

# Plasmid-driven strategies for clone success in *Escherichia coli*

Received: 10 September 2024

Accepted: 28 February 2025

Published online: 03 April 2025



Sergio Arredondo-Alonso<sup>1,2,3</sup>, Anna K. Pöntinen<sup>1,2,23</sup>, João A. Gama<sup>3,4,23</sup>, Rebecca A. Gladstone<sup>1,2,3</sup>, Klaus Harms<sup>3</sup>, Gerry Tonkin-Hill<sup>1,5,6,7</sup>, Harry A. Thorpe<sup>1</sup>, Gunnar S. Simonsen<sup>2,8</sup>, Ørjan Samuelsen<sup>2,3,24</sup>, Pål J. Johnsen<sup>3,24</sup>, Jukka Corander<sup>1,9,10,24</sup> ✉ & Norwegian E. coli BSI Study Group\*

*Escherichia coli* is the most widely studied microbe in history, but the population structure and evolutionary trends of its extrachromosomal elements known as plasmids remain poorly delineated. Here we used long-read technology to high-resolution sequence the entire plasmidome and the corresponding host chromosomes from an unbiased longitudinal survey covering two decades and over 2000 *E. coli* isolates. We find that some plasmids have persisted in lineages even for centuries, demonstrating strong plasmid-lineage associations. Our analysis provides a detailed map of recent vertical and horizontal evolutionary events involving plasmids with key antibiotic resistance, competition and virulence determinants. We present genomic evidence of both chromosomal and plasmid-driven success strategies adopted by distant lineages by independently inheriting the same genomic elements. Further, we use in vitro experiments to verify the importance of key bacteriocin-producing plasmids for clone success. Our study has general implications for understanding plasmid biology and bacterial evolutionary strategies.

The extra-intestinal pathogenic *E. coli* (ExPEC) have been of particular interest to both eco-evolutionary and epidemiological research communities since they near-universally colonize the healthy human gut in an asymptomatic fashion, while also causing a high global burden of opportunistic infections, predominantly in the urinary tract and bloodstream<sup>1</sup>. ExPEC possess an arsenal of iron acquisition mechanisms, toxins, and adhesins that enhance their persistence in this niche<sup>2–5</sup>.

The ExPEC population is composed of a high diversity of clones, but a relatively small subset of lineages contributes disproportionately to the majority of urinary tract and bloodstream infections<sup>6</sup>, where increasing levels of antibiotic resistance represent a challenge for effective treatment. Epidemiological investigations have uncovered an intriguing feature of ExPEC, showing stable maintenance of pandemic lineages over longer timescales and the emergence of novel successful clones in clinical surveillance, many with foodborne associations<sup>6–9</sup>.

<sup>1</sup>Department of Biostatistics, University of Oslo, Oslo, Norway. <sup>2</sup>Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway. <sup>3</sup>Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway. <sup>4</sup>Department of Microbiology, Ramón y Cajal University Hospital, Ramón y Cajal Institute for Health Research (IRYCIS), Madrid, Spain. <sup>5</sup>Peter MacCallum Cancer Centre, Melbourne, Australia. <sup>6</sup>Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Australia. <sup>7</sup>Department of Microbiology and Immunology, The University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia. <sup>8</sup>Department of Medical Biology, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway. <sup>9</sup>Parasites and Microbes, Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK. <sup>10</sup>Helsinki Institute for Information Technology HIIT, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland. <sup>23</sup>These authors contributed equally: Sergio Arredondo-Alonso, Anna K. Pöntinen, João A. Gama, Rebecca A. Gladstone. <sup>24</sup>These authors jointly supervised this work: Ørjan Samuelsen, Pål J. Johnsen, Jukka Corander. \*A list of authors and their affiliations appears at the end of the paper. ✉e-mail: [jukka.corander@medisin.uio.no](mailto:jukka.corander@medisin.uio.no)

The new clones typically expand rapidly initially, but are then constrained within only a few years such that the population transits to another equilibrium still maintaining a considerable diversity of lineages. Unbiased longitudinal genomic surveys have demonstrated that the transient disruption of the equilibrium bears the hallmarks of negative frequency-dependent selection (NFDS)<sup>10–12</sup>, and further modeling work suggested this selection is acting on accessory genomic elements rather than on core genome variation<sup>13</sup>.

Plasmids, initially described as transmissible agents carrying antibiotic resistance and competitor-inhibiting determinants in *E. coli* in the 1950s<sup>14–17</sup>, play a key role in accessory genome dynamics. Numerous traits involved in antibiotic resistance, iron-acquisition mechanisms, bacterial competition, immune evasion and metabolism are encoded by genes residing on plasmids. In particular, some *E. coli* carry colicinogenic plasmids that produce bacteriocins: toxic proteins typically active against close relatives<sup>18</sup>. Bacteriocins play an active role in intra-species competition<sup>19,20</sup> and have been proposed as one of the mechanisms maintaining NFDS<sup>21,22</sup>. Despite the strong evidence that plasmids have played a critical role in the emergence and dissemination of successful clones<sup>23–26</sup>, the structure of the plasmidome, and the horizontal and vertical evolutionary dynamics of plasmids and ExPEC hosts remain largely unknown. This is due to the paucity of scalable computational approaches to type complete plasmid sequences, and the lack of large-scale longitudinal long-read sequencing studies, which has hindered developing a firm understanding of the plasmidome structure and prevalence of particular plasmids.

Except in a recent study<sup>27</sup>, ExPEC plasmids have previously been explored on a small-scale, most often in association with specific selective criteria, e.g. isolates carrying a particular resistance locus such as extended-spectrum  $\beta$ -lactamase (ESBL), and using mostly short-read sequencing data<sup>24,28–30</sup>. To circumvent this, we long-read sequenced 2045 *E. coli* isolates representing the genomic diversity of earlier short-read sequenced 3245 bloodstream infection isolates from a 16-year national longitudinal study<sup>11</sup>. These isolates were sampled regardless of their clonal background, antibiotic resistance profile or any other bacterial genotypic or phenotypic characteristic, making it ideal for the study of evolution, expansion and persistence of the plasmidome.

Our approaches resulted in a total of 4485 circularized plasmid sequences and through inferred evolutionary rates and dated phylogenies we estimated the timing of plasmid emergence/acquisition into major ExPEC clones. We demonstrate multiple cases of acquisition of the same plasmid-driven phenotypic traits in distant genetic backgrounds, supporting the conclusion that plasmids play key roles in initial clonal establishment and later endemic maintenance in the host population. In particular, we found that some ExPEC clones were strongly associated with pColV-like plasmids encoding microcin V, a narrow-spectrum bacterial toxin involved in bacterial competition<sup>17</sup>. By exploring the bacteriocin content present in the ExPEC population, we propose that plasmid-encoded bacteriocins play a pivotal role in shaping and maintaining the NFDS dynamics in *E. coli*. Through experimental approaches, we further demonstrate microcin V activity in different clonal backgrounds to inhibit the growth of multi-drug resistant ExPEC clones, thus contributing to the clonal success and maintenance of non-resistant clones in the population.

## Results

### Typing the ExPEC plasmidome: a network-based approach

To elucidate the ExPEC plasmid repertoire and its stability over time, we focused on a unique Norwegian epidemiological longitudinal collection (NORM collection) spanning a period of 16 years and involving 3245 bloodstream infection isolates. This study included isolates regardless of their sequence type (ST) or antibiotic resistance background (e.g. presence of ESBL genes) and showed that almost all STs found in Norway were internationally circulating by a comparison to

earlier systematic longitudinal studies conducted elsewhere<sup>10–12</sup>. To ensure the inclusion of isolates representative of the whole ExPEC population for long-read sequencing, we first selected 1085 isolates (33.44%, 1,085/3,245) using a statistical method informed by the accessory genome diversity of the isolates (Fig. S1) based on short-read sequences<sup>31</sup>. Next, we included all the remaining 960 isolates belonging to the four major STs, referred to as pandemic clones (ST69, ST73, ST95 and ST131) that caused nearly half of the bloodstream infections (44.19%, 1438/3254)<sup>11</sup>. Thus, a total of 2045 isolates (62.85% of the NORM collection) were selected for long-read sequencing, representing 216 distinct STs covering the entire ExPEC phylogeny (Fig. S2).

Highly contiguous assemblies were obtained for 1999 genomes (Fig. S3A) with median N50 chromosome length 4.98 Mbp (Fig. S3B) and an average L50 chromosome count 1.09. A total of 5417 contigs were identified as plasmid-derived, including 4485 circular plasmid sequences, collectively here referred to as the plasmidome. Chromosome lengths differed significantly between the major clones (Fig. S3C,  $p$ -value < 0.05), where ST69 and ST73 showed a larger chromosome size on average compared to ST95 and ST131. The total length and the number of plasmid sequences observed per isolate was significantly lower for ST73 ( $p$ -value < 0.05) compared to the other major clones (Fig. S3D). This finding confirms previous reports showing that ST73 frequently carries lower plasmid load, demonstrated by lower conjugation frequencies and higher fitness costs<sup>32,33</sup>.

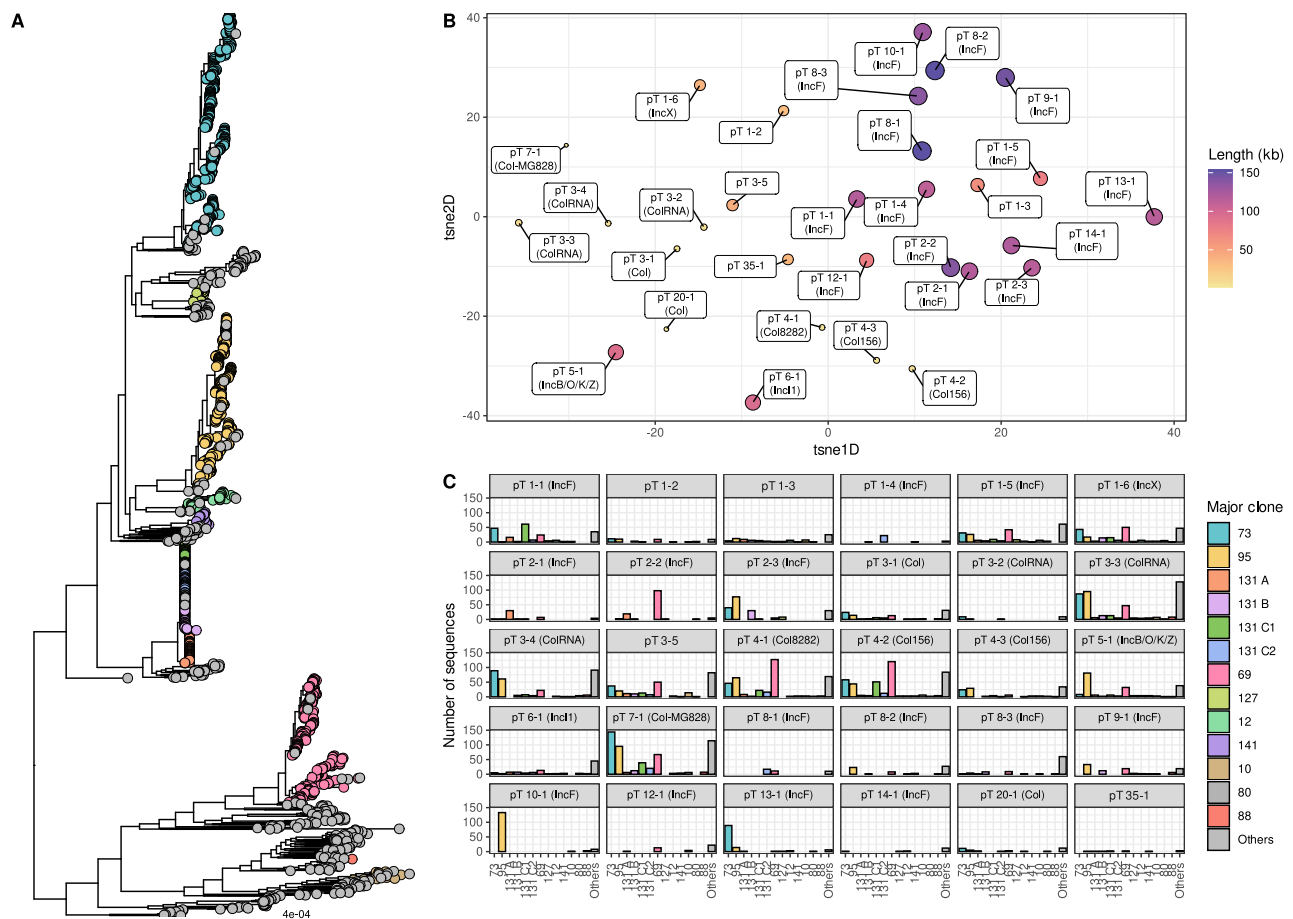
To type the circular plasmid sequences ( $n = 4485$ ) identified from the hybrid assemblies, we first removed redundancy with *cd-hit-est* resulting in a set of 2560 non-redundant plasmid sequences (Fig. S4). Next, we applied our previously developed method ‘mge-cluster’, that is well-suited for typing plasmid sequences derived from large studies<sup>34</sup>. To ensure robust typing results, we clustered plasmids using a weighted undirected network based on multiple runs of *mge-cluster* with distinct parameter combinations. In the resulting network, nodes correspond to plasmids and edges to the association strength among plasmids co-clustering together in the distinct solutions (see Methods).

The network consisted of 15 components (connected subgraphs) encompassing 2392 non-redundant plasmid sequences (93%) while 168 sequences remained as singletons (7%) (Fig. S4). Next, we used a popular community detection algorithm based on modularity optimization (Louvain method) to assign highly connected communities into non-overlapping groups (bins)<sup>35</sup>. After quality control (see Methods), 2285 non-redundant plasmids were clustered into 30 distinct groups referred to as plasmid types (pTs) labeled with two digits representing their network component and the community bin defined by the Louvain algorithm (Fig. S4). To obtain a consistent typing scheme and a compact visualization (Fig. 1), we selected the *mge-cluster* solution (perplexity = 62, cluster size = 49) closest (adjusted Rand index 0.99) to the network typing system.

Finally, the redundant sequences ( $n = 1925$ ) discarded by *cd-hit* were assigned to the same pT as their representative sequence (Fig. S4). Using the existing operational mode of *mge-cluster*, we could further assign 498 non-circular plasmid sequences to specific pTs without the need to re-consider the entire network and to modify the typing scheme. In total, 4569 plasmid sequences (84%, 4569/5417: 4071 circular and 498 non-circular) were assigned to the 30 defined pTs (Fig. S4) representing the ExPEC plasmidome.

### Clonal spread of key plasmid types is interspersed with horizontal transfer

Figure 1A shows the maximum likelihood phylogeny of the 1999 ExPEC isolates with hybrid assemblies, based on core genome variation. To avoid visual overloading of the phylogeny, only the ten most prevalent STs are labeled, including the subclades of ST131 (A, B, C1, C2) which have been of wide interest to the research community given their rapid diversification and distinct evolutionary trajectories. Joint inspection



**Fig. 1 | Plasmid diversity and typing scheme on the set of predicted plasmid sequences obtained from 1999 ExPEC hybrid assemblies.** **A** Maximum-likelihood phylogeny of the 1999 ExPEC isolates with an associated hybrid assembly. Isolates corresponding to the ten most prevalent sequence types (ST) in the NORM study are indicated with a distinct color and their position marked in the phylogeny. In the case of ST131, the distinct clades present in the clone (A, B, C1, C2) are represented. **B** The closest mge-cluster solution (perplexity = 62, minimum cluster size = 49) to the network approach was considered to visualize the position of the pTs in

the 2D embedding space. For visualization purposes, we computed the mean tSNE1D and tSNE2D coordinates of all sequences belonging to each pT and considered those coordinates to represent the position and anchor each pT in the embedding. Each pT is labeled with its associated bin consisting of two digits indicating their graph component and associated bin and its size is proportional to the average plasmid length. **C** Barplot indicating the total number of plasmid sequences assigned to each pT and belonging to the main STs found in the ExPEC collection.

of the pT map (Fig. 1B) and the corresponding frequency distribution of individual pTs (Fig. 1C) reveals that both the smallest plasmids (e.g. pT3-3, pT4-1, pT7-1, pT20-1) and large plasmids (e.g. pT1-1, pT1-3, pT5-1, pT8-2) can be widely disseminated across the phylogeny, indicating their persistence in the population by successful horizontal transfer. In a stark contrast, only some large plasmids display a much narrower phylogenetic distribution and evidence of overwhelmingly clonal expansion aided by their host lineage (pT1-4, pT8-1, pT10-1). Comparison of the plasmidome structure for the genomes sequenced in this study (Fig. 1B) with that of the published plasmid sequences of *E. coli* as delineated recently<sup>34</sup> revealed a very high degree of congruence between the two (adjusted Rand Index 0.94; a standard measure of congruence between two partitions of data). This suggests that plasmid evolution in ExPEC is more constrained than previously considered, and that there are only a limited number of successful plasmid backbones which can serve as stable vehicles for the dissemination of advantageous genes.

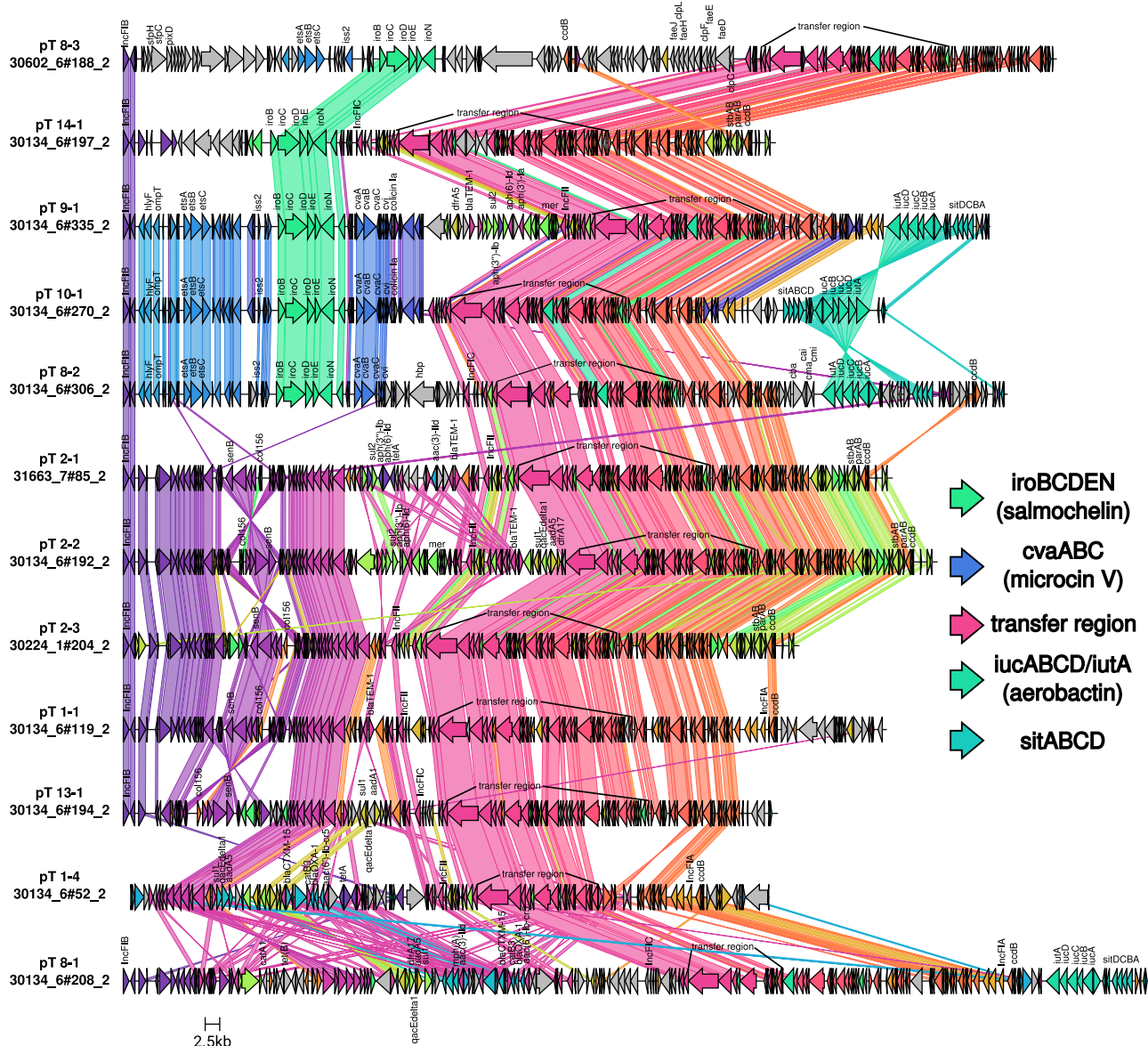
### High levels of gene sharing and genetic relatedness of IncF plasmids

To link our assigned pTs with existing plasmid classification and typing schemes, we identified the dominating replicons (PlasmidFinder), plasmid MLST schemes (pmlst) and checked their congruence with

another well-established plasmid typing tool (MOB-suite) (Supplementary Data 2). In a previous publication<sup>34</sup>, we showed that mge-cluster and MOB-suite were concordant on the typing of *E. coli* plasmids, but mge-cluster was generally better at grouping together sequences with a shared plasmid backbone. For each pT, we further determined the degree of gene sharing, predicted mobility (conjugative, mobilizable or non-mobilizable) and the presence of antibiotic and virulence genes (Supplementary Data 2).

Our analyses revealed that all pTs with a large average plasmid size (>100 kb) contained IncF variant replicons, harbored antibiotic and/or virulence determinants, and displayed high degrees of gene sharing (Fig. S5) as measured by their containment index<sup>36</sup>. Our results support large plasmids as a key source of bacterial adaptive traits.

Gene synteny analysis (minimum identity 0.8) among representative sequences of large IncF pTs (>100 kb) (Fig. 2) revealed the presence of large genomic blocks in common including siderophore-operons (*iroBCDEN*, *iutA-iucABCD* and *sitABCD* operons), antibiotic resistance genes and the prototypical conjugation F-type transfer system. By increasing the stringency to create a link between genes in synteny (minimum identity 0.99) (Fig. S6), we observed that the F-type transfer region among pTs 8-2/9-1/10-1 and pTs 2-1/2-2/2-3 was highly conserved and these pTs differ mainly by the acquisition of antibiotic resistance genes. Our data initially suggested that these pTs have



**Fig. 2 | Synteny analysis and visualization of the main IncF large plasmid types (pT).** For each pT, we selected a plasmid sequence (tagged with its sequence ID) representing the most predominant features present in the cluster (as summarized in Supplementary Data 2). The gene synteny analysis was performed using

clinker<sup>88</sup> considering a minimum identity threshold of 0.8 to draw a link between genes. Virulence and antibiotic resistance genes were identified using the EcoH database<sup>81</sup> (indexed in Abricate) and AMRFinderPlus<sup>82</sup> and their names curated according to well-studied ExPEC reference plasmids.

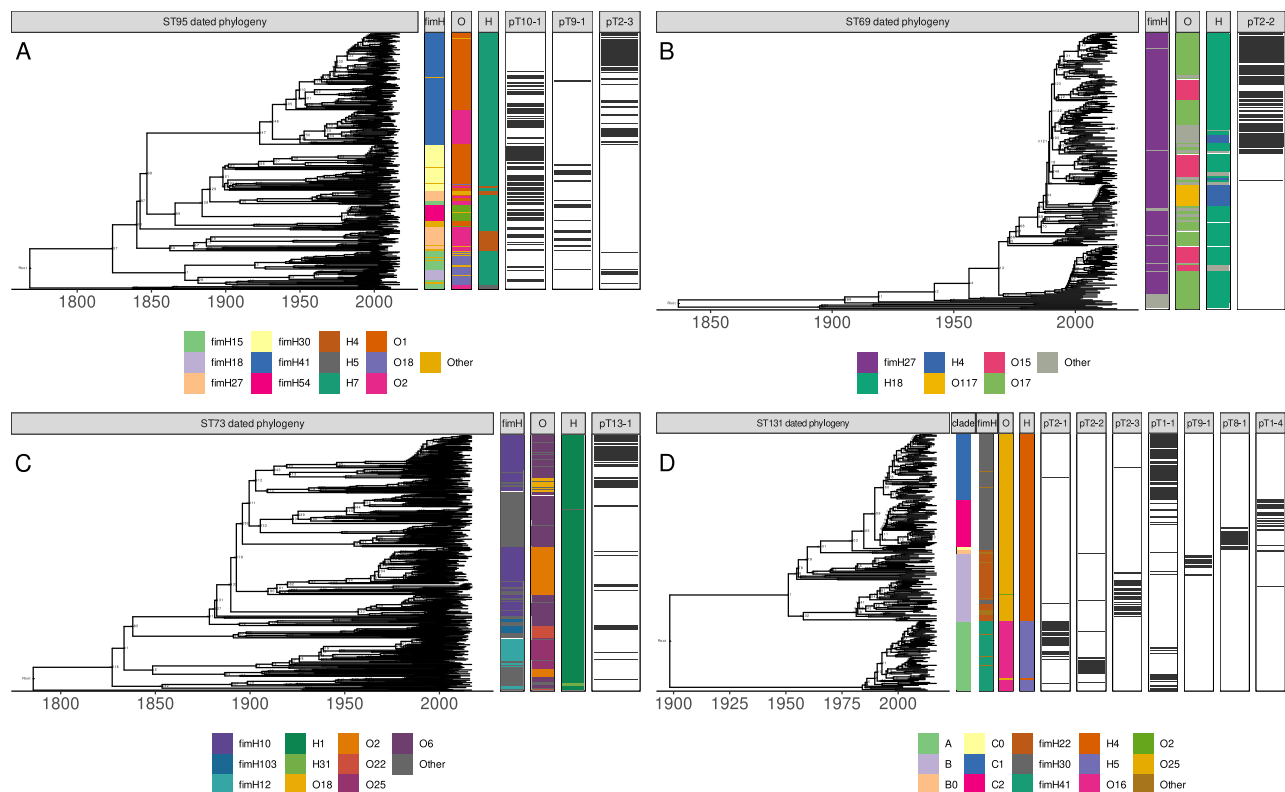
recently evolved from the same ancestral pT and we confirmed this by reconstructing a recombination-free phylogeny using the aligned plasmid sequences from each combination of plasmid types (Fig. S7), which showed that pT 8-2 and pT 2-3 were the ancestral plasmid types. Furthermore, pT 9-1 and pT 2-2 arose respectively from pT 10-1 and pT 2-1. In addition, these pTs were clustered in the same group by MOB-suite confirming their relatedness (Supplementary Data 2).

### Unraveling the shared evolutionary history of plasmid and clones

After elucidating and typing the ExPEC plasmidome, we explored whether particular pTs could have contributed to the success of pandemic ExPEC clones. We focused our analyses on pTs corresponding to medium and large plasmids (>10 kb) and in more detail on pColV-like and pUTI89-like plasmids which have previously been linked to: (i) ExPEC virulence, (ii) adaptive traits influencing host range, (iii) zoonotic potential, and (iv) antibiotic resistance carriage<sup>37–40</sup>. To this end, we inferred the evolutionary rates and dated the phylogenies of the

four major clones (ST69, ST73, ST95, ST131) using BactDating<sup>41</sup> and further detected probable clonal expansions associated with the acquisition of pTs using CaveDive<sup>42</sup>. This permitted us to estimate when ExPEC clones may have acquired specific pTs and whether a clonal expansion was followed by that association.

ST95, an ExPEC clone mostly susceptible to antibiotics, showed a high prevalence (37%) of the pColV-like plasmid pT 10-1 (Fig. S8). As illustrated in Fig. 2, this plasmid possesses all features representative of pColV-like plasmids<sup>37</sup> including several iron-acquisition mechanisms (salmochelin, aerobactin), bacteriocin genes involved in clone competition (microcin V, colicin Ia)<sup>38,43</sup>, immune evasion genes conferring resistance to serum (*iss*)<sup>39</sup> and virulence genes (e.g. *hlyF*)<sup>44</sup>. The distribution of pT 10-1 in the phylogeny of ST95 (Fig. 3A) suggests that the plasmid was introduced in the most recent common ancestor (MRCA) of ST95 dating back to 1768 (95% CI, 1721–1806). Furthermore, the major branch leading from the root of ST95 was identified as a clonal expansion occurring in 1823 (95% CI, 1789–1851). By comparing plasmid versus core-genome distances, we observed a positive correlation



**Fig. 3 | Dated phylogenies of the four main ExPEC clones and presence/absence pattern of the most prevalent plasmid types (pT) found in each sequence type (ST). A** Dated phylogeny of ST95 with clades indicated based on fimH/O/H typing and presence/absence of the pColV-like plasmids (pTs 10-1 and 9-1) and the pUTI89-like plasmid (pT 2-3). **B** Dated phylogeny of ST69 with clades indicated

based on fimH/O/H typing and presence/absence of the pUTI89-like plasmid (pT 2-2). **C** Dated phylogeny of ST73 with clades indicated based on fimH/O/H typing and presence/absence of the pT 13-1. **D** Dated phylogeny of ST131 with clades indicated based on existing nomenclature (A, B, B0, C0, C1, C2), fimH/O/H typing and presence/absence of multiple associated pTs.

(0.78) between pT 10-1 and ST95 phylogenies further supporting that the plasmid has been vertically transmitted in ST95 instead of being repeatedly acquired (Fig. S9A). However, we cannot discard the possibility of a joined transfer evolution event for which pT 10-1 has been horizontally acquired several times and clonally expanded afterwards. In some instances, we observed the replacement of pT 10-1 with pT 9-1, the alternate version of the pColV-like plasmid characterized by the presence of an antibiotic and heavy-metal resistance gene cassette. A particular clade of ST95 (fimH41:O1/O2:H7) showed multiple and independent replacements of pT 10-1 with pT 2-3 (Fig. S8, Fig. S9B), a pUTI89-like plasmid labeled by the pmlst scheme as F29:A-B10. This plasmid has been characterized by its immune evasion properties<sup>45</sup>, and by playing an important role in virulence and biofilm regulation (*senB* and *parAB* genes)<sup>46,47</sup>.

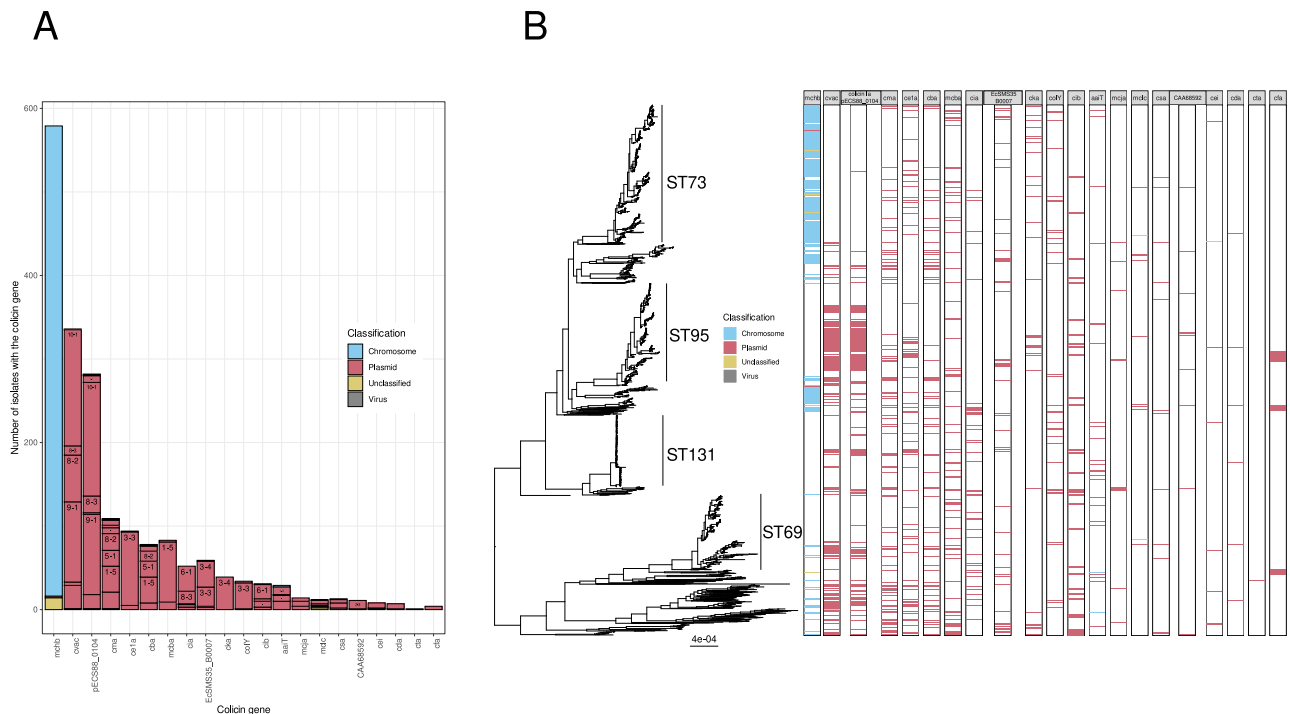
The basal part of the ST69 phylogeny represents isolates notably void of larger plasmid carriage, suggesting that such plasmids did not play a central role in the early evolution of this lineage (Fig. 3B). However, a particular clade of ST69 (fimH27:O17/O15:H18/H4) showed the ubiquitous presence of the plasmid type pT 2-2, a pUTI89-like plasmid version highly similar to the pT 2-3 described in the specific clade of ST95. In contrast, this plasmid harbors several antibiotic resistance genes providing resistance to multiple antibiotic groups (e.g.  $\beta$ -lactams, aminoglycosides, sulfonamides and trimethoprim) and to heavy-metals (mercury) (Fig. 2). The plasmid type was estimated to be acquired around -1990 (95% CI, 1986–1992) and notably this acquisition was nested with a major expansion event occurring two nodes earlier in the tree (1987, 95% CI, 1984–1990).

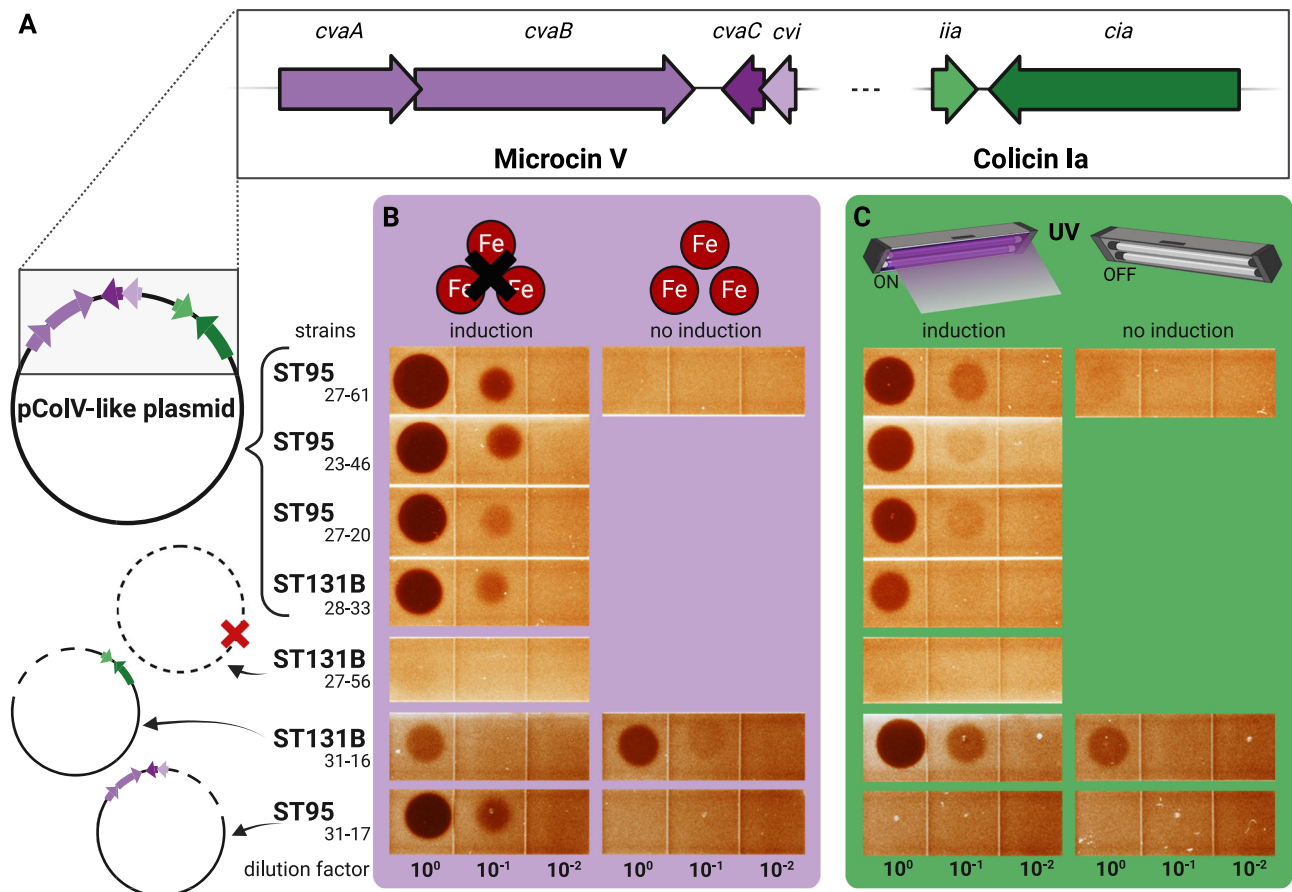
Consistent with its generally smaller inferred plasmidome size (Fig. S3D), ST73 only carries large plasmids in a minority of isolates (181/560, 32%). Out of these, the plasmid type pT 13-1 (prevalence 18%)

(Fig. S8) encoding similar virulence traits as pUTI89-like plasmids (e.g. *senB*) has been introduced multiple times into ST73. In particular, we detected a clonal expansion event in 1983 (95% CI, 1977–1989) after the plasmid had acquired the  $\beta$ -lactamase *bla*<sub>SHV-1</sub> gene, displaying increased ampicillin resistance, around 80 years after the initial pT 13-1 acquisition (1900, 95% CI 1877–1919) into the ST73 fimH10:O6:H1 clade (Supplementary Data 3, Fig. 3C). This IncF plasmid (pmlst F51:A-B10) frequently encoded several antibiotic resistance genes including  $\beta$ -lactam (*bla*<sub>SHV-1</sub>, *bla*<sub>TEM-1</sub>), aminoglycoside (*aadA1*, *aph(3')-Ib*, *aph(6)-Id*) and sulfonamide (*sulI*, *sul2*) resistance genes (Supplementary Data 1) and has been shown to be associated with a higher carriage of resistance genes within ST73<sup>48</sup>. Despite the general trend of susceptibility for this clone, this demonstrates that particular subclades of ST73 can become multi-drug resistant by incorporating this plasmid type.

In agreement with previous reports based on short-read sequencing data<sup>24,28</sup>, we observed strong plasmid-clade associations between particular IncF pTs and ST131 clades. In clade A (fimH41:O16:H5), the most basal clade of ST131, pT 1-1 was present in the MRCA of the clade (1984, 95% CI 1978–1989) and later replaced around 1991 (95% CI, 1986–1996) by pUTI89-like plasmids (pTs 2-1, 2-2). The branch where this plasmid-type replacement occurred was inferred to correspond to a clonal expansion event estimated to have occurred around 1992 (95% CI, 1988–1996), recapitulating the detected association between plasmid and clonal expansion observed for ST69.

In clade B (fimH22:O25:H4), which is generally strongly associated with antibiotic susceptibility<sup>41</sup>, we observed a similar plasmid distribution as described above for ST95. The pColV-like plasmid, pT 9-1, was introduced in a sublineage of clade B around 1955 (95% CI, 1944–1964) while the pUTI89-like plasmid (pT 2-3) entered between 1957 and





**Fig. 5 | Schematic representation of the bacteriocin susceptibility assay.**

**A** Description of bacteriocinogenic plasmids from pT 9-1 (ST131 clade B) and pT 10-1 (ST95). These archetypical pColV-like plasmids encode two bacteriocin gene clusters: microcin V and colicin Ia. Darker and lighter arrows indicate respectively the toxin (*cvaC*, *cia*) and immunity genes (*cvi*, *iia*) of each cluster. Control isolates either harbor no plasmid (ST131 clade B isolate 27-56) or harbor plasmid versions encoding only a single bacteriocin (colicin Ia for ST131 B isolate 31-16, microcin V for ST95 isolate 31-17). **B** Susceptibility of *E. coli* MG1655 to microcin V. Production of

microcin V is induced by iron limitation (simulated experimentally with the addition of the chelating agent 2,2'-bipyridyl). Bacteriocinogenic activity is demonstrated by growth inhibition in the spotted area (in a dilution-dependent way and without formation of individual plaques). **C** Susceptibility of *E. coli* MG1655 to colicin Ia. Production of colicin Ia is induced through SOS induction (stimulated experimentally with UV irradiation). Bacteriocinogenic activity is demonstrated as in **B**. Created in BioRender. Gama, J. (2025) <https://BioRender.com/e79b562>.

major STs, as well as against ST10, which is an ancestral clone ubiquitously present in the gut of mammals and avian species<sup>49,50</sup>. As expected, the bacteriocin-producing isolates mentioned above all displayed resistance to any of the corresponding four batches of microcin V, while 39 of the remaining 47 isolates were sensitive (Fig. 6, Supplementary Data 4). Sensitive isolates were generally inhibited by all four batches, except in a few cases where inhibition was only observed for the more concentrated bacteriocin batches (produced by isolates 23-46 and 27-61). Overall, ST69 and ST131 isolates were sensitive to the action of microcin V, while isolates belonging to the other STs displayed a more heterogeneous behavior. The supernatants of uninduced 27-61 and induced 27-56 (lacking colicinogenic plasmid) did not affect growth of the clinical isolates, as expected.

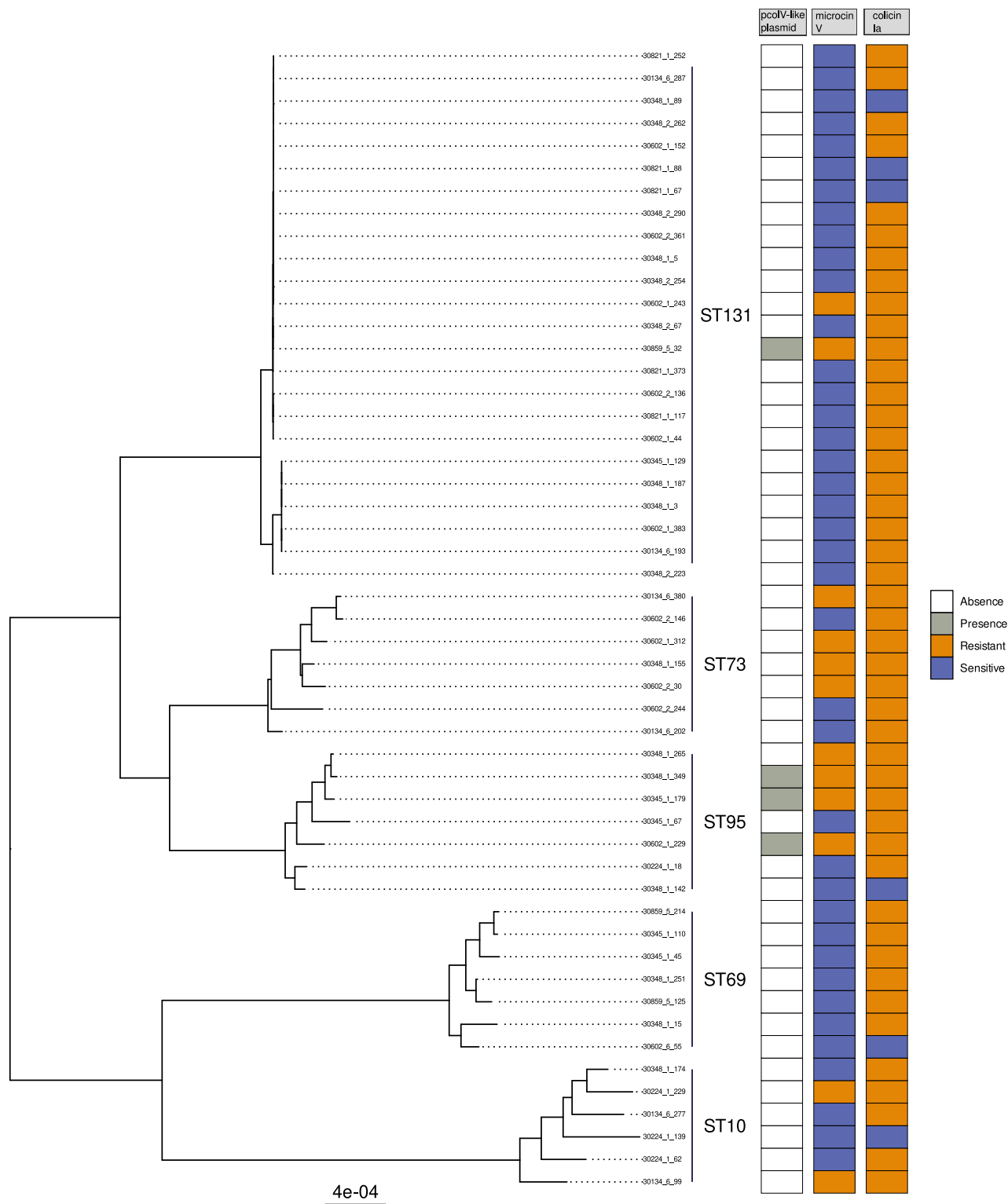
The colicin Ia gene cluster is composed of two genes (Fig. 5A): *ciaA* encoding the respective toxin, and *iia* that encodes immunity. This colicin also uses the Cir receptor, and is induced by the SOS response<sup>43</sup>. The same four plasmid-harboring isolates used above were exposed to UV irradiation to induce the SOS response and consequently production of colicin Ia. All four culture supernatants inhibited the growth of *E. coli* MG1655 in a dose-dependent manner with no observable phage plaques (Fig. 5C). As shown for microcin V, the uninduced ST95 isolate 27-61 produced a weaker inhibition zone, while the induced ST131-B isolate 27-56 not harboring a colicinogenic plasmid was not affected. Testing the four supernatants

against the 51 isolates mentioned above, only six displayed sensitivity (Fig. 6, Supplementary Data 4). Supernatants of isolates 27-56 (induced, no plasmid) and 27-61 (uninduced) did not display inhibitory activities.

Altogether, these results show that pColV-like plasmids present in ST95 and ST131-B encode functional bacteriocins, and in particular we show the microcin V gene cluster inhibits the growth of a wide range of *E. coli* isolates from different genetic backgrounds supporting their role in clone competition and possibly NFDS maintenance.

## Discussion

Plasmids are generally considered as the ultimate vehicle of rapid evolution in bacterial genomes, while also sometimes regarded as parasitic elements to their host cells<sup>14-16,51</sup>. Our analysis indicated that larger plasmids in *E. coli* are frequently associated with traits falling under bacterial competition, pathogenesis and antibiotic resistance among other traits which could not be annotated. A recent study comparing plasmids in *E. coli* and *K. pneumoniae* isolated from bloodstream infections at a single hospital in the UK with those circulating globally showed that key antibiotic resistance loci such as extended-spectrum  $\beta$ -lactamases and carbapenemases were likely to “piggyback” common successful plasmid backbones, which increases the risk of their further dissemination<sup>27</sup>. In our study even many of the small plasmids were found to incorporate likely beneficial bacteriocin-producing systems, and were scattered



**Fig. 6 | Bacteriocin susceptibility of NORM isolates.** Pruned phylogeny of the 51 *E. coli* NORM isolates tested experimentally for bacteriocin susceptibility (right). Susceptibility to microcin V and colicin Ia is individually indicated in the heatmap in

blue, while yellow indicates bacteriocin insensitivity (no effect detected). Isolates harboring the archetypical pColV-like bacteriocinogenic plasmids are indicated in gray.

around the population phylogeny, indicating their repeated horizontal acquisition, but lacking evidence of further clonal expansion. Only two co-occurring plasmid-encoded bacteriocins (microcin V and colicin Ia) were frequent enough and showed evidence of clonal expansion in different genetic backgrounds to indicate a major role in colonization competition and contribution towards maintaining stable

population equilibrium. To further examine their role in competition, we experimentally confirmed the ability of microcin V to inhibit the majority of frequently observed clones in the population that lack this system. Notably, this bacteriocin shows a stable association with lineages/clades displaying a notable degree of susceptibility to antibiotics (ST95, ST131-B) and is able to inhibit isolates from the most successful multi-drug

resistant lineages/clades (ST69, ST131-C1, ST131-C2), suggesting it is an important factor contributing to the continued maintenance of the susceptible lineages in the population. Similarly, only a single dominant chromosomal bacteriocin, the microcin H47, was identified in our collection. Characteristics of these bacteriocins thus stand in stark contrast with the substantial bacteriocin diversity previously found to contribute to the maintenance of a stable equilibrium in the *Streptococcus pneumoniae* population over time<sup>22</sup>. However, two ST10 isolates were found insensitive to microcin V, despite lacking this bacteriocin-producing system. This is well aligned with the frequently observed colonization ability of ST10 and its lower virulence potential<sup>49</sup> that was recently quantified at the population level by systematically comparing neonatal colonization rates with those observed in bloodstream infections<sup>5</sup>. Why a subset of ST10 population would remain insensitive to microcin V is currently unclear and warrants further investigation into evasion of the effects of these inhibition mechanisms.

The unique characteristics of our dataset allowed detection of multiple events of phenotypic evolution following the same path across distant lineages, associated with acquisition and stable maintenance of certain plasmid types over time. In addition to the same plasmid-based bacteriocin-producing systems observed in ST95 and ST131-B discussed above, these two distant lineages display congruent phenotypes also in other clades, that host plasmids different from the pColV-like plasmids. These correspond to the widely studied reference plasmid pUTI89, encoding virulence traits archetypal for uropathogenesis<sup>45</sup>. In both evolutionary backgrounds, these clades were non-overlapping with the clades hosting pColV-like plasmid, and have been found to be strongly associated with avian pathogenic *E. coli*<sup>25,38,52</sup>. ST131-B has been previously widely indicated as a foodborne uropathogen, in particular related to consumption of poultry meat<sup>53,54</sup>. Combined, this evidence indicates that ST95 and ST131-B lineages circulating in human hosts are stably divided into human vs. avian adapted subclades assisted by their plasmid-associated traits, and that the latter likely reflects a constant spillover from the avian niche via foodborne transmission<sup>24,25</sup>. Interestingly, the estimated timing of the pColV-like plasmid acquisition in ST131-B during 1950–60 s coincides with a rapidly intensifying poultry production in many countries after WW2, which may have provided an opportunity for *E. coli* equipped with particular plasmid-derived traits to quickly expand in this niche.

Similar to the previously discussed observations, ST69 and ST131-A display clades characterized by a shared plasmid content, which we showed arose from the pUTI89-like virulence plasmids and absorbed additional resistance genes that remain in a stable association with the virulence determinants. This again exemplifies plasmid-associated phenotypic evolution following the same trajectory in two lineages belonging to different phylogroups (D, B2). ST69 was originally detected in an outbreak in California<sup>55</sup>, and later epidemiological investigation suggested it emerged in the late 1990s and then became endemic globally within 10 years<sup>56</sup>. Our dating analysis gives the narrow time interval between 1987–1992 for the acquisition of the plasmid which ultimately led to a global dissemination of this clone with the specific virulence and antibiotic resistance characteristics. Further, as shown by our results, this evolutionary process closely reflects the parallel acquisition of the same plasmid by ST131-A subclade, which is again associated with remarkably similar epidemiological characteristics, i.e. a rapid expansion<sup>11,12</sup> and global endemicity over time. These two evolutionary trajectories across the successful lineages of *E. coli* bear the same hallmarks as the convergence of hypervirulence and multi-drug resistance in *Klebsiella pneumoniae*, which were first identified as non-overlapping traits<sup>57</sup>, but later studies pinpointed to plasmid-driven merging of the two in particular genetic backgrounds<sup>58,59</sup>.

Our observations of multiple cases of highly similar phenotypic evolution in distant genetic backgrounds by plasmids as a vehicle, suggests that they function as a significant means for new clones to first establish, and then later endemically sustain themselves in a host

population. Future studies, including mechanistic modeling of plasmid evolution and maintenance, could shed further light on the factors that clone success ultimately depends on.

Despite the strong association between specific plasmid types, we observed that some isolates across the phylogeny have lost their associated plasmid. The normalized coverage of those plasmid types in the hybrid assemblies suggests that a proportion of the cells sequenced have lost the plasmid as they are present in a copy number lower than the chromosome. Such lower coverage could be explained by deletions of large plasmid regions, which can be adaptive by rendering plasmid variants less costly relative to the original plasmid<sup>23,60</sup>. Plasmid stability within its bacterial host is dependent on the ecological context, and some of the isolates could lose the associated plasmid when cultivated in the lab environment<sup>61</sup>. In addition, some *E. coli* plasmids carrying ESBL genes have lower fitness costs for the population when present in low frequencies<sup>62</sup>. Altogether, these factors could thus hide the true prevalence of some of the plasmid types we identified in the ExPEC clones. It should further be noted that the current analysis does not prove the causal roles of the plasmids that were found to be associated with the inferred clonal expansions, despite that the combined evidence from functional characterization and population genomic modeling is strong. Indeed, predicting the stability of a plasmid in a host population is a difficult endeavor<sup>63–66</sup>. Plasmid-host associations result from multiple factors such as the level of benefit provided by the plasmid<sup>67</sup>, the ability to acquire compensatory events during co-evolution<sup>65</sup>, the pre-existence of host genotypes that favor the acquisition/maintenance of a plasmid<sup>67–70</sup>, as well as the ecological context<sup>71,72</sup>. Furthermore, the highlighted ExPEC plasmids might provide essential bacterial traits for the survival, evolution and/or competition of the distinct lineages which could result in oversampling isolates carrying those particular plasmids and potentially affecting their true prevalence in the population. Definite causal quantification of the impact of these plasmids is challenging to obtain since the conditions for their expansions in the human host population are very difficult to replicate in any experimental setting and a counterfactual analysis remains practically an impossibility in this ecological context.

## Methods

### Isolate selection for long-read sequencing

To resolve the accessory genome and mobile genetic elements of the extensive Norwegian Surveillance System for Resistant Microbes (NORM) ExPEC population<sup>11</sup>, we selected 2045 isolates (2045/3254, 62.85%) for long-read Oxford Nanopore Technologies (ONT) sequencing and, subsequently, for hybrid assembly. This selection was performed following a two-step strategy. First, to ensure that the resulting genomes represented the accessory genome diversity inherent in the NORM collection, 1085 isolates (1085/2045, 53.06%) were selected regardless of their clonal complex using an unbiased statistical approach specifically designed for selecting representative isolates for long-read sequencing<sup>31</sup> (Fig. S1A). Briefly, we considered the presence/absence matrix of the orthologous genes computed by Panaroo (version 1.2.3)<sup>73</sup> on the 3254 Illumina assemblies<sup>11</sup> and color the isolates based on their previously described PopPUNK (version 2.0.2) groups<sup>11,74</sup>. From these, we used the k-means algorithm to select 1085 isolates on the dimensionally reduced matrix computed by t-sne (perplexity = 30) (R stats package Rtsne, version 0.15) considering Jaccard distances as input.

Second, we focused on the four major ExPEC clones (ST73, ST95, ST131 and ST69) and sequenced all their remaining isolates not selected within the first step (960/2045, 46.94%). This permitted us to directly estimate the prevalence of particular plasmid types among the four major ExPEC clones.

### DNA isolation and ONT sequencing

Long-read sequencing of the 2045 isolates was performed using a high-throughput multiplexing approach based on ONT reads<sup>31</sup>. All the isolates were separately grown on MacConkey agar No. 3 (Oxoid Ltd.,

Thermo Fisher Scientific Inc., Waltham, MA, USA) at 37 °C overnight. Individual colonies were picked for overnight growth at 37 °C in 1.6 mL LB (Miller) broth (BD, Franklin Lakes, NJ, USA) each. Genomic high-molecular-weight DNA was extracted from cell pellets using MagAttract R HMW DNA Kit (Qiagen, Hilden, Germany) to a final elution volume of 100 µL. Output concentration and DNA integrity were measured using NanoDropOne spectrophotometer (Thermo Scientific) and the Qubit dsDNA HS assay kit (Thermo Fisher Scientific) on a CLARIOstar microplate reader (BMG Labtech, Ortenberg, Germany). The samples were then adjusted to 400 ng for long-read sequencing. The ONT libraries were prepared using SQK-LSK109(-XL) or SQK-NBD110-96 barcoding kit for 24- and 96-barcoding runs, respectively, and 40 fmol per sample was loaded onto FLO-MIN106 flow cells. Sequencing was run for 72 h on GridION (Oxford Nanopore Technologies, Oxford, UK) using MinKNOW Core versions 3.6.0 to 4.2.5. Basecalling and demultiplexing were performed using Guppy versions 3.2.8 to 4.3.4, with fast basecalling for the 24-barcoding and high-accuracy basecalling for the 96-barcoding runs.

### Hybrid assemblies

ONT long reads were combined with the existing short-read data<sup>11</sup> using a previously published hybrid assembly pipeline<sup>75</sup> publicly available at [https://gitlab.com/sirarredondo/hybrid\\_assembly](https://gitlab.com/sirarredondo/hybrid_assembly). This pipeline is largely based on Unicycler (version 0.4.7) and was designed to automate the generation of (near-)complete genomes.

In total, we could obtain a hybrid assembly for 1999 genomes (1999/2045, 97.75%). The information derived from Unicycler regarding the length, circularity and depth of each contig was extracted and considered for the downstream analyses. The hybrid assemblies were split into individual contigs and circlator (version 1.5.5)<sup>76</sup> was used with the command `fixstart` to rotate the starting position of each contig. A phylogeny displaying the species wide coverage of the hybrid data and complete sampling of the four major lineages was created by mapping illumina reads for each of isolates with hybrid assemblies to the ExPEC EC958 (GCF\_000285655) reference and building a SNP-sites phylogeny with IQtree<sup>77</sup>.

Each contig was classified either as chromosome-derived, plasmid-derived or virus-derived (bacteriophage) considering both the predictions of mlplasmids (species 'Escherichia coli') (version 2.1.0) and geNomad (command `end-to-end`) (version 1.5.0)<sup>78</sup>. First, contigs were labeled as chromosomal if they either had a size larger than 500 kbp or a mlplasmids chromosome probability higher than 0.7 and geNomad chromosome score higher than 0.7. Second, contigs were labeled as plasmids if they had a minimum mlplasmids plasmid probability of 0.3 and geNomad plasmid score of 0.7. These settings were used to compensate for the number of false negative predictions reported previously for mlplasmids using the *E. coli* model<sup>79</sup>. Third, contigs not meeting any of the previous criteria were classified as virus-derived if they had a minimum geNomad virus score of 0.7. Fourth, the remaining contigs were considered as unclassified.

Prokka (version 1.14.6)<sup>80</sup> using the genus *Escherichia* and species *coli* was considered to annotate each of the contigs resulting from the hybrid assemblies. Abricate (version 1.0.1) (<https://github.com/tseemann/abicate>) coupled with the databases of PlasmidFinder<sup>81</sup>, *ecoli\_vf* ([https://github.com/phac-nml/ecoli\\_vf](https://github.com/phac-nml/ecoli_vf)), and EcOH<sup>82</sup> was used to assign the presence (minimum identity and coverage of 80%) of plasmid replicon sequences, *E. coli* known virulence factors (including colicin genes) and *E. coli* O-H serotype predictions respectively. AMRFinderPlus (version 3.10.18)<sup>83</sup> using the flag `--plus` and *Escherichia* as organism was also considered to identify antibiotic resistance, stress-response and virulence genes. The presence of plasmid replicon sequences was assessed using PlasmidFinder<sup>81</sup>, while relaxases, mate-pair formation types and the predicted mobility of the contig was computed using the module `mob_typer` (version 3.0.3) of MOB-suite<sup>84,85</sup>. The latter tool was used to extract the 'primary\_cluster\_id' which clusters sequences based

on a fixed Mash distance threshold<sup>85</sup>. Plasmid multi-locus sequence typing (pMLST) was computed using the IncF, IncA/C, IncHI1, IncHI2, IncI1, IncN and pbsb1-family schemes using the `pmlst.py` script available at <https://bitbucket.org/genomicpidemiology/pmlst><sup>81</sup>. For the plasmids with an IncF scheme annotation, we reported the FAB formula to contextualize them with previous literature.

### Typing the plasmid-derived sequences using a network approach based on mge-cluster

We removed the redundancy within the set of circular plasmid sequences using `cd-hit-est` (version 4.8.1)<sup>86,87</sup> using a sequence identity threshold of 0.99, alignment coverage of 0.9 and length difference cutoff of 0.9. `Cd-hit-est` computed a total of 2560 clusters from which a single representative sequence was chosen.

The set of non-redundant plasmid sequences ( $n=2560$ ) was considered as input for performing the plasmid clustering based on `mge-cluster` (version 1.1.0)<sup>34</sup>. This tool relies on two main parameters: perplexity and minimum cluster size. To avoid choosing an arbitrary parameter selection, we explored the hyperparameter space with: (i) perplexity values ranging from 10 to 100 and (ii) minimum cluster size values ranging from 10 to 50. By selecting 100 random combinations of these two parameters, we generated 100 distinct models with `mge-cluster` using the `--create` operational mode considering a unitig filtering variance of 0.01. For obtaining a robust plasmid assignment, we only considered plasmids with a minimum membership probability of belonging to a specific cluster ('Standard Cluster') of 0.8. The distinct `mge-cluster` models were combined into a single solution using a co-occurrence network approach. In this case, the network consisted of nodes belonging to plasmids ( $n=2560$ ) and edges to indirect connections with a weight ranging from 0 (plasmids are never part of the same cluster in the solutions) to 100 (plasmids are part of the same cluster in all solutions).

This network was scanned with the Louvain algorithm available in the R `igraph` package (version 1.2.6) to detect highly-connected communities within each non-singleton graph component, establishing a modularity threshold of 0.2. These communities were considered as plasmid types (referred as pTs) and are labeled using two digits representing their network component and their inferred community bin. For each pT, we discarded any sequences differing by two standard-deviations of the average plasmid length. This criteria was included to ensure that only plasmids with a similar size were part of the same pT.

To validate and contextualize the resulting pTs, the circular and non-redundant plasmid sequences of each pT were sketched ( $k=31$ , `scaled=1000`, `noabund`) with `sourmash` (version 4.8.2)<sup>36</sup> and a containment matrix computed using the command `compare` and the flag `--containment`. This matrix contains containment values corresponding to the fraction of a particular sketch found in a second sketch. The containment values were averaged within and between pTs. For each pT, we reported the most predominant (prevalence > 0.3) (i) replicon found by PlasmidFinder, (ii) pMLST scheme, (iii) mobility (conjugative, mobilizable, non-mobilizable) as reported by MOB-suite, (iv) cluster id ('primary\_cluster\_id') reported by MOB-suite, (v) antibiotic resistance genes detected by AMRFinderPlus, and (vi) virulence genes detected by *ecoli\_vf* (part of *abicate*). This information was summarized in Supplementary Data 2.

The redundant sequences discarded by `cd-hit` ( $n=1925$ ) were assigned to the same pT as their representative sequence.

The results of the plasmid typing approach based on the co-occurrence network were compared to the 100 distinct solutions provided by `mge-cluster` using the adjusted Rand Index implemented in the R package `mclust` (version 5.4.7). The most similar solution provided by `mge-cluster` (perplexity = 62, minimum cluster size = 49) was used to visualize the clustering of the network approach by anchoring the sequences in a 2D embedding space. In addition, we used this solution in combination with the `--existing` operational mode

to predict the position in the 2D embedding space of non-circular plasmid sequences ( $n = 932$ ) and the following well-studied ExPEC plasmids: CP000244 (pUT189), CP007150 (pRS218), CU928146 (pECOS88), CU928148 (pIESCUM), DQ381420 (pAPEC-OI-ColBM), EU330199 (pVM01), LQHK01000003 (pG150\_1) and NC\_022651 (pJJ1886\_5). These sequences were assigned to the same pT as their closest (Euclidean distance) circular plasmid sequence. We only considered predictions differing by less than two standard-deviations of the average pT plasmid length.

### Plasmid synteny and phylogenetic analyses

For the pTs corresponding to large ( $>100$  kbp) IncF plasmids pTs (1-1, 1-4, 2-1, 2-2, 2-3, 8-1, 8-2, 8-3, 9-1, 10-1, 13-1, 14-1), we selected a plasmid sequence representative of the most common pMLST annotation and antibiotic resistance/virulence gene content of the cluster and with the earliest year of isolation (Supplementary Data 2). The starting position of these sequences was changed to their IncFIB replicon using circulator (version 1.5.5)<sup>76</sup> with the command `fixstart`. Clinker (version 0.0.21)<sup>88</sup> with default settings was used to perform a gene synteny analysis and visualize which genomic blocks were shared among pTs, considering a minimum sequence identity of 0.8 (Fig. 2) and 0.99 (Fig. S6) to draw a link between genes.

We performed recombination-free phylogenies using the circular plasmid sequences from pTs 8-2, 9-1, 10-1 and 2-1, 2-2, 2-3. To create a core-genome alignment from these pairs of pTs, we used snippy (version 4.6.0) <https://github.com/tseemann/snippy> which mapped the sequences from pTs 8-2, 9-1, 10-1 and 2-1, 2-2, 2-3 against the reference genomes 30348\_1#293\_2 (pT 10-1) and 30224\_1#245\_5 (pT 2-3), respectively. The resulting core genome alignment was polished using the module from snippy (`snippy-clean_full_aln`). Gubbins (version 3.1.3)<sup>89</sup> was used to compute a recombination-free phylogeny of the pTs using 100 bootstrap replicates, 50 algorithm iterations, and RAXML<sup>90</sup> as the application for model fitting and other default settings (GTRGAMMA as nucleotide substitution model). The branch lengths of the resulting trees are measured as substitutions per site multiplied by the number of recombination-filtered SNPs.

### Dating the phylogenies of the four major ExPEC clones

The four major ExPEC clones were each mapped to a reference genome belonging to that lineage and recombination was removed using Gubbins v2.4.1<sup>89</sup>. Gubbins output was supplied to the R package BactDating v1.1.1<sup>41</sup> in three replicates and one with randomized tip dates. The R CaveDive package v0.1.1 was used to detect probable expansion events in the dated phylogeny<sup>42</sup>. To report the dates of the distinct plasmid acquisition or introduction dates, we indicate in the text the lower and upper bounds of the confidence intervals (CI). For instance, pT 8-1 was acquired by ST131 from 1979 (95% CI 1971-1986) to 1984 (95% CI 1978-1989) (Supplementary Data 2). This is reported in the text as acquired around -1981 (95% CI 1971-1989).

### Determination of bacteriocin activity

We aimed to investigate whether pColV-like plasmids had a role in mediating bacterial competition, and selected four isolates carrying pColV-like plasmids (predicted to encode both microcin V and colicin Ia) for experimental work. When selecting the isolates we aimed to minimize potential confounding factors (e.g., presence of additional bacteriocins), and include strains that could serve as controls as well. The features of these isolates are summarized in Supplementary Data 4: pT of the pColV-like plasmid, bacteriocins encoded and respective genomic location.

We produced four independent batches of microcin V by growing plasmid-harboring isolates 23-46, 27-20, 27-61 (ST95) and 28-33 (ST131 clade B) for 20 h at 37 °C with aeration in 10 mL Lysogeny Broth (LB; LB Broth, Miller, Difco) with 0.2 mM 2,2'-bipyridyl (to limit iron availability; Sigma-Aldrich). The cultures were then centrifuged (4000 rpm,

10 min), filtered (0.2  $\mu$ m polyethersulfone filter; VWR), and stored at 4 °C. As controls, a plasmid-free ST131 clade B isolate 27-56 was treated in the same conditions, while isolate 27-61 was further treated without induction (no 2,2'-bipyridyl added).

To produce batches of colicin Ia, the four plasmid-harboring isolates mentioned above were grown in 10 mL LB overnight (15–16 h) at 37 °C with aeration, centrifuged (4000 rpm, 10 min) and then the pellets were resuspended in 10 mL PBS. The suspensions were UV-irradiated (254 nm wavelength) with a dose of 0.36 mW/cm<sup>2</sup>/second for 10 s. 500  $\mu$ L from these irradiated suspensions were inoculated into 9.5 mL LB and incubated 4 h at 37 °C with aeration (covered in aluminum foil to prevent photoreactivation). The cultures were then centrifuged (4000 rpm, 10 min), filtered (0.2  $\mu$ m polyethersulfone filter; VWR), and stored at 4 °C. As controls, the plasmid-free isolate 27-56 was treated in the same conditions, while isolate 27-61 was further treated without induction (no UV-irradiation).

51 NORM isolates (Supplementary Data 4), as well as *E. coli* MG1655, were grown in 1 mL LB in 96-deep-well plates overnight at 37 °C with aeration. Indicator plates were prepared by inoculating 15  $\mu$ L of each overnight culture into 15 mL LB top agar (LB with 0.75% agar; Sigma-Aldrich) with 0.1 mM 2,2'-bipyridyl (to limit iron availability and promote expression of the colicin receptor Cir). 10  $\mu$ L of each bacteriocin and control batch were spotted on indicator plates, which were incubated overnight at 30 °C after drying. These assays were performed three independent times. *E. coli* MG1655 was further exposed to 10-fold dilutions of the bacteriocin batches to verify the characteristic dose-dependent inhibition and exclude phage activity.

During pilot experiments to optimize the conditions for these assays we generally observed a lack of inhibitory effect during the assay if the indicator strain was not growing in iron limitation conditions (data not shown). We argue that the probability of the pColV-carrying isolates to encode undetected inhibitory factors that behave exactly (e.g., same induction conditions and required cell receptor) like the predicted bacteriocins would be quite rare. Especially considering the conserved patterns observed with both ST95 and ST131-B isolates. Thus we can likely exclude the effect of bacteriocins not predicted from the sequencing data. However, while most of the sensitive NORM isolates displayed sensitivity only against microcin V, the six sensitive to colicin Ia were sensitive to both bacteriocins. Because the supernatants could contain both bacteriocins (since at least one seems to be spontaneously produced at low concentrations), and they have a common receptor in target cells, we added a further control to unequivocally test the sensitivity of these isolates to each of the bacteriocins. From genomic data, a few isolates were predicted to encode only one of the bacteriocins and we used two of those (isolates 31-16 and 31-17) for further controls. These two isolates were treated, with and without specific induction for each bacteriocin, as described above. The ST95 isolate 31-17 encodes only microcin V, and produces inhibition only when grown under iron-limiting conditions (Fig. 5B). Alternatively, ST131-B isolate 31-16 encodes only colicin Ia and always produces inhibition, albeit at stronger levels upon SOS induction (Fig. 5C). Combined, this shows that the spontaneously produced bacteriocin is colicin Ia. This isolate is sensitive against the supernatant of 31-17, verifying the loss of the microcin V genes (lack of the immunity genes). Each of these two supernatants was tested against the six isolates mentioned above, and both were inhibitory to all six (as displayed in Fig. 4, Supplementary Data 4).

### Quantification and statistical analyses

**Choosing long-read isolates spanning the genomic diversity present in the NORM study.** To select isolates for long-read sequencing, we used the presence/absence matrix inferred by Panaroo (version 1.2.3) on the short-read genomes of the 3254 isolates included in the NORM study. This matrix was embedded into 2D using the R package Rtsne (version 0.15). The k-means algorithm part of the R stats package (version 3.6.3) was used to allocate 1085 centroids and the ratio of the

between-cluster sum of squares (betweenss) and the total sum of squares (totss) was considered to assess if the chosen isolates represented the diversity present in the 2D embedding.

**Hybrid assembly evaluation.** The contiguity of the hybrid assemblies was given based on the N50 value reporting the sequence length of the shortest contig representing 50% of the total assembly length. This value was reported using exclusively contigs predicted as chromosome-derived which better reflects the contiguity status of a hybrid assembly as previously reported<sup>31</sup>.

**Comparing chromosome and plasmidome sizes of ExPEC clones.** The total sequence length of contigs predicted as chromosome and plasmid (termed plasmidome) was compared among ExPEC clones using the pairwise.wilcox.test function in the R package stats (version 3.6.3) using the Benjamini-Hochberg method for accounting for multiple testing correction. In addition, the length of each plasmid contig was multiplied by its copy number as reported by Unicycler which normalizes the depth of contigs in the assembly graph to the median value. This was performed to evaluate the total number of base-pairs present in each genome devoted to plasmid sequences.

**Differentiating vertical versus horizontal plasmid transmission in the phylogenies.** The cophenetic.phylo function from R package ape (version 5.5) was used to extract pairwise distances between the pairs of tips of the ExPEC phylogenies and specific pT phylogenies. Pearson correlation implemented in the R package stats (version 3.6.3) was used to estimate the correlation between plasmid-based (recombination-free) and ExPEC phylogenies (core-genome based) distances. A positive correlation among the two distances was considered suggestive of vertical transmission of the plasmid in the ExPEC phylogeny indicating a shared evolutionary history. In contrast, the lack of correlation between the two distances was indicative of multiple acquisitions of the same pT in the ExPEC phylogeny.

**Assessing the convergence of the dated phylogenies.** The R package Bactdating (version 1.1.1) was used to date the phylogenies of the ST69, ST73, ST95 and ST131 clones. We considered three independent models (replicates) and a model with randomized dates to infer the significance of the given dates. Each model ran through Markov chain Monte Carlo (MCMC) chains of 100,000,000 generations sampled every 1000 states with a 10,000,000 burn-in using the Additive Relaxed Clock (ARC) model<sup>41</sup>. The three replicate MCMC chains were deemed to have converged with Gelman diagnostic of approximately 1 for mu, sigma and alpha using the coda R package<sup>91</sup>. We assessed whether the effective sample size (ESS) on the first replicate model was greater than 200 using the effectiveSize function of the coda R package<sup>91</sup> and determined that the true dates model was better than the randomized dates model.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The nucleotide plasmid sequences listed in Supplementary Data 1 are available at the permanent Figshare link <https://doi.org/10.6084/m9.figshare.24302884.v1>. In addition, their corresponding genome accessions are listed in Supplementary Data 5 and are publicly available at the European Nucleotide Archive under the studies IDs PRJEB45354 and PRJEB57633.

## Code availability

An Rmarkdown document with the code and raw data required to reproduce the figures and results presented in the manuscript is publicly available at [https://gitlab.com/sirarredondo/expec\\_plasmidome](https://gitlab.com/sirarredondo/expec_plasmidome).

## References

- Donnenberg, M. *Escherichia Coli: Pathotypes and Principles of Pathogenesis* (Academic Press, 2013).
- Massot, M. et al. Phylogenetic, virulence and antibiotic resistance characteristics of commensal strain populations of *Escherichia coli* from community subjects in the Paris area in 2010 and evolution over 30 years. *Microbiology* **162**, 642–650 (2016).
- Nowrouzian, F. L. et al. *Escherichia coli* B2 phylogenetic subgroups in the infant gut microbiota: predominance of uropathogenic lineages in Swedish infants and enteropathogenic lineages in Pakistani infants. *Appl. Environ. Microbiol.* **85**, e01681-19 (2019).
- Denamur, E., Clermont, O., Bonacorsi, S. & Gordon, D. The population genetics of pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **19**, 37–54 (2021).
- Mäklin, T. et al. Strong pathogen competition in neonatal gut colonisation. *Nat. Commun.* **13**, 7417 (2022).
- Riley, L. W. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin. Microbiol. Infect.* **20**, 380–390 (2014).
- Cummins, E. A., Snaith, A. E., McNally, A. & Hall, R. J. The role of potentiating mutations in the evolution of pandemic *Escherichia coli* clones. *Eur. J. Clin. Microbiol. Infect. Dis.* <https://doi.org/10.1007/s10096-021-04359-3> (2021).
- Dunn, S. J., Connor, C. & McNally, A. The evolution and transmission of multi-drug resistant *Escherichia coli* and *Klebsiella pneumoniae*: the complexity of clones and plasmids. *Curr. Opin. Microbiol.* **51**, 51–56 (2019).
- Manges, A. R. & Johnson, J. R. Food-borne origins of *Escherichia coli* causing extraintestinal infections. *Clin. Infect. Dis.* **55**, 712–719 (2012).
- Kallonen, T. et al. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res.* **27**, 1437–1449 (2017).
- Gladstone, R. A. et al. Emergence and dissemination of antimicrobial resistance in *Escherichia coli* causing bloodstream infections in Norway in 2002-17: a nationwide, longitudinal, microbial population genomic study. *Lancet Microbe* **2**, e331–e341 (2021).
- Pöntinen, A. K. et al. Modulation of multi-drug resistant clone success in *Escherichia coli* populations: a longitudinal multi-country genomic and antibiotic usage cohort study. *Lancet Microbe* **5**, E142–E150 (2024).
- McNally, A. et al. Diversification of colonization factors in a multidrug-resistant *Escherichia coli* lineage evolving under negative frequency-dependent selection. *MBio* **10** (2019).
- Watanabe, T. Infective heredity of multiple drug resistance in bacteria. *Bacteriol. Rev.* **27**, 87–115 (1963).
- Helinski, D. R. A brief history of plasmids. *EcoSal* **10**, eESP00282021 (2022).
- Rankin, D. J., Rocha, E. P. C. & Brown, S. P. What traits are carried on mobile genetic elements, and why? *Heredity* **106**, 1–10 (2011).
- Baquero, F., Lanza, V. F., Baquero, M.-R., Del Campo, R. & Bravo-Vázquez, D. A. Microcins in Enterobacteriaceae: peptide antimicrobials in the eco-active intestinal chemosphere. *Front. Microbiol.* **10**, 2261 (2019).
- Cascales, E. et al. Colicin biology. *Microbiol. Mol. Biol. Rev.* **71**, 158–229 (2007).
- Dawid, S., Roche, A. M. & Weiser, J. N. The blp bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both in vitro and in vivo. *Infect. Immun.* **75**, 443–451 (2007).
- Kjos, M. et al. Expression of *Streptococcus pneumoniae* Bacteriocins Is Induced by Antibiotics via Regulatory Interplay with the Competence System. *PLoS Pathog.* **12**, e1005422 (2016).
- Majeed, H., Gillor, O., Kerr, B. & Riley, M. A. Competitive interactions in *Escherichia coli* populations: the role of bacteriocins. *ISME J.* **5**, 71–81 (2011).

22. Corander, J. et al. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nat. Ecol. Evol.* **1**, 1950–1960 (2017).
23. Johnson Timothy, J. Role of plasmids in the ecology and evolution of ‘high-risk’ extraintestinal pathogenic *Escherichia coli* clones. *EcoSal Plus* **9**, eESP–0013–2020 (2021).
24. Johnson, T. J. et al. Separate F-type plasmids have shaped the evolution of the H30 subclone of *Escherichia coli* sequence type 131. *mSphere* **1**, (2016).
25. Cummins, M. L., Reid, C. J. & Djordjevic, S. P. F plasmid lineages in *Escherichia coli* ST95: implications for host range, antibiotic resistance, and zoonoses. *mSystems* **7**, e0121221 (2022).
26. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C. & San Millán, Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat. Rev. Microbiol.* **19**, 347–359 (2021).
27. Lipworth, S. et al. The plasmidome associated with Gram-negative bloodstream infections: a large-scale observational study using complete plasmid assemblies. *Nat. Commun.* **15**, 1612 (2024).
28. Kondratyeva, K., Salmon-Divon, M. & Navon-Venezia, S. Meta-analysis of pandemic *Escherichia coli* ST131 plasmidome proves restricted plasmid-clade associations. *Sci. Rep.* **10**, 36 (2020).
29. Mahmud, B. et al. Epidemiology of plasmid lineages mediating the spread of extended-spectrum beta-lactamases among clinical *Escherichia coli*. *mSystems* **7**, e0051922 (2022).
30. Cave, R., Ter-Stepanyan, M. M. & Mkrtchyan, H. V. Short- and long-read sequencing reveals the presence and evolution of an IncF plasmid harboring blaCTX-M-15 and blaCTX-M-27 Genes in *Escherichia coli* ST131. *Microbiol. Spectr.* **11**, e0035623 (2023).
31. Arredondo-Alonso, S. et al. A high-throughput multiplexing and selection strategy to complete bacterial genomes. *Gigascience* **10** (2021).
32. Bengtsson, S., Naseer, U., Sundsfjord, A., Kahlmeter, G. & Sundqvist, M. Sequence types and plasmid carriage of uropathogenic *Escherichia coli* devoid of phenotypically detectable resistance. *J. Antimicrob. Chemother.* **67**, 69–73 (2012).
33. Gama, J. A., Kloos, J., Johnsen, P. J. & Samuelsen, Ø. Host dependent maintenance of a blaNDM-1-encoding plasmid in clinical *Escherichia coli* isolates. *Sci. Rep.* **10**, 9332 (2020).
34. Arredondo-Alonso, S. et al. Mge-cluster: a reference-free approach for typing bacterial plasmids. *NAR Genom. Bioinform* **5**, lqad066 (2023).
35. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
36. Pierce, N. T., Irber, L., Reiter, T., Brooks, P. & Brown, C. T. Large-scale sequence comparisons with sourmash. *F1000Res.* **8**, 1006 (2019).
37. Johnson, T. J. & Nolan, L. K. Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **73**, 750–774 (2009).
38. Johnson, T. J., Siek, K. E., Johnson, S. J. & Nolan, L. K. DNA sequence of a ColV plasmid and prevalence of selected plasmid-encoded virulence genes among avian *Escherichia coli* strains. *J. Bacteriol.* **188**, 745–758 (2006).
39. Nolan, L. K. et al. Resistance to serum complement, iss, and virulence of avian *Escherichia coli*. *Vet. Res. Commun.* **27**, 101–110 (2003).
40. Rozwandowicz, M. et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J. Antimicrob. Chemother.* **73**, 1121–1137 (2018).
41. Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R. & Wilson, D. J. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* **46**, e134 (2018).
42. Helekal, D., Ledda, A., Volz, E., Wyllie, D. & Didelot, X. Bayesian inference of clonal expansions in a dated phylogeny. *Syst. Biol.* **71**, 1073–1087 (2022).
43. Jeziorowski, A. & Gordon, D. M. Evolution of microcin V and colicin Ia plasmids in *Escherichia coli*. *J. Bacteriol.* **189**, 7045–7052 (2007).
44. Chagneau, C. V. et al. HlyF, an underestimated virulence factor of uropathogenic *Escherichia coli*. *Clin. Microbiol. Infect.* <https://doi.org/10.1016/j.cmi.2023.07.024> (2023).
45. Cusumano, C. K., Hung, C. S., Chen, S. L. & Hultgren, S. J. Virulence plasmid harbored by uropathogenic *Escherichia coli* functions in acute stages of pathogenesis. *Infect. Immun.* **78**, 1457–1467 (2010).
46. Wijetunge, D. S. S. et al. Complete nucleotide sequence of pRS218, a large virulence plasmid, that augments pathogenic potential of meningitis-associated *Escherichia coli* strain RS218. *BMC Microbiol.* **14**, 203 (2014).
47. Song, N., De Greve, H., Wang, Q., Hernalsteens, J.-P. & Li, Z. Plasmid parB contributes to uropathogenic *Escherichia coli* colonization in vivo by acting on biofilm formation and global gene regulation. *Front. Mol. Biosci.* **9**, 1053888 (2022).
48. Li, D. et al. Dominance of *Escherichia coli* sequence types ST73, ST95, ST127 and ST131 in Australian urine isolates: a genomic analysis of antimicrobial resistance and virulence linked to F plasmids. *Microb. Genom* **9** (2023).
49. Köhler, C.-D. & Dobrindt, U. What defines extraintestinal pathogenic *Escherichia coli*? *Int. J. Med. Microbiol.* **301**, 642–647 (2011).
50. Bojesen, A. M., Ahmed, U., Skaarup, H. & Espinosa-Gongora, C. Recurring outbreaks by the same *Escherichia coli* ST10 clone in a broiler unit during 18 months. *Vet. Res.* **53**, 2 (2022).
51. Bergstrom, C. T., Lipsitch, M. & Levin, B. R. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**, 1505–1519 (2000).
52. Johnson, T. J., Johnson, S. J. & Nolan, L. K. Complete DNA sequence of a ColBM plasmid from avian pathogenic *Escherichia coli* suggests that it evolved from closely related ColV virulence plasmids. *J. Bacteriol.* **188**, 5975–5983 (2006).
53. Liu, C. M. et al. *Escherichia coli* ST131-H22 as a Foodborne Uropathogen. *MBio* **9** (2018).
54. Liu, C. M. et al. Using source-associated mobile genetic elements to identify zoonotic extraintestinal *E. coli* infections. *One Health* **16**, 100518 (2023).
55. Manges, A. R. et al. Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. *N. Engl. J. Med.* **345**, 1007–1013 (2001).
56. Johnson, J. R. et al. Global distribution and epidemiologic associations of *Escherichia coli* clonal group A, 1998–2007. *Emerg. Infect. Dis.* **17**, 2001–2009 (2011).
57. Holt, K. E. et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc. Natl Acad. Sci. USA* **112**, E3574–E3581 (2015).
58. Lam, M. M. C. et al. Convergence of virulence and MDR in a single plasmid vector in MDR *Klebsiella pneumoniae* ST15. *J. Antimicrob. Chemother.* **74**, 1218–1222 (2019).
59. Wyres, K. L. et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive *Klebsiella pneumoniae* from South and Southeast Asia. *Genome Med.* **12**, 11 (2020).
60. Zhang, X., Deatherage, D. E., Zheng, H., Georgoulis, S. J. & Barrick, J. E. Evolution of satellite plasmids can prolong the maintenance of newly acquired accessory genes in bacteria. *Nat. Commun.* **10**, 5809 (2019).
61. Hall, J. P. J., Wright, R. C. T., Guymer, D., Harrison, E. & Brockhurst, M. A. Extremely fast amelioration of plasmid fitness costs by multiple functionally diverse pathways. *Microbiology* **166**, 56–62 (2020).
62. Dimitriu, T. et al. Negative frequency dependent selection on plasmid carriage and low fitness costs maintain extended spectrum  $\beta$ -lactamases in *Escherichia coli*. *Sci. Rep.* **9**, 17211 (2019).

63. De Gelder, L., Ponciano, J. M., Joyce, P. & Top, E. M. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiology* **153**, 452–463 (2007).
64. Turner, P. E. et al. Antibiotic resistance correlates with transmission in plasmid evolution. *Evolution* **68**, 3368–3380 (2014).
65. Porse, A., Schønning, K., Munck, C. & Sommer, M. O. A. Survival and evolution of a large multidrug resistance plasmid in new clinical bacterial hosts. *Mol. Biol. Evol.* **33**, 2860–2873 (2016).
66. Benz, F. & Hall, A. R. Host-specific plasmid evolution explains the variable spread of clinical antibiotic-resistance plasmids. *Proc. Natl Acad. Sci. USA* **120**, e2212147120 (2023).
67. Alonso-Del Valle, A. et al. Antimicrobial resistance level and conjugation permissiveness shape plasmid distribution in clinical enterobacteria. *Proc. Natl Acad. Sci. USA* **120**, e2314135120 (2023).
68. Kloos, J., Gama, J. A., Hegstad, J., Samuelsen, Ø. & Johnsen, P. J. Piggybacking on niche adaptation improves the maintenance of multidrug-resistance plasmids. *Mol. Biol. Evol.* **38**, 3188–3201 (2021).
69. Sota, M. & Top, E. M. Host-specific factors determine the persistence of IncP-1 plasmids. *World J. Microbiol. Biotechnol.* **24**, 1951–1954 (2008).
70. Low, W. W. et al. Mating pair stabilization mediates bacterial conjugation species specificity. *Nat. Microbiol.* **7**, 1016–1027 (2022).
71. Ma, K., Feng, Y. & Zong, Z. Fitness cost of a mcr-1-carrying IncHI2 plasmid. *PLoS ONE* **13**, e0209706 (2018).
72. Bottery, M. J. Ecological dynamics of plasmid transfer and persistence in microbial communities. *Curr. Opin. Microbiol.* **68**, 102152 (2022).
73. Tonkin-Hill, G. et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
74. Lees, J. A. et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res.* **29**, 304–316 (2019).
75. Pöntinen, A. K. et al. Apparent nosocomial adaptation of *Enterococcus faecalis* predates the modern hospital era. *Nat. Commun.* **12**, 1523 (2021).
76. Hunt, M. et al. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* **16**, 294 (2015).
77. Nguyen, L.-M. et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
78. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2023).
79. Paganini, J. A., Plantinga, N. L., Arredondo-Alonso, S., Willems, R. J. L. & Schürch, A. C. Recovering *Escherichia coli* plasmids in the absence of long-read sequencing data. *Microorganisms* **9** (2021).
80. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
81. Carattoli, A. et al. In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
82. Ingle, D. J. et al. In silico serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Micro. Genom.* **2**, e000064 (2016).
83. Feldgarden, M. et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728 (2021).
84. Robertson, J. & Nash, J. H. E. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb. Genom.* **4** (2018).
85. Robertson, J., Bessonov, K., Schonfeld, J. & Nash, J. H. E. Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. *Microb. Genom.* **6** (2020).
86. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
87. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
88. Gilchrist, C. L. M. & Chooi, Y.-H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btab007> (2021).
89. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
90. Stamatakis, A. RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
91. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R. N.* **6**, 7–11 (2006).

## Acknowledgements

The authors acknowledge the support from the Genomic Support Centre Tromsø, UiT The Arctic University of Norway for Oxford Nanopore sequencing and François Cléon for excellent technical assistance. The project was funded by the Trond Mohn Foundation (grant identifier TMS2019TMT04 to A.K.P., R.A.G., Ø.S., P.J.J., and J.C.). The presented work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions (grant No. 801,133 to S.A.-A. and A.K.P.) and has also been supported by the European Research Council (grant No. 742158 to J.C.).

## Author contributions

J.C., P.J., Ø.S. designed and sought funding for the study. A.P. performed bacteria culturing, DNA extraction and quality control for long-read sequencing. S.A.-A. performed the hybrid assemblies, and plasmid clustering, statistical analyses. R.G. performed the computational work corresponding to the phylogenetic dating and clonal expansion assessment. H.T. and G.T.H. assisted in genomic analyses and plasmid clustering. J.A.G. performed the experimental and functional assays on plasmid-encoded bacteriocins, K.H. supervised these experiments. J.C., P.J., Ø.S., S.A.-A., A.P., J.A.G., R.G., G.S. wrote the manuscript which was further improved and revised by all named authors. All Norwegian *E. coli* BSI Study Group members contributed isolates for the study.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-57940-1>.

**Correspondence** and requests for materials should be addressed to Jukka Corander.

**Peer review information** *Nature Communications* thanks Alvaro San Millan and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

---

## Norwegian E. coli BSI Study Group

---

**Nina Handal<sup>11</sup>, Nils Olav Hermansen<sup>12</sup>, Anita Kanestrøm<sup>13</sup>, Hege Elisabeth Larsen<sup>14</sup>, Paul Christoffer Lindemann<sup>15</sup>, Iren Høyland Löhr<sup>16</sup>, Åshild Marvik<sup>17</sup>, Einar Nilsen<sup>18</sup>, Marcela Zamudio<sup>19</sup>, Elisabeth Sirnes<sup>20</sup>, Ståle Tofteland<sup>21</sup> & Kyriakos Zaragkoulias<sup>22</sup>**

---

<sup>11</sup>Akershus University Hospital, Dept. of Microbiology, Lørenskog, Norway. <sup>12</sup>Oslo University Hospital Ullevål, Dept. of Microbiology, Oslo, Norway. <sup>13</sup>Østfold Hospital, Centre for Laboratory Medicine, Sections for Microbiology, Kalnes, Norway. <sup>14</sup>Nordland Hospital, Dept. of Microbiology, Bodø, Norway. <sup>15</sup>Haukeland University Hospital, Dept. of Microbiology, Bergen, Norway. <sup>16</sup>Stavanger University Hospital, Dept. of Medical Microbiology, Stavanger, Norway. <sup>17</sup>Vestfold Hospital, Dept. of Microbiology, Tønsberg, Norway. <sup>18</sup>Molde Hospital and Ålesund Hospital, Møre and Romsdal Hospital Trust, Molde, Norway. <sup>19</sup>Oslo University Hospital Rikshospitalet, Institute of Medical Microbiology, Oslo, Norway. <sup>20</sup>Førde Hospital, Førde Health Trust, Førde, Norway. <sup>21</sup>Sørlandet Hospital, Dept. of Medical Microbiology, Kristiansand, Norway. <sup>22</sup>Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway.