Mapping the evolutionary path towards multi-drug resistance in the pandemic *Escherichia coli* ST131 lineage

Anna K. Pöntinen[1], Nguyen Vinh Trung[2,3], Sudaraka Mallawaarachchi[1,4,5], Rebecca A. Gladstone[1], Juri Kuronen[1], Ørjan Samuelsen[6], Huynh Xuan Yen[2], Phung Le Kim Yen[2], Nguyen Phu Huong Lan[7], Nguyen Thanh Dung[7], Nguyen Van Vinh Chau[8], Julian Parkhill[9], Gerry Tonkin-Hill[4,5,10], Ngo Thi Hoa[2,11,12], Jukka Corander[1,13,14]

Affiliations

1 Department of Biostatistics, Faculty of Medicine, University of Oslo, Oslo, Norway
2 Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam
3 Faculty of Veterinary Medicine, College of Agriculture, Can Tho University, Can Tho, Vietnam
4 Peter MacCallum Cancer Centre, Melbourne, Australia
5 Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Australia
6 Norwegian Centre for Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway
7 Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam
8 Department of Health, Ho Chi Minh City, Vietnam
9 Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom
10 Department of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia
11 Nuffield Department of Medicine, Centre for Tropical Medicine, University of Oxford, Oxford, United Kingdom
12 Center for BioMedicine Research, Pham Ngoc Thach University of Medicine, Ho Chi Minh City, Vietnam
13 Parasites and Microbes, Wellcome Sanger Institute, Cambridge, United Kingdom
14 Helsinki Institute of Information Technology, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

Corresponding authors: Anna K. Pöntinen (a.k.pontinen@medisin.uio.no) and Jukka Corander (jukka.corander@medisin.uio.no)

Abstract

Escherichia coli sequence type (ST) 131 is the most widely studied genetic lineage of the species to date, originally identified in the early 2000s as an increasingly common cause of human urinary tract and bloodstream infections worldwide. This lineage is subdivided into four extant main subclades A, B, C1 and C2 that exhibit distinct features in terms of invasiveness, antibiotic resistance and transmissibility. However, the evolutionary pathway from the generally susceptible ST131-B to the drug-resistant ST131-C clades remains poorly mapped. To fill this knowledge gap, we analysed in detail human clinical isolates obtained in Vietnam, designated as belonging to the generally neglected minor clade ST131-B0. We sequenced them using both short- and long-read technology, and combined these data with a recently published high-resolution genomic collection to provide further insight into the evolutionary process and its timeline. Extensive genomic analyses established ST131-B0 as an intermediary progenitor in the evolutionary path leading from ST131-B to the ST131-C clades, associated with multiple type I pili switches, as well as the loss and gain of specific chromosomal genes representing diverse core functions such as metabolism, transcription, DNA binding and type II toxin-antitoxin systems. Furthermore, all Vietnamese isolates of ST131-B0 unprecedentedly harboured $bla_{\text{CTX-M}}$ genes encoding extended-spectrum β-lactamases, a trait dominant in ST131-C clades and not previously observed in ST131-B0. Our study supports the hypothesis that the ST131-C clades have gradually evolved from ST131-B by reducing the host range with better adaptation to colonising humans under selective conditions.

Introduction

Escherichia coli is a leading causative agent of human bloodstream infections (BSI) and urinary tract infections (UTI) globally with high associated morbidity and mortality (Foxman 2014; Kern and Rieg 2020; Bonten et al. 2021). The clinical concern is exacerbated by the high prevalence of multi-drug resistant (MDR) *E. coli*, reducing therapeutic options and, thereby, prolonging hospitalisation and increasing mortality. In particular, the increase in the prevalence of extended-spectrum β-lactamase (ESBL)-producing *E. coli* remains a global concern (Naylor et al. 2019; Shamsrizi et al. 2020; WHO 2024). Among extraintestinal pathogenic *E. coli* (ExPEC), sequence type (ST) 131 has generally held a predominant position worldwide among ESBL-producers since the early 2000s (Nicolas-Chanoine et al. 2007; Nicolas-Chanoine et al. 2014; Kallonen et al. 2017; Gladstone et al. 2021). Based on representative genomic surveillance, the prevalence of ST131 rapidly increased, such that it became one of the leading causes of invasive ExPEC infections, followed by a new equilibrium within the ExPEC population (Kallonen et al. 2017; Gladstone et al. 2021; Pöntinen et al. 2024).

Based on representative longitudinal sampling and modelling, antimicrobial resistance (AMR) appears not to be the sole driver of the epidemiological success of ST131 (Gladstone et al. 2021; Pöntinen et al. 2024). It has been suggested that a combination of negative frequency-dependent selection (NFDS) within a wider commensal niche and particular selection pressures deriving from non-penicillin β-lactam usage have resulted in differing equilibrium prevalences of ST131 in separate geographical regions (Kallonen et al. 2017; McNally et al. 2019; Gladstone et

al. 2021; Pöntinen et al. 2024). Furthermore, the sublineages of ST131, here referred to as ST131 clades A, B, C1, and C2, have been characterised by notably distinct associations with AMR traits. Of note, ST131-C1 and -C2 have been described as frequently carrying CTX-M-type ESBLs (Price et al. 2013; Petty et al. 2014; Brodrick et al. 2017; Gladstone et al. 2021). Typically, ST131-C2 has been associated with CTX-M-15, while ST131-C1 more commonly carries CTX-M-14 (Nicolas-Chanoine et al. 2014; Bevan et al. 2017; Komori et al. 2024). While found integrated into the chromosome (Coque et al. 2008; Hirai et al. 2013; Hamamoto et al. 2016), ESBL-encoding genes are also frequently carried on plasmids in *E. coli* (Cave et al. 2023; Arredondo-Alonso et al. 2025). Therefore, concern caused by increasing AMR levels is further exacerbated by the global dissemination of readily mobile plasmid-borne AMR determinants. Furthermore, a study by Molari et al. (2025) indicated a broader role of the accessory genes in *E. coli* evolution by extensive accessory genome variation within *E. coli* ST131, as well as structural modifications at rates comparable to individual nucleotide substitutions (Molari et al. 2025).

Human colonisation represents the main reservoir for ExPECs, and ST131 is, accordingly, capable of asymptomatically residing in the human intestines (Sarkar et al. 2018; Johnson et al. 2022). However, ST131-B, in particular, has been shown to represent a zoonotic pathogen (Platell et al. 2011; Liu et al. 2018; Reid et al. 2019). Animal ST131-B isolates are particularly associated with poultry (Saidenberg et al. 2020), and the contaminated products thereof are proposed as potential foodborne vehicles for human infections caused by *E. coli* ST131 (Manges 2016; Liu et al. 2018; Roer et al. 2019). Although generally reported as more susceptible to antimicrobials compared to the other ST131 clades, ST131-B (ST131-*H*22) isolates with both human BSI and poultry origin have been reported to sporadically carry plasmid-mediated colistin resistance (*mcr*) traits (Roer et al. 2019; Saidenberg et al. 2020). However, in a study of uropathogenic *E. coli* infections, the clinical presentation within the human urinary tract did not differ significantly between ST131-B and ST131-C, as the clades were equally likely manifested as asymptomatic bacteriuria (Liu et al. 2018). Furthermore, an investigation of the gut colonisation in neonates and the relative invasiveness of *E. coli* lineages showed that ST131-B was more associated with asymptomatic human colonisation compared with ST131-C clades (Mäklin et al. 2022; Pöntinen et al. 2024). Combined, these findings suggest that ST131-B has adapted to two different primary niches, one representing asymptomatic circulation among humans and the other colonisation of poultry, both of which then contribute towards the human infection burden. The genetic determinants behind these adaptations are potentially partly overlapping, but not currently fully resolved, see the discussion in Arredondo-Alonso et al. (Arredondo-Alonso et al. 2025). Finally, it is worth noticing that the estimated transmissibility for ST131-C1 and ST131-C2 is significantly lower compared with ST131-A (sufficient data not available for ST131-B), suggesting that ST131-C clades may be more adept at transmission within the healthcare settings aided by antibiotic selection pressure than in community settings (Ojala et al. 2025).

The current evidence suggests that a branch of ST131-B has over time evolved to the minor subclade ST131-C0 and further to ST131-C1 and ST131-C2 clades by gradual *fimH* transition and by acquiring fluoroquinolone resistance mutations and plasmid-borne $bla_{\text{CTX-M}}$ genes in the

ST131-C clades (Petty et al. 2014; Ben Zakour et al. 2016; Matsumura et al. 2016; Stoesser et al. 2016; Pitout and Finn 2020). However, due to the limited number of intermediary isolates, the full evolutionary path from ST131-B ancestor to ST131-C clades has remained elusive. Here, we introduce a collection of clinical ST131-B0 isolates, sequenced by using both short- and long-read techniques, and combine these data with ST131 genomes from a recent high-resolution sequenced *E. coli* BSI cohort (Gladstone et al. 2021; Arredondo-Alonso et al. 2025) to fill this prevailing knowledge gap. We show that ST131-B0 unprecedentedly carries CTX-M-type ESBLs both on plasmids and chromosomes and, combined with high-resolution dating analyses in context, we map in detail how ST131-B0 arose from ST131-B, and further evolved into the present predominantly MDR pathogen lineage ST131-C.

Results

Isolate collection and population structure

In total, 212 *E. coli* ST131-B and -C clinical isolates were analysed within the study (Supp. Table 1, Microreact project [https://microreact.org/project/ecoli-st131-evolution]), of which 9 were from the Oxford University Clinical Research Unit (OUCRU), Vietnam (sequenced in this study) and 203 from the Norwegian surveillance programme on resistant microbes (NORM), Norway (Gladstone et al. 2021). Of note, a total of 206 hybrid assemblies was available for the combined collection. All OUCRU isolates belonged to the clade ST131-B0, while the NORM isolates were distributed between ST131-B (n=74), ST131-B0 (n=4), ST131-C0 (n=4), ST131-C1 (n=70), and ST131-C2 (n=51). Aligning the *fimH*-encoded type I pili with the dated phylogeny identified an apparent switch in the predominant *fimH* type per clade, from predominantly *fimH22* in ST131-B to *fimH30* in the ST131-C1 and ST131-C2, while *fimH54* was the most common type in ST131-B0 (Fig. 1). When screening for flagellar H and polysaccharide O antigens, all isolates were of H4 and, apart from one O2 in ST131-B, of O25 serotype (Fig. 1). The most abundant capsule type was K5, covering 52.8% (112/212) of the isolates across all ST131 clades. It has previously been found to be the ancestral capsule type in ST131-B and -C clades (Gladstone et al. 2025). Of note, 8 out of 9 of the OUCRU ST131-B0 isolates carried the K5 capsule, supporting their position as an intermediary progenitor to ST131-C where substantial capsule diversity has emerged later (Gladstone et al. 2025). The second most common capsule type overall was K2, covering 10.8% (23/212) of the total isolates and disseminated in all of the ST131 clades apart from ST131-C1. K100 was found in 4.7% (10/212) of the total isolates and exclusively in ST131-C2, and K16 in 3.8% (8/212), only in ST131-B. K1 type was found in 4 ST131-B and K20 type in one ST131-C2 isolate (Fig. 1). 25.5% (54/212) isolates remained with untypeable or unknown capsule loci.

Evolutionary timeline from ST131-B to ST131-C

Dating analyses further support ST131-B0 as an intermediary progenitor of the ST131-C clades in the evolutionary path deriving from ST131-B (Fig. 1). The most recent common ancestor (MRCA) for the combined OUCRU and NORM ST131-B and ST131-C clades was estimated to

date back to 1943 (95% highest posterior density [HPD] interval 1924-1957). ST131-B0 and ST131-C clades were estimated to diverge from the ancestral ST131-B clade in 1952 (95% HPD 1940-1964). Shortly after, estimated in 1960 (95% HPD 1958-1970), an older branch of ST131-B0 clade diverged from the rest of the clade, creating a deep split between the two branches. Finally, ST131-C clades were estimated to diverge from the more recent branch of ST131-B0 approximately two decades later in 1979 (95% HPD 1971-1985). Splits in the ST131-C clades were seen in more rapid succession, with ST131-C0 estimated to diverge in 1984 (95% HPD 1977-1989), and ST131-C1 and ST131-C2 clades in 1990 (95% HPD 1984-1993) (Fig. 1).

CTX-M ESBL association with ST131 clades

In total, 6 different CTX-M type ESBLs were found in the combined collection: CTX-M-14, CTX-M-15, CTX-M-19, CTX-M-24, CTX-M-27 and CTX-M-127 (Fig. 2). The most common type was CTX-M-15, covering 20.9% (43/206) of the total isolates (hybrid assemblies), and was found in the OUCRU ST131-B0 and NORM ST131-C1 and ST131-C2. It was found equally often carried on plasmids (21/43) and chromosomes (22/43), and 6 of these isolates also carried more than one copy of the gene on their chromosomes. However, those located in chromosomes were most commonly (19/22) found in ST131-C2 (also 1 in ST131-C1 and 3 in OUCRU ST131-B0). CTX-M-27 was found in 9.7% (20/206) of the isolates, in ST131-C1 (15/20) and the OUCRU ST131-B0 (5/20). It was mostly plasmid-borne but 3 out of 20 isolates carried it on chromosomes. CTX-M-14 was less common and found only in 5 ST131-C1 isolates, carried on both chromosomes and plasmids. CTX-M-19 and -127 were found only in one ST131-C1 plasmid and ST131-C2 chromosome, respectively, while two copies of CTX-M-24 were found in a single ST131-C1 chromosome.

Comparative pangenome analysis and genome-wide co-selection

Of the 10,781 total genes within the combined ST131 collections (206 hybrid assemblies), 3964 (36.8%) were estimated as core genes (present in ≥99% of isolates), amounting to 80.7% (3964/4909) of the average genome size. Comparison of gene gain and loss rates between clades revealed the largest estimate of gene exchange relative to core branch length in ST131-C2 (Supp. Fig. 1A), statistically significantly different from both ST131-B0 ($p=0.0455$) and ST131-B ($p=0.0049$). Gene exchange rates of ST131-C1 were also significantly higher than those of ST131-B ($p=0.0392$). These differences were also seen in the cumulative gene gain and loss events (Supp. Fig. 1B).

Comparison of gene presence and absence between ST131-B0 and ST131-C clades (C1, C2) revealed specific genes and potential homologues gained and lost in the transition between these clades (Supp. Table 2). Of note, pangenomes were estimated with a minimum pairwise sequence-identity threshold of 70% between genes, and additional homologous genes might exist at lower sequence identities. In total, six unique genes were only present in ST131-B0. These were all predicted to be located on the chromosome and included genes functioning in arginine (*arcA*), ornithine (*argF*) and carbamate (*arcC*) metabolism and encoding a putative

membrane transporter (*yfcC*). The eight unique genes gained by the ST131-C clades (absent in ST131-B0), included genes encoding oxidoreductase and transcription repressor (*bdcAR*), type II toxin-antitoxin system (*chpBS*), DNA-binding proteins, endoribonuclease (*yjgH*) and membrane protein (*yjgN*). These were also predominantly chromosomal as at least 95.9% (117/122) of the isolates harboured them on chromosome-predicted contigs, supporting the importance of both the lost and gained genes as essential to the core functionality of their respective subclades.

Investigation of potential co-selection and pangenome-spanning epistasis of genomic variation by LDWeaver (Mallawaarachchi et al. 2024) revealed several short- and long-range links between genomic variation, illustrated in network plots (Supp. Fig. 2). The top-ranking region among short-range linkage of variation with most significant links was found to be a site in *sitD*, encoding an iron/manganese ABC transporter permease subunit (Supp. Fig. 2A). The network analysis indicated potential co-selection of *sitD* amongst others with *ydaT*, encoding a bacterial toxin, and *bet,* encoding a phage recombination protein. The highest number of links was found between *sitD* and replication protein P (DOHBNP_06365). The top-ranking long-range linkage of variation revealed a complex network of multiple linked loci associated especially with metabolic processes (Supp. Fig. 2B). Amongst them were sites in genes such as *speC,* encoding an ornithine decarboxylase, *dgoT*, encoding a D-galactonate transporter and *ghrB*, encoding a glyoxylate/hydroxypyruvate reductase.

Outcome of the coselection and epistasis analysis was further visualised and investigated by using GWES-Explorer (https://github.com/jurikuronen/GWES-Explorer) (Supp. Fig. 3). Within the short-range links, ST131-B presented apparent loss events of many of the phage-related sites, including e.g. prophage host-nuclease inhibitor and phage tail proteins, which were maintained in ST131-B0 and ST131-C clades. Most ST131-B still harbouring the sites also showed an allele change compared to the loci in ST131-B0 and ST131-C. Oppositely, a distinct subclade in ST131-B showed a shared allele in phage recombination protein Bet to those of the other ST131 clades. As for the long-range links, most sites had shared alleles between ST131-B0 and ST131-C clades, differentiating them from ST131-B. An exception was the network of variants between leucyl aminopeptidase PepA and iron-sulfur stress protein YtfE and YtfB, where shared alleles differentiated most ST131-B0 and ST131-B isolates from the ST131-C clades.

Plasmidome analyses

Altogether, 32 contigs of the OUCRU ST131-B0 hybrid assemblies were predicted as plasmid-derived. In addition, 578 plasmid-derived contigs were retrieved from the NORM collection (Arredondo-Alonso et al. 2025). These contigs were assigned to 15 different plasmid types (pT) (Fig. 3, Supp. Table 3), according to the existing mge-cluster *E. coli* plasmid typing scheme (Arredondo-Alonso et al. 2023; Arredondo-Alonso et al. 2025). While most pTs were disseminated across the phylogeny, some plasmid-clade association was seen in particular pTs (Supp. Table 3). Most of the plasmid contigs of pT2 (84.6%, 104/123), pT4 (77.6%, 59/76) and pT8 (95.0%, 76/80) were found in ST131-C1 and -C2, while pT5 included exclusively ST131-B

243 plasmids. Also pT11 were harboured almost solely by ST131-B, with only 1 of 31 plasmids
244 being in ST131-C1. pT10 corresponded to a singleton type with one ST131-B plasmid.
245 Altogether seven of the pTs were harboured by OUCRU ST131-B0 isolates (pT0, pT2, pT3,
246 pT4, pT7, pT8 and pT15). The remaining five types (pT1, pT12, pT13, pT14, pT16) were a
247 mixture of ST131-B and ST131-C plasmids from the NORM collection (Fig. 3). In total, 110
248 plasmid-derived contigs remained unassigned to any plasmid type (pT-1) (Supp. Table 3).
249
250 Of all the pTs, pT7 was found in most OUCRU ST131-B0 plasmid-predicted contigs in total
251 (n=6), while pT2 was found in most OUCRU isolates (n=5) (Figs 1 and 3). Plasmids within the
252 pT7 shared an IncFII replicon (Fig. 4A) and were average in size with most variation of all the
253 pTs (mean 73.9 kbp, SD 35.1 kbp). Only two pT7 plasmids harboured CTX-M type β-
254 lactamases: CTX-M-15 in an OUCRU ST131-B0 and CTX-M-19 in a NORM ST131-C1 plasmid
255 (Fig. 4A), whereas six pT7 plasmids harboured $bla_{\text{TEM-1}}$ β-lactamase genes. Of these, the
256 OUCRU ST131-B0 plasmid carried both CTX-M-15 and TEM-1, notably colocalised with type II
257 toxin-antitoxin systems. The synteny analyses within pT7 further depicted the loss of multiple
258 colicin loci in the path from ST131-B to the ST131-C clades, which had already disappeared in
259 the ST131-B0 (Fig. 4A). Of note, none of the ST131-B0 plasmids harboured microcin (colicin V)
260 that was found widely present in ST131-B plasmids, particularly in pT5 (Supp. Table 3). The pTs
261 corresponding to the largest subsets of plasmids in total were pT2 (n=123) and pT8 (n=80), the
262 latter of which corresponds to the previously described plasmid type with high prevalence in
263 ST131-C1 and frequently harbouring β-lactamase genes (Arredondo-Alonso et al. 2025). The
264 pT8 plasmids also shared an IncFII replicon type (Fig. 4B), were amongst the largest in length
265 (mean 115.9 kbp, SD 25.4 kbp) and 37.5% (30/80) harboured CTX-M-14, -15 or -27. pT2
266 plasmids, on the contrary, were amongst the smallest ones (mean 4.9 kbp, SD 0.8 kbp) and
267 harboured no known AMR genes.
268
269 To investigate potential co-selection of genomic variation within pT7 and pT8 plasmids that were
270 evaluated to share evolutionary origin, the plasmid contigs were collectively analysed using the
271 alignment-free PAN-GWES pipeline (Kuronen et al. 2024). Annotated identified links between
272 variation included essential mobile genetic element functions under potential coselection, such
273 as type IV conjugative transfer system coupling protein-encoding *traD* and type IV secretion
274 system protein-encoding *traC* (Supp. Fig. 4). In addition, *traC* was found to pair with a partial
275 overlap on a hypothetical protein-encoding *yhfA*.

276 Discussion

277 The rapid rise of the *E. coli* ST131 lineage in clinical settings from the early 2000s onwards has
278 inspired a considerable body of evolutionary research. The current consensus asserts that the
279 success of the extantly circulating major ST131 subclades cannot be pinned down to a single
280 genetic factor. Substantial evidence has accumulated to support the roles of metabolic and
281 colonisation competition factors, as well as virulence and AMR determinants (McNally et al.
282 2019; Cummins et al. 2022; Arredondo-Alonso et al. 2025). Systematic population-based
283 comparison of ST131 in asymptomatic colonisation versus bloodstream infections has further
284 revealed that the ST131-A and ST131-B clades are relatively benign compared with ST131-C1

285 and ST131-C2 (Mäklin et al. 2022; Pöntinen et al. 2024), which are over-represented among
286 infection cases. Further, recent study quantified that this is not primarily driven by any specific
287 virulent capsule types, but corresponds instead to a lineage effect stemming from a larger set of
288 genetic factors (Gladstone et al. 2025).
289
290 ST131-B is largely characterised as susceptible to antimicrobials compared to the ST131-C
291 clades (Gladstone et al. 2021; Pöntinen et al. 2024; Arredondo-Alonso et al. 2025), although
292 has in some cases been found to carry AMR traits (Reid et al. 2019; Roer et al. 2019;
293 Saidenberg et al. 2020). ST131-B has been shown to be zoonotic, associated in particular with
294 poultry meat (Liu et al. 2018), while in addition adapted to asymptomatic circulation among
295 humans (Mäklin et al. 2022). The loss of microcin (colicin V) from the plasmids we observed
296 here during evolutionary transition from ST131-B to ST131-B0 is well aligned with a restriction
297 of the primary colonising host range to humans, since this microcin has been shown to be an
298 important fitness determinant against Salmonella (Baquero et al. 2019). Expectedly, ST131-B
299 would frequently compete with Salmonella when colonizing the gut of poultry. ST131-B0 has in
300 turn been rarely observed in ST131 studies to date, but similarly to ST131-B, identified isolates
301 are mainly from high-income countries and predominantly susceptible to antimicrobials. Overall,
302 overrepresentation in high-income countries would be expected with more genomes sequenced.
303 However, sequenced ST131-B0 isolates from China and Thailand were also found to lack
304 $bla_{CTX\text{-}M}$ genes (McNally et al. 2016; Stoesser et al. 2016; Gladstone et al. 2021).
305
306 The current consensus is that the ST131-C clades evolved from ST131-B through a *fimH* switch
307 and by acquisition of mutational and plasmid-borne resistance traits (Ben Zakour et al. 2016;
308 Matsumura et al. 2016; Pitout and Finn 2020). However, the lack of intermediary ST131-B0
309 isolates with complete genomic data, and particularly those CTX-M-positive, has resulted in a
310 notable knowledge gap considering the more detailed evolutionary pathway of ST131 into MDR.
311 In this study, combining high-quality sequence data and hybrid assemblies of clinical ST131-B0
312 isolates with a high-resolution *E. coli* ST131 cohort (Gladstone et al. 2021; Arredondo-Alonso et
313 al. 2025), we were able to more thoroughly map the evolutionary path from ST131-B to ST131-
314 C. Dating analyses of the combined cohorts placed ST131-B0 as an intermediary progenitor of
315 ST131-C clades, as has been previously suggested (Ben Zakour et al. 2016; Pitout and
316 DeVinney 2017). Moreover, our analyses presented a potential intermediary switch to and from
317 the most common type *fimH54* in ST131-B0. For comparison, no ST131 isolates in the NORM
318 cohort carried the *fimH54* allele, while in the entire Norwegian BSI cohort (n=3254) by
319 (Gladstone et al. 2021), *fimH54* was detected in 141 isolates, majority of them in ST393
320 (26.2%), ST10 (21.3%) and ST95 (19.9%). Furthermore, most of the ST131-B0 isolates carried
321 the ancestral K5 capsule type, preceding the higher capsule diversity seen in the ST131-C
322 clades (Gladstone et al. 2025), and thereby further supporting their position as an intermediary
323 ST131 clade.
324
325 In the combined collection of ST131 isolates from Vietnam and Norway, we found CTX-M-15
326 and CTX-M-27 to be the most common types of the CTX-M β-lactamases. We detected CTX-M-
327 27 predominantly on plasmids while CTX-M-15 resided equally on plasmids and chromosomes,
328 even with multiple copies on the same chromosome. Indeed, CTX-M-15 has been found in

329 varying genetic contexts both on plasmids and chromosomes, particularly within *E. coli* ST131
330 (Lipworth et al. 2024). Of note, resistance levels conferred by β-lactamases are shown to be
331 gene dosage-dependent (San Millan et al. 2016; Rodríguez-Beltrán et al. 2021). Thereby,
332 carrying these genes on plasmids, which despite being relatively large in size can be harboured
333 in multiple copies in a subset of cells in the colony, could provide a higher level of resistance to
334 the bacteria regarded as a within-host community (Rodríguez-Beltrán et al. 2021). In the same
335 line, upon chromosomal integration, like in the case of CTX-M-15 here, multiple copies may be
336 beneficial for the pathogen due to gene dosage effects. Current evidence also shows AMR-
337 carrying plasmids to be more plastic than others (Coluzzi and Rocha 2025) and recombination
338 in general to be more frequent in genes carried on plasmids than those on chromosomes
339 (Rodríguez-Beltrán et al. 2015), further suggesting that plasmid-borne genes evolve at a faster
340 rate than chromosomal ones. That could also support the mutational evolution of CTX-M-27
341 from the CTX-M-14 background, possibly resulting in its position as one of the dominant alleles
342 in our data.
343
344 Notably, we found all of the Vietnamese ST131-B0 isolates harboured CTX-M-type ESBLs, a
345 trait previously thought of as distinctive for ST131-C clades. Indeed, none of the Norwegian
346 ST131-B0 nor ST131-B clade had any $bla_{\text{CTX-M}}$ genes (Fig. 2). The CTX-M types in the
347 Vietnamese ST131-B0 isolates were either CTX-M-15 or CTX-M-27 which have been most
348 commonly seen in ST131-C2 and ST131-C1, respectively (Johnson et al. 2016; Matsumura et
349 al. 2016; Gladstone et al. 2021). Furthermore, bar one chromosomal CTX-M-15 with 98.63%
350 sequence identity, both types were each found with fully identical sequences on chromosomes
351 as well as on plasmids in the Vietnamese ST131-B0 (Fig. 2). This indicates that, in addition to
352 being readily mobile in ST131-B0, CTX-M-type β-lactamases may have already been
353 established in the ST131 lineage prior to the branching event ultimately leading to the ST131-C
354 clades, unless these were acquired in a separate event later.
355
356 We also found $bla_{\text{CTX-M}}$ genes co-localising on plasmids with toxin-antitoxin systems (Fig. 4),
357 possibly aiding in the distribution of AMR traits in the ST131-B0. Toxin-antitoxin systems are
358 important mechanisms in plasmid stability and maintenance, and co-localisation of these
359 systems together with AMR traits on conjugative plasmids are of concern due to the potential for
360 dissemination and fixation of the resistance determinants in the bacterial populations. Of the
361 type II TA systems, the ParDE super family has been identified as distributed in IncF-type
362 plasmids in *E. coli*, also contributing to the stress tolerance of the pathogen (Kamruzzaman and
363 Iredell 2019). This system was also present in the CTX-M-15-carrying IncF-type plasmid in a
364 Vietnamese ST131-B0 isolate. While ParDE was absent in the Vietnamese CTX-M-27 plasmid,
365 another two TA systems common in Enterobacteriaceae, CcdAB and PemIK (Wu et al. 2020),
366 were found on ST131-B0 plasmids harbouring both CTX-M-15 and CTX-M-27.
367
368 Regarding unique genes found exclusively in ST131-B0 while absent in ST131-C clades,
369 prevalence of the arginine deiminase (ADI) pathway (*arc* operon) has been shown to be higher
370 in ESBL-producing *E. coli*, in particular those carrying CTX-M-type β-lactamases (Billard-
371 Pomares et al. 2019). However, the study found it more common in IncF plasmids than
372 chromosomes. As per *in vivo* competition experiments in a mouse model, it was suggested to

be of advantage for the strain in UTI, while being costly in intestinal colonization. Our findings of gene gain and loss in the stepwise evolution towards ST131-C clades thus highlight further targets for experimental work to elucidate their underlying exact functional drivers. Similarly, recent evidence supports that the ST131-C clade genomes continue to undergo rapid diversification (Gladstone et al. 2025; Molari et al. 2025), and further research is needed to elucidate how much of this represents drift versus selection.

Materials and Methods

Genomic datasets

In total, 9 *E. coli* human clinical isolates from blood stream infections (BSI; n = 7) and urinary tract infections (UTI; n = 2) from Oxford University Clinical Research Unit (OUCRU), Ho Chi Minh City, Vietnam, collected in 2017 and 2018, were sequenced within this study by short and long-read techniques. In order to use data for comparison with complete genomes and both short- and long-read sequences available, a collection of BSI isolates from the Norwegian surveillance programme on resistant microbes (NORM) study, collected 2002–17 in Norway (Gladstone et al. 2021), was selected for comparative genomics analyses. Whole-genome sequences, short-read assemblies, *fimH* types and clade designations of 203 *E. coli* clonal complex (CC) 131 clades, here referred to as ST131-B, -B0, -C0, -C1 and -C2, were included in the analyses. For comparative pangenome and plasmidome analyses, available hybrid assemblies (n=197) and plasmid-predicted contigs (n=578) of the NORM isolates were retrieved from (Arredondo-Alonso et al. 2025).

For short-read sequencing of the OUCRU collection, whole-genome DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, Madison, USA) according to the manufacturer's instructions and sequencing was performed at the Wellcome Sanger Institute, UK, using Illumina HiSeq X Ten paired end platform. Mixed strains and species contamination were excluded by using Kraken v.0.10.6 (Wood and Salzberg 2014). Sequencing quality was controlled by 30x minimum depth of coverage, and maximum of 100 contigs, 0.20 highest minor allele frequency in multi-locus sequence type (MLST) genes and 500 heterozygous SNP sites. Short-read sequence data was assembled by using a published pipeline with Velvet optimisation (Page et al. 2016) to coincide with the NORM dataset.

For long-read sequencing on the Oxford Nanopore Technologies (ONT) platform, whole-genome DNA of the OUCRU isolates was extracted using an in house phenol:chloroform extraction procedure. Concentration and DNA integrity were measured using NanoDropOne spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and the Qubit dsDNA HS assay kit (Thermo Fisher Scientific) on a CLARIOstar microplate reader (BMG Labtech, Ortenberg, Germany). The ONT libraries were prepared using SQK-NBD112-24 barcoding kit for 24-barcoding runs, on FLO-MIN106 flow cells. Sequencing was run for 72 hours on GridION (Oxford Nanopore Technologies, Oxford, UK) using MinKNOW Core v.5.8.6. Basecalling and demultiplexing were performed using Guppy v.6.4.6 with high-accuracy basecalling. ONT long

413 reads were combined with the short-read data using Hybracter v.0.10.0 with default parameters
414 (read length >1 kbp and read quality score of >9 for long reads) and recommended estimated
415 genome size of 4 Mbp for *E. coli* (Bouras et al. 2024), resulting in 6 complete and 3 incomplete
416 hybrid assemblies.

417 Sequence typing

418 ST131 sequence type of the OUCRU isolates was confirmed by using SRST2 v.0.2.0
419 (https://github.com/katholt/srst2) by comparing short-read data to the *Escherichia* species MLST
420 database (https://pubmlst.org/organisms/escherichia-spp, retrieved 9th May 2025). *fimH* typing
421 was performed on short-read assemblies by using Abricate v.1.0.1
422 (https://github.com/tseemann/abricate) on the fimtyper database
423 (https://bitbucket.org/genomicepidemiology/fimtyper_db/downloads/, retrieved 9th May 2025),
424 and ST131 clade B0 was further assigned by clade-specific SNPs (Price et al. 2013; Ben
425 Zakour et al. 2016) and phylogenetic comparison to the NORM ST131.
426 *E. coli* O-H serotypes of both the OUCRU and NORM isolates were assigned on short-read
427 assemblies (n=212) by using Abricate v.1.0.1 (https://github.com/tseemann/abricate) on the
428 EcOH database (Ingle et al. 2016). Capsule typing of both collections was performed on the
429 short-read assemblies (n=212) by using Kaptive v.3.1.0 (Stanton et al. 2025) against the *E. coli*
430 group 2 and 3 capsule database v.2.0.0 (Gladstone et al. 2025). Only typeable loci without any
431 identified problems were accepted.

432 *Bla*$_{\text{CTX-M}}$ screening

433 To investigate presence of and differentiating chromosomal *vs* plasmid loci and number of
434 chromosomal copies of the *bla*$_{\text{CTX-M}}$-type ESBL genes in the OUCRU and NORM isolates, the
435 resistance genes were screened from the hybrid assemblies (n=206) using AMRFinderPlus
436 v.4.0.22 (Feldgarden et al. 2021) with database v.2025-03-25.1 and *Escherichia* taxonomy
437 group (--organism Escherichia). Minimum identity was set to 75% and minimum coverage to
438 80%.

439 Dating analyses

440 For dating analyses of the combined OUCRU and NORM isolates (n=212), short reads were
441 mapped against the *E. coli* ST131 reference genome EC958 (HG941718.1) using Snippy
442 v.4.6.0 (https://github.com/tseemann/snippy) and recombination was removed by using Gubbins
443 v.3.4.0 (Croucher et al. 2015). Dated tree was inferred with BactDating v.1.1.1 (Didelot et al.
444 2018) in 3 separate runs, using the additive relaxed clock (ARC) model, 100,100,100 MCMC
445 chain length and 100,000 thin, and verified for estimated sample size of over 200. The model fit
446 was compared to that of random dates for significance of the temporal signal.

Comparative genomics

For comparison of the pangenomes as well as the gene gain and loss rates between the ST131 clades in the combined collection, the hybrid assemblies (n=206) were annotated by using Bakta v.1.10.3 with the database v.5.0 and *Escherichia* genus (Schwengers et al. 2021). Pangenomes were estimated by using Panaroo v.1.5.2 (Tonkin-Hill et al. 2020) with sensitive mode, by removing invalid genes and merging paralogs, and with a minimum pairwise sequence-identity threshold of 70%. Gene gain and loss rates between ST131-B0, ST131-B, ST131-C1 and ST131-C2 were estimated and compared using Panstripe v.0.3.3 assuming a Gaussian distribution (Tonkin-Hill et al. 2023). The combined dated phylogeny was used to partition ST131-C1 and ST131-C2 for the analysis, while a separately dated phylogeny was used in partitioning ST131-B and ST131-B0.

Potential coselection and pangenome-spanning epistasis of genomic variation was investigated from multiple-sequence alignment against the *E. coli* ST131 reference genome EC958 (HG941718.1) by using LDWeaver v.1.5.2 (Mallawaarachchi et al. 2024) and further visually analysed by using GWES-Explorer (https://github.com/jurikuronen/GWES-Explorer). EC958 was annotated using Bakta v.1.9.3 with the database v.5.1 and *Escherichia* genus (Schwengers et al. 2021) and used in mapping and annotation of the top hits.

Plasmidome analyses

For comparative plasmidome analyses of the combined OUCRU and NORM datasets, plasmid prediction of the OUCRU collection was performed to coincide with the NORM analyses (Arredondo-Alonso et al. 2025): Hybrid assembly contigs smaller than 1 kbp and larger than 300 kbp were filtered out. Plasmid contigs were then predicted by using mlplasmids v.2.1.0 with *Escherichia coli* species model and minimum plasmid probability threshold of 0.3 (Arredondo-Alonso et al. 2018) and by using geNomad v.1.7.0 end-to-end command with minimum plasmid score of 0.7 (Camargo et al. 2024). Remaining contigs were classified as plasmid-predicted if they contained a replication type as defined by mob-typer v.3.1.4 (Robertson and Nash 2018; Robertson et al. 2020). Predicted plasmid contigs were then delineated into plasmid types (pT) by using mge-cluster v.1.1.0 (Arredondo-Alonso et al. 2023) against the existing *E. coli* scheme with a perplexity value of 62 (Arredondo-Alonso et al. 2025).

Representative plasmids of each available ST131 clade, opting for ESBL β-lactamases when present, were selected for within-pT comparison in pT7 and pT8. The starting position of the sequences was rotated to IncFII by using Circlator v.1.5.5 (Hunt et al. 2015) with the fixstart command. The plasmids were then screened for AMR, virulence and stress resistance genes using AMRFinderPlus v.4.0.22 (Feldgarden et al. 2021) with database v.2025-03-25.1 and *Escherichia* taxonomy group (--organism Escherichia), with minimum identity of 75% and minimum coverage of 80%. The plasmids were further annotated using Bakta v.1.7.0 with the database v.5.0 and *Escherichia* genus (Schwengers et al. 2021) and gene synteny analysis performed using Clinker v.0.0.31 (Gilchrist and Chooi 2021) with minimum alignment sequence identity of 80%.

Potential coselection and epistasis of genomic variation within pT7 and pT8 plasmids was investigated directly from predicted plasmid contigs (n=103) by using phenotype- and alignment-free PAN-GWES pipeline (Kuronen et al. 2024), implementing cuttlefish v.2.2.0 (Khan and Patro 2021; Khan et al. 2022) and SpydrPick v.1.2.0 (Puranen et al. 2018; Pensar et al. 2019). K-mer length for PAN-GWES was set to 81 bp that was estimated to capture fine-scale mutations, while allowing for enough sequence diversity and capturing differences at gene level. Unitig pairs with an average distance exceeding the standard deviation of their distances and pairs present in less than 5% of the plasmid contigs were filtered out. Selected outlier top hits were annotated using annotate_hits_pyseer of the pyseer tool v.1.3.12 (Lees et al. 2018) against selected reference plasmid (28328_1#6_plasmid00001; 106718 bp) that was confirmed to cover for all of the hits.

All visualisations in the study were performed in R v.4.5.0 or higher (R Core Team 2020), Clinker v.0.0.31 (Gilchrist and Chooi 2021) and Microreact (Argimón et al. 2016).

Data availability

Short-read Illumina and long-read ONT sequence reads generated within this study are publicly available at the European Nucleotide Archive (ENA) under the study ID PRJEB29743 and PRJNA1370503, respectively. The Norwegian short-read data is publicly available at the ENA under the study ID PRJEB32059 and long-read data under the study IDs PRJEB45354 and PRJEB57633. Descriptive data on the isolates together with a dated phylogeny is publicly available as a Microreact project [https://microreact.org/project/ecoli-st131-evolution].

Acknowledgements

We acknowledge the support from the Wellcome Sanger Institute sequencing facility and the Genomics Support Center Tromsø.

Funding

This work was supported by Trond Mohn Foundation (BATTALION grant to AKP, RAG, ØS and JC) and the Wellcome Trust (grant 225167/A/22/Z).

Author contributions

Conceptualisation - AKP, NTD, NVVC, NTH, JC
Funding acquisition and resources - NVVC, JP, NTH, JC,
Methodology, formal analysis - AKP, NVT, SM, JK, HXY, PLKY, NPHL, JC
Validation - AKP, SM, RAG, ØS, GT-H, JC
Visualisation - AKP, SM
Writing, original draft - AKP with contributions from JC
Writing, review and editing - all authors

Main figures

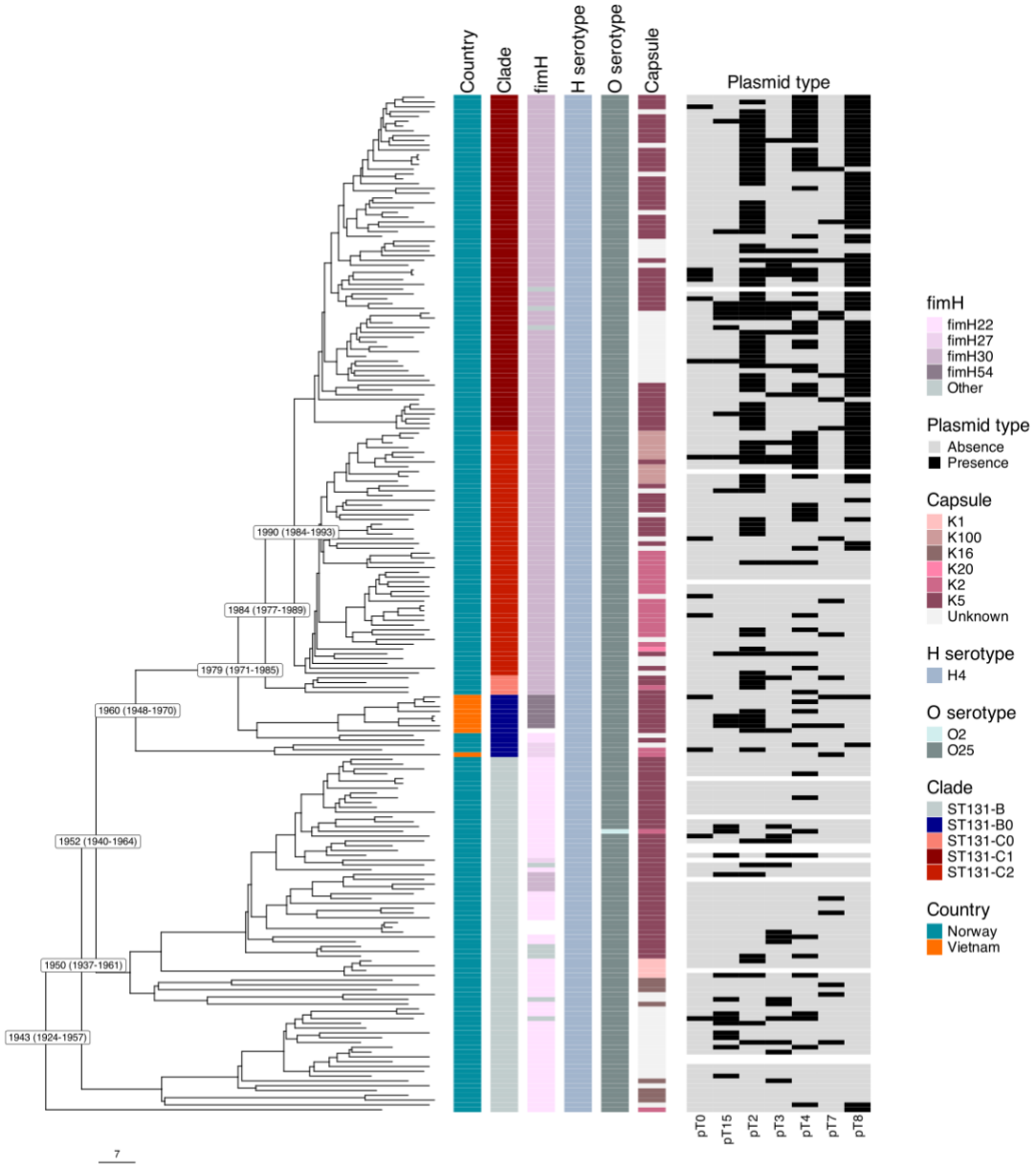

Figure 1. Dated phylogeny of the Vietnamese ST131-B0 (n=9) and Norwegian ST131-B (n=74), ST131-B0 (n=4), ST131-C0 (n=4), ST131-C1 (n=70) and ST131-C2 (n=51), aligned with *fimH*/O/H/K typing and presence/absence patterns of the selected plasmid types (pT) found in the Vietnamese ST131-B0, as indicated in the legend. Isolates lacking any plasmid-predicted contigs are shown as empty in the plasmid type panel. The tree scale indicates years. Time of the most recent common ancestor of each clade with 95% highest posterior density intervals is indicated within the inferred dated phylogeny.

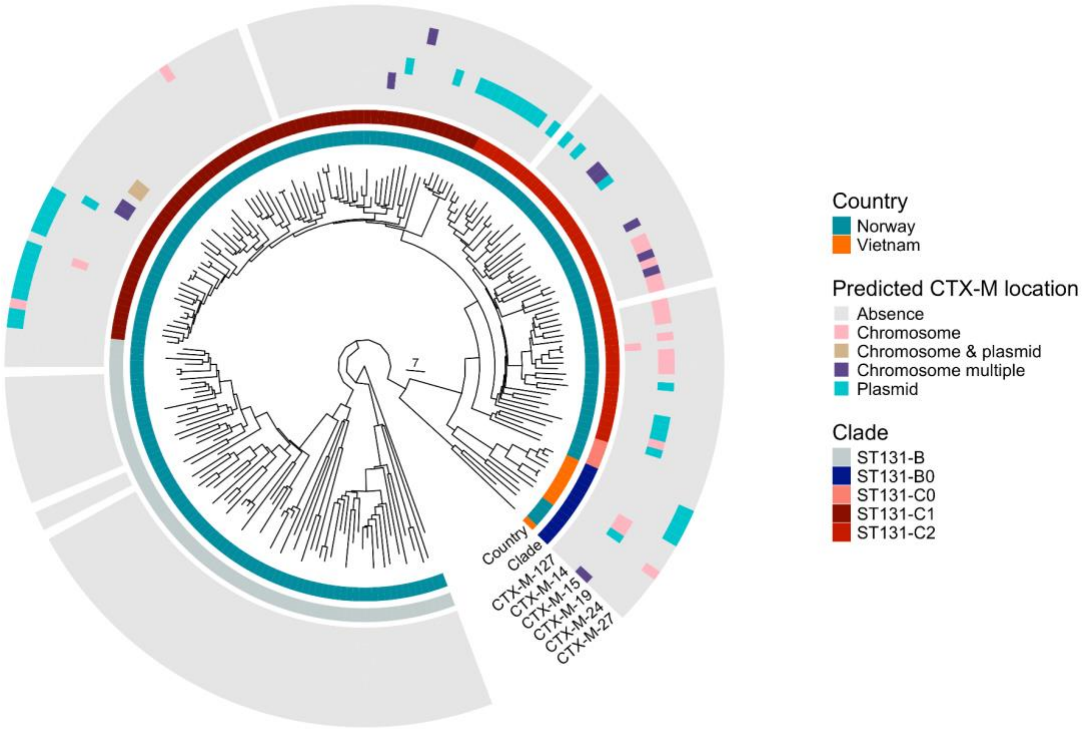

Figure 2. Predicted plasmid and chromosome locations of the CTX-M-type ESBLs in the Vietnamese and Norwegian ST131 isolates, aligned with the dated phylogeny where the tree scale indicates years. Presence of the *bla*$_{\text{CTX-M}}$ genes is indicated by chromosome/plasmid affinity and by CTX-M type as colour coded in the legend. Isolates lacking hybrid assemblies are shown as empty.

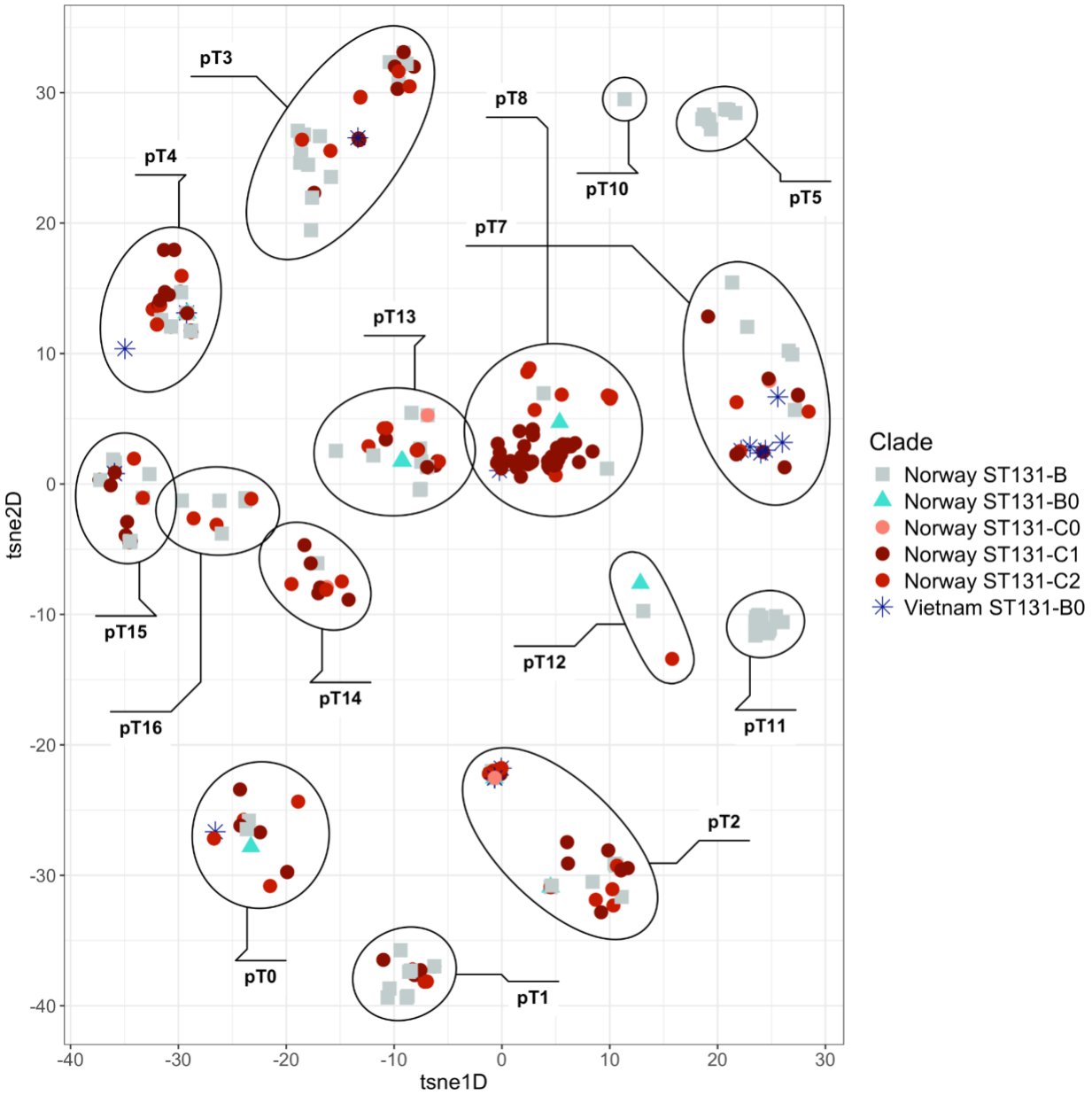

Figure 3. OpenTSNE embedding of 500 *E. coli* plasmid-predicted contigs of the Vietnamese ST131-B0 and Norwegian ST131-B, -B0, -C0, -C1 and -C2. Each point corresponds to a plasmid sequence and their assigned cluster is labelled by the plasmid type (n=15) defined by using mge-cluster with the existing cluster scheme (perplexity=62) (Arredondo-Alonso et al. 2025). Sequences are coloured and shaped by the country of origin and ST131 clade as indicated in the legend. The ellipses (in black) delimit the cluster coordinates. Plasmid-predicted contigs unassigned to any plasmid type (pT-1) were removed.

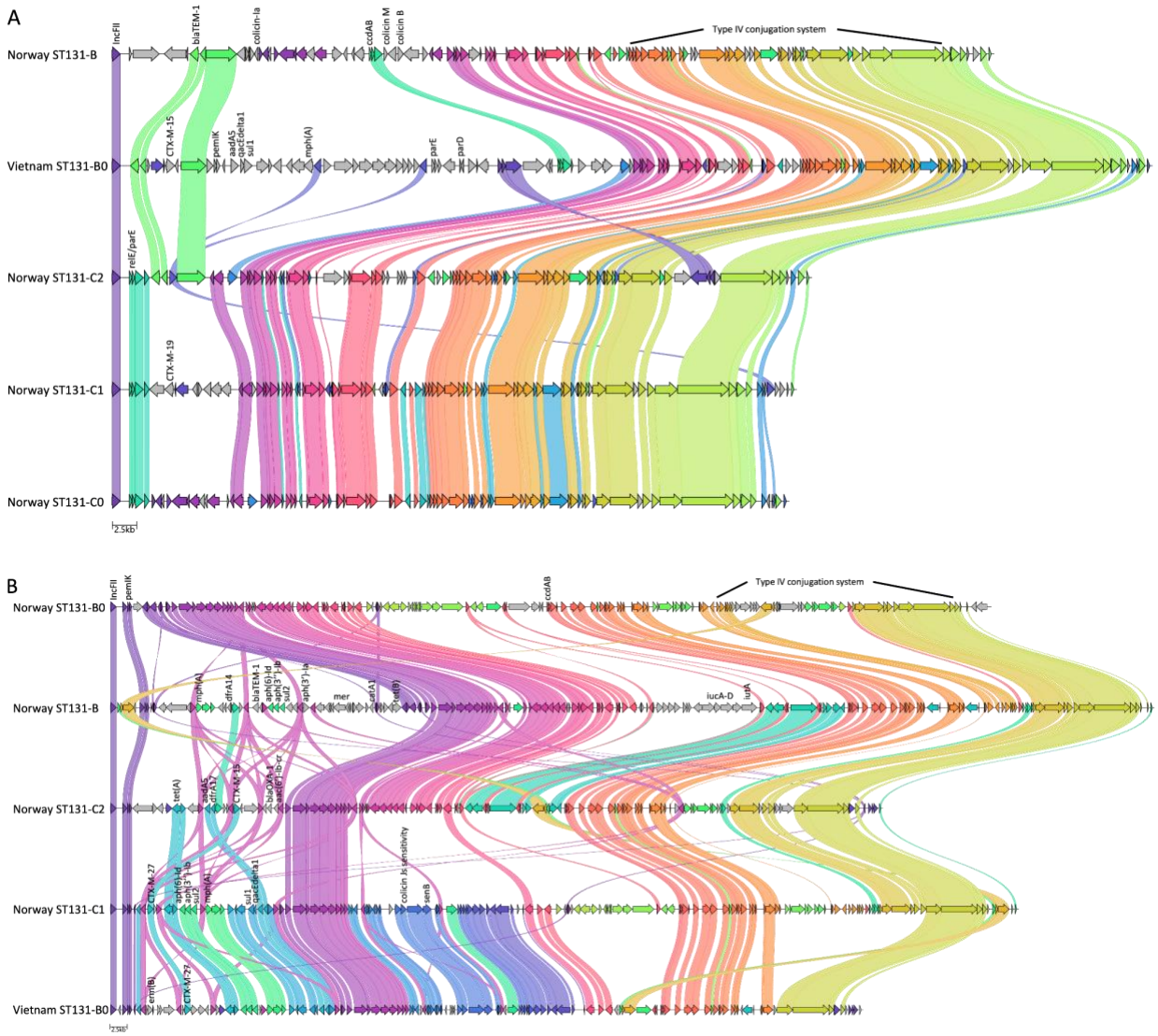

Figure 4. Synteny analysis of representative plasmid-predicted contigs of each ST131 clade within pT7 (A) and pT8 (B) plasmid clusters aligned by using Clinker (Gilchrist and Chooi 2021) with 80% sequence identity threshold. AMR, virulence and stress resistance genes were annotated using AMRFinderPlus (Feldgarden et al. 2021). Replication initiation proteins, conjugation machinery, colicins and type II toxin-antitoxin systems were retrieved from Bakta annotations (Schwengers et al. 2021). Within shared sequences, only the topmost annotations are noted.

References

Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, et al. 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom.* 2:e000093.

Arredondo-Alonso S, Gladstone RA, Pöntinen AK, Gama JA, Schürch AC, Lanza VF, Johnsen PJ, Samuelsen Ø, Tonkin-Hill G, Corander J. 2023. Mge-cluster: a reference-free approach for typing bacterial plasmids. *NAR Genom Bioinform.* 5:lqad066.

Arredondo-Alonso S, Pöntinen AK, Gama JA, Gladstone RA, Harms K, Tonkin-Hill G, Thorpe HA, Simonsen GS, Samuelsen Ø, Johnsen PJ, et al. 2025. Plasmid-driven strategies for clone success in Escherichia coli. *Nat Commun.* 16:2921.

Arredondo-Alonso S, Rogers MRC, Braat JC, Verschuuren TD, Top J, Corander J, Willems RJL, Schürch AC. 2018. Mlplasmids: A user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. *Microb Genom.* 4:e000224.

Baquero F, Lanza VF, Baquero M-R, Del Campo R, Bravo-Vázquez DA. 2019. Microcins in Enterobacteriaceae: Peptide Antimicrobials in the Eco-Active Intestinal Chemosphere. *Front Microbiol.* 10:2261.

Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M, Schembri MA, Beatson SA. 2016. Sequential Acquisition of Virulence and Fluoroquinolone Resistance Has Shaped the Evolution of Escherichia coli ST131. *Mbio.* 7:e00347–16.

Bevan ER, Jones AM, Hawkey PM. 2017. Global epidemiology of CTX-M β-lactamases: temporal and geographical shifts in genotype. *J Antimicrob Chemother.* 72:2145–2155.

Billard-Pomares T, Clermont O, Castellanos M, Magdoud F, Royer G, Condamine B, Fouteau S, Barbe V, Roche D, Cruveiller S, et al. 2019. The arginine deiminase operon is responsible for a fitness trade-off in extended-spectrum-β-lactamase-producing strains of Escherichia coli. *Antimicrob Agents Chemother.* 63:10-1128.

Bonten M, Johnson JR, van den Biggelaar AHJ, Georgalis L, Geurtsen J, de Palacios PI, Gravenstein S, Verstraeten T, Hermans P, Poolman JT. 2021. Epidemiology of Escherichia coli Bacteremia: A Systematic Literature Review. *Clin Infect Dis.* 72:1211–1219.

Bouras G, Houtak G, Wick RR, Mallawaarachchi V, Roach MJ, Papudeshi B, Judd LM, Sheppard AE, Edwards RA, Vreugde S. 2024. Hybracter: enabling scalable, automated, complete and accurate bacterial genome assemblies. *Microb Genom.* 10:001244.

Brodrick HJ, Raven KE, Kallonen T, Jamrozy D, Blane B, Brown NM, Martin V, Török ME, Parkhill J, Peacock SJ. 2017. Longitudinal genomic surveillance of multidrug-resistant Escherichia coli carriage in a long-term care facility in the United Kingdom. *Genome Med.*

594 9:70.

595 Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S, Kyrpides NC.
596 2024. Identification of mobile genetic elements with geNomad. *Nat Biotechnol.* 42:1303–
597 1312.

598 Cave R, Ter-Stepanyan MM, Mkrtchyan HV. 2023. Short- and long-read sequencing reveals the
599 presence and evolution of an IncF Plasmid harboring blaCTX-M-15 and blaCTX-M-27
600 genes in Escherichia coli ST131. *Microbiol Spectr.* 11:e0035623.

601 Coluzzi C, Rocha EPC. 2025. The spread of antibiotic resistance is driven by plasmids among
602 the fastest evolving and of broadest host range. *Mol Biol Evol.* 42:msaf060.

603 Coque TM, Novais A, Carattoli A, Poirel L, Pitout J, Peixe L, Baquero F, Cantón R, Nordmann
604 P. 2008. Dissemination of clonally related Escherichia coli strains expressing extended-
605 spectrum beta-lactamase CTX-M-15. *Emerg Infect Dis.* 14:195–200.

606 Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR.
607 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome
608 sequences using Gubbins. *Nucleic Acids Res.* 43:e15.

609 Cummins EA, Hall RJ, Connor C, McInerney JO, McNally A. 2022. Distinct evolutionary
610 trajectories in the Escherichia coli pangenome occur within sequence types. *Microb*
611 *Genom.* 8:000903.

612 Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. 2018. Bayesian inference of
613 ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* 46:e134.

614 Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M,
615 Pettengill JB, Prasad AB, Tillman GE, et al. 2021. AMRFinderPlus and the Reference Gene
616 Catalog facilitate examination of the genomic links among antimicrobial resistance, stress
617 response, and virulence. *Sci Rep.* 11:12728.

618 Foxman B. 2014. Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk
619 factors, and disease burden. *Infect Dis Clin North Am.* 28:1–13.

620 Gilchrist CLM, Chooi Y-H. 2021. Clinker & clustermap.Js: Automatic generation of gene cluster
621 comparison figures. *Bioinformatics* 37:2473–2475.

622 Gladstone RA, McNally A, Pöntinen AK, Tonkin-Hill G, Lees JA, Skytén K, Cléon F, Christensen
623 MOK, Haldorsen BC, Bye KK, et al. 2021. Emergence and dissemination of antimicrobial
624 resistance in Escherichia coli causing bloodstream infections in Norway in 2002-17: a
625 nationwide, longitudinal, microbial population genomic study. *Lancet Microbe* 2:e331–e341.

626 Gladstone RA, Pesonen M, Pöntinen AK, Mäklin T, MacAlasdair N, Thorpe H, Shao Y,
627 Mallawaarachchi S, Arredondo-Alonso S, Parcell BJ, et al. 2025. Group 2 and 3 ABC-
628 transporter-dependent capsular K-loci contribute significantly to variation in the estimated

629 invasive potential of *Escherichia coli*. *medRxiv*
630 https://doi.org/10.1101/2024.11.22.24317484

631 Hamamoto K, Ueda S, Toyosato T, Yamamoto Y, Hirai I. 2016. High Prevalence of
632 Chromosomal blaCTX-M-14 in Escherichia coli Isolates Possessing blaCTX-M-14.
633 *Antimicrob. Agents Chemother.* 60:2582–2584.

634 Hirai I, Fukui N, Taguchi M, Yamauchi K, Nakamura T, Okano S, Yamamoto Y. 2013. Detection
635 of chromosomal blaCTX-M-15 in Escherichia coli O25b-B2-ST131 isolates from the Kinki
636 region of Japan. *Int J Antimicrob Agents.* 42:500–506.

637 Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated
638 circularization of genome assemblies using long sequencing reads. *Genome Biol.* 16:294.

639 Ingle DJ, Valcanis M, Kuzevski A, Tauschek M, Inouye M, Stinear T, Levine MM, Robins-
640 Browne RM, Holt KE. 2016. In silico serotyping of E. coli from short read data identifies
641 limited novel O-loci but extensive diversity of O:H serotype combinations within and
642 between pathogenic lineages. *Microb Genom.* 2:e000064.

643 Johnson JR, Clabots C, Porter SB, Bender T, Johnston BD, Thuras P. 2022. Intestinal
644 persistence of colonizing Escherichia coli strains, especially ST131-H30, in relation to
645 bacterial and host factors. *J Infect Dis.* 225:2197–2207.

646 Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, Flores-Figueroa C,
647 Aziz M, Stoesser N, Sokurenko E, et al. 2016. Separate F-Type Plasmids Have Shaped the
648 Evolution of the H30 Subclone of Escherichia coli Sequence Type 131. *mSphere.* 1:10-128.

649 Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ, Parkhill J.
650 2017. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates
651 a stable population structure only transiently disturbed by the emergence of ST131.
652 *Genome Res.* 27:1437–1449.

653 Kamruzzaman M, Iredell J. 2019. A ParDE-family toxin antitoxin system in major resistance
654 plasmids of Enterobacteriaceae confers antibiotic and heat tolerance. *Sci Rep.* 9:9872.

655 Kern WV, Rieg S. 2020. Burden of bacterial bloodstream infection-a brief update on
656 epidemiology and significance of multidrug-resistant pathogens. *Clin Microbiol Infect.*
657 26:151–157.

658 Khan J, Kokot M, Deorowicz S, Patro R. 2022. Scalable, ultra-fast, and low-memory
659 construction of compacted de Bruijn graphs with Cuttlefish 2. *Genome Biol.* 23:190.

660 Khan J, Patro R. 2021. Cuttlefish: fast, parallel and low-memory compaction of de Bruijn graphs
661 from large-scale genome collections. *Bioinformatics.* 37:i177–i186.

662 Komori K, Aoki K, Harada S, Ishii Y, Tateda K. 2024. Plasmid-mediated acquisition and
663 chromosomal integration of blaCTX-M-14 in a subclade of Escherichia coli ST131-H30

664 clade C1. *Antimicrob Agents Chemother.* 68:e0081724.

665 Kuronen J, Horsfield ST, Pöntinen AK, Mallawaarachchi S, Arredondo-Alonso S, Thorpe H,
666 Gladstone RA, Willems RJL, Bentley SD, Croucher NJ, et al. 2024. Pangenome-spanning
667 epistasis and coselection analysis via de Bruijn graphs. *Genome Res.* 34:1081–1088.

668 Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. 2018. pyseer: a comprehensive tool
669 for microbial pangenome-wide association studies. *Bioinformatics.* 34:4310–4312.

670 Lipworth S, Matlock W, Shaw L, Vihta K-D, Rodger G, Chau K, Barker L, George S, Kavanagh
671 J, Davies T, et al. 2024. The plasmidome associated with Gram-negative bloodstream
672 infections: A large-scale observational study using complete plasmid assemblies. *Nat*
673 *Commun.* 15:1612.

674 Liu CM, Stegger M, Aziz M, Johnson TJ, Waits K, Nordstrom L, Gauld L, Weaver B, Rolland D,
675 Statham S, et al. 2018. Escherichia coli ST131-H22 as a Foodborne Uropathogen. *MBio.*
676 9:10-128.

677 Mäklin T, Thorpe HA, Pöntinen AK, Gladstone RA, Shao Y, Pesonen M, McNally A, Johnsen
678 PJ, Samuelsen Ø, Lawley TD, et al. 2022. Strong pathogen competition in neonatal gut
679 colonisation. *Nat Commun.* 13:7417.

680 Mallawaarachchi S, Tonkin-Hill G, Pöntinen AK, Calland JK, Gladstone RA, Arredondo-Alonso
681 S, MacAlasdair N, Thorpe HA, Top J, Sheppard SK, et al. 2024. Detecting co-selection
682 through excess linkage disequilibrium in bacterial genomes. *NAR Genom Bioinform*
683 6:lqae061.

684 Manges AR. 2016. Escherichia coli and urinary tract infections: the role of poultry-meat. *Clin*
685 *Microbiol Infect.* 22:122–129.

686 Matsumura Y, Pitout JDD, Gomi R, Matsuda T, Noguchi T, Yamamoto M, Peirano G, DeVinney
687 R, Bradford PA, Motyl MR, et al. 2016. Global Escherichia coli Sequence Type 131 Clade
688 with blaCTX-M-27 Gene. *Emerg Infect Dis.* 22:1900–1907.

689 McNally A, Kallonen T, Connor C, Abudahab K, Aanensen DM, Horner C, Peacock SJ, Parkhill
690 J, Croucher NJ, Corander J. 2019. Diversification of Colonization Factors in a Multidrug-
691 Resistant Escherichia coli Lineage Evolving under Negative Frequency-Dependent
692 Selection. *MBio.* 10: 10-1128.

693 McNally A, Oren Y, Kelly D, Pascoe B, Dunn S, Sreecharan T, Vehkala M, Välimäki N, Prentice
694 MB, Ashour A, et al. 2016. Combined Analysis of Variation in Core, Accessory and
695 Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of
696 Bacterial Populations. *PLoS Genet.* 12:e1006280.

697 Molari M, Shaw LP, Neher RA. 2025. Quantifying the evolutionary dynamics of structure and
698 content in closely related E. coli genomes. *Mol Biol Evol.* 42:msae272.

699 Naylor NR, Pouwels KB, Hope R, Green N, Henderson KL, Knight GM, Atun R, Robotham JV,
700 Deeny SR. 2019. The health and cost burden of antibiotic resistant and susceptible
701 Escherichia coli bacteraemia in the English hospital setting: A national retrospective cohort
702 study. *PLoS One.* 14:e0221944.

703 Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. 2014. Escherichia coli ST131, an intriguing
704 clonal group. *Clin Microbiol Rev.* 27:543–574.

705 Nicolas-Chanoine M-H, Blanco J, Leflon-Guibout V, Demarty R, Alonso MP, Canica MM, Park
706 Y-J, Lavigne J-P, Pitout J, Johnson JR. 2007. Intercontinental emergence of Escherichia
707 coli clone O25:H4-ST131 producing CTX-M-15. *J Antimicrob Chemother.* 61:273–281.

708 Ojala F, Pesonen H, Gladstone RA, Mäklin T, Tonkin-Hill G, Marttinen P, Corander J. 2025.
709 Basic reproduction number varies markedly between closely related pandemic *Escherichia*
710 *coli* clones. *Nat Commun.* 16:9490.

711 Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, Otto TD, Keane JA. 2016. Robust
712 high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data.
713 *Microb Genom.* 2:e000083.

714 Pensar J, Puranen S, Arnold B, MacAlasdair N, Kuronen J, Tonkin-Hill G, Pesonen M, Xu Y,
715 Sipola A, Sánchez-Busó L, et al. 2019. Genome-wide epistasis and co-selection study
716 using mutual information. *Nucleic Acids Res.* 47:e112.

717 Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan M-D,
718 Gomes Moriel D, Peters KM, Davies M, et al. 2014. Global dissemination of a multidrug
719 resistant Escherichia coli clone. *Proc Natl Acad Sci U. S. A.* 111:5694–5699.

720 Pitout JDD, DeVinney R. 2017. Escherichia coli ST131: a multidrug-resistant clone primed for
721 global domination. *F1000Res.* 6:195.

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

724 Platell JL, Johnson JR, Cobbold RN, Trott DJ. 2011. Multidrug-resistant extraintestinal
725 pathogenic Escherichia coli of sequence type ST131 in animals and foods. *Vet Microbiol.*
726 153:99–108.

727 Pöntinen AK, Gladstone RA, Pesonen H, Pesonen M, Cléon F, Parcell BJ, Kallonen T,
728 Simonsen GS, Croucher NJ, McNally A, et al. 2024. Modulation of multidrug-resistant clone
729 success in Escherichia coli populations: a longitudinal, multi-country, genomic and antibiotic
730 usage cohort study. *Lancet Microbe.* 5:e142–e150.

731 Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M,
732 Chattopadhyay S, Stegger M, et al. 2013. The epidemic of extended-spectrum-β-
733 lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic
734 subclone, H30-Rx. *MBio.* 4:e00377–13.

735 Puranen S, Pesonen M, Pensar J, Xu YY, Lees JA, Bentley SD, Croucher NJ, Corander J.
736 2018. SuperDCA for genome-wide epistasis analysis. *Microb Genom.* 4:e000184.

737 R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for
738 Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/

739 Reid CJ, McKinnon J, Djordjevic SP. 2019. Clonal ST131-H22 Escherichia coli strains from a
740 healthy pig and a human urinary tract infection carry highly similar resistance and virulence
741 plasmids. *Microb Genom.* 5:e000295.

742 Robertson J, Bessonov K, Schonfeld J, Nash JHE. 2020. Universal whole-sequence-based
743 plasmid typing and its utility to prediction of host range and epidemiological surveillance.
744 *Microb Genom*. 6:e000435.

745 Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction and
746 typing of plasmids from draft assemblies. *Microb Genom*. 4:e000206.

747 Rodríguez-Beltrán J, DelaFuente J, León-Sampedro R, MacLean RC, San Millán Á. 2021.
748 Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat Rev*
749 *Microbiol.* 19:347–359.

750 Rodríguez-Beltrán J, Tourret J, Tenaillon O, López E, Bourdelier E, Costas C, Matic I, Denamur
751 E, Blázquez J. 2015. High recombinant frequency in extraintestinal pathogenic Escherichia
752 coli strains. *Mol Biol Evol.* 32:1708–1716.

753 Roer L, Overballe-Petersen S, Hansen F, Johannesen TB, Stegger M, Bortolaia V,
754 Leekitcharoenphon P, Korsgaard HB, Seyfarth AM, Mossong J, et al. 2019. ST131 fimH22
755 Escherichia coli isolate with a blaCMY-2/IncI1/ST12 plasmid obtained from a patient with
756 bloodstream infection: highly similar to E. coli isolates of broiler origin. *J Antimicrob*
757 *Chemother.* 74:557–560.

758 Saidenberg ABS, Stegger M, Price LB, Johannesen TB, Aziz M, Cunha MPV, Moreno AM,
759 Knöbl T. 2020. Mcr-positive Escherichia coli ST131-H22 from poultry in Brazil. *Emerg Infect*
760 *Dis.* 26:1951–1954.

761 San Millan A, Escudero JA, Gifford DR, Mazel D, MacLean RC. 2016. Multicopy plasmids
762 potentiate the evolution of antibiotic resistance in bacteria. *Nat Ecol Evol.* 1:10.

763 Sarkar S, Hutton ML, Vagenas D, Ruter R, Schüller S, Lyras D, Schembri MA, Totsika M. 2018.
764 Intestinal colonization traits of pandemic multidrug-resistant Escherichia coli ST131. *J Infect*
765 *Dis.* 218:979–990.

766 Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. 2021. Bakta: rapid
767 and standardized annotation of bacterial genomes via alignment-free sequence
768 identification. *Microb Genom*. 7:000685.

769 Shamsrizi P, Gladstone BP, Carrara E, Luise D, Cona A, Bovo C, Tacconelli E. 2020. Variation

770 of effect estimates in the analysis of mortality and length of hospital stay in patients with
771 infections caused by bacteria-producing extended-spectrum beta-lactamases: a systematic
772 review and meta-analysis. *BMJ Open.* 10:e030266.

773 Stanton TD, Hetland MAK, Löhr IH, Holt KE, Wyres KL. 2025. Fast and accurate in silico
774 antigen typing with Kaptive 3. *Microb Genom.* 11:001428.

775 Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW,
776 Kasarskis A, Batty EM, et al. 2016. Evolutionary history of the global emergence of the
777 Escherichia coli epidemic clone ST131. *MBio.* 7:e02162.

778 Tonkin-Hill G, Gladstone RA, Pöntinen AK, Arredondo-Alonso S, Bentley SD, Corander J. 2023.
779 Robust analysis of prokaryotic pangenome gene gain and loss rates with Panstripe.
780 *Genome Res.* 33:129–140.

781 Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S,
782 Beaudoin C, Floto RA, et al. 2020. Producing polished prokaryotic pangenomes with the
783 Panaroo pipeline. *Genome Biol.* 21:180.

784 WHO. 2024. WHO bacterial priority pathogens list, 2024: bacterial pathogens of public health
785 importance, to guide research, development, and strategies to prevent and control
786 antimicrobial resistance. World Health Organization.

787 Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using
788 exact alignments. *Genome Biol.* 15:R46.

789 Wu AY, Kamruzzaman M, Iredell JR. 2020. Specialised functions of two common plasmid
790 mediated toxin-antitoxin systems, ccdAB and pemIK, in Enterobacteriaceae. *PLoS One.*
791 15:e0230652.