

Parallel loss of type VI secretion systems in two multi-drug-resistant *Escherichia coli* lineages

Elizabeth A. Cummins¹, Robert A. Moran¹, Ann E. Snaith¹, Rebecca J. Hall¹, Christopher H. Connor², Steven J. Dunn¹ and Alan McNally^{1,*}

Abstract

The repeated emergence of multi-drug-resistant (MDR) *Escherichia coli* clones is a threat to public health globally. In recent work, drug-resistant *E. coli* were shown to be capable of displacing commensal *E. coli* in the human gut. Given the rapid colonization observed in travel studies, it is possible that the presence of a type VI secretion system (T6SS) may be responsible for the rapid competitive advantage of drug-resistant *E. coli* clones. We employed large-scale genomic approaches to investigate this hypothesis. First, we searched for T6SS genes across a curated dataset of over 20000 genomes representing the full phylogenetic diversity of *E. coli*. This revealed large, non-phylogenetic variation in the presence of T6SS genes. No association was found between T6SS gene carriage and MDR lineages. However, multiple clades containing MDR clones have lost essential structural T6SS genes. We characterized the T6SS loci of ST410 and ST131 and identified specific recombination and insertion events responsible for the parallel loss of essential T6SS genes in two MDR clones.

DATA SUMMARY

The genome sequence data generated in this study are publicly available from NCBI under BioProject PRJNA943186, alongside a complete assembly in GenBank under accessions CP120633–CP120634. All other sequence data used in this paper have been taken from ENA with the appropriate accession numbers listed within the Methods. The *Escherichia coli* genome data sets used in this work are from a previous publication, the details of which can be found in the corresponding supplementary data files 10.6084/m9.figshare.21360108 [1].

INTRODUCTION

Escherichia coli has been ranked globally as the number one causal pathogen of deaths associated with bacterial antimicrobial resistance (AMR) [2] and AMR has been declared by the World Health Organisation as a top ten global public health threat (www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance). The emergence and evolution of multi-drug-resistant (MDR) *E. coli* is a pressing and relevant issue for global healthcare as their extensive resistance profiles result in diminishing therapeutic options for treating infections.

A trait increasingly ascribed to MDR *E. coli* lineages is the ability to rapidly, and asymptotically, colonize the human intestinal tract [3]. Multiple travel studies have now shown that people travelling from areas of low AMR incidence to AMR-endemic regions become colonized by extended-spectrum beta-lactam-resistant or MDR *E. coli* during travel [4–6]. Furthermore, genomic analysis has shown that the gain of MDR *E. coli* was due to the acquisition of a new MDR strain and not the preceding commensal *E. coli* becoming MDR [5]. The ability to displace and colonize may be attributed to the MDR phenotype itself, but longitudinal studies from the UK and Norway [7, 8] have shown that multi-drug resistance alone is not a sufficient driver for epidemiological

Received 18 July 2023; Accepted 31 October 2023; Published 16 November 2023

Author affiliations: ¹Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK; ²Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne 3000, Australia.

*Correspondence: Alan McNally, a.mcnally.1@bham.ac.uk

Keywords: type VI secretion system; *Escherichia coli*; multi-drug resistance; ST410; ST131.

Abbreviations: AMR, antimicrobial resistance; CC, clonal complex; HGT, horizontal gene transfer; IS, insertion sequence; MDR, multi-drug-resistant; ML, maximum-likelihood; ST, sequence type; TSD, target site duplication; T6SS, type VI secretion system.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Three supplementary figures and one data sheet are available with the online version of this article.

001133 © 2023 The Authors

 This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

Impact Statement

Escherichia coli is a globally significant pathogen that causes the majority of urinary tract infections. Treatment of these infections is exacerbated by increasing levels of drug resistance. Pandemic multi-drug-resistant (MDR) clones, such as ST131-C2/H30Rx, contribute significantly to global disease burden. MDR *E. coli* clones are able to colonize the human gut and displace the resident commensal *E. coli*. It is important to understand how this process occurs to better understand why these pathogens are so successful. Type VI secretion systems (T6SSs) may be one of the antagonistic systems employed by *E. coli* in this process. Our findings provide the first detailed characterization of the T6SS loci in ST410 and ST131 and shed light on events in the evolutionary pathways of the prominent MDR pathogens ST410-B4/H42RxC and ST131-C2/H30Rx.

success of *E. coli*. Recent metagenomic analysis has also revealed that colonization by MDR *E. coli* does not disrupt the wider gut microbiome's composition or diversity [9].

One possible reason for the ability of drug-resistant *E. coli* to displace resident commensal *E. coli* so rapidly is a result of the drug-resistant *E. coli* possessing a type VI secretion system (T6SS) that allows contact-dependent killing of the resident commensal *E. coli*. The T6SS is a multi-functional apparatus that some Gram-negative bacteria possess to facilitate nutrient uptake [10], manipulation of host cells [11] and the killing of competing bacteria [12–15]. T6SS distribution varies by environment [16, 17] and the presence of T6SSs varies on all taxonomic levels [16, 18, 19] with over-representation in the *Gammaproteobacteria* [20]. T6SS genes can be gained and lost via horizontal gene transfer (HGT) [12, 18] and T6SS presence may be influenced by genomic incompatibilities, either between mobile genetic elements, donor and recipient bacteria, or via assembly of multiple T6SSs [16]. T6SSs are particularly prevalent in complex microbial communities that are host-associated [17, 18] and are commonly found in pathogens, including *E. coli* [16, 21–23].

In *E. coli* T6SSs are classified into three distinct phylogenetic groups (T6SS-1 to T6SS-3) and all have been shown to be directly involved in pathogenesis and antibacterial activity [21]. Sub-types T6SS-1 and T6SS-2 are more commonly found in chromosomes than T6SS-3. The majority of genes encode structural components, alongside *vgrG* that encodes a spike protein and regions of variable effector proteins, some of which include recombination hotspot (*rhs*) genes (T6SS-2 only). Research on *E. coli* T6SSs have often been pathotype-specific [22–24]. We instead chose to focus on MDR lineages of *E. coli* due to their clinical relevance and the current lack of knowledge in this area. For instance, the T6SS of MDR ST131 has not been characterized, despite its global prevalence and dominance in clinical settings. The existing connections between T6SS and multi-drug resistance in *Acinetobacter baumannii* [25, 26] makes T6SSs in *E. coli* an intriguing avenue of exploration in the context of pandemic MDR clones and how they become successful. Here, we assess the prevalence of T6SS genes across *E. coli* and identify no association with MDR or indeed any given lineages. We characterize the T6SS loci within MDR lineages ST410 and ST131 and show that successful MDR clones ST410-B4/H24RxC and ST131-C2/H30Rx have lost a functional T6SS through deletion and insertion events respectively.

METHODS**Genome collection**

The 20577 *E. coli* genomes used here were curated from Enterobase [27] and were fully characterized and described in a previous study [1]. Briefly, they were chosen to represent the phylogenetic, genotypic and phenotypic diversity of the species. We include commensal, pathogenic and generalist lineages and a spectrum of resistance profiles from susceptible to highly resistant. The data set is sampled from a variety of source niches. The assemblies covered six phylogroups and 21 different sequence types (STs) of *E. coli*: ST3, ST10, ST11, ST12, ST14, ST17, ST21, ST28, ST38, ST69, ST73, ST95, ST117, ST127, ST131, ST141, ST144, ST167, ST372, ST410 and ST648. Enterobase employs quality filters when adding draft assemblies to the database: ≤ 800 contigs, $>70\%$ contigs assigned to species using Kraken, genome length 3.7–6.4 Mb and a minimum N50 value of 20 kb [27].

T6SS database creation and interrogation

The SecReT6 database is an open-access archive of known and putative T6SS proteins [28, 29]. The experimentally validated SecReT6 database was retrieved and reformatted so that sequence identifiers were a suitable format for a custom ABRicate database [github.com/lillycummins/ReformattedSecReT6]. ABRicate (v0.9.8) was run using the default 75% minimum nucleotide sequence identity to blast all 20577 genomes individually against the experimentally validated T6SS gene database to assess presence across our 21 *E. coli* STs. The results were combined into a single csv file using the abricate summary command, and partial hits (instances where a gene hit was split over multiple contigs) were accounted for and processed with a custom Python script [https://github.com/lillycummins/InterPangenome/blob/main/process_partial_hits.py]. We utilized methodology identical to a previous study by our group studying the distribution of T6SS across *Klebsiella pneumoniae* [30]. A gene is determined as present in a genome if a hit exceeded 85% coverage with a nucleotide sequence identity above the default threshold of 75%. Any matches to genes

annotated as effector or immunity protein-encoding were removed from the search results. These presence–absence patterns alongside N50 value, length and number of contigs per genome are available within the Supplementary Data. Phylogenies for each lineage were built using raxmlHPC-PTHREADS-AVX(v8.2.12) with the GTRGAMMA model, rapid bootstrap analysis and search for best-scoring maximum-likelihood (ML) tree in one program run.

Genomic screening for insertion sequences and deletion events

Gene Construction Kit (v4.5.1) (Textco Biosoftware) was used to annotate and visualize DNA sequences. Insertion sequences (ISs) were identified using the ISFinder database [31]. Target site duplications (TSDs) flanking ISs were identified manually. Two junction sequences specific for a given insertion were generated by taking 100 bp contiguous sequences that span the left and right ends of the insertion. Each junction sequence comprised 50 bp of the target IS and 50 bp of its adjacent sequence. To generate an ancestral sequence *in silico*, the interrupting IS and one copy of its associated TSD were removed manually. A 100 bp sequence that spanned the insertion point was taken to represent the naive site. A database of all ST131 genomes in the collection was generated for screening with standalone BLASTn [32]. The database was queried with insertion-junction sequences and the naive site sequence to determine whether an insertion was present, with only complete and identical matches to the 100 bp query sequences considered positive matches. The same 100 bp indicator sequence method was used to identify deletion events.

Genome sequencing

To the best of our knowledge, no complete reference genome for ST131-A/H41 was publicly available so we generated sequence data for the strain USVAST219 [33]. DNA was extracted using the Monarch Genomic DNA Extraction Kit (New England Biolabs) before sequencing with MinION (Oxford Nanopore Technology) using the ligation kit LSK-SQK109 and R9.4.1 flow cells. The data were basecalled with Guppy (v6.0.1) (<https://github.com/nanoporetech/pyguppyclient>) and adapters were trimmed with qcat (v1.1.0) (<https://github.com/nanoporetech/qcat>). Short-read genome sequencing was provided by MicrobesNG (microbesng.com) where DNA was prepared using the Nextera XT kit (Illumina) and sequenced on the Illumina HiSeq platform (Illumina). Long- and short-read data are both available within BioProject PRJNA943186. A hybrid assembly using both the long- and short-read sequencing data was created with Unicycler (v0.4.8) [34] using the default settings.

RESULTS

Structural T6SS gene presence varies between and within phylogroups and STs

All 20577 *E. coli* genomes interrogated contained at least one structural T6SS gene from the SecReT6 experimental database [28, 29]. There was variation in the average number of T6SS genes present per ST and the average number of genes per ST within phylogroups (Fig. 1). ST410 exhibited the largest range (0–17) of T6SS genes and the highest number of T6SS genes in a single genome ($n=17$) alongside ST17, ST12, ST127, ST14 and ST117. The highest average number of T6SS genes ($n=17$) was found in B2 lineages ST12, ST127 and ST14. The distribution of the number of T6SS genes present per genome did not align with phylogroups. STs known to contain MDR lineages, ST131, ST167, ST410 and ST648, did not possess a similar range or average of T6SS genes (Fig. 1). A heatmap of gene distributions across STs is shown in Fig. S1 (available in the online version of this article). Grouping genes into functional units is therefore necessary to gain insight into how many potentially functional T6SSs are present within STs and whether this may correlate with multi-drug resistance.

Presence of structural T6SS genes in lineages containing MDR clones

We examined T6SS gene units in STs containing MDR clones in greater detail. The MDR clones ST131-C2/H30Rx and ST410-B4/H24RxC were selected for further investigation due to their global distribution and clinical burden [7, 8, 35, 36]. Fig. 2 displays the presence of structural T6SS genes within ST410 (Fig. 2a) and ST131 (Fig. 2b) with the corresponding clades highlighted on the phylogenies. Structural genes were the focus of this search because they are well characterized and vary less in comparison to effector/immunity proteins [21]. In both STs, we found a correlation between clades possessing MDR clones and the absence of structural T6SS genes (Fig. 2). The absence of T6SS genes does not correlate with genome size or number of contigs for either ST (Fig. S2).

Presence of structural T6SS genes in ST410

Within ST410, *tssABCDEFGHIJKLM* (henceforth referred to as the *tss* region) was present in 94.59% ($n=34$) of clade A/H53 (Fig. 2a). The organization of genes in this region classifies the system as T6SS-2. Of the three genomes that did not contain a full *tss* region, one genome was only missing *tssH* due to poor sequence coverage (coverage <10× for the three contigs spanning *tssH*). In the two other genomes only *vgrG* was missing, caused by a combination of low coverage and the presence of multiple *vgrG* genes, which is known to lead to sequence fragmentation. We conclude that the native T6SS of ST410-A/H53 is conserved across the clade.

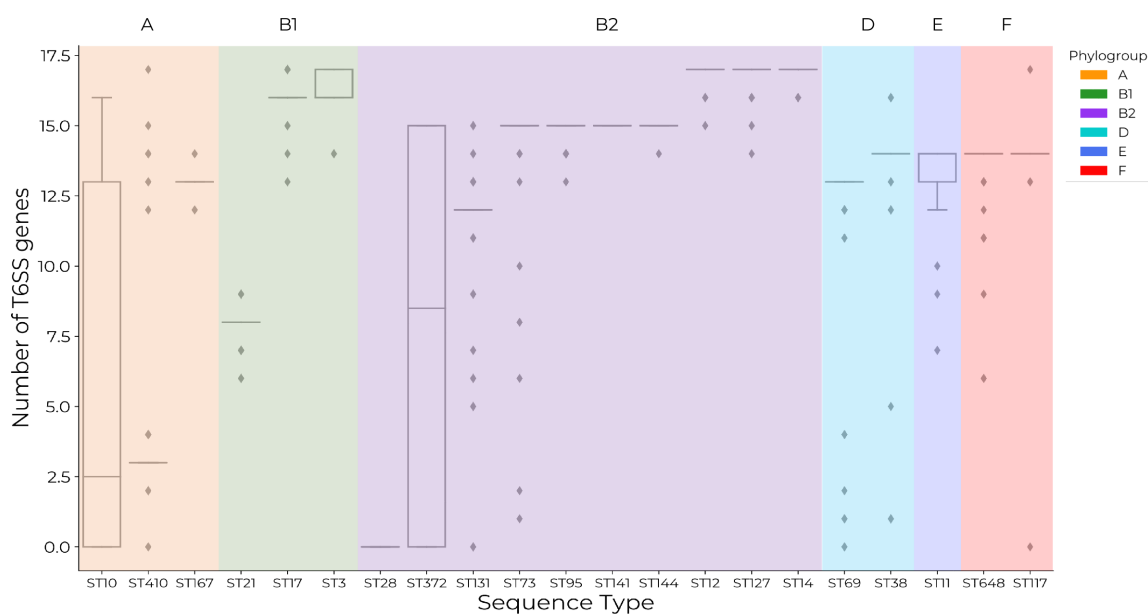


Fig. 1. Distributions of the number of structural type VI secretion system (T6SS) genes from the SecReT6 experimentally validated database [28, 29] across 21 sequence types of *E. coli* show that gene carriage varies by sequence type and phylogroup. Sequence types are colour coded according to their respective phylogroups: orange=A; green=B1; purple=B2; light blue=D; blue=E; red=F. A gene is determined as present if it exceeds a DNA sequence identity of 75% and a coverage of 85%.

The *tss* region was present in just 0.83% ($n=8$) of clade B/H24 genomes. Two ST410-B/H24 genomes did not contain *tssI/vgrG* and four genomes did not contain *tssLM*, but these otherwise contained all *tss* region genes. While isolates from clade A/H53 are largely drug-susceptible, clade B/H24 contains the globally distributed MDR clone ST410-B4/H24RxC, which is highlighted in Fig. 2(a) [34, 36]. In the eight rare instances where a full *tss* region was present within the B/H24 clade, five genomes contained *tss* regions most closely related ($\geq 98\%$ nucleotide identity) to that of the T6SS-2 region of F-type plasmid pSTEC299_1 (GenBank accession: CP022280) when compared to the NCBI non-redundant database. Two of the remaining three genomes contained *tss* regions that were closely related ($\geq 98\%$ nucleotide identity) to chromosomal T6SS-1 regions from other *E. coli* (GenBank accessions: CP062901, CP091020). These findings together suggest that, although rare, multiple acquisitions of distinct *tss* regions have occurred within ST410 clade B/H24.

Characterization of the ST410 type VI secretion locus

We resolved the structure of the complete 28.3 kb T6SS locus of ST410-A/H53 using a draft ST410-A/H53 assembly (Enterobase assembly barcode: LB4500AA_AS), partially scaffolded (Fig. S3) using the complete genome of ST88 *E. coli* RHBSTW-00313 (GenBank accession: CP056618). ST88 belongs to the same clonal complex (CC) as ST410 (CC23) and was used because there is no complete reference genome for ST410-A/H53 and we were unable to access a strain to generate a closed genome. The complete T6SS region contained 18 ORFs, 16 of which could be assigned functions (Fig. 3). Among these were representatives of all 13 *tss* genes required for synthesis of a type VI secretion apparatus (*tssA* to *tssM*), along with two additional different *tssA* and *tssD* genes. The recombination hotspot *rhsG* gene, which is known to encode a toxin [37], is located downstream of *vgrG* and is the probable determinant for this system's effector protein. The locus also contains a gene for an FHA-domain-containing protein, which is known to be involved in T6SS activation in *Pseudomonas aeruginosa* [38]. The structure of this region is indicative of a T6SS-2 type system [21] and we conclude that this region of the ST410-A/H53 chromosome includes all determinants necessary for the production of a functional T6SS.

Deletion events account for the loss of T6SS in ST410-B/H24

To explain the absence of T6SS genes in ST410-B/H24, we annotated the chromosomal region of reference strain SCEC020026 (GenBank accession CP056618) that corresponds to where the *tss* region characterized in ST410-A/H53 is located. Adjacent to the *tssA2* and *tssD2* genes, which are present in the majority of ST410-B/H24 genomes (Fig. 4a), we found remnants of *rhsG* and *tssM* (Fig. 4b). Relative to the complete *tss* region in clade A/H53, clade B/H24 has lost 21.0 kb between *rhsG* and *tssM* in a deletion event. As there are no mobile genetic elements present at the deletion site, which is located within a recombination hotspot, the deletion event was probably the result of recombination. To determine whether this deletion event was responsible for the loss of the *tss* region in other ST410-B/H24 isolates, we queried their genomes with the unique 100 bp sequence found at

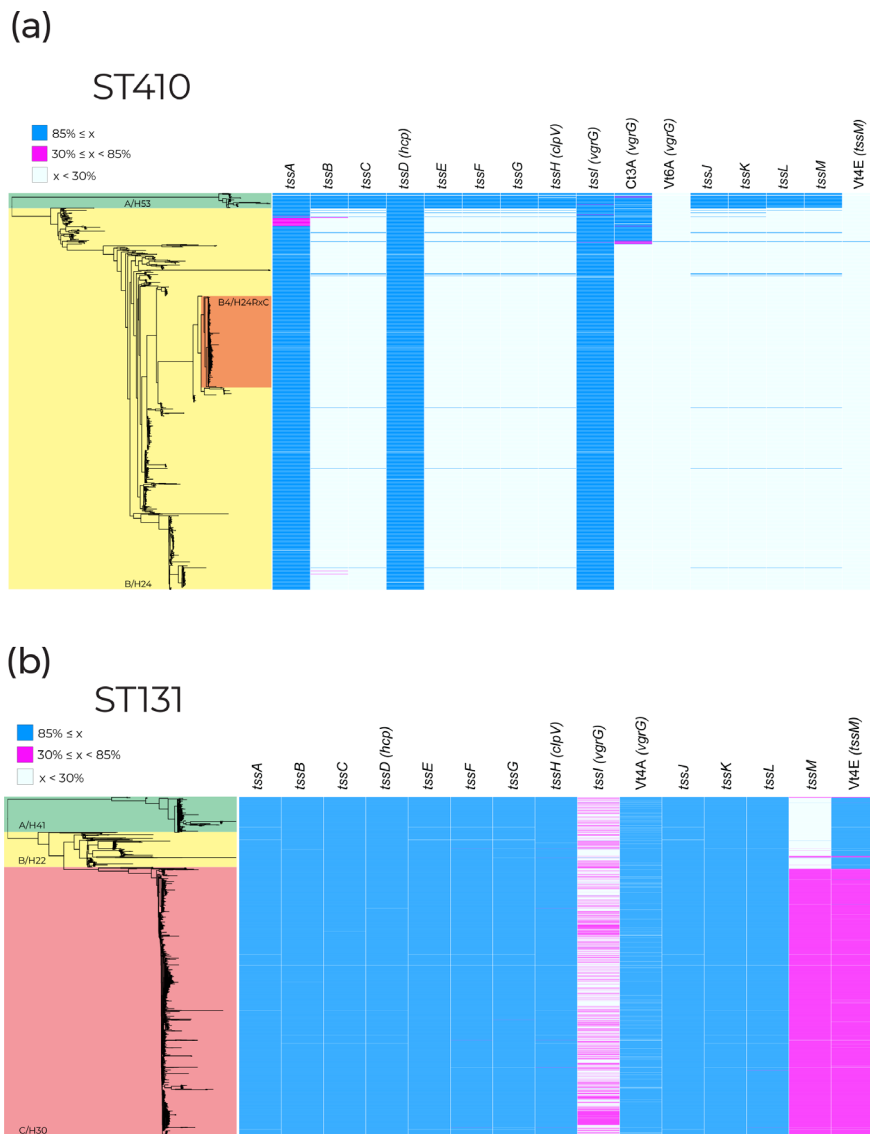


Fig. 2. Presence of structural experimentally validated type VI secretion genes taken from the SecReT6 database [28, 29] across (a) 1006 ST410 genomes and (b) 3186 ST131 genomes. Phylogenetic clades are labelled clade name/*fimH* allele. Blue blocks represent gene coverage values of >85%, which was used as a cut-off to indicate gene presence. White blocks represent gene coverage values of <30%, indicating gene absence. Pink blocks denote sequences with coverage values of 30–85%; these intermediate values are indicative of partial sequences that are not expected to represent functional genes. ST131 and ST410 clade A: green; ST131 and ST410 clade B: yellow; MDR subclade ST410 B4/H24RxC: orange; ST131 clade C: red.

the *rhsG-tssM* junction in SCEC020026. The junction sequence (G–M) was present in the majority of ST410-B/H24 genomes ($n=941$ out of 969; Fig. 4a), indicating that this deletion event occurred as a single event early in the evolution of ST410-B/H24 and not as multiple independent events within this clade.

Of the 28 ST410-B/H24 genomes that did not contain the *rhsG-tssM* junction sequence, 20 were clustered in the phylogeny and also lacked *tssA* (Fig. 4a). This suggests that a different or additional deletion event may have occurred in a small sub-clade of ST410-B/H24. To determine this, we annotated a contig from one of these 20 genomes (EnteroBase assembly barcode: ESC_HA8479AA_AS), which contained *tssD2* and exhibited a different *tss* region configuration. To confirm the structure of this region, we used the ESC_HA8479AA_AS *tss* contig to query GenBank and found an identical sequence in a complete ST410 genome (strain E94, accession: CP199740) that had been published after our data set was assembled. We then used the E94 genome as a representative for the 20 clustered genomes. This revealed the absence of a 19.8 kb segment of the *tss* region between *tssD1* and *tssA2* (Fig. 4c). The presence of the same *tssD-tssA* junction sequence (D–A) in all 20 clustered genomes indicated that this deletion event was responsible for the absence of *tss* genes in this sub-clade of ST410-B/H24.

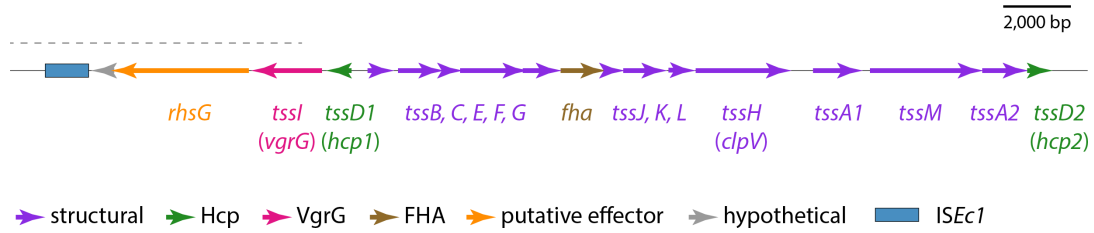


Fig. 3. The ST410-A/H53 chromosomal T6SS locus. The extents and orientations of ORFs are indicated by labelled arrows, with colours corresponding to protein types as outlined in the key below. The part of the sequence for which sequence from ST88 was used as a scaffold is indicated by the dotted black line above. FHA: forkhead-associated. Drawn to scale from EnteroBase assembly barcode: LB4500AA_AS, and GenBank accession: CP056618.

Presence of structural T6SS genes in ST131

Within ST131, the ancestral and less drug-resistant clades A/H41 and B/H22 possess a complete gene labelled Vt4E from the SecReT6 database, which is not complete in clade C/H30 (Fig. 2b). To deduce the function of Vt4E, we used BLASTn [31] to search the GenBank non-redundant nucleotide database, which revealed that Vt4E shared 100% sequence identity with *vasK*. The *vasK* gene is a homologue of *tssM/icmF* [37, 39] which is a component of the inner membrane complex of T6SSs [21]. The absence of a functional Vsk protein has been shown to render the T6SS non-functional in *Vibrio cholerae* [39].

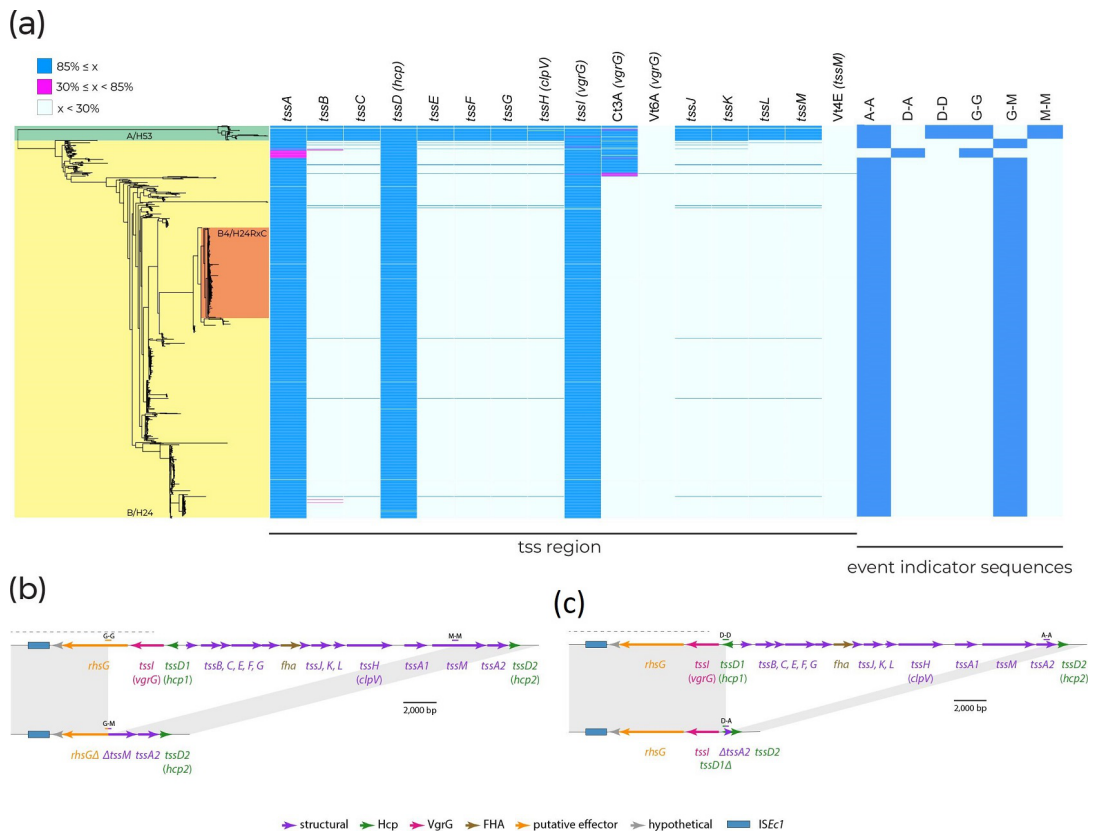


Fig. 4. (a) ST410 phylogeny. Genomes are shaded according to their clade and *fimH* allele. The presence of *tss* region genes and 100 bp deletion event indicator sequences is shown by blue shading to the right of the phylogeny where only 100% matches are shown. Blue blocks represent gene coverage values of >85%, which was used as a cut-off to indicate gene presence. White blocks represent gene coverage values of <30%, indicating gene absence. Pink blocks denote sequences with coverage values of 30–85%; these intermediate values are indicative of partial sequences that are not expected to represent functional genes. (b) Deletion of a 21.0 kb segment of the *tss* region between *rhsG* and *tssM*. (c) Deletion of 19.8 kb segment of the *tss* region between *tssD1* and *tssA2*. Chromosomal sequence is shown as thin black lines, with ORFs indicated by arrows beneath. Dotted black lines indicate sections of the region where sequence from ST88 was used as a scaffold.

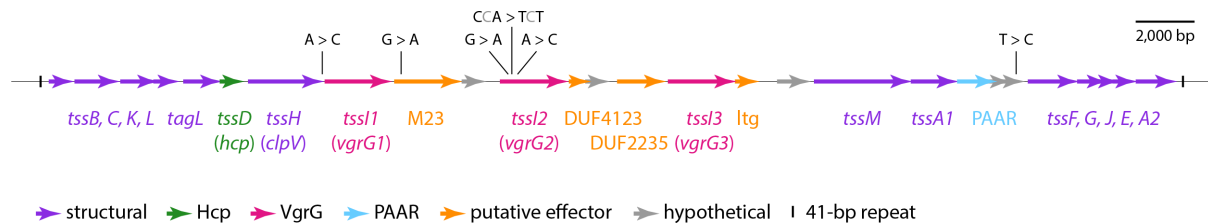


Fig. 5. The ST131 chromosomal T6SS locus. The extents and orientations of ORFs are indicated by labelled arrows, with colours corresponding to protein types as outlined in the key below. SNPs that differentiate clade C strain EC958 from clade A strain USVAST219 are shown in black text above. M23: M23-family peptidase, Itg: lytic transglycosylase. Drawn to scale from GenBank accession CP120633.

Characterization of the ST131 type VI secretion locus

To facilitate comparative analyses, we characterized the T6SS-determining region of the ancestral ST131 lineage, ST131-A/H41. First, we mapped the complete and uninterrupted T6SS locus found in a 39211 bp region of the USVAST219 chromosome (GenBank accession: CP120633) which was flanked by copies of a perfect 41 bp repeat sequence (Fig. 5). The region contained 27 ORFs, 22 of which could be assigned functions (Fig. 5). Among these were representatives of all 13 *tss* genes required for synthesis of a type VI secretion apparatus (*tssA* to *tssM*), along with the structural gene *tagL* and a second, different *tssA* gene. Genes for Hcp (*tssD*) and PAAR-domain containing proteins, along with three different *vgrG* (*tssI*) genes, encode the core and spike of the secretion apparatus in ST131-A/H41. Downstream of each *vgrG*, we found ORFs that encode putative T6SS effector proteins: an M23-family peptidase, a lytic transglycosylase and proteins with DUF4123 and DUF2235 domains, which have been associated with T6SS effectors [40]. This aligns with prior research by Ma and colleagues that identified diverse effector/immunity modules within *vgrG* modules [37]. Together, the presence and order of these components classify this locus as encoding a T6SS-1 type system [21]. We conclude that this region of the ST131 chromosome probably includes all determinants sufficient for production of a functional T6SS.

Reference strain EC958 (GenBank accession: HG941718) [40] was used to represent ST131-C2/H30Rx in comparisons with ST131-A/H41. The T6SS regions of USVAST219 and EC958 were almost identical, differing by just seven SNPs (shown in Fig. 5). The only SNP located in a structural gene (*tssH*) was silent. We therefore conclude that this chromosomal region encodes the native T6SS of ST131, and is conserved in clades A/H41, B/H22 and C/H30, consistent with the presence/absence data shown in Fig. 2(b). However, the EC958 T6SS region does not encode TssM.

The *tssM* gene is interrupted by *ISEc12* in ST131-C/H30

To determine the cause of *tssM* loss in ST131-C/H30, we examined the complete genome of reference strain EC958 [41] in further detail. We found that *tssM* had been interrupted by a copy of *ISEc12*, which was flanked by the 5 bp target site duplication ACTGC (Fig. 6). The *tssM* reading frame was split approximately in half by the insertion, with 1570 bp located to the left of *ISEc12*, and 1768 bp to the right. The interruption of *tssM* by *ISEc12*, and the resultant two *tssM* fragments, explains the <85% coverage hits to *tssM/Vt4E* obtained in the SecReT6 database screen (Fig. 2b). To deduce whether other *tssM* homologues were present in EC958, and could encode replacements for the interrupted *tssM*, we used tBLASTn to query the complete EC958 chromosome with the TssM sequence encoded by USVAST219.

This query did not return any hits, apart from the fragmented sequences produced by the *ISEc12*-interrupted *tssM*. It has been proven that the deletion of *tssM* prevents T6SS activity in *E. coli* [39] and the importance of *tssM* homologues for T6SS functionality in other species is well described [42, 43]. The *ISEc12* insertion has almost certainly rendered *tssM* non-functional in EC958, and in the absence of *tssM* redundancy, it appears most likely that the entire T6SS itself is also non-functional.

From the EC958 genome, we generated 100bp sequences that span the left and right ends of the *ISEc12* insertion (marked M–I and I–M in Fig. 6b), and from the USVAST219 genome a 100 bp sequence that represents an ancestral *tssM* uninterrupted by this insertion (marked M–M in Fig. 6b). To determine whether the same insertion event was responsible for the absence of *tssM* in all clade C/H30 genomes, we screened the ST131 collection for the three 100bp sequences. The naive M–M sequence was absent in just 2.97% ($n=21$) of clade A/H41 and B/H22 genomes, confirming that almost all of them contained uninterrupted *tssM* (Fig. 6a). In total, 2462 out of the 2478 (99.48%) clade C/H30 genomes lacked the M–M sequence but contained one or both of the left and right *ISEc12* junction sequences M–I and I–M (Fig. 3). Of the 16 remaining clade C/H30 genomes, three genomes contained M–M, and 13 genomes did not include either M–M or either of the *ISEc12* junction sequences (Fig. 6a). Manual inspection of the 13 genomes that lacked indicator sequences revealed that in six the *ISEc12* insertion was present, but low assembly quality had resulted in truncated contigs that were too short for recognition with the strict threshold used for junction sequence detection. The remaining seven genomes were excluded from further analysis, as low coverage appears to have resulted in highly fragmented assemblies such that the structure of the T6SS region was impossible to determine accurately. Thus, while *tssM* is uninterrupted in 97.03% ($n=687$) of ST131 clade A/H41 and B/H22

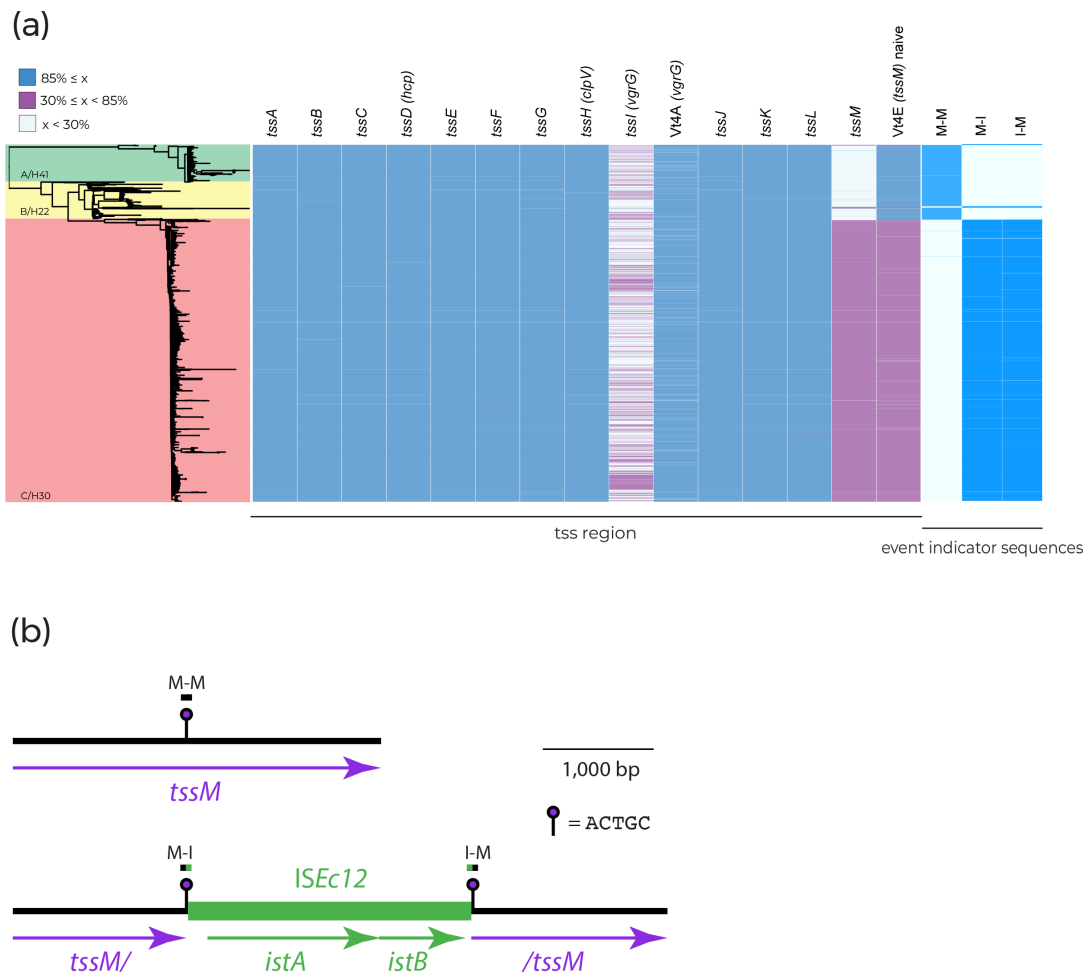


Fig. 6. (a) ST131 phylogeny. Genomes are shaded according to their clade and *fimH* allele. The presence of the *tss* region and 100 bp *ISEc12* insertion event indicator sequences is indicated by blue shading to the right of the phylogeny where only 100% matches are shown. Blue blocks represent gene coverage values of >85%, which was used as a cut-off to indicate gene presence. White blocks represent gene coverage values of <30%, indicating gene absence. Pink blocks denote sequences with coverage values of 30–85%; these intermediate values are indicative of partial sequences that are not expected to represent functional genes. (b) Interruption of *tssM* by *ISEc12*. Chromosomal sequence is shown as a thin black line, and *ISEc12* as a thicker green line, with ORFs indicated by arrows beneath. The positions of target site duplication sequences and extents of indicator sequences are shown above. Drawn to scale using sequence obtained from GenBank accessions: CP120633 and HG94718.

genomes, it has been interrupted by *ISEc12* at exactly the same position in 99.78% ($n=2465$) of clade C/H30 genomes. We conclude that the loss of *tssM* in ST131-C/H30 was the result of a single *ISEc12* insertion event that occurred in an ancestral strain.

The *tss* region is conserved in other phylogroup B2 lineages

Other pathogenic lineages of *E. coli* were examined to investigate whether an incomplete *tss* region was specific to clades containing MDR clones or a general feature of successful extraintestinal pathogenic *E. coli* (ExPEC). ST73 and ST95 were selected as comparators due to their documented prevalence in clinical settings in the UK and lack of multi-drug resistance [7, 44]. They and ST131 belong to phylogroup B2, which would also allow us to address phylogenetic signal influencing the presence of the *tss* region. The *tss* region was found to be consistently present in both lineages (Fig. 7a, b), demonstrating that T6SS presence varies between pathogenic *E. coli* lineages within phylogroup B2.

DISCUSSION

Extensive work has been done to characterize the T6SS in terms of its structure, function, variability and prevalence in various genera. However, this is the first study, to the best of our knowledge, to determine the prevalence of the *tss* region within MDR ExPEC lineages. We have particularly focused on ST410 and ST131 as they are known to cause infections worldwide [35, 36, 45, 46]. We have identified the evolutionary events that led to the loss of key structural T6SS components in these lineages via recombination or insertion events.

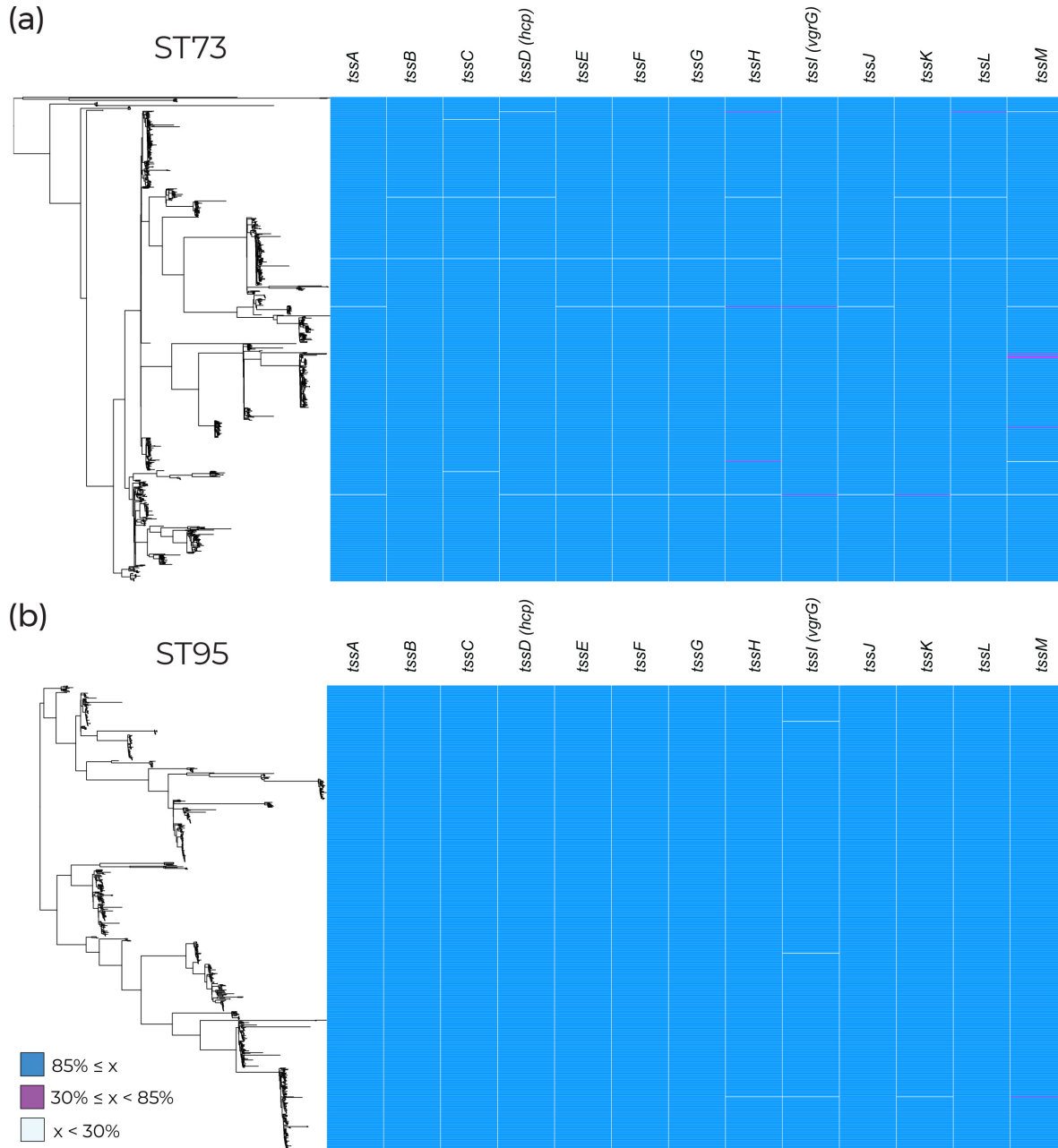


Fig. 7. Presence of structural experimentally validated type VI secretion genes taken from the SecReT6 database [28, 29] across (a) 873 ST73 genomes and (b) 758 ST95 genomes. Blue blocks represent gene coverage values of $>85\%$, which was used as a cut-off to indicate gene presence. White blocks represent gene coverage values of $<30\%$, indicating gene absence. Pink blocks denote sequences with coverage values of $30\text{--}85\%$; these intermediate values are indicative of partial sequences that are not expected to represent functional genes.

Our study has yielded new insights into the evolution of T6SS functionality in two of the most clinically relevant MDR *E. coli* lineages, and has provided a broader overview of T6SS determinants in other lineages.

In both ST410 and ST131, we have observed a loss of T6SS that appears to coincide with the acquisition of MDR. This suggests we may have uncovered a common evolutionary trajectory in two MDR lineages. We have previously described the important role of potentiating mutations involved in colonization factor determinants, anaerobic metabolism genes and intergenic regions as being common evolutionary trajectories in MDR clone formation [47]. The loss or degradation of a T6SS could perhaps integrate into the stepwise evolutionary course of ST410-B4/H24RxC and ST131-C2/H30Rx. The structural components for a complete T6SS were chromosomal and conserved within the ancestral clades ST410-A/H53, ST131-A/H41 and ST131-B/H22. We found that the T6SS in

ST410-B/H24 was inactivated by deletion events and that the T6SS in ST131-C/H30 was inactivated by an insertion event. In ST131, the rare instances where *tssM* is interrupted in clade A/H41 and B/H22, and uninterrupted in clade C/H30 (Fig. 6a), might be explained by inter-lineage HGT and homologous recombination, but the T6SS region is too well conserved (>99.99% ID between lineages) to accurately detect recombinant sequences. In both ST131 and ST410, the inactivation of T6SS occurring prior to or during the formation of MDR clades implies that these events may have contributed to their successful emergence and expansion, but we cannot generalize to all MDR lineages as further analyses of the common MDR lineages ST167 and ST648, not shown here, showed intact T6SS regions.

It has previously been suggested that acquiring MDR plasmids is a vital step in the evolution of pandemic clones [47]. We speculate that the lack of a functional T6SS allowed both ST410-B/H24 and ST131-C/H30 to become more receptive to cell-to-cell contact, and therefore conjugative transfer. The deleterious effect of T6SS on plasmid conjugation has been demonstrated in *Acinetobacter baumannii*, where a conjugative plasmid has been shown to repress the host T6SS in order to increase conjugation rates [26]. While the *A. baumannii* experiments involved inhibition of T6SS in donor cells, the absence of T6SS in potential recipient cells seems likely to increase their permissiveness to plasmid transfer. The acquisition of conjugative MDR plasmids by ST410-B/H24 and ST131-C/H30 might therefore have been facilitated by their loss of functional T6SS.

Our genomic analyses, combined with existing literature, suggest that the MDR clone ST131-C2/H30Rx does not have a functional T6SS due to the interruption of a single gene, *tssM*. However, the functionality of the uninterrupted *tss* region in the ancestral ST131 clades A/H41 and B/H22 has not been verified experimentally. Further work to determine the functionality of the T6SS in all three ST131 clades is therefore required. Final experimental validation would require complementation of *tssM* to restore function of the T6SS in ST131-C/H30.

The data presented here do not support the hypothesis that the ability of drug-resistant *E. coli* to displace resident commensal *E. coli* is due to the production of a T6SS. Phage, colicins or other diffusible elements are alternative explanations that should be considered. By focusing on T6SS, we have uncovered parallel evolutionary outcomes in ST410-B4/H24RxC and ST131-C2/H30Rx, where T6SS determinants have been lost by MDR clones. Our work suggests that the loss of a functional T6SS might have played a role in the evolution and success of the pandemic MDR *E. coli* clones in ST131 and ST410.

Funding information

E.A.C. was funded by the Wellcome Antimicrobial and Antimicrobial Resistance (AAMR) DTP (108876B15Z).

Acknowledgements

We would like to thank Professor Jose Bengoechea for helpful discussion and all members of the McNally group for their support.

Author contributions

E.A.C.: data curation, methodology, formal analysis, visualization, writing – original draft preparation and review and editing. R.A.M.: methodology, formal analysis, writing – original draft preparation and review and editing. A.E.S.: methodology. R.J.H.: writing – review and editing. C.H.C.: data curation. S.J.D.: data curation. A.M.: conceptualization, supervision, writing – review and editing.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Cummins EA, Hall RJ, Connor C, McInerney JO, McNally A. Distinct evolutionary trajectories in the *Escherichia coli* pangenome occur within sequence types. *Microb Genom* 2022;8:11.
- Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* 2022;399:629–655.
- Connor CH, Zucoloto AZ, Yu I-L, Corander J, McDonald B, et al. Multi-drug resistant *E. coli* displace commensal *E. coli* from the intestinal tract, a trait associated with elevated levels of genetic diversity in carbohydrate metabolism genes. *bioRxiv* 2022.
- Arcilla MS, van Hattem JM, Haverkate MR, Bootsma MCJ, van Genderen PJJ, et al. Import and spread of extended-spectrum β -lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. *Lancet Infect Dis* 2017;17:78–85.
- Bevan ER, McNally A, Thomas CM, Piddock LJV, Hawkey PM. Acquisition and loss of CTX-M-producing and non-producing *Escherichia coli* in the fecal microbiome of travelers to South Asia. *mBio* 2018;9:e02408-18.
- Kantele A, Kuenzli E, Dunn SJ, Dance DAB, Newton PN, et al. Dynamics of intestinal multidrug-resistant bacteria colonisation contracted by visitors to a high-endemic setting: a prospective, daily, real-time sampling study. *Lancet Microbe* 2021;2:e151–e158.
- Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, 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* 2017;27:1437–1449.
- Gladstone RA, McNally A, Pöntinen AK, Tonkin-Hill G, Lees JA, 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* 2021;2:e331–e341.
- Davies M, Galazzo G, van Hattem JM, Arcilla MS, Melles DC, et al. Enterobacteriaceae and Bacteroidaceae provide resistance to travel-associated intestinal colonization by multi-drug resistant *Escherichia coli*. *Gut Microbes* 2022;14.
- Mariano G, Trunk K, Williams DJ, Monlezun L, Strahl H, et al. A family of type VI secretion system effector proteins that form ion-selective pores. *Nat Commun* 2019;10:5484.
- Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci U S A* 2007;104:15508–15513.
- Salomon D, Klimko JA, Trudgian DC, Kinch LN, Grishin NV, et al. Type VI secretion system toxins horizontally shared between marine bacteria. *PLoS Pathog* 2015;11:e1005128.

13. Bernal P, Allsopp LP, Filloux A, Llamas MA. The *Pseudomonas putida* T6SS is a plant warden against phytopathogens. *ISME J* 2017;11:972–987.
14. Troselj V, Treuner-Lange A, Sogaard-Andersen L, Wall D. Physiological heterogeneity triggers sibling conflict mediated by the type VI secretion system in an aggregative multicellular bacterium. *mBio* 2018;9:e01645-17.
15. Tang L, Yue S, Li G-Y, Li J, Wang X-R, et al. Expression, secretion and bactericidal activity of type VI secretion system in *Vibrio anguillarum*. *Arch Microbiol* 2016;198:751–760.
16. Unni R, Pintor KL, Diepold A, Unterweger D. Presence and absence of type VI secretion systems in bacteria. *Microbiology* 2022;168.
17. Kempnich MW, Sison-Mangus MP. Presence and abundance of bacteria with the Type VI secretion system in a coastal environment and in the global oceans. *PLoS One* 2020;15:e0244217.
18. García-Bayona L, Coyne MJ, Comstock LE. Mobile Type VI secretion system loci of the gut bacteroidales display extensive intra-ecosystem transfer, multi-species spread and geographical clustering. *PLoS Genet* 2021;17:e1009541.
19. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, et al. Identification of protein secretion systems in bacterial genomes. *Sci Rep* 2016;6:23080.
20. Bingle LE, Bailey CM, Pallen MJ. Type VI secretion: a beginner's guide. *Curr Opin Microbiol* 2008;11:3–8.
21. Journet L, Cascales E. The Type VI secretion system in *Escherichia coli* and related species. *EcoSal Plus* 2016;7.
22. Ma J, Bao Y, Sun M, Dong W, Pan Z, et al. Two functional type VI secretion systems in avian pathogenic *Escherichia coli* are involved in different pathogenic pathways. *Infect Immun* 2014;82:3867–3879.
23. de Pace F, Nakazato G, Pacheco A, de Paiva JB, Sperandio V, et al. The type VI secretion system plays a role in type 1 fimbria expression and pathogenesis of an avian pathogenic *Escherichia coli* strain. *Infect Immun* 2010;78:4990–4998.
24. Brunet YR, Espinosa L, Harchouni S, Mignot T, Cascales E. Imaging type VI secretion-mediated bacterial killing. *Cell Rep* 2013;3:36–41.
25. Weber BS, Ly PM, Irwin JN, Pukatzki S, Feldman MF. A multidrug resistance plasmid contains the molecular switch for type VI secretion in *Acinetobacter baumannii*. *Proc Natl Acad Sci U S A* 2015;112:9442–9447.
26. Di Venanzio G, Moon KH, Weber BS, Lopez J, Ly PM, et al. Multidrug-resistant plasmids repress chromosomally encoded T6SS to enable their dissemination. *Proc Natl Acad Sci U S A* 2019;116:1378–1383.
27. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M. The Enterobase user's guide, with case studies on salmonella transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res* 2020;30:138–152.
28. Li J, Yao Y, Xu HH, Hao L, Deng Z, et al. SecReT6: a web-based resource for type VI secretion systems found in bacteria. *Environ Microbiol* 2015;17:2196–2202.
29. Zhang J, Guan J, Wang M, Li G, Djordjevic M, et al. SecReT6 update: a comprehensive resource of bacterial Type VI secretion systems. *Sci China Life Sci* 2023;66:626–634.
30. Storey D, McNally A, Åstrand M, sa-Pessoa Graca Santos J, Rodriguez-Escudero I, et al. *Klebsiella pneumoniae* type VI secretion system-mediated microbial competition is PhoPQ controlled and reactive oxygen species dependent. *PLoS Pathog* 2020;16:e1007969.
31. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006;34:D32–6.
32. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
33. Colpan A, Johnston B, Porter S, Clabots C, Anway R, et al. *Escherichia coli* sequence type 131 (ST131) subclone H30 as an emergent multidrug-resistant pathogen among US veterans. *Clin Infect Dis* 2013;57:1256–1265.
34. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
35. Roer L, Overballe-Petersen S, Hansen F, Schønning K, Wang M, et al. *Escherichia coli* sequence type 410 is causing new international high-risk clones. *mSphere* 2018;3:00337–18.
36. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci U S A* 2014;111:5694–5699.
37. Ma J, Sun M, Pan Z, Lu C, Yao H. Diverse toxic effectors are harbored by vgrG islands for interbacterial antagonism in type VI secretion system. *Biochim Biophys Acta Gen Subj* 2018;1862:1635–1643.
38. Ma J, Sun M, Dong W, Pan Z, Lu C, et al. PAAR-Rhs proteins harbor various C-terminal toxins to diversify the antibacterial pathways of type VI secretion systems. *Environ Microbiol* 2017;19:345–360.
39. Kostiuik B, Santoriello FJ, Diaz-Satizabal L, Bisaro F, Lee K-J, et al. Type VI secretion system mutations reduced competitive fitness of classical *Vibrio cholerae* biotype. *Nat Commun* 2021;12:6457.
40. Salomon D, Kinch LN, Trudgian DC, Guo X, Klimko JA, et al. Marker for type VI secretion system effectors. *Proc Natl Acad Sci U S A* 2014;111:9271–9276.
41. Forde BM, Ben Zakour NL, Stanton-Cook M, Phan M-D, Totsika M, et al. The complete genome sequence of *Escherichia coli* EC958: A high quality reference sequence for the globally disseminated multidrug resistant *E. coli* O25b:H4-ST131 clone. *PLoS One* 2014;9:e104400.
42. Ma LS, Narberhaus F, Lai EM. IcmF family protein TssM exhibits ATPase activity and energizes type VI secretion. *J Biol Chem* 2012;287:15610–15621.
43. Zhang L, Xu J, Xu J, Chen K, He L, et al. TssM is essential for virulence and required for type VI secretion in *Ralstonia solanacearum*. *J Plant Dis Prot* 2012;119:125–134.
44. Lipworth S, Vihta K-D, Chau K, Barker L, George S, et al. Ten-year longitudinal molecular epidemiology study of *Escherichia coli* and *Klebsiella* species bloodstream infections in Oxfordshire, UK. *Genome Med* 2021;13:144.
45. Feng Y, Liu L, Lin J, Ma K, Long H, et al. Key evolutionary events in the emergence of a globally disseminated, carbapenem resistant clone in the *Escherichia coli* ST410 lineage. *Commun Biol* 2019;2:322.
46. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, et al. Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. *mBio* 2016;7:e02162.
47. Cummins EA, Snaith AE, McNally A, Hall RJ. The role of potentiating mutations in the evolution of pandemic *Escherichia coli* clones. *Eur J Clin Microbiol Infect Dis* 2021.