

Title: *Escherichia coli* extra-intestinal population translocation in leukemia patients

Running Head: *E. coli* translocation in leukemia patients

Authors: Julie Marin^{a#}, Violaine Walewski^{b,*}, Samira Dziri^b, Mélanie Magnan^{c,⌘}, Erick Denamur^c, Etienne Carbonnelle^{a,b,*}, Antoine Bridier-Nahmias^{c#}.

^a Université Sorbonne Paris Nord, Inserm, IAME, F-93000 Bobigny, France

^b APHP, HUPSSD, Hôpital Avicenne, Service de Microbiologie clinique, F-93009 Bobigny, France

^c Université Paris Cité, INSERM, IAME, 75018 Paris, France

Corresponding authors: Julie Marin (julie.marin@sorbonne-paris-nord.fr) and Antoine Bridier-Nahmias (antoine.bridier-nahmias@inserm.fr)

* present address: Violaine Walewski, Hôpital Henri Mondor, Laboratoire de Biologie Médicale, 15000 Aurillac, France.

⌘ present address: Mélanie Magnan, Université Paris Cité, CNRS, Inserm, Institut Cochin, F-75014 Paris, France.

* present address: Etienne Carbonnelle, Pôle Biopathologie, Centre Hospitalier Universitaire d'Orléans, 14 Avenue de l'Hôpital, 45100 Orléans.

Abstract

Escherichia coli, a commensal species of the human gut, is an opportunistic pathogen which can reach extra-intestinal compartments, including the bloodstream and the bladder, among others.

In non-immunosuppressed patients, purifying or neutral evolution of *E. coli* populations has been reported in the gut. Conversely, it has been suggested that when migrating to extra-intestinal compartments, *E. coli* genomes undergo diversifying selection as supported by strong evidence for adaptation. The level of genomic polymorphism and the size of the populations translocating from the gut to extra-intestinal compartments is largely unknown.

To gain insights in the pathophysiology of these translocations, we investigated the level of polymorphism and the evolutionary forces acting on the genomes of 77 *E. coli* isolated from various compartments in three immunosuppressed patients. We detected a unique strain for each patient across the blood, the urine and the gut. In one case, all isolates recovered were mutators i.e. isolates with a very high mutation rate. In all instances, we observed that translocation encompasses the majority of the genomic diversity present in the gut. The same signature of selection, whether purifying or diversifying, and as anticipated, neutral for mutator isolates, was observed in both the gut and bloodstream. Additionally, we found a limited number of non-specific mutations among compartments for non-mutator isolates. In all cases, urine isolates were dominated by neutral selection. These findings indicate that substantial proportions of populations are undergoing translocation and that they present a complex compartment-specific pattern of selection at the patient level.

Importance (150 words)

It has been suggested that intra and extra-intestinal compartments differentially constrain the evolution of *E. coli* strains. Whether host particular conditions, such as immunosuppression, could affect the strain evolutionary trajectories remain understudied. We found that, in immunosuppressed patients, large fractions of *E. coli* gut populations are translocating with variable modifications of the signature of selection for commensal and pathogenic isolates according to the compartment and/or the patient. Such multiple site sampling should be

performed in large cohorts of patients to get a better understanding of *E. coli* extra-intestinal diseases.

Keywords *Escherichia coli*, evolution, whole-genome sequencing, infection, immunosuppression, adaptation, blood stream infection (BSI), genomics

Introduction

Bloodstream infections (BSIs) are still a major concern among onco-hematologic patients, depending on the type of pathogen, level of host immunodeficiency and status of underlying disease. Despite advancements in the clinical management of hematological malignancies, BSIs remain life-threatening complications in the clinical course of these patients, with reported crude mortality rate up to 40% (1–3). A clear shift of bacterial species causing BSIs in patients with hematological malignancies has been recently reported, transitioning from Gram-positives to Gram-negatives. In the first place Enterobacteriaceae, and in particular *Escherichia coli*, represent the most frequently involved bacterial species, together with the worrisome and growing phenomenon of multiresistant bacteria (4, 5).

E. coli is a commensal species of the lower intestine of humans (6). The gut is the primary habitat of *E. coli*, and probably the main ecological context of selection. Virulence genes, for instance, are thought to be primarily selected in the intestine as a by-product of commensalism (7, 8). *E. coli* is also an opportunistic pathogen, frequently responsible for intestinal and extra-intestinal infections (9). When reaching a new compartment, such as the bladder or the bloodstream, *E. coli* faces new challenges and new opportunities for adaptation.

In the past, only a few studies have investigated the signatures of selection in commensal *E. coli* isolates and in isolates sampled from extra-intestinal infections, revealing different scenarios. The evolution of commensal *E. coli* has been found to be governed by purifying selection, whether it implies the entire genome (10) or specific genes such as the H7 flagellin genes (11). However, another study, following the evolution of a clone in a single individual over almost a year, did not find any evidence for selection in the gut (12). On the contrary, it was found

that adaptation is at play during chronic and acute infections (13). In particular, it was shown that *E. coli* isolates colonizing extra-intestinal sites were adapting under strong selective pressure, with an excess of non-synonymous mutations and patterns of convergence at the gene level (including for H7 flagellin genes). Evidence for adaptation during chronic infection have been emphasized for other bacteria, such as *Burkholderia dolosa* (14) or *Pseudomonas aeruginosa* (15–17), with genotypically and phenotypically diversifying lineages during cystic fibrosis infections for instance (18–21). Adaptation to the human host (22, 23), evasion from the immune response (24, 25) and acquisition of antibiotic resistances (26, 27) are favored by this accumulation of mutations. In addition, niche adaptation shapes the allelic diversity among compartments with specific mutations occurring in the gut or in the bladder associated with functions increasing *E. coli* fitness in one or the other compartment (28).

What happens when the intestinal barrier and the immune response are weakened? Patient's deficiencies, such as immunosuppression, weakening of the intestinal barrier by antibiotic therapy or chemotherapy, can increase the risk of *E. coli* extra-intestinal infection (29–31). Anticancer chemotherapy drugs have a direct impact on the intestinal microbiota (32). The resulting dysbiosis can lead to intestinal mucositis which increases the risk of bacterial translocation to the bloodstream (33). Moreover, among patients undergoing chemotherapy, leukemia patients are highly prone to extra-intestinal infections and relapse (34). The weakening of intestinal barriers and the lack of immune defense could modify the adaptive conditions described for commensal and pathogenic isolates of *E. coli* in non-immunosuppressed patients. Modifications in the signature of selection are therefore expected for commensal and pathogenic isolates of *E. coli* in immunosuppressed patients. In addition, the level of genomic polymorphism and the proportion of the population translocating is largely unknown.

Here, we evaluated the selective forces acting on *E. coli* evolution in three immunosuppressed patients among three compartments, the bladder, the bloodstream and the gut. For each patient, we analyzed whole genome sequences of the isolates found in each compartment to evaluate genetic diversity and assess the strength of the various selective processes at play.

Material and methods

Sampling

Clinical isolates were isolated from three patients with hematologic disease and *E. coli* sepsis hospitalized in Avicenne (Seine-Saint-Denis, France). Patient A had three infectious episodes while patient B and C had one each. For each episode, isolates were sampled from positive blood culture along from bladder and feces samples on the same or subsequent day. The initial blood culture was obtained before any antibiotic therapy. All procedures performed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national), validated by the ethics committee of Avicenne's hospital (Comité Local d'Éthique d'Avicenne). Information on the type and status of hematologic disease, presence of neutropenia ($<0.5 \times 10^9/L$), previous exposure to any antibiotic therapy, including prophylaxis or treatment of prior infectious episodes, type of infection, microbiological isolate, and outcome was collected in a database.

Patient characteristics

The three patients (A to C) studied had hematological pathologies: two acute myeloid leukemia and one with Hodgkin's lymphoma. Patient A was diagnosed with acute myeloid leukemia and myelofibrosis in January 2014 (Acute myeloblastic leukemia with minimal maturation M1). He was treated with allograft and experienced relapse in March 2015, with febrile neutropenia and presence of circulating blasts. During the BSI, the patient presented an inflammatory syndrome, a severe immunosuppression and had received antibiotics in the previous 30 days. Patient B also suffered from undifferentiated acute myeloblastic leukemia (M0) diagnosed and allografted in 2015. At the time of BSI, he was receiving chemotherapy, had severe immunosuppression, an inflammatory syndrome, and had received antibiotics within the preceding 30 days. Patient C was diagnosed with Hodgkin's lymphoma in 2009. Relapsed in 2017 during the BSI episode, the patient exhibited inflammatory syndrome, severe immunosuppression and had previously received antibiotics.

Prior in vitro phylogroup typing

We determined the phylogroup of each sample (blood, urinary or feces) using quadruplex PCR (35) to select isolates belonging to the same phylogroup for each infection episode.

Genome sequencing and assembling

We whole-genome sequenced each sample with Illumina technology (MiSeq and HiSeq 2500) using Nextera XT library preparation kits as instructed by the manufacturer (Illumina, San Diego, CA). Fastq files (raw sequencing data) were submitted to the European nucleotide archive (see **Table S1** for accession numbers). Genome assembly was performed with SPAdes v.3.15.5 (36) .

For each patient, one strain was chosen as a reference and was sequenced with the Oxford Nanopore Technologies MinION platform using an R9.4 flow cell. We prepared the samples with kits LSK-108 and NBD-104 for library preparation and barcoding. The Nanopore reads were filtered with Filtlong (37) using the following parameters: minimum length of 1000 and 95% of the best reads kept. High-quality assemblies of the three reference genomes were obtained through a hybrid strategy using both Illumina and Nanopore reads with Unicycler V0.4.4 (38).

Core and accessory genome

The 77 SPAdes assemblies were annotated with Prokka (39). Plasmid sequences were predicted by PlaScope (40). We then performed pan-genome analysis from annotated assemblies with Roary on the 77 genomes using default parameters (41). The core genome alignment and the list of genes of the accessory genome were generated for the three patients.

Variant calling (SNPs and deletions)

We first discarded bases with a low-quality score (< 30) and removed the adapters with Trim Galore [a wrapper of the Cutadapt program (42)]. To detect SNPs, we aligned the reads of each patient to the corresponding reference sequence (**Table S1**) with Snippy 4.4.0 (43) with the following parameters: the nucleotide minimum quality to be analyzed (basequal) equal to 20, the minimum number of reads covering a site (mincov) equal to 10 and the minimum proportion of

those reads different from the reference (minfrac) equal to 0.9. To detect structural variants (deletions) we first mapped the reads to the corresponding reference assembly with BWA-MEM (44) and then analyzed the obtained Sequence Alignment Map (sam) files with Whamg (45). As recommended, we remove calls smaller than 50 bp and larger than 2 Mbs. For each strain, we also removed calls with less than 5 supporting reads.

Typing and genotypic antibiotic resistance and virulence

We used an in-house script, petanc (46), that integrates several existing bacterial genomic tools to perform the typing of isolates with several genotyping schemes using the genomic tool SRST2 (47). Sequence types (STs) were defined using the Warwick MLST scheme and the Pasteur scheme (48, 49). We only used the Warwick scheme for the analyses described thereafter. We also determined the O:H serotypes and the FimH alleles (Ingle et al. 2016; Roer et al. 2017). The phylogroups were confirmed using the ClermonTyping method (50).

The resistome and virulome were first established using the in-house script Petanc (46). They were defined by BlastN with Abricate ([https:// github.com/tseemann/abricate](https://github.com/tseemann/abricate)) using the ResFinder (version 4.2.2) database (51), a custom database including the VirulenceFinder database and VFDB (52, 53), to which we added selected genes (54). We set the threshold for minimum identity to 80% with a minimum coverage of 90%. Next, we built a pan-resistome and pan-virulome including all the antibiotic resistance and virulence associated genes of all isolates. We mapped the reads to the corresponding pan-resistome and pan-virulome with BWA-MEM (44). We considered a gene as present when we found more than 80% coverage and at least one read. We also tested more conservative thresholds with 80% coverage and more than 5 reads.

Genomic diversity and traces of selection

We compared the rates of nonsynonymous and synonymous mutation by computing dN/dS ratios (R language (55)) from the gene alignments obtained with Roary, to evaluate the genomic traces of selection. First, we removed the identical sequences. Next, we calculated the number of non-synonymous substitutions (dN) and synonymous substitutions (dS) as the observed number of substitutions of each type divided by the number of potential substitutions of the same type in the

considered sequence. For each codon, the number of potential non-synonymous or synonymous substitutions is determined by the genetic code. We then computed the dN/dS ratio for each pair of sequences, we determined the median dN/dS ratio and assess whether it significantly differed from 1 (Wilcoxon test), where 1 represents perfect balance between diversifying and purifying selection indicating no visible selection. When in a pair of sequences 1 or more non synonymous and 0 synonymous mutations were encountered, an infinite value is returned. Those were replaced by the maximum non-infinite value obtained for the considered set of sequences. 'Not a number' (NaN) values, when there was no non-synonymous mutation, were replaced by 0.

Results

We evaluated the genomic diversity (SNPs and deletions) and the genomic traces of selection in 77 *E. coli* isolates sampled concomitantly in the gut, blood and urine compartments of three patients (**Figure 1**). The distribution of isolates for each patient and compartment is detailed in **Table 1**.

Phylogroup, ST, serotype and fimH allele diversity

Within each infection episode, all isolates belonged to the same phylogroup and ST and had the same serotype and *fimH* allele (**Table 1**). Isolates belonged to the phylogroup A (patient A), B1 (patients B) and B2 (patient C).

Global genomic diversity

We first looked at the SNP diversity. At the exception of patient B, a low number of mutated genes (between 8 and 11) and no deletion was recovered when comparing isolates of each patient between them (**Table S2**). For patient B, we found 328 mutated genes. Indeed, for patient B isolates we noted a deletion of *mutS* (see below) and non-synonymous mutations in *mutL*. These inactivations of the DNA mismatch repair system conferred to them a mutator phenotype (56) (**Table S3**).

In line with those results, we found a low number of SNPs for patients A and C (mean number of 5.07 [95% confidence interval: 4.20-5.95]) compared to patient B (mean number of

117.13 [82.16-152.10], p-value = 1.237e-08 (Wilcoxon test)) (**Figure 2**). Similarly, we did not find any difference between intra- and extra-intestinal isolates among variant categories (**Table 2**) (chi-squared test after correcting for sample size, p-value = 0.99). No common gene with SNPs was identified among patients (**Table S2 and S4**).

Within patient compartment diversity

All isolates from a given patient were highly similar (except for patient B with mutator isolates) (**Figures 2-3**). As a consequence, no large deletions were detected because we used one of these isolates as the reference strain for each patient (**Table S2**).

The majority of SNPs were shared among compartments (**Figure 4**). For patient A, 13 over 20 SNPs were identical among at least 2 compartments and 13 over 22 for patient C. For patient B (mutator strain), despite a high level of polymorphism, most of the SNPs were shared among stool and blood isolates. Furthermore, most of the stool polymorphism was found in other compartments, 87%, 77% and 87% for patient A, B and C respectively, suggesting a large population translocation from the gut.

Antibiotic resistance and virulence associated gene content

Regarding genes associated with antibiotic resistance, isolates within each infection were highly similar. We only found few differences among isolates of patient B (mutator isolates) and A which were sampled at different times (**Figures 1 and 5**). For patient A, we found gene presence/absence discrepancies for three resistance genes, *sul2*, *tet(A)* and *dfrA5*. The gene *sul2* was consistently predicted as plasmidic across all isolates (**Table S5**). The genes *dfrA5* and *tet(A)* were found on the same contig in 20 out of 23 isolates possessing both genes, and were predicted as plasmidic for 96.67% and 86.96% of the isolates respectively. For patient B, we found presence/absence discrepancies for all the 7 resistance genes detected. Specifically, the genes *qnrS1*, *blaCTX-M-15*, *blaTEM-1B*, *aph(6)-Id*, *aph(3'')-Ib* and *sul2* were always found on the same contig and predicted as plasmidic. The gene *dfrA14* was predicted as plasmidic for all isolates. Similar results were obtained using a more conservative threshold (at least 5 reads covering more

than 80% of the gene) (**Figure S1**), with the exception of the gene *mdf(A)*, predicted as chromosomic, which was missing for three isolates of patient B.

We also found high gene content similarity for virulence associated genes (**Figure 6**). An identical gene content was found in patient A isolates. For patient C, there was a discrepancy for a single gene, *iss11*, consistently predicted as chromosomic, and absent in four isolates. Slightly more differences were found for patient B (mutator strain). The genes *espX5*, *espX1* and *gad20* were missing for 1, 1 and 5 isolates respectively, all predicted as chromosomic (**Table S6**). We found more discrepancies with a more conservative threshold (at least 5 reads covering more than 80% of the gene), highlighting the lower sequencing depth of two isolates in particular (**Figure S2**). In patient A, all the isolates had the same gene content as the exception of 2 isolates (2ana3 and 9b4c8). We found 6 missing genes for 2ana3 and one missing gene for 2c1c5. These genes were all predicted as chromosomic. In addition, *entA*, *entB* genes were always adjacent on the same contig (**Table S6**). In patient C, the gene *iss11*, always predicted as chromosomic, was missing for 5 isolates. We found more differences for patient B. Over the 47 virulence-associated genes detected, all predicted as chromosomic, only 15 were present in all isolates.

Genomic traces of selection

We computed dN/dS ratios to test whether gene sequences evolved neutrally or were under purifying or diversifying selection (**Figure 7, Table S7**). For all patients, the same pattern of selection was found for blood and stool isolates, diversifying selection for patient A, neutral selection for patient B and purifying selection for patient C. The dN/dS estimation for patient B isolates, which are all mutators, is not indicative of the action of selection. However, the presence of mutators at high-frequency in the population (100 % in patient B) can be considered as a signature of sustained adaptation (57, 58). We also found that all isolates sampled in urine evolved under neutral selection (patients A and C). It should be noted that the number of structural variants (missense and synonymous) cannot be directly compared to dN/dS ratios. Structural variants were computed with reference to a single sequence whereas ratios were computed pairwise.

Discussion

E. coli is an opportunistic pathogen which may cross the intestinal barrier to reach extra-intestinal compartments causing infections. While evidence for purifying and neutral selection has been detected in commensal *E. coli* (10–12), extra-intestinal infection isolates have been shown to be under strong adaptive selection (13). Here we investigated both commensal and extra-intestinal isolates of leukemia patients undergoing chemotherapy. We observed the same strain in all compartments with a very limited number of SNPs, at the exception of the mutator strain. Interestingly, while each patient had a different selection signature, it was identical in both the bloodstream and the gut. In urine, in all cases, neutral selection was at play.

Gut origin of urine and blood immunosuppressed patient isolates

For each patient, the isolates collected from the gut, urine, and blood exhibited highly similar sequences. They share the same phylogroup, ST, serotype, and *fimH* allele (**Table 1**), indicating that the strain most likely originated from the gut, as previously shown (34, 59). Moreover, with the exception of the mutator isolates, there were less than 15 SNPs between pairs of isolates of the same patient (**Figure 2**), indicating low genetic diversity, and the resistance and virulence profiles were stable (**Figures 5-6**). Interestingly, a previous study, on four patients, also showed a low diversity (ranging from 0 to 4 SNPs) when comparing urine isolates translocating to the blood (60). The phylogenetic distribution of isolates, together with the SNPs distribution among compartments, suggested the translocation of the gut population diversity to extra-intestinal compartments for all patients (**Figures 3-4**).

Same adaptation forces at play in the gut and bloodstream of immunosuppressed patient

In non-immunosuppressed patients, purifying (or neutral) evolution of *E. coli* populations has been reported in the gut (12) but not in extra-intestinal compartments where bacteria DNA sequence are under diversifying selection with strong evidence for gene level convergence (13).

Here, we found various signatures of selection in gut isolates. In patient A, diversifying selection is at play, whereas purifying selection shaped the isolates of patient C. Differences in the microbiota perturbation of these patients undergoing a specific antibiotic therapy and

chemotherapy could explain this discrepancy. However, the limited number of patients in this study prevents us from drawing such conclusions. The presence of mutators (all isolates of patent B) is indirect evidence for ongoing selection (57, 58). Some anticancer chemotherapy drugs enhance the bacterial mutagenesis, thus promoting the emergence of a mutator clone (61). Then, if the amount of beneficial mutations is large enough, a mutator clone can reach a high frequency.

Strikingly, the same adaptive process was at play in the bloodstream and in the gut for each patient. Moreover, unlike in extra-intestinal infections of non-immunosuppressed patients (13), no convergence among patients was detected. Finally, isolates were almost identical with a very low number of SNPs differentiating them (five on average) unspecific to the compartment. Whereas, in non-immunosuppressed patients, niche adaptation was evidenced by specific mutations associated with isolates from the gut or the bladder (28). These results suggest that the intestinal barrier does not act as a barrier anymore and that the adaptive constraints were homogenized among both compartments. A weakened immune system and a porous intestinal barrier, due to chemotherapy and antibiotic treatments, is the most obvious explanation for this phenomenon.

Neutral selection in the bladder of immunosuppressed patient

For both patients with urine samples, isolates evolved neutrally. Fluctuating environment or a small population size could explain the weak strength of selection observed in the commensal habitat of *E. coli*. Indeed, bacteria in the bladder face continuous adaptive challenges with fluctuations in exposure to the host immune system, to antibiotic treatments, to nutrients and to a diverse microbial community in the host due to an alternation between storage and voiding phases (62).

Limitations and future directions

Our work obviously presents a major limitation as we evaluated a very limited number of patients, which prevents us from drawing general conclusions. Collecting stool before antibiotic treatment is indeed very difficult to achieve in clinical practices. Nevertheless, the evaluation of larger cohorts with a greater number of samples is necessary to associate evolution patterns and the

environment (patient follow-up and patient-dependent factors) and to decipher the factors linked to the selection patterns observed here.

Conclusion

Our results highlight the complexity of the pathophysiology of *E. coli* BSIs. Without an efficient immune system and strong intestinal walls, these bacteria can easily cross the intestinal barrier and colonize extra-intestinal compartments. Here, we showed that these particular conditions homogenized the adaptive conditions in the gut and the bloodstream, further reinforcing the ease of large bacterial population translocation between them.

Data availability

The data generated in this study have been submitted to the NCBI BioProject database under the accession number PRJEB69525 [<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB69525>].

Code availability

The R scripts used to analyze the data and produce the figures are available on GitHub (https://github.com/j-marin/hemato_AVC) under a permissive licence (MIT).

Acknowledgments

We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing computing resources. We would also like to thank the Molecular Biology and Data Processing Platform of the Sorbonne Paris Nord University. This study was partly financed by two grants: one awarded by the research commission of the Université Sorbonne Paris Nord (BQR) and the other by the IFRB, UFR SMBH, Université Sorbonne Paris Nord.

Table 1. Patient sampling and strain typing.

Patient	Number of isolates (stool – blood – urine) †	Phylogroup	Sequence type	Serotype	<i>fimH</i> allele
Patient A	30 (13 -12 - 5)	A	10	O101:H9	<i>fimH54</i>
Patient B*	17 (10 - 7)	B1	1431	O8:H19	<i>fimH32</i>
Patient C	30 (10 - 10 - 10)	B2	131	O25b:H4	<i>fimH30</i>

Phylogroups, MLSTs (Warwick and Pasteur), serotypes and *fimH* alleles are indicated. All the isolates of patient B were mutators (*).

(†) Patients B and C were sampled at a single time point whereas samples of patient A correspond to 4 timepoints.

Table 2. Number of structural variant types among isolates of a patient's infection episode found with snippy.

	Site	conservative inframe insertion	disruptive inframe deletion	frameshift variant	missense variant	non coding transcript variant	stop gained	stop lost	synonymous variant
Patient A	blood (n=12)	9	0	1	3	0	1	0	9
	stool (n = 13)	12	1	0	6	0	0	1	17
	urine (n=5)	0	5	0	20	0	0	5	20
Patient B*	blood (n=7)	0	0	216	541	5	11	20	292
	stool (n=10)	0	0	86	177	1	4	4	146
Patient C	blood (n=10)	0	0	9	7	0	0	0	2
	stool (n=10)	0	0	16	26	0	0	0	14
	urine (n=10)	0	0	12	20	0	0	0	6
Intra-intestinal isolates	stool (n=23)	12	1	16	32	0	0	1	31
Extra-intestinal isolates	blood/urine (n=37)	9	5	22	50	0	1	5	37

We excluded patient B from intra- and extra-intestinal isolates because all the 17 isolates were mutators (*). n: number of isolates.

Figure legends

Figure 1. Patient follow-up with their main clinical characteristics and the sampling schemes. Absence (light gray) or presence (dark grey) of each criterion are shown. Patient A had four infectious episodes, patients B and C had one infectious episode.

Figure 2. Genomic diversity of *E. coli* isolates among samples and compartments for the three patients. (A-C) Heatmaps showing the number of SNPs between each isolate of patients (A, B and C). Samples are ordered by site horizontally and clustered according to their SNP similarities vertically (method: complete). All isolates of patient B are mutators (*). The isolate corresponding to the reference sequence is indicated (#). We used the same color scale for all patients. Note that the number of SNPs for patients A and C is between 0 and 14 and between 0 and 400 for patient B.

Figure 3. Unrooted trees of *E. coli* isolates from patients A, B and C. Trees were built using neighbor joining from the substitution presence/absence matrix. The scale indicates the number of substitutions. All isolates of patient B are mutators (*). We zoomed in on a clade of the patient B to highlight the scale difference. The bicolor points (patients A and C) denote the presence of isolates sampled in different sites with identical sequences (0 SNP).

Figure 4. Venn diagrams showing the SNPs distribution among compartments for patients A, B and C. The ellipses are proportional to the number of SNPs.

Figure 5. Presence/absence heatmaps of antibiotic resistance genes of *E. coli* isolates when compared to the pan-resistome (including the resistance genes of all isolates). We considered a gene as present when at least 80% of its length was covered by at least one read. Genes are ordered by synteny on contigs. All isolates of patient B are mutators (*). The prevailing predicted localization of genes by PlaScope (chromosomal or plasmidic) is indicated (full list in supplementary Table S5). Note that chromosomal genes are not mobile.

Figure 6. Presence/absence heatmaps of virulence associated genes of *E. coli* isolates when compared to the pan-virulome (including the virulence genes of all isolates). We considered a gene as present when at least 80% of its length was covered by at least one read. Genes are ordered by synteny on contigs. All isolates of patient B are mutators (*). The prevailing predicted localization of genes by PlaScope (chromosomic or plasmidic) is indicated (full list in supplementary Table S6). Note that plasmidic genes are not mobile, at the opposite of resistance genes (see Figure 5).

Figure 7. Action of selection on sequences (dN/dS) of *E. coli* isolates for patients A, B and C. Significant results are framed in black (see Table S7). All isolates of patient B are mutators (*). Neutrality (dN/dS not significantly different from 0) is indicated in pale yellow whereas purifying and diversifying selection are indicated in green and red, respectively.

References

1. Pagano L, Caira M, Rossi G, Tumbarello M, Fanci R, Garzia MG, Vianelli N, Filardi N, De Fabritiis P, Beltrame A, Musso M, Piccin A, Cuneo A, Cattaneo C, Aloisi T, Riva M, Rossi G, Salvadori U, Brugiattelli M, Sannicolò S, Morselli M, Bonini A, Viale P, Nosari A, Aversa F, for the Hema e-Chart Group I. 2012. A prospective survey of febrile events in hematological malignancies. *Ann Hematol* 91:767–774.
2. Klastersky J, Ameye L, Maertens J, Georgala A, Muanza F, Aoun M, Ferrant A, Rapoport B, Rolston K, Paesmans M. 2007. Bacteraemia in febrile neutropenic cancer patients. *International Journal of Antimicrobial Agents* 30:51–59.
3. Tumbarello M, Spanu T, Caira M, Trecarichi EM, Laurenti L, Montuori E, Fianchi L, Leone F, Fadda G, Cauda R, Pagano L. 2009. Factors associated with mortality in bacteremic patients with hematologic malignancies. *Diagnostic Microbiology and Infectious Disease* 64:320–326.
4. Trecarichi EM, Pagano L, Candoni A, Pastore D, Cattaneo C, Fanci R, Nosari A, Caira M, Spadea A, Busca A, Vianelli N, Tumbarello M. 2015. Current epidemiology and antimicrobial resistance data for bacterial bloodstream infections in patients with hematologic malignancies: an Italian multicentre prospective survey. *Clinical Microbiology and Infection* 21:337–343.
5. Montassier E, Batard E, Gastinne T, Potel G, de La Cochetière MF. 2013. Recent changes in bacteremia in patients with cancer: a systematic review of epidemiology and antibiotic resistance. *Eur J Clin Microbiol Infect Dis* 32:841–850.
6. Tenailon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8:207–217.

7. Diard M, Garry L, Selva M, Mosser T, Denamur E, Matic I. 2010. Pathogenicity-Associated Islands in Extraintestinal Pathogenic *Escherichia coli* Are Fitness Elements Involved in Intestinal Colonization. *Journal of Bacteriology* 192:4885–4893.
8. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, Tenaillon O. 2007. Extraintestinal Virulence Is a Coincidental By-Product of Commensalism in B2 Phylogenetic Group *Escherichia coli* Strains. *Molecular Biology and Evolution* 24:2373–2384.
9. Denamur E, Clermont O, Bonacorsi S, Gordon D. 2021. The population genetics of pathogenic *Escherichia coli*. *Nat Rev Microbiol* 19:37–54.
10. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, Waller A, Mende DR, Kultima JR, Martin J, Kota K, Sunyaev SR, Weinstock GM, Bork P. 2013. Genomic variation landscape of the human gut microbiome. *Nature* 493:45–50.
11. Hu D, Fuller NR, Caterson ID, Holmes AJ, Reeves PR. 2022. Single-gene long-read sequencing illuminates *Escherichia coli* strain dynamics in the human intestinal microbiome. *Cell Rep* 38:110239.
12. Ghalayini M, Launay A, Bridier-Nahmias A, Clermont O, Denamur E, Lescat M, Tenaillon O. 2018. Evolution of a Dominant Natural Isolate of *Escherichia coli* in the Human Gut over the Course of a Year Suggests a Neutral Evolution with Reduced Effective Population Size. *Applied and Environmental Microbiology* 84:e02377-17.
13. Bridier-Nahmias A, Launay A, Bleibtreu A, Magnan M, Walewski V, Chatel J, Dion S, Robbe-Saule V, Clermont O, Norel F, Denamur E, Tenaillon O. 2021. *Escherichia coli* Genomic Diversity within Extraintestinal Acute Infections Argues for Adaptive Evolution at Play. *mSphere* 6:e01176-20.

14. Lieberman TD, Michel J-B, Aingaran M, Potter-Bynoe G, Roux D, Davis MR, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, McAdam AJ, Priebe GP, Kishony R. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. 12. Nat Genet 43:1275–1280.
15. Yang L, Jelsbak L, Marvig RL, Damkiær S, Workman CT, Rau MH, Hansen SK, Folkesson A, Johansen HK, Ciofu O, Høiby N, Sommer MOA, Molin S. 2011. Evolutionary dynamics of bacteria in a human host environment. Proceedings of the National Academy of Sciences 108:7481–7486.
16. Marvig RL, Sommer LM, Molin S, Johansen HK. 2015. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. 1. Nat Genet 47:57–64.
17. Jorth P, Staudinger BJ, Wu X, Hisert KB, Hayden H, Garudathri J, Harding CL, Radey MC, Rezayat A, Bautista G, Berrington WR, Goddard AF, Zheng C, Angermeyer A, Brittnacher MJ, Kitzman J, Shendure J, Fligner CL, Mittler J, Aitken ML, Manoil C, Bruce JE, Yahr TL, Singh PK. 2015. Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. Cell Host & Microbe 18:307–319.
18. Ashish A, Paterson S, Mowat E, Fothergill JL, Walshaw MJ, Winstanley C. 2013. Extensive diversification is a common feature of *Pseudomonas aeruginosa* populations during respiratory infections in cystic fibrosis. Journal of Cystic Fibrosis 12:790–793.
19. Darch SE, McNally A, Harrison F, Corander J, Barr HL, Paszkiewicz K, Holden S, Fogarty A, Crusz SA, Diggle SP. 2015. Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection. 1. Sci Rep 5:7649.
20. Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, Kishony R. 2014. Genetic

variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. 1. Nat Genet 46:82–87.

21. Williams D, Evans B, Haldenby S, Walshaw MJ, Brockhurst MA, Winstanley C, Paterson S. 2015. Divergent, Coexisting *Pseudomonas aeruginosa* Lineages in Chronic Cystic Fibrosis Lung Infections. Am J Respir Crit Care Med 191:775–785.
22. Suerbaum S, Josenhans C. 2007. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. 6. Nat Rev Microbiol 5:441–452.
23. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D’Argenio DA, Miller SI, Ramsey BW, Speert DP, Moskowitz SM, Burns JL, Kaul R, Olson MV. 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. Proceedings of the National Academy of Sciences 103:8487–8492.
24. Moxon ER, Rainey PB, Nowak MA, Lenski RE. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Current Biology 4:24–33.
25. van der Woude MW, Bäumlér AJ. 2004. Phase and Antigenic Variation in Bacteria. Clinical Microbiology Reviews 17:581–611.
26. Musher DM, Dowell ME, Shortridge VD, Flamm RK, Jorgensen JH, Le Magueres P, Krause KL. 2002. Emergence of Macrolide Resistance during Treatment of Pneumococcal Pneumonia. New England Journal of Medicine 346:630–631.
27. Wong A, Kassen R. 2011. Parallel evolution and local differentiation in quinolone resistance in *Pseudomonas aeruginosa*. Microbiology 157:937–944.
28. Thänert R, Choi J, Reske KA, Hink T, Thänert A, Wallace MA, Wang B, Seiler S, Cass C, Bost MH, Struttman EL, Iqbal ZH, Sax SR, Fraser VJ, Baker AW, Foy KR, Williams B, Xu B,

Capocci-Tolomeo P, Lautenbach E, Burnham C-AD, Dubberke ER, Kwon JH, Dantas G. 2022.

Persisting uropathogenic *Escherichia coli* lineages show signatures of niche-specific within-host adaptation mediated by mobile genetic elements. *Cell Host & Microbe* 30:1034-1047.e6.

29. Kessler R, Nisa S, Hazen TH, Horneman A, Amoroso A, Rasko DA, Donnenberg MS. 2015. Diarrhea, bacteremia and multiorgan dysfunction due to an extraintestinal pathogenic *Escherichia coli* strain with enteropathogenic *E. coli* genes. *Pathogens and Disease* 73.
30. Maslow JN, Mulligan ME, Arbeit RD. 1994. Recurrent *Escherichia coli* bacteremia. *Journal of Clinical Microbiology* 32:710–714.
31. Berg RD. 1995. Bacterial translocation from the gastrointestinal tract. *Trends in Microbiology* 3:149–154.
32. van Vliet MJ, Tissing WJE, Dun CAJ, Meessen NEL, Kamps WA, de Bont ESJM, Harmsen HJM. 2009. Chemotherapy Treatment in Pediatric Patients with Acute Myeloid Leukemia Receiving Antimicrobial Prophylaxis Leads to a Relative Increase of Colonization with Potentially Pathogenic Bacteria in the Gut. *Clinical Infectious Diseases* 49:262–270.
33. van Vliet MJ, Harmsen HJM, Bont ESJM de, Tissing WJE. 2010. The Role of Intestinal Microbiota in the Development and Severity of Chemotherapy-Induced Mucositis. *PLOS Pathogens* 6:e1000879.
34. Samet A, Śledzińska A, Krawczyk B, Hellmann A, Nowicki S, Kur J, Nowicki B. 2013. Leukemia and risk of recurrent *Escherichia coli* bacteremia: genotyping implicates *E. coli* translocation from the colon to the bloodstream. *Eur J Clin Microbiol Infect Dis* 32:1393–1400.
35. Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli*

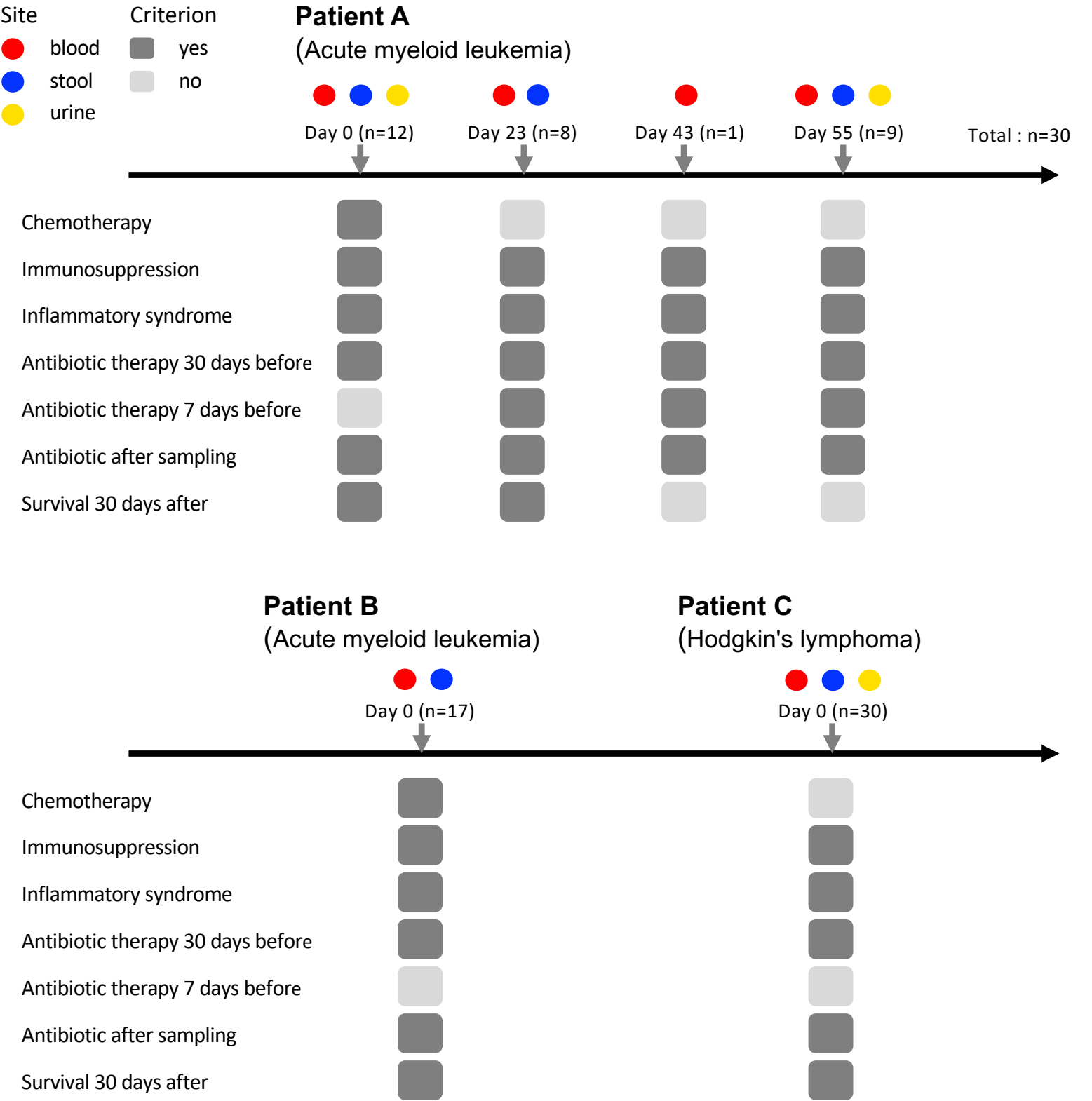
phylo-typing method revisited: improvement of specificity and detection of new
phylo-groups. Environmental Microbiology Reports 5:58–65.

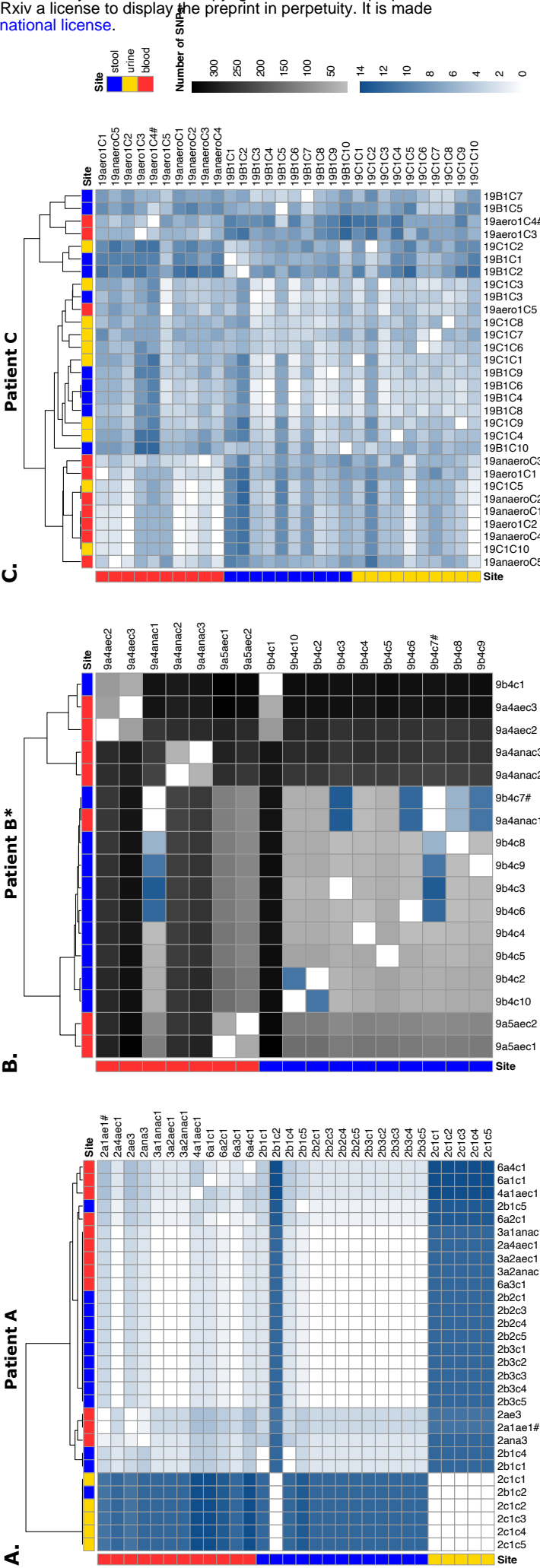
36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19:455–477.
37. Wick R. 2022. rrwick/Filtlong. C++.
38. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology 13:e1005595.
39. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.
40. Royer G, Decousser JW, Branger C, Dubois M, Médigue C, Denamur E, Vallenet D. 2018. PlaScope: a targeted approach to assess the plasmidome from genome assemblies at the species level. Microbial Genomics 4:e000211.
41. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693.
42. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 1. EMBnet.journal 17:10–12.
43. Seemann T. 2015. Snippy: fast bacterial variant calling from NGS reads.
44. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio].

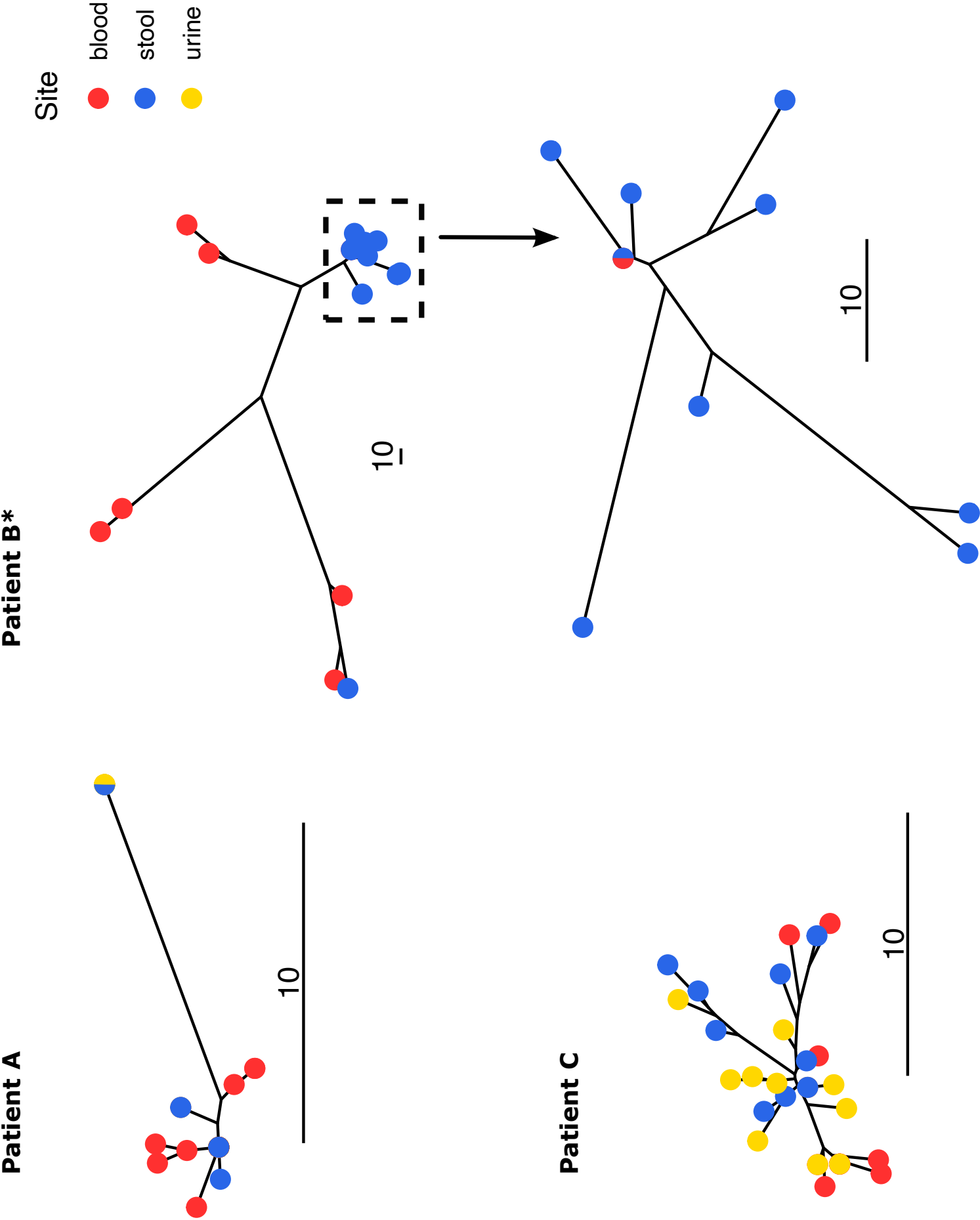
45. Kronenberg ZN, Osborne EJ, Cone KR, Kennedy BJ, Domyan ET, Shapiro MD, Elde NC, Yandell M. 2015. Wham: Identifying Structural Variants of Biological Consequence. *PLOS Computational Biology* 11:e1004572.
46. Bourrel AS, Poirel L, Royer G, Darty M, Vuillemin X, Kieffer N, Clermont O, Denamur E, Nordmann P, Decousser J-W, IAME Resistance Group. 2019. Colistin resistance in Parisian inpatient faecal *Escherichia coli* as the result of two distinct evolutionary pathways. *Journal of Antimicrobial Chemotherapy* 74:1521–1530.
47. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine* 6:90.
48. Jauregui F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbonnelle E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S. 2008. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 9:560.
49. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Molecular Microbiology* 60:1136–1151.
50. Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. 2018. ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb Genom* 4:e000192.
51. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy* 67:2640–2644.

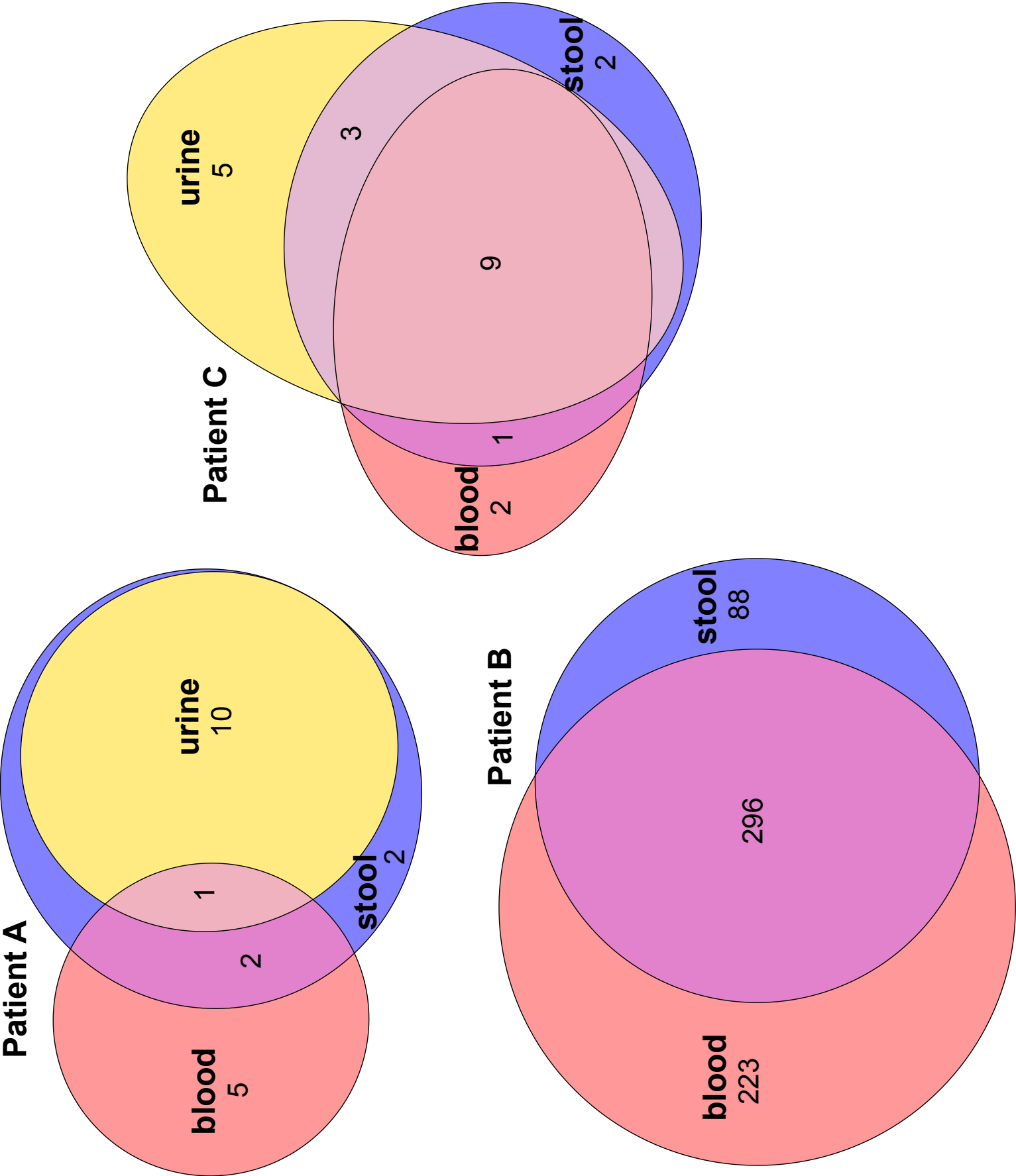
52. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Research* 44:D694–D697.
53. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic *Escherichia coli*. *Journal of Clinical Microbiology* 52:1501–1510.
54. Marin J, Clermont O, Royer G, Mercier-Darty M, Decousser JW, Tenaillon O, Denamur E, Blanquart F. 2022. The Population Genomics of Increased Virulence and Antibiotic Resistance in Human Commensal *Escherichia coli* over 30 Years in France. *Applied and Environmental Microbiology* 0:e00664-22.
55. R Development Core Team. 2010. R: A language and environment for statistical computing. (No Title).
56. Oliver A, Cantón R, Campo P, Baquero F, Blázquez J. 2000. High Frequency of Hypermutable *Pseudomonas aeruginosa* in Cystic Fibrosis Lung Infection. *Science* 288:1251–1253.
57. Tenaillon O, Toupance B, Le Nagard H, Taddei F, Godelle B. 1999. Mutators, Population Size, Adaptive Landscape and the Adaptation of Asexual Populations of Bacteria. *Genetics* 152:485–493.
58. Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. 6634. *Nature* 387:703–705.
59. Krawczyk B, Śledzińska A, Szemiako K, Samet A, Nowicki B, Kur J. 2015. 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* 34:1135–1143.

60. McNally A, Alhashash F, Collins M, Alqasim A, Paszckiewicz K, Weston V, Diggle M. 2013. Genomic analysis of extra-intestinal pathogenic *Escherichia coli* urosepsis. *Clinical Microbiology and Infection* 19:e328–e334.
61. Hobson CA, Bonacorsi S, Hocquet D, Baruchel A, Fahd M, Storme T, Tang R, Doit C, Tenaillon O, Birgy A. 2020. Impact of anticancer chemotherapy on the extension of beta-lactamase spectrum: an example with KPC-type carbapenemase activity towards ceftazidime-avibactam. *Sci Rep* 10:589.
62. Vahabi B, Drake MJ. 2015. Physiological and pathophysiological implications of micromotion activity in urinary bladder function. *Acta Physiologica* 213:360–370.









Resistance genes

