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

Elizabeth A. Cummins[1], Robert A. Moran[1], Ann E. Snaith[1],
Rebecca J. Hall[1], Chris H. Connor[2], Steven J. Dunn[1], Alan McNally[1]

[1] Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

[2]Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne 3000, Australia

Corresponding author: a.mcnally.1@bham.ac.uk

Key words: type VI secretion system, *Escherichia coli*, multi-drug resistance, ST410, ST131

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 colonisation 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 20,000 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 characterised 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 is 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 has been taken from ENA with the appropriate accession numbers listed within the methods section. The *E. 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].

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 colonise 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 may be one of the antagonistic systems employed by *E. coli* in this process. Our findings provide the first detailed characterisation 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.

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.

63

64 A trait increasingly ascribed to MDR *E. coli* lineages is the ability to rapidly, and
65 asymptomatically, colonise the human intestinal tract [3]. Multiple travel studies have now
66 shown that people travelling from areas of low AMR incidence to AMR endemic regions
67 become colonised by extended spectrum beta-lactam (ESBL)-resistant or MDR *E. coli* during
68 travel [4; 5; 6]. Furthermore, genomic analysis has shown that the gain of MDR *E. coli* was
69 due to the acquisition of a new MDR strain and not the preceding commensal *E. coli*
70 becoming MDR [5]. The ability to displace and colonise may be attributed to the MDR
71 phenotype itself,
72 but longitudinal studies from the UK and Norway [7; 8] have shown that multi-drug
73 resistance alone is not a sufficient driver for epidemiological success of *E. coli*. Recent
74 metagenomic analysis has also revealed that colonisation by MDR *E. coli* does not disrupt
75 the wider gut microbiome's composition or diversity [9].
76
77 One possible reason for the ability of drug-resistant *E. coli* to displace resident commensal *E.*
78 *coli* so rapidly is a result of the drug-resistant *E. coli* possessing a type VI secretion system
79 that allows contact-dependent killing of the resident commensal *E. coli*. The type VI
80 secretion system (T6SS) is a multi-functional apparatus that some Gram-negative bacteria
81 possess to facilitate nutrient uptake [10], manipulation of host cells [11], and the killing of
82 competing bacteria [12; 13; 14; 15]. T6SS distribution varies by environment [16; 17] and
83 the presence of T6SSs varies on all taxonomic levels [16; 18; 19] with over-representation in
84 Gammaproteobacteria [20]. T6SS genes can be gained and lost via horizontal gene transfer
85 (HGT) [12; 18] and T6SS presence may be influenced by genomic incompatibilities, either
86 between mobile genetic elements, donor and recipient bacteria, or via assembly of multiple
87 T6SSs [16]. T6SSs are particularly prevalent in complex microbial communities that are host-
88 associated [17; 18] and are commonly found in pathogens, including *E. coli* [16; 21; 22; 23].
89
90 *E. coli* is a relatively understudied organism within the research field of T6SSs, with many
91 studies only focusing on specific pathotypes [22; 23; 24]. In *E. coli* T6SSs are classified into
92 three distinct phylogenetic groups (T6SS-1 to T6SS-3) and all have been shown to be directly
93 involved in pathogenesis and antibacterial activity [21]. The existing connections between
94 T6SS and multi-drug resistance in other species [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 characterise 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 20,577 *E. coli* genomes used here were curated from EnteroBase [27] and were fully characterised 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 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.

T6SS database creation and interrogation

The SecReT6 [28; 29] experimentally validated database was retrieved and reformatted so that sequence identifiers were a suitable format for a custom ABRicate database. ABRicate (v0.9.8) was run using the default 75% minimum nucleotide sequence identity to obtain a list of T6SS genes that are present across our 21 *E. coli* sequence types. The results were summarised and partial hits (instances where a gene hit was split over multiple contigs) were accounted for and processed with a custom Python script (github.com/lillycummins/InterPangenome/blob/main/process_partial_hits.py). We chose to determine a gene as present in a genome if at least 85% of the gene was covered with a nucleotide sequence identity above the default 75% threshold. For the focused functional

126 gene unit analysis, any matches to genes annotated as effector or immunity protein-
127 encoding were removed from the search results.
128

129 Genomic screening for insertion sequences and deletion events

130

131 Gene Construction Kit (v4.5.1) (Textco Biosoftware, Raleigh, USA) was used to annotate and
132 visualise DNA sequences. Insertion sequences (ISs) were identified using the ISFinder
133 database [30]. Target site duplications (TSDs) flanking ISs were identified manually. Two
134 junction sequences specific for a given insertion were generated by taking 100 bp
135 contiguous sequences that span the left and right ends of the insertion. Each junction
136 sequence was comprised of 50 bp of the target IS and 50 bp of its adjacent sequence. To
137 generate an ancestral sequence in silico, the interrupting IS and one copy of its associated
138 TSD were removed manually. A 100 bp sequence that spanned the insertion point was taken
139 to represent the naive site. A database of all ST131 genomes in the collection was generated
140 for screening with standalone BLASTn [31]. The database was queried with insertion-
141 junction sequences and the naive site sequence to determine whether an insertion was
142 present, with only complete and identical matches to the 100 bp query sequences
143 considered positive matches. The same 100 bp indicator sequence method was used to
144 identify deletion events.
145

146 Genome sequencing

147

148 To the best of our knowledge, no complete reference genome for ST131-A/H41 was publicly
149 available so we generated sequence data for the strain USVAST219 [32]. DNA was extracted
150 using the Monarch Genomic DNA Extraction Kit (New England Biolabs, Massachusetts, USA)
151 before sequencing with MinION (Oxford Nanopore Technology, UK) using R9.4.1 flow cells.
152 The data was basecalled with Guppy (v6.0.1) (github.com/nanoporetech/pyguppyclient) and
153 adapters were trimmed with qcat (v1.1.0) (github.com/nanoporetech/qcat). Short-read
154 genome sequencing was provided by MicrobesNG (microbesng.com) where DNA was
155 prepared using the Nextera XT kit (Illumina, San Diego, CA, USA) and sequenced on the
156 Illumina HiSeq platform (Illumina). Long and short-read data are both available within

157 BioProject PRJNA943186. A hybrid assembly using both the long and short-read sequencing
158 data was created with Unicycler (v0.4.8) [33].
159

160 Results

161

162 T6SS gene presence varies between and within phylogroups and sequence 163 types

164

165 All 20,577 *E. coli* genomes interrogated contained at least one T6SS gene from the SecReT6
166 experimental database [28; 29]. There was variation in the average number of T6SS genes
167 present per ST and the average number of genes per ST within phylogroups (Figure 1).
168 ST131 exhibited the largest range (2 to 29) of T6SS genes, highest number of T6SS genes in a
169 single genome (n = 29), and highest average number of T6SS genes (joint with ST12, ST14,
170 ST17, and ST127). The distribution of the number of T6SS genes present per genome did not
171 align with phylogroups. STs known to contain MDR lineages, ST131, ST167, ST410, and
172 ST648, did not possess a similar range or average of T6SS genes (Figure 1). Grouping genes
173 into functional units is therefore necessary to gain insight into how many potentially
174 functional T6SSs are present within STs and whether this may correlate with multi-drug
175 resistance.

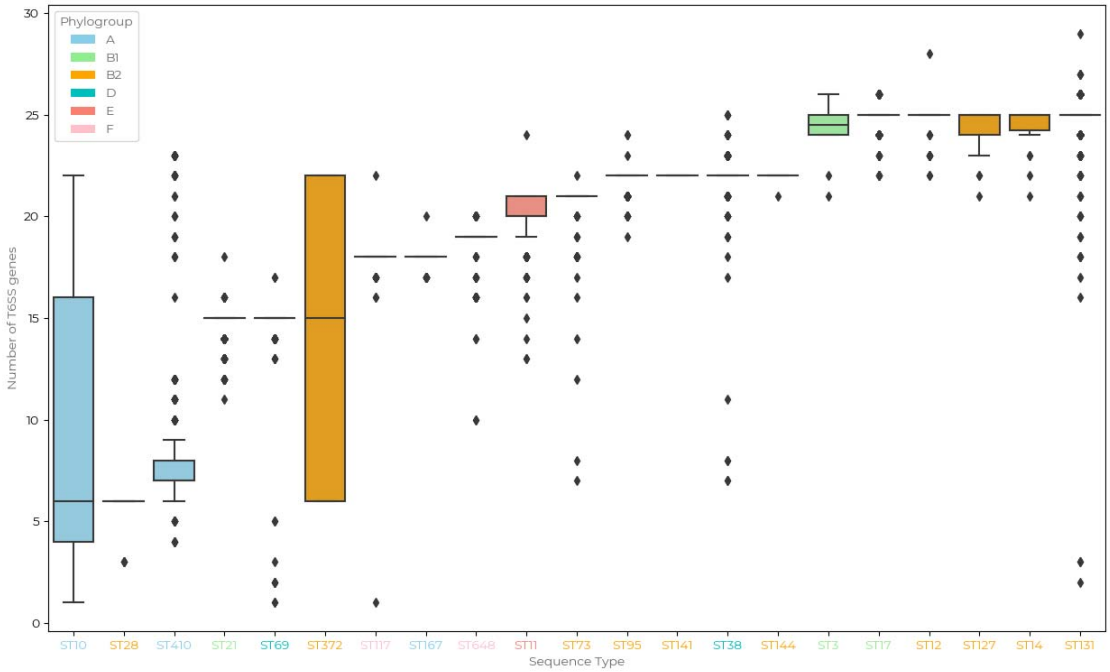


Fig. 1 Distributions of the number *of type VI secretion system (T6SS) genes from the SecReT6 experimentally validated database [28; 29] across 21 sequence types of E. coli shows gene carriage varies by sequence type and phylogroup. Sequence types are colour coded according to their respective phylogroups: 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%.

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; 34; 35]. Figure 2 displays the presence of structural T6SS genes within ST410 (Figure 2a) and ST131 (Figure 2b) with the corresponding clades highlighted on the phylogenies. Structural genes were the focus of this search because they are well characterised 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 (Figure 2).

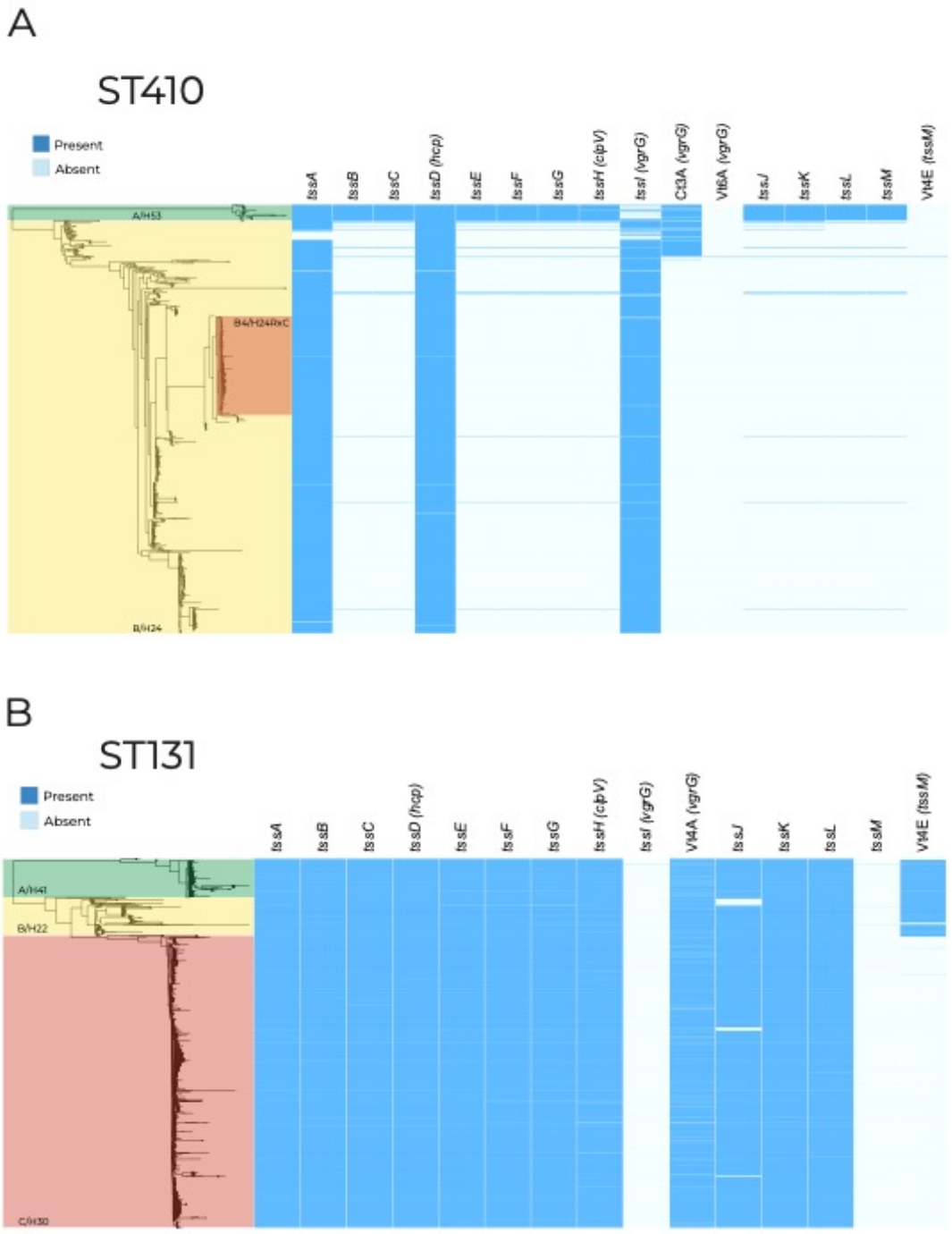
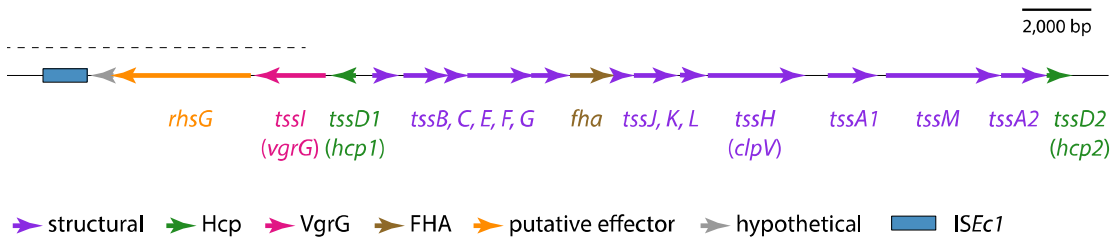


Fig. 2 Presence of structural experimentally validated type VI secretion genes taken from the SecReT6 database [28; 29] across (A) 1,006 ST410 genomes and (B) 3,186 ST131 genomes. Phylogenetic clades are labelled clade name/fimH allele.

Presence of structural T6SS genes in ST410

197 Within ST410, *tssABCDEFGHIJKLM*, (henceforth referred to as the *tss* region) was present in
198 94.59% (n = 34) of clade A/H53 (Figure 2a). The organisation of genes in this region classifies
199 the system as T6SS-2. Of the three genomes that did not contain a full *tss* region, one
200 genome was only missing *tssH* due to poor sequence coverage (coverage <10x for the three
201 contigs spanning *tssH*). In the two other genomes only *vgrG* was missing, caused by a
202 combination of low coverage and the presence of multiple *vgrG* genes, which is known to
203 lead to sequence fragmentation. We conclude that the native T6SS of ST410-A/H53 is
204 conserved across the clade.
205
206 The *tss* region was present in just 0.83% (n = 8) of clade B/H24 genomes. Two ST410-B/H24
207 genomes did not contain *tssI/vgrG* and four genomes did not contain *tssLM*, but these
208 otherwise contained all *tss* region genes. While isolates from clade A/H53 are largely drug
209 susceptible, clade B/H24 contains the globally distributed multi-drug resistant clone ST410-
210 B4/H24RxC, which is highlighted in Figure 2a [34; 36]. In the eight rare instances where a full
211 *tss* region was present within the B/H24 clade, five genomes contained *tss* regions most
212 closely related (≥98% nucleotide identity) to that of the T6SS-2 region of F-type plasmid
213 pSTEC299_1 (GenBank accession: CP022280) when comparing to the NCBI non-redundant
214 database. Two of the remaining three genomes contained *tss* regions that were closely
215 related (≥98% nucleotide identity) to chromosomal T6SS-1 regions from other *E. coli*
216 (GenBank accessions: CP062901, CP091020). These findings together suggest that, although
217 rare, multiple acquisitions of distinct *tss* regions have occurred within ST410 clade B/H24.
218
219
220 **Characterisation of the ST410 type VI secretion locus**
221
222 We resolved the structure of the complete 28.3 kb T6SS locus of ST410-A/H53 using a draft
223 ST410-A/H53 assembly (EnteroBase assembly barcode: LB4500AA_AS), partially scaffolded
224 (Fig. S1) using the complete genome of ST88 *E. coli* RHBSTW-00313 (GenBank accession:
225 CP056618). ST88 belongs to the same clonal complex as ST410 (CC23) and was used
226 because there is no complete reference genome for ST410-A/H53 and we were unable to
227 access a strain to generate a closed genome. The complete T6SS region contained 18 open

228 reading frames (ORFs), 16 of which could be assigned functions (Figure 3). Among these
229 were representatives of all 13 *tss* genes required for synthesis of a type VI secretion
230 apparatus
231 (*tssA* to *tssM*), along with two additional different *tssA* and *tssD* genes. The recombination
232 hotspot *rhsG* gene, which is known to encode a toxin [39], is located downstream of *vgrG*
233 and is the likely determinant for this systems' effector protein. The locus also contains a
234 gene for an FHA-domain containing protein, which is known to be involved in T6SS
235 activation in *Pseudomonas aeruginosa* [40]. The structure of this region is indicative of a
236 T6SS-2 type system [21] and we conclude that this region of the ST410-A/H53 chromosome
237 includes all determinants necessary for the production of a functional T6SS.



238

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

243

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

245

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

257 majority of ST410-B/H24 genomes (n = 941 out of 969; Figure 4a), indicating that this
258 deletion event occurred as a single event early in the evolution of ST410-B/H24 and not as
259 multiple independent events within this clade.
260
261 Of the 28 ST410-B/H24 genomes that did not contain the *rhsG-tssM* junction sequence, 20
262 were clustered in the phylogeny and also lacked *tssA* (Figure 4a). This suggests that a
263 different or additional deletion event may have occurred in a small sub-clade of ST410-
264 B/H24. To determine this, we annotated a contig from one of these 20 genomes
265 (EnteroBase assembly barcode: ESC_HA8479AA_AS), which contained *tssD2* and exhibited a
266 different *tss* region configuration. To confirm the structure of this region, we used the
267 ESC_HA8479AA_AS *tss* contig to query GenBank and found an identical sequence in a
268 complete ST410 genome (strain E94, accession: CP199740) that had been published after
269 our data set was assembled. We then used the E94 genome as a representative for the 20
270 clustered genomes. This revealed the absence of a 19.8 kb segment of the *tss* region
271 between *tssD1* and *tssA2* (Figure 4c). The presence of the same *tssD-tssA* junction sequence
272 (D-A) in all 20 clustered genomes indicated that this deletion event was responsible for the
273 absence of *tss* genes in this sub-clade of ST410-B/H24.

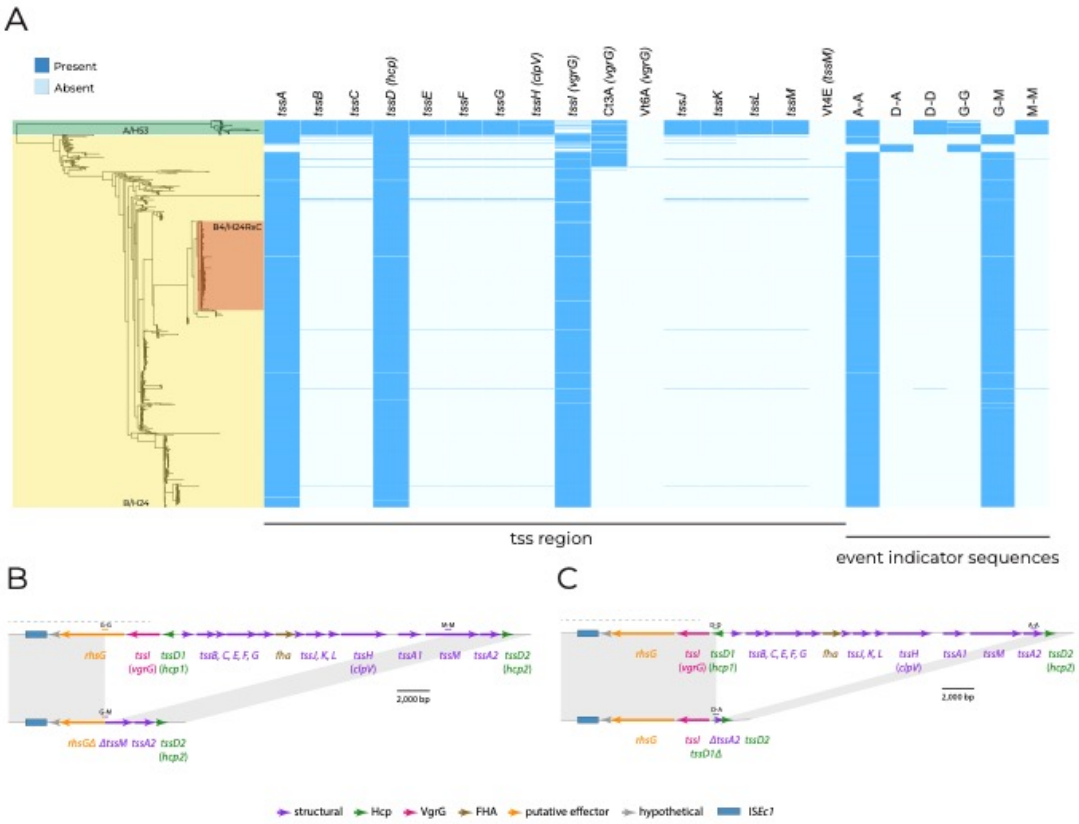


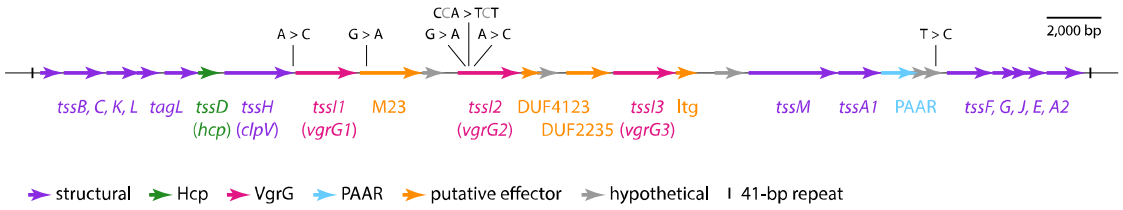
Fig 4 (A) ST410 phylogeny. Genomes are shaded according to their clade and fimH allele. The presence of tss region genes and 100bp deletion event indicator sequences is shown by blue shading to the right of the phylogeny. (B) Deletion of 21.0 kb segment of tss region between rhsG and tssM. (C) Deletion of 19.8 kb segment of tss region between tssD1 and tssA2. Chromosomal sequence is shown as thin black lines, with open reading frames indicated by arrows beneath. Dotted black lines indicate sections of region where sequence from ST88 was used as a scaffold.

Presence of structural T6SS genes in ST131

Within ST131, the ancestral and less drug-resistant clades A/H41 and B/H22 possess a gene labelled Vt4E from the SecReT6 database, which is not present in clade C/H30 (Figure 2b). To deduce the function of Vt4E, we used BLASTn to search [31] 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; 38] which is a component of the inner membrane complex of T6SSs [21]. The absence of a functional VasK protein has been shown to render the T6SS non-functional in *Vibrio cholerae* [37].

Characterisation of the ST131 type VI secretion locus

292 To facilitate comparative analyses, we characterised the T6SS-determining region of the
293 ancestral ST131 lineage, ST131-A/H41. First, we mapped the complete and uninterrupted
294 T6SS locus found in a 39,211 bp region of the USVAST219 chromosome (GenBank accession:
295 CP120633) which was flanked by copies of a perfect 41 bp repeat sequence (Figure 5). The
296 region contained 27 ORFs, 22 of which could be assigned functions (Figure 5). Among these
297 were representatives of all 13 *tss* genes required for synthesis of a type VI secretion
298 apparatus (*tssA* to *tssM*), along with the structural gene *tagL* and a second, different *tssA*
299 gene. Genes
300 for Hcp (*tssD*) and PAAR-domain containing proteins, along with three different *vgrG* (*tssI*)
301 genes, encode the core and spike of the secretion apparatus in ST131-A/H41. Downstream
302 of each *vgrG*, we found ORFs that encode putative T6SS effector proteins; a M23-family
303 peptidase, a lytic transglycosylase and proteins with DUF4123 and DUF2235 domains, which
304 have been associated with T6SS effectors [41]. This aligns with prior research by Ma and
305 colleagues that identified diverse effector/immunity modules within *vgrG* modules [38].
306 Together, the presence and order of these components classify this locus as encoding a
307 T6SS-1 type system [21]. We conclude that this region of the ST131 chromosome likely
308 includes all determinants sufficient for production of a functional T6SS.
309
310 Reference strain EC958 (GenBank accession: HG941718) [42] was used to represent ST131-
311 C2/H30Rx in comparisons with ST131-A/H41. The T6SS regions of USVAST219 and EC958
312 were almost identical, differing by just seven single nucleotide polymorphisms (SNPs)
313 (shown in Figure 5). The only SNP located in a structural gene (*tssH*) was silent. We
314 therefore conclude that this chromosomal region encodes the native T6SS of ST131, and is
315 conserved in clades A/H41, B/H22, and C/H30, consistent with the presence/absence data
316 shown in Figure 2b. However, the EC958 T6SS region does not encode TssM.



317
318 *Fig. 5 The ST131 chromosomal T6SS locus. The extents and orientations of open reading frames are indicated by labelled*
319 *arrows, with colours corresponding to protein types as outlined in the key below. SNPs that differentiate clade C strain*
320 *EC598 from clade A strain USVAST219 are shown in black text above. M23: M23-family peptidase, ltg: lytictransglycosylase.*
321 *Drawn to scale from GenBank accession CP120633.*

322

323 ## *tssM* is interrupted by IS*Ec12* in ST131-C/H30

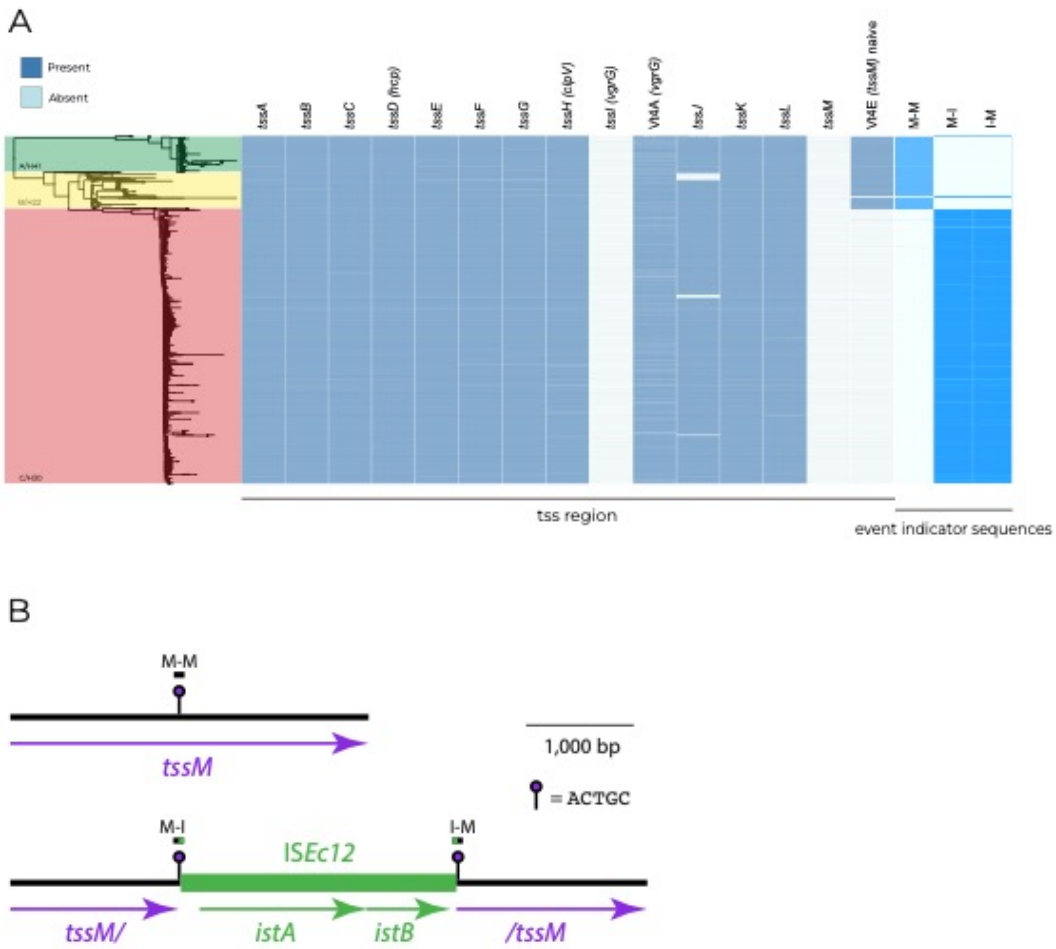
324

325 To determine the cause of *tssM* loss in ST131-C/H30, we examined the complete genome of
326 reference strain EC958 [42] in further detail. We found that *tssM* had been interrupted by a
327 copy of IS*Ec12*, which was flanked by the 5 bp target site duplication ACTGC (Figure 6). The
328 *tssM* reading frame was split approximately in half by the insertion, with 1,570 bp located to
329 the left of IS*Ec12*, and 1,768 bp to the right. To deduce whether other *tssM* homologues
330 were present in EC958, and could encode replacements for the interrupted *tssM*, we used
331 tBLASTn to query the complete EC958 chromosome with the TssM sequence encoded by
332 USVAST219.
333 This query did not return any hits, apart from the fragmented sequences produced by the
334 IS*Ec12*-interrupted *tssM*. It has been proven that the deletion of *tssM* prevents T6SS activity
335 in *E. coli* [38] and the importance of *tssM* homologues for T6SS functionality in other species
336 is well described [43; 44]. The IS*Ec12* insertion has almost certainly rendered *tssM* non-
337 functional in EC958, and in the absence of *tssM* redundancy, it appears most likely that the
338 entire T6SS itself is also non-functional.

339

340 From the EC958 genome, we generated 100 bp sequences that span the left and right ends
341 of the IS*Ec12* insertion (marked M-I and I-M in Figure 6b), and from the USVAST219 genome
342 a 100 bp sequence that represents an ancestral *tssM* uninterrupted by this insertion
343 (marked M-M in Figure 6b). To determine whether the same insertion event was
344 responsible for the absence of *tssM* in all clade C/H30 genomes, we screened the ST131
345 collection for the three 100 bp sequences. The naive M-M sequence was absent in just
346 2.97% (n = 21) of clade
347 A/H41 and B/H22 genomes, confirming that almost all of them contained uninterrupted
348 *tssM* (Figure 6a). 2,462 out of the 2,478 (99.48%) clade C/H30 genomes lacked the M-M
349 sequence but contained one or both of the left and right IS*Ec12* junction sequences M-I and
350 I-M (Figure 3). Of the 16 remaining clade C/H30 genomes, three genomes contained M-M,
351 and 13 genomes did not include either M-M or either of the IS*Ec12* junction sequences
352 (Figure 6a). Manual inspection of the 13 genomes that lacked indicator sequences revealed

353 that in six the IS*Ec12* insertion was present, but low assembly quality had resulted in
354 truncated contigs that were too short for recognition with the strict threshold used for
355 junction sequence detection. The remaining seven genomes were excluded from further
356 analysis, as low coverage appears to have resulted in highly-fragmented assemblies such
357 that the structure of the T6SS region was impossible to determine accurately. Thus, while
358 *tssM* is uninterrupted in 97.03% (n = 687) of ST131 clade A/H41 and B/H22 genomes, it has
359 been interrupted by IS*Ec12* at exactly the same position in 99.78% (n = 2465) of clade C/H30
360 genomes. We conclude that the loss of *tssM* in ST131-C/H30 was the result of a single
361 IS*Ec12* insertion event that occurred in an ancestral strain.



362

363 *Fig. 6 (A) ST131 phylogeny. Genomes are shaded according to their clade and fimH allele. The presence of the tss region and*
364 *100 bp ISEc12 insertion event indicator sequences is dictated by blue shading to the right of the phylogeny. (B) Interruption*
365 *of tssM by ISEc12. Chromosomal sequence is shown as a thin black line, and ISEc12 as a thicker green line, with open*
366 *reading frames indicated by arrows beneath. The positions of target site duplication sequences and extents of indicator*
367 *sequences are shown above. Drawn to scale using sequence obtained from GenBank accessions: CP120633 and HG94718.*

368

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

370

371 Other pathogenic lineages of *E. coli* were examined to investigate whether an incomplete
372 *tss* region was specific to clades containing MDR clones or a general feature of successful
373 extraintestinal pathogenic *E. coli* (ExPEC). ST73 and ST95 were selected as comparators due
374 to their documented prevalence in clinical settings in the UK [7; 45]. They and ST131 belong
375 to phylogroup B2, which would also allow us to address phylogenetic signal influencing the
376 presence of the *tss* region. The *tss* region was found to be consistently present in both
377 lineages (Figure 7a and b), demonstrating that T6SS presence varies between pathogenic *E.*
378 *coli* lineages within phylogroup B2.

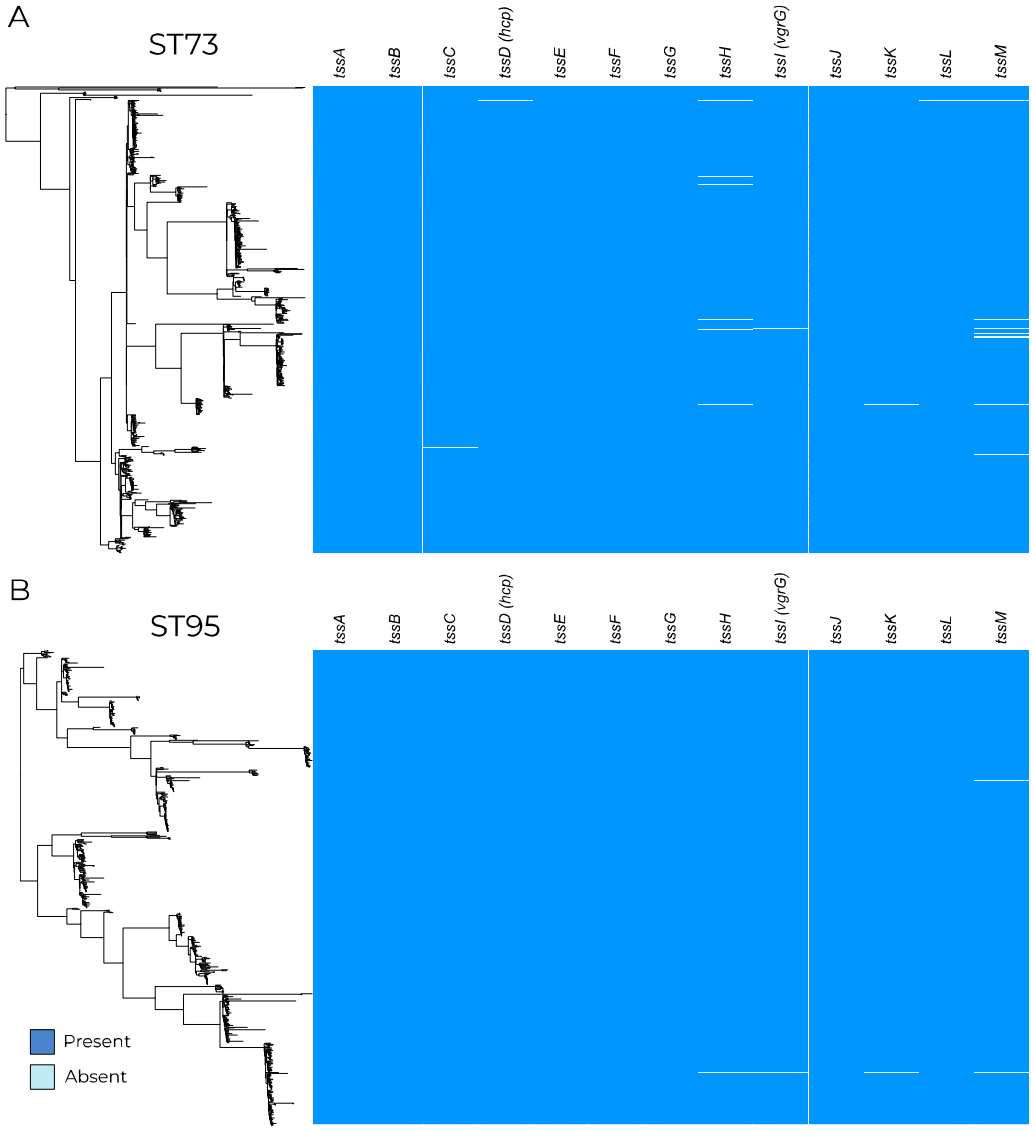


Fig. 7 Presence of structural experimentally validated type VI secretion genes taken from the SecReT6 database [28; 29] across (A) 873 ST73 genomes and (B) 758 ST95 genomes.

Discussion

Extensive work has been done to characterise 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 [34; 35; 36; 46]. We have identified the evolutionary events that led to the loss

390 of key structural T6SS components in these lineages via recombination or insertion events.
391 Our study has yielded new insights on the evolution of T6SS functionality in two of the most
392 clinically-relevant MDR *E. coli* lineages, and has provided a broader overview of T6SS
393 determinants in other lineages.
394
395 In both ST410 and ST131, we have observed a loss of T6SS that appears to coincide with the
396 acquisition of MDR. This suggests we may have uncovered a common evolutionary
397 trajectory in two MDR lineages. We have previously described the important role of
398 potentiating mutations involved in colonisation factor determinants, anaerobic metabolism
399 genes, and intergenic regions as being common evolutionary trajectories in MDR clone
400 formation [47]. The loss or degradation of a T6SS could perhaps integrate into the stepwise
401 evolutionary course of ST410-B4/H24RxC and ST131-C2/H30Rx. The structural components
402 for a complete T6SS were chromosomal and conserved within the ancestral clades ST410-
403 A/H53, ST131-A/H41 and ST131-B/H22. We found that the T6SS in ST410-B/H24 was
404 inactivated by deletion events and that the T6SS in ST131-C/H30 was inactivated by an
405 insertion event. In ST131, the rare instances where *tssM* is interrupted in clade A/H41 and
406 B/H22, and uninterrupted in clade C/H30 (Figure 6a), might be explained by inter-lineage
407 horizontal gene transfer and homologous recombination, but the T6SS region is too well-
408 conserved (>99.99% ID between lineages) to accurately detect recombinant sequences. In
409 both ST131 and ST410, the inactivation of T6SS occurring prior to or during the formation of
410 MDR clades implies that these events may have contributed to their successful emergence
411 and expansion, but we cannot generalise to all MDR lineages as further analyses of the
412 common MDR lineages ST167 and ST648, not shown here, showed intact T6SS regions.
413
414 It has previously been suggested that acquiring MDR plasmids is a vital step in the evolution
415 of pandemic clones [47]. We speculate that the lack of a functional T6SS allowed both
416 ST410-B/H24 and ST131-C/H30 to become more receptive to cell-to-cell contact, and
417 therefore conjugative transfer. The deleterious effect of T6SS on plasmid conjugation has
418 been demonstrated in *Acinetobacter baumannii*, where a conjugative plasmid has been
419 shown to repress the host T6SS in order to increase conjugation rates [26]. While the *A.*
420 *baumannii* experiments involved inhibition of T6SS in donor cells, the absence of T6SS in
421 potential recipient cells seems likely to increase their permissiveness to plasmid transfer.

422 The acquisition of conjugative MDR plasmids by ST410-B/H24 and ST131-C/H30 might
423 therefore have been facilitated by their loss of functional T6SS.
424
425 Our genomic analyses, combined with existing literature, suggest that the MDR clone ST131-
426 C2/H30Rx does not have a functional T6SS due to the interruption of a single gene, *tssM*.
427 However, the functionality of the uninterrupted *tss* region in the ancestral ST131 clades
428 A/H41 and B/H22 has not been verified experimentally. Further work to determine the
429 functionality of the T6SS in all three ST131 clades is therefore required. Final experimental
430 validation would require complementation of *tssM* to restore function of the T6SS in ST131-
431 C/H30.
432
433 The data presented here does not support the hypothesis that the ability of drug resistant *E.*
434 *coli* to displace resident commensal *E. coli* is due to the production of a T6SS. Phage,
435 colicins, or other diffusible elements are alternative explanations that should be considered.
436 By focusing on T6SS, we have uncovered parallel evolutionary outcomes in ST410-
437 B4/H24RxC and ST131-C2/H30Rx, where T6SS determinants have been lost by MDR clones.
438 Our work suggests that the loss of functional T6SS might have played a role in the evolution
439 and success of the pandemic MDR *E. coli* clones in ST131 and ST410.
440

441 # Author Statements

442

443 ## Author Contributions

444 E.A.C.: data curation, methodology, formal analysis, visualisation, writing – original draft
445 preparation and review and editing. R.A.M.: methodology, formal analysis, writing – original
446 draft preparation and review and editing. A.E.S: methodology. R.J.H.: writing - review and
447 editing. C.H.C.: data curation. S.J.D.: data curation. A.M.: conceptualisation, supervision,
448 writing – review and editing.
449

450 ## Conflicts of interest

451

452 The authors declare that there are no conflicts of interest.

Funding Information

EAC 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.

References

[1] Cummins EA, Hall RJ, Connor C, McInerney JO, McNally A. Distinct evolutionary trajectories in the *Escherichia coli* pangenome occur within sequence types. Microbial Genomics. 2022;8(11).

[2] Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet. 2022 2;399(10325):629-55.

[3] Connor CH, Zucoloto AZ, Yu IL, Corander J, McDonald B, McNally A, 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.

[4] Arcilla MS, van Hattem JM, Haverkate MR, Bootsma MCJ, van Genderen PJJ, Goorhuis A, et al. Import and spread of extended-spectrum β-lactamase producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. The Lancet Infectious Diseases. 2017 1;17(1):78-85.

[5] 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 11;9(6).

484

485 [6] Kantele A, Kuenzli E, Dunn SJ, Dance DAB, Newton PN, Davong V, et al. Dynamics of
486 intestinal multidrug-resistant bacteria colonisation contracted by visitors to a high-endemic
487 setting: a prospective, daily, real-time sampling study. The Lancet Microbe.
488 2021;2(4):e1518.
489

490 [7] Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, et al. Systematic
491 longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population
492 structure only transiently disturbed by the emergence of ST131. Genome Research.
493 2017;27(8):1437-49.
494

495 [8] Gladstone RA, McNally A, Pontinen AK, Tonkin-hill G, Lees JA, Skyten K, et al. Emergence
496 and dissemination of antimicrobial resistance in *Escherichia coli* causing bloodstream
497 infections in Norway in 2002 – 17 : a nationwide, longitudinal , microbial population
498 genomic study. The Lancet Microbe. 2021;5247(21):1-11.
499

500 [9] Davies M, Galazzo G, van Hattem JM, Arcilla MS, Melles DC, de Jong MD, et al.
501 Enterobacteriaceae and Bacteroidaceae provide resistance to travel-associated intestinal
502 colonization by multi-drug resistant *Escherichia coli.* Gut Microbes. 2022;14(1).
503

504 [10] Mariano G, Trunk K, Williams DJ, Monlezun L, Strahl H, Pitt SJ, et al. A family of Type VI
505 secretion system effector proteins that form ion-selective pores. Nature Communications.
506 2019 12;10(1).
507

508 [11] Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ. Type VI secretion system
509 translocates a phage tail spike-like protein into target cells where it cross-links actin.
510 Proceedings of the National Academy of Sciences of the United States of America. 2007
511 9;104(39):15508-13.
512

513 [12] Salomon D, Klimko JA, Trudgian DC, Kinch LN, Grishin NV, Mirzaei H, et al. Type VI
514 Secretion System Toxins Horizontally Shared between Marine Bacteria. PLoS Pathogens.
515 2015 8;11(8).

516

517 [13] Bernal P, Allsopp LP, Filloux A, Llamas MA. The *Pseudomonas putida* T6SS is a plant
518 warden against phytopathogens. ISME Journal. 2017 4;11(4):972-87.
519

520 [14] Troselj V, Treuner-Lange A, S.gaard-Andersen L, Wall D. Physiological Heterogeneity
521 Triggers Sibling Conflict Mediated by the Type VI Secretion System in an Aggregative
522 Multicellular Bacterium. mBio. 2018;9(1).
523

524 [15] Tang L, Yue S, Li GY, Li J, Wang XR, Li SF, et al. Expression, secretion and bactericidal
525 activity of type VI secretion system in *Vibrio anguillarum*. Archives of Microbiology. 2016
526 10;198(8):751-60.
527

528 [16] Unni R, Pintor K, Diepold A, Unterweger D. Presence and absence of type VI secretion
529 systems in bacteria. Microbiology. 2022; in print:1-13.
530

531 [17] KempnichMW, Sison-Mangus MP. Presence and abundance of bacteria with the Type VI
532 secretion system in a coastal environment and in the global oceans. PLoS ONE. 2020
533 12;15(12 December 2020).
534

535 [18] Garcia-Bayona L, Coyne MJ, Comstock LE. Mobile Type VI secretion system loci of the
536 gut Bacteroidales display extensive intra-ecosystem transfer, multi-species spread and
537 geographical clustering. PLoS Genetics. 2021 4;17(4).
538

539 [19] Abby SS, Cury J, Guglielmini J, N.ron B, Touchon M, Rocha EPC. Identification of protein
540 secretion systems in bacterial genomes. Scientific Reports. 2016 3;6.
541

542 [20] Bingle LE, Bailey CM, Pallen MJ. Type VI secretion: a beginner's guide. Current Opinion
543 in Microbiology. 2008 2;11(1):3-8.
544

545 [21] Journet L, Cascales E. The Type VI Secretion System in *Escherichia coli* and Related
546 Species. EcoSal Plus. 2016;7(1).
547

548 [22] Ma J, Bao Y, Sun M, Dong W, Pan Z, Zhang W, et al. Two functional type VI secretion
549 systems in avian pathogenic *Escherichia coli* are involved in different pathogenic pathways.
550 Infection and Immunity. 2014;82(9):3867-79.
551
552 [23] De Pace F, Nakazato G, Pacheco A, De Paiva JB, Sperandio V, Da Silveira WD. The type VI
553 secretion system plays a role in type 1 fimbria expression and pathogenesis of an avian
554 pathogenic *Escherichia coli* strain. Infection and Immunity. 2010 12;78(12):4990-8.
555
556 [24] Brunet YR, Espinosa L, Harchouni S, Mignot T, Cascales E. Imaging Type VI Secretion-
557 Mediated Bacterial Killing. Cell Reports. 2013;3(1):36-41.
558
559 [25] Weber BS, Ly PM, Irwin JN, Pukatzki S, Feldman MF. A multidrug resistance plasmid
560 contains the molecular switch for type VI secretion in *Acinetobacter baumannii*. Proceedings
561 of the National Academy of Sciences of the United States of America. 2015;112(30):9442-7.
562
563 [26] Venanzio GD, Moon KH, Weber BS, Lopez J, Ly PM, Potter RF, et al. Multidrug-resistant
564 plasmids repress chromosomally encoded T6SS to enable their dissemination. Proceedings
565 of the National Academy of Sciences of the United States of America. 2019;116(4):1378-83.
566
567 [27] Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M. The EnteroBase user's guide, with
568 case studies on Salmonella transmissions, *Yersinia pestis* phylogeny, and Escherichia core
569 genomic diversity. Genome Research. 2020;30(1):138-52.
570
571 [28] Li J, Yao Y, Xu HH, Hao L, Deng Z, Rajakumar K, et al. SecReT6: A web-based resource for
572 type VI secretion systems found in bacteria. Environmental Microbiology. 2015
573 7;17(7):2196-202.
574
575 [29] Zhang J, Guan J, Wang M, Li G, Djordjevic M, Tai C, et al. SecReT6 update: a
576 comprehensive resource of bacterial Type VI Secretion Systems. Science China Life Sciences.
577 2022 11.
578

579 [30] Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre
580 for bacterial insertion sequences. Nucleic acids research. 2006;34(Database issue).
581
582 [31] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
583 Architecture and applications. BMC Bioinformatics. 2009 12;10.
584
585 [32] Colpan A, Johnston B, Porter S, Clabots C, Anway R, Thao L, et al. *Escherichia coli*
586 sequence type 131 (ST131) subclone h30 as an emergent multidrug-resistant pathogen
587 among US Veterans. Clinical Infectious Diseases. 2013 11;57(9):1256-65.
588
589 [33] Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies
590 from short and long sequencing reads. PLoS Computational Biology. 2017 6;13(6).
591
592 [34] Roer L, Overballe-Petersen S, Hansen F, Schønning K, Wang M, Røder BL, et al.
593 *Escherichia coli* Sequence Type 410 Is Causing New International High-Risk Clones. mSphere.
594 2018;3(4):00337-18.
595
596 [35] Petty NK, Zakour NLB, Stanton-Cook M, Skippington E, Totsika M, Forde BM, et al.
597 Global dissemination of a multidrug resistant *Escherichia coli* clone. Proceedings of the
598 National Academy of Sciences of the United States of America. 2014;111(15):5694-9.
599
600 [36] Feng Y, Liu L, Lin J, Ma K, Long H, Wei L, et al. Key evolutionary events in the emergence
601 of a globally disseminated, carbapenem resistant clone in the *Escherichia coli* ST410 lineage.
602 Communications Biology. 2019;2(1):1-13.
603
604 [37] Kostiuk B, Santoriello FJ, Diaz-Satizabal L, Bisaro F, Lee KJ, Dhody AN, et al. Type VI
605 secretion system mutations reduced competitive fitness of classical *Vibrio cholerae* biotype.
606 Nature Communications. 2021;12(1):1-11.
607
608 [38] Ma J, Sun M, Pan Z, Lu C, Yao H. Diverse toxic effectors are harbored by vgrG islands for
609 interbacterial antagonism in type VI secretion system. Biochimica et Biophysica Acta -
610 General Subjects. 2018 7;1862(7):1635-43.

611

612 [39] Ma J, Sun M, Dong W, Pan Z, Lu C, Yao H. PAAR-Rhs proteins harbor various C-terminal
613 toxins to diversify the antibacterial pathways of type VI secretion systems. Environmental
614 Microbiology. 2017 1;19(1):345-60.

615

616 [40] Silverman JM, Austin LS, Hsu F, Hicks KG, Hood RD, Mougous JD. Separate inputs
617 modulate phosphorylation-dependent and -independent type VI secretion activation.
618 Molecular Microbiology. 2011 12;82(5):1277-90.

619

620 [41] Salomon D, Kinch LN, Trudgian DC, Guo X, Klimko JA, Grishin NV, et al. Marker for type
621 VI secretion system effectors. Proceedings of the National Academy of Sciences of the
622 United States of America. 2014 6;111(25):9271-6.

623

624 [42] Forde BM, Ben Zakour NL, Stanton-Cook M, Phan MD, Totsika M, Peters KM, et al. The
625 complete genome sequence of *Escherichia coli* EC958: A high quality reference sequence for
626 the globally disseminated multidrug resistant *E. coli* O25b:H4-ST131 clone. PLoS ONE. 2014
627 8;9(8).

628

629 [43] Ma LS, Narberhaus F, Lai EM. IcmF family protein TssM exhibits ATPase activity and
630 energizes type VI secretion. Journal of Biological Chemistry. 2012 5;287(19):15610-21.

631

632 [44] Zhang L, Xu J, Xu J, Chen K, He L, Feng J. TssM is essential for virulence and required for
633 type VI secretion in *Ralstonia solanacearum*; 2012. 4. A

634

635 [45] Lipworth S, Vihta KD, Chau K, Barker L, George S, Kavanagh J, et al. Ten-year
636 longitudinal molecular epidemiology study of *Escherichia coli* and *Klebsiella* species
637 bloodstream infections in Oxfordshire, UK. Genome Medicine. 2021 12;13(1).

638

639 [46] Stoesser N, Sheppard AE, Pankhurst L, de Maio N, Moore CE, Sebra R, et al.
640 Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131.
641 mBio. 2016;7(2):1-15.

642

643 [47] Cummins EA, Snaith AE, McNally A, Hall RJ. The role of potentiating mutations in the
644 evolution of pandemic *Escherichia coli* clones. European Journal of Clinical Microbiology &
645 Infectious Diseases. 2021.