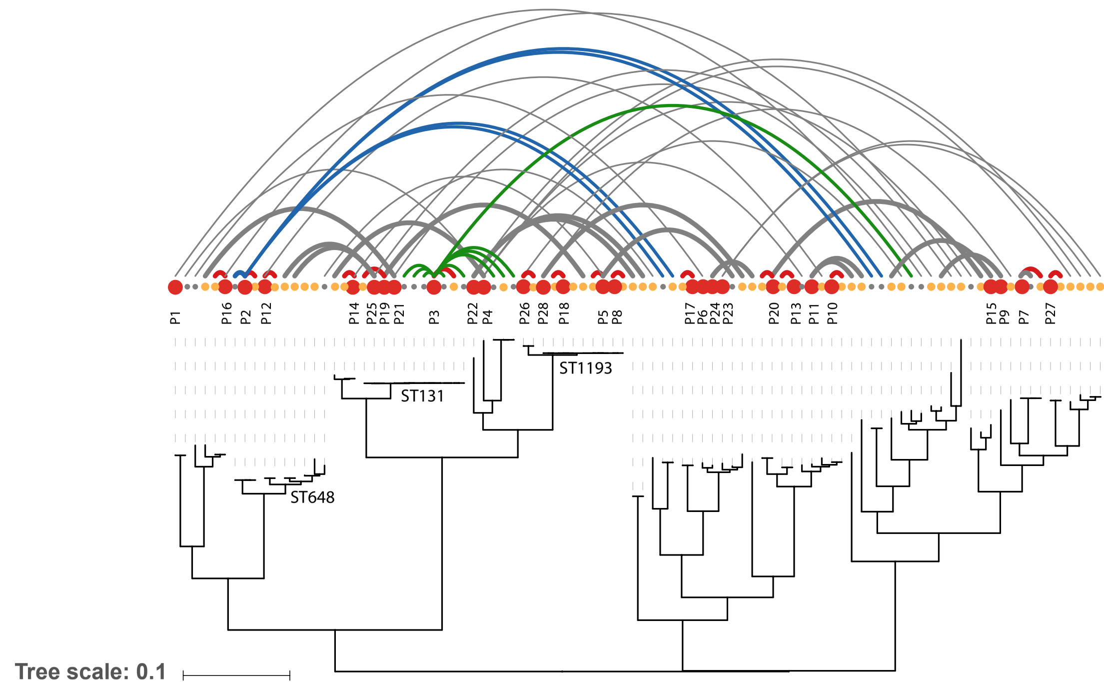

Figure 4.1: Phylogenetic diversity among rectal swab and blood derived *E. coli* isolates from HTD.

A phylogenetic tree constructed from the core genome of 94 isolates. Each node (coloured spot) represents one isolate and the arcs connecting the nodes link rectal swab isolates to their respective blood isolate within the same patient. Red nodes are isolates from blood (n=28), yellow nodes are baseline rectal swab isolates, while grey nodes are longitudinal rectal swab isolates. The red arc indicates the blood and rectal swab isolates were the same ST. The blue arc connects all the isolates taken from patient P2, the green arc connects isolates taken from patient P3 as detailed in Figure 4.3.

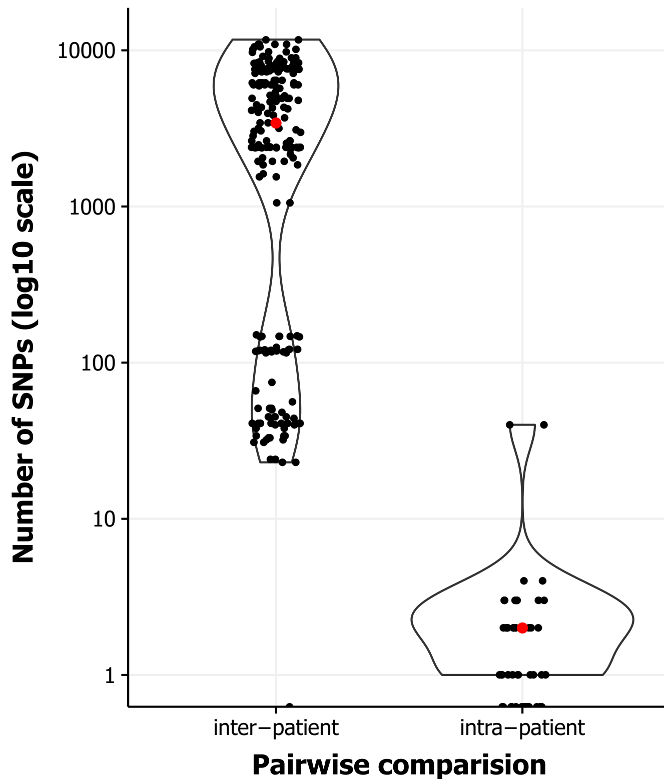

Figure 4.2: Pairwise SNP distances within and between patients based on the variant calls from the core gene alignment. A violin plot for SNP distances between isolates taken from the same patient (intra-patient) and in different patients (inter-patient) is reported. Only those paired samples separated by fewer than 12,000 SNPs are reported, since these capture genomic diversity within an MLST, while isolates between different MLST types had a median of 65,000 SNPs. Red dot represents the median SNP differences from each group. Fifteen pairs that are separated by 0 SNPs (*i.e.*, are identical) are reported in this figure as being separated by fewer than one SNP.

In an attempt to answer this question, we showed that of the remaining 12 patients where the blood and rectal swab isolates belonged to different STs, 6 patients were infected with UPEC clones such as ST73, ST69 and ST131. Looking through the hospital records these patients also had confirmed UTIs (while 2 had ascites, 1 with cholecystitis and 1 unknown), suggesting that the other route of infection possibly was through UTIs. Combined, this shows that for the majority of paired isolates the dominant isolate in the gut (and perhaps the bladder) belonged to the same ST and, from the core genome alignment, was almost indistinguishable from the isolates causing BSIs in that patient on admission to hospital.

4.3.2. Longitudinal diversity

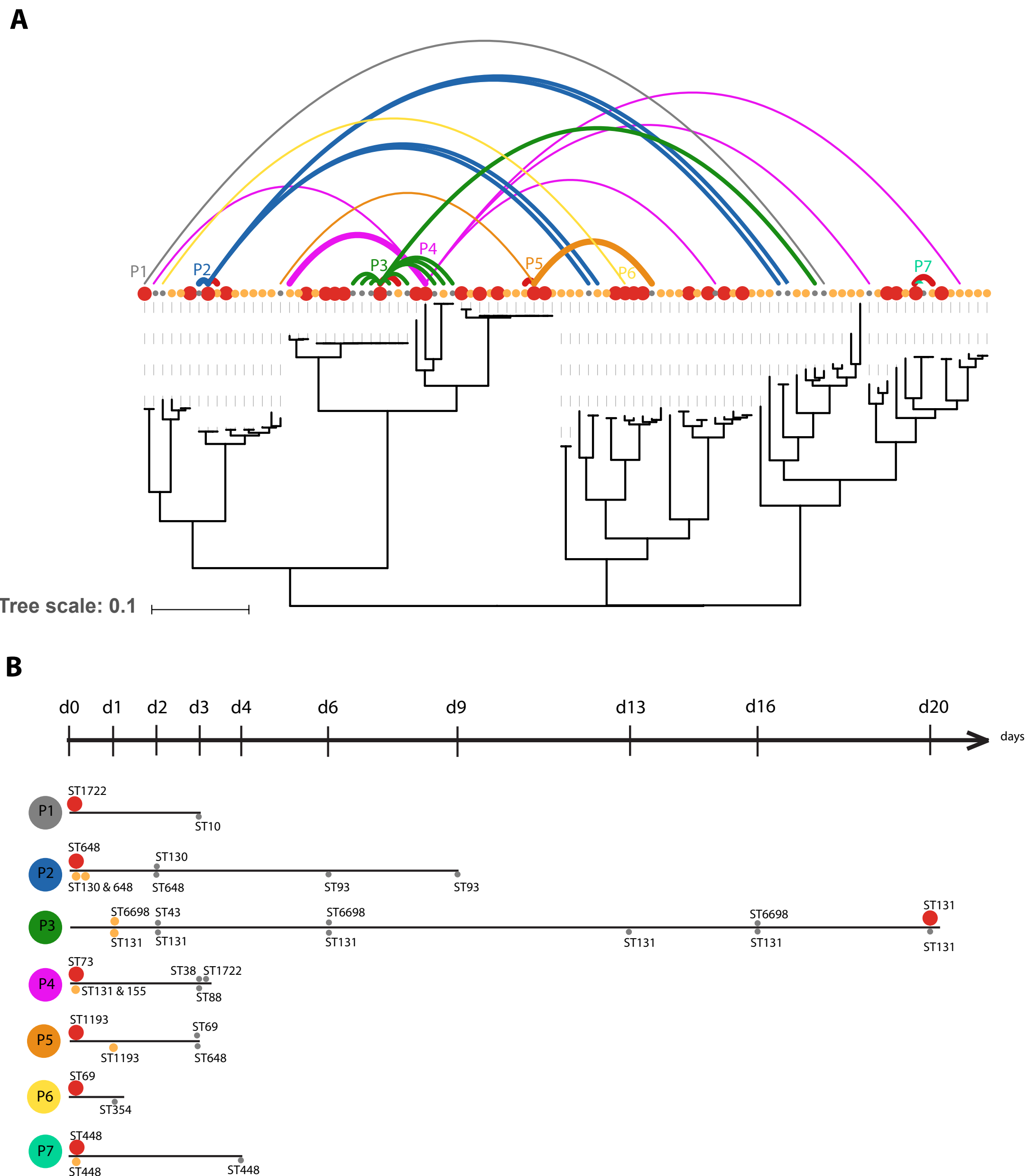

Figure 4.3: Twenty longitudinal isolates and from seven patients. The tree presented in this figure is the same as Figure 4.1 except that, for ease of viewing, only arcs from seven of the patients with multiple rectal swab samples are shown. Each arc with a different colour represents blood and rectal swab isolates taken on admission (red and orange nodes, respectively), or isolates from longitudinal samples (grey nodes) belonging to the same patient. The lower diagram shows the timeline of these longitudinal isolates and their STs.

How does the population structure of *E. coli* change over time during ICU hospitalisation and antibiotic treatment?

As mentioned above, within our sample collection there were samples that belonged to 7 patients who had been sampled multiple times after admission to HTD. The sections below attempt to bring all of the genomic and patient clinical meta data described in this and previous chapters together for these seven patients in order to describe how the population of *E. coli* in these patients has changed over time and highlight what factors may have influenced that. The phenotypic antibiotic resistance data discussed below is summarised in Supplementary Table S1 and S2 and was determined at the HTD Vietnam. The genes encoding the resistance were investigated in Chapter 3 (Supplementary Table S6). Figure 4.3 describes the STs of the *E. coli* taken from these patients plotted over the duration of their stay in the hospital with the sampling intervals labelled on it.

Patient P1 (03-0048) and P6 (03-0209)

Unfortunately, patient P1 (03-0048) and P6 (03-0209) did not have baseline isolates (strains could not be recovered from storage). However, looking at the longitudinal isolates collected from these patients after admission showed they were different to their blood isolates (ST10 and ST354) (Figure 4.3). These two patients both had liver disease (cirrhosis and ascites), and both suffered from community-acquired BSIs, infected with ESBL-producing *E. coli* belonging to ST1722 (P1) and ST69 (P6) respectively. P1 was treated with ceftriaxone and had a fatal outcome after three days. P6 was treated with imipenem for five days after admission, therefore this patient survived the BSIs although they were admitted for a total of 29 days in HTD. ST354 colonised P6 on day 1 and is an ESBL-producing *E. coli*, that carried both *bla*$_{\text{CTX-M-24}}$ and *bla*$_{\text{CMY-42}}$, and was phenotypically resistant to all tested antibiotics except carbapenem.

Patient P2 (03-0314)

Patient P2 (03-0314) was a patient suffering from cirrhosis. Initially, this patient's rectal swab sample was dominated by two STs, ST130 and ST648, both sensitive to 12 out of 13 tested antimicrobials (Supplementary Table S2). The blood isolate matched the rectal swab isolate with only one SNP difference. *E. coli* isolates were also cultured with the same AMR profile from peritoneal fluid (data not shown). It is very likely that this BSI emerged from translocation of *E. coli* from the gut. Patient P2 was treated with ceftriaxone. However, on day 6 and day 9, only MDR ESBL-producing *E. coli* of ST93 were seen in culture. These isolates harboured IncX and IncFII plasmid replicon types (result from Chapter 3). This might reflect either lineage replacement whereby only when sensitive STs such as ST130 and ST648 were killed off by exposure to ceftriaxone, we were then able to pick up drug resistant ST93. The other possibility was this patient acquired a hospital-associated MDR *E. coli*, although we have no means to identify transmission route unless we had sequenced all carriage isolates from other patients admitted to ICU around that time. The patient unfortunately did not survive.

Patient P4 and P5

Longitudinal rectal swab isolates taken from patients P4 (03-0559) and P5 (03-0710) were also found to be different STs from the original rectal swab isolate taken on admission. P4 was treated with ceftriaxone, since the admission BSI isolate, ST73, was found to be resistant to cotrimoxazole (SXT) and also had reduced susceptibility to ticarcillin/clavulanic acid (TCC) and amoxicillin/clavulanic acid (AMC). However, the rectal swab sample was dominated by ESBL-producing ST131 *bla*$_{CTX-M-14}$ and SXT resistant *E. coli* ST155 cultured from the rectal swab collected on admission. On day three post-admission, further rectal swab samples were taken and *E. coli* isolates of ST38, ST1722 and ST88 were grown. Interestingly, they were all phenotypically resistant to third- and fourth-generation cephalosporins and were ESBL-producing isolates (Supplementary Table S2). Screening for AMR genes using ARIBA and ResFinder revealed (See sections 3.3) that they all harboured *bla*$_{CTX-M}$ genes. ST38 isolates contained *bla*$_{CTX-M-27}$, ST1722 had both *bla*$_{CTX-M-27}$ and *bla*$_{DHA-1}$ while ST88 harboured *bla*$_{CTX-M-55}$.

Patient P5 had diabetes with confirmed UTIs, while carriage isolates were both the same ESBL-producing *E. coli* ST1193 harboured *bla*$_{CTX-M-27}$ so it was very likely that BSIs arose from the gut colonization then contaminated UTIs. After treatment with both ceftriaxone and amikacin, on day 3, two different STs including sensitive ST69 and MDR ST648 isolates with *aac(6')-Ib-cr*, *bla*$_{CTX-M-15}$, *bla*$_{OXA-1}$, *bla*$_{TEM-1B}$ correlating with the presence of a IncFII plasmid (data visualised in Bandage - not shown) were cultured from rectal swabs. This patient's condition did not improve after treatment, and they were transferred to another hospital, hence the opportunity to follow this up was lost.

Patient P7 (03-0514)

Patient P7 (03-0514) was one of three patients that were classified as HAIs in 06EI study since all three patients developed BSIs while staying in the hospital when treated for cirrhosis or tetanus. Since this patient was infected with the same ESBL-producing ST448 in the blood and rectal swab samples, they were treated with norfloxacin and ertapenem; nevertheless, ST448 was also shown to be present in rectal samples on day 4, with zero SNPs separating it from the original two isolates. This patient was discharged to die at home. ResFinder results revealed that these isolates carried *bla*$_{CMY-2}$.

Patient P3 (03-283)

Patient P3 (03-283) was admitted to the hospital suffering from tetanus. This patient did not have a bacteraemia on admission. This patient was swabbed six times during a 19-day stay in the ICU and *E. coli* isolates were cultured and retained from the rectal swabs on each occasion. The patient was screened for *E. coli* upon admission, and was treated with metrodinazole (as per standard guidelines for tetanus treatment). Interestingly, the cultured isolates showed multiple colony morphologies on each occasion and so two colonies were preserved from each of the rectal cultures, including those from day one. In addition, this patient developed a BSI on day 20 with the infection being recorded as a hospital-

acquired infection. It is clear (Figure 4.3) that patient P3 (03-283) rectal samples were dominated by two different *E. coli* MLST clones (ST131, ST6698), while ST43 seems likely to have colonised this patient transiently. It is also clear that the day 20 BSI isolate belonged to the same ST and phylogroup as the majority of the rectal isolates from this patient. This ST131 clone was picked up every sampling time because it is a β-haemolytic *E. coli*, hence it shows a phenotype that is distinct from other lactose fermenting clone (see Supplementary Table 3).

To see if there was evidence of variation between the longitudinally sampled ST131 isolates we re-drew the ST131 cluster in Phandango to include all seven isolates from this patient (Figure 4.4 carriage isolates denoted as LF (lactose-fermented) or NLF (non-lactose fermented) or Bhae(β-hemolytic) proceeding their sample ID, e.g. EI2793NLF, EI2793LF2/EI3892_Bhae; the day 20 BSI isolate is labelled as 010615-16405). An additional three BSIs isolates, and one carriage isolate from three different patients were added for context (Figure 4.4; Labelled as P19, P21, P25). Within seven longitudinal isolates taken from patient P3 03-0283, the blood isolate is most genetically closely related to isolate EI1605 β-haemolytic isolate from day 6, which is identical with 0 SNPs difference in the core genome alignment. The maximum SNP difference between these seven longitudinal isolates was 3 SNPs. In order to understand if there were any differences in the accessory genome between the blood and rectal swab isolates included here, we mapped sequencing reads against 205 genes known to be involved in invasive disease to look for gene presence or absence (Figure 4.4). On first analysis, this seemed to reveal something interesting. The blood isolates differed from the other five carriage isolates from the same patient by possessing two genes, *kspM*-1 and *papX* genes. PapX is important because it is a protein which represses the synthesis of P fimbriae, which we showed was highly correlated with invasive disease at HTD (described in Chapter 3). The impact of this gene would be to attenuate attachment and bacterial motility. Whilst *kspM*-1 mediated K1 capsules help *E. coli* escape host immunity response [154]. However, looking across all isolates included in Figure 4.4 it is clear that the presence or absence of these genes, although interesting, does not correlate with sample type: blood or rectal swab although it does highlight some within-host variation between isolates.

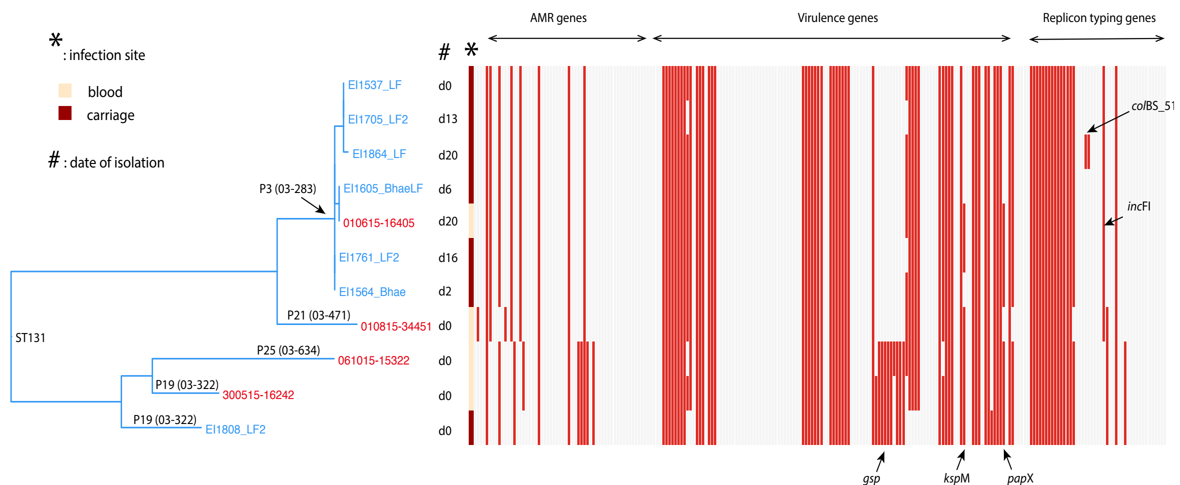

Figure 4.4: Graph shows phylogenetic relationship between isolates in patient P3 (03-283). The presence and absence of AMR and virulence genes was related to the core genome phylogeny of seven isolates from patient P3 and two other patients infected with ST131 BSIs.

4.4. Discussion

In our study, by using whole genome sequencing as a high-resolution method to dissect the complexity of *E. coli* populations, we found here that there is a substantial diversity of *E. coli* co-existing in each patient's gut; each patient carried at least two different STs from different phylogenetic lineages. Hence, our results are consistent with the study by Stoesser *et al* on multiple single *E. coli* isolates isolated from Cambodian children, in which it was found that one healthy individual can carry up to 10 different STs and a range of virulence and AMR determinants located on multiple plasmids [155].

It is also clear that 57 % of BSIs collected from our patients for whom we had matched rectal swab isolates were of the same ST and differed by 0 or only a small number of SNPs from their rectal swab isolates, which suggests that they come from their own microbiota. This has been observed in other ICU studies for *Klebsiella pneumoniae* [156, 157]. However, although we were not able to infer the route of infection for the other cases a significant number of these other patients carried isolates strongly associated with UTIs in their blood, were recorded as having UTI's and were largely female. This may suggest that this is the other major source of BSIs in these patients and is consistent with the findings of Chapter 2 and 3 showing a strong gender bias for the invasive isolates and a link to UTI's.

The diversity of multiple STs living in the same gut flora lead to an interesting point as to whether drug-sensitive *E. coli* could acquire AMR genes through plasmid transfer from AMR strains under antibiotic pressure. Although gut inflammation has been shown to boost the conjugation of a colicin-resistance plasmid between *Salmonella* Typhimurium and commensal *E. coli* in a mouse model [158], our finding in a small subset of patients rather show the lineage replacement of sensitive STs by those resistant to

the drug used for treatment (as shown in patients P2, P4 and P5). The repeated isolation of the same *E. coli* carried *bla*$_{CMY-27}$ in patient P7 (03-514) (Figure 4.3) show the potential of treatment failure of ertapenem on *bla*$_{CMY-27}$-producing *E. coli*.

Our study has limitation. Due to the number of strains that needed to be identified and stored during the study, we assumed that all *E. coli* isolates that were of the same phenotype (colony morphology, colour on MacConkey agar), and of the same AMR pattern, were of the same genotype. This meant that only one or two representative isolates from each phenotype was kept and sequenced. This could lead to an underestimation of the real diversity of *E. coli* present in the gut and might explain why 43% of our cases harbour different STs between blood and rectal swab on admission.