bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Emergence of enteroaggregative *Escherichia coli* within the ST131 lineage as a cause
2 of extraintestinal infections

3

4 Erik J. Boll[1*], Marc Stegger[1,2*#], Henrik Hasman[1], Louise Roer[1], Søren Overballe-Petersen[1], Kim
5 Ng[1], Flemming Scheutz[1], Anette M. Hammerum[1], Arnold Dungu[3], Frank Hansen[1], Berit Lilje[1],
6 Dennis Schrøder Hansen[3], Karen A. Krogfelt[1], Lance B. Price[2,4], James R. Johnson[5], Carsten Struve[1]
7 and Bente Olesen[3]
8 ** These authors contributed equally to this work.*

9

10 [1] Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Copenhagen, Denmark
11 [2] Department of Environmental and Occupational Health, Milken Institute School of Public Health,
12 The George Washington University, Washington, DC, USA
13 [3] Department of Clinical Microbiology, Herlev and Gentofte Hospital, University of Copenhagen,
14 Denmark
15 [4] Division of Pathogen Genomics, Translational Genomics Research Institute, Flagstaff, Arizona,
16 USA
17 [5] VA Medical Center, Minneapolis, Minnesota, USA; University of Minnesota, Minneapolis,
18 Minnesota, USA

19

20 Running title: Emergence of ESBL-producing EAEC ST131

21

22 **Corresponding author:**
23 [#] Marc Stegger. Mail: mtg@ssi.dk

24

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25 Word count: 250 abstract, 4849 main article.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26 **Abstract**

27 *Escherichia coli* sequence type 131 (ST131) is a major cause of urinary and bloodstream infections
28 and its association with extended-spectrum β-lactamases (ESBL) significantly complicates treatment.
29 Most notorious is its rapidly expanding *H*30-Rx clade (named for containing allele 30 of the type-1
30 fimbrial adhesin gene *fimH* and extensive antimicrobial resistance), which appears to have emerged
31 in the United States due in part due to the acquisition of the ESBL-encoding $bla_{\text{CTX-M-15}}$ gene and
32 resistance to fluoroquinolones. However, non-*H*30 ST131 lineages with acquired CTX-M-type
33 resistance genes also are emerging. Based on whole-genome analyses, we describe here the presence
34 of an (*fimH*) *H*27 *E. coli* ST131 lineage that currently is causing an outbreak of community-acquired
35 bacteremia and recurrent urinary tract infections (UTIs) in Denmark. This lineage has acquired both
36 a virulence plasmid (pAA) that defines the enteroaggregative *E. coli* (EAEC) diarrheagenic
37 pathotype and multiple genes associated with extraintestinal *E. coli* (ExPEC) that combined has
38 made this particular ST131 lineage highly successful at colonizing its human host and cause
39 recurrent UTI. Moreover, using a historic World Health Organization *E. coli* collection and
40 publically available genome sequences, we identify a global *H*27 EAEC ST131 lineage dating back
41 as far as 1998. Most *H*27 EAEC ST131 isolates harbor pAA or pAA-like plasmids, which analysis
42 strongly imply was caused by a single ancestral acquisition. These findings illustrate the profound
43 plasticity of this important pathogenic *E. coli H*27 lineage in general, and the genetic acquisitions of
44 EAEC-specific virulence traits that likely confer an enhanced ability to cause intestinal colonization.

45

46 **Importance**

47 The *E. coli* ST131 lineage is a notorious extraintestinal pathogen. A signature characteristic of
48 ST131 is its ability to asymptomatically colonize the gastrointestinal tract and then opportunistically
49 cause extraintestinal infections, such as cystitis, pyelonephritis and urosepsis. In this study, we report
50 a novel ST131 sublineage that has acquired the enteroaggregative diarrheagenic phenotype, spread

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

51 across multiple continents and has been associated with multiple outbreaks of community-acquired
52 bloodstream infections in Denmark. The strain's ability to both cause diarrhea and colonize the
53 human gastrointestinal tract may facilitate its dissemination and establishment in the community,
54 whereas the strain's clonal nature may facilitate targeted control strategies, such as vaccination.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

55 **Introduction**

56 *Escherichia coli* sequence type 131 (ST131) is the dominant multidrug-resistant (MDR)
57 extraintestinal pathogenic *E. coli* (ExPEC) lineage worldwide, causing a wide range of infections
58 including bloodstream and urinary tract infections (BSIs/UTIs) (*1-3*). Its rise to global dominance
59 and pathogenicity is thought to have been primed by sequential acquisition of virulence-associated
60 genes followed by development of antibiotic resistance (*4, 5*). ST131 predominantly exhibits
61 serotype O25:H4 and is closely associated with fluoroquinolone resistance and the production of the
62 CTX-M-15 extended-spectrum β-lactamase (ESBL) (*2, 6*). The expansion of *E. coli* ST131 in the
63 United States has been driven mainly by a single clade, designated *H*30 because of its tight
64 association with allele 30 of the type-1 fimbrial adhesin gene, *fimH*. *H*30 has a prominent MDR-
65 associated clade, *H*30R, which accounts for most fluoroquinolone resistance within ST131. *H*30R in
66 turn has two main sub-clades: *H*30R, which accounts for almost all ST131-associated CTX-M-14
67 and CTX-M-27 production (although most members are ESBL-negative), and *H*30Rx, which
68 accounts for almost all ST131-associated CTX-M-15 production (*6*). In addition to *H*30, other
69 distinct clades of ST131 are also circulating worldwide, most commonly associated to the 22 and 41
70 *fimH* alleles. While these non-*H*30 allelic variants are normally not associated with carriage of CTX-
71 M-genes, cases of $bla_{CTX\text{-}M\text{-}14}$, $bla_{CTX\text{-}M\text{-}15}$ and $bla_{CTX\text{-}M\text{-}27}$ acquisition by ST131 lineages carrying
72 *fimH*22 or –*H*41 have been reported (*4, 7*).

73 ST131 isolates typically carry multiple ExPEC-associated virulence genes encoding
74 adhesins, toxins, and siderophores, whereas virulence genes typical of diarrheagenic *E. coli* (DEC)
75 rarely have been reported (*8, 9*). However, we recently surveyed an international World Health
76 Organization (WHO) collection of historic *E. coli* for ST131 O25 isolates, assessing them for
77 temporal trends of antibiotic resistance and virulence traits. Among a total of 128 ST131 isolates we
78 found 12 (9%) that fulfilled molecular criteria for the enteroaggregative *E. coli* (EAEC) pathotype
79 (*10*). Pulsed-field gel electrophoresis (PFGE) analysis revealed a cluster comprising seven of these

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

80 EAEC isolates. Of these, six – including two urine isolates from patients with UTI, three fecal
81 isolates from patients with diarrhea, and one lower respiratory tract isolate (associated symptoms
82 unknown) – were from Danish patients (1998-2000), supporting the occurrence of an unrecognized
83 EAEC ST131-associated outbreak of UTI and possibly also diarrhea in this time period in Denmark
84 (*10*).

85 EAEC strains cause endemic diarrheal illness in developing countries and foodborne
86 outbreaks in developed countries, and have been associated with extraintestinal infections (*11*). This
87 pathotype gained particular attention following a major foodborne outbreak in Germany in 2011 that
88 was caused by a Shiga toxin (Stx)-producing O104:H4 EAEC strain and resulted in 3,842 confirmed
89 cases and 54 deaths (*12*). Additionally, a worryingly high prevalence of multidrug resistance among
90 EAEC strains has been reported, and reports of CTX-M-type ESBL-producing EAEC have appeared
91 recently from around the world (*13-16*). EAEC pathogenesis involves adherence to human intestinal
92 mucosa by virtue of aggregative adherence fimbriae (AAF) and subsequent biofilm formation. The
93 AAF are encoded on large plasmids designated pAA that also encodes a suite of other EAEC
94 virulence factors, including AggR, a global regulator of EAEC virulence; dispersin, required for
95 proper dispersal of AAFs on the bacterial surface; the AatPABCD transporter system, which
96 mediates dispersin secretion; and Aar, a recently described negative regulator of AggR (*17, 18*).

97 To understand the emergence and underlying genetic acquisitions leading to ESBL-
98 producing EAEC ST131, we investigated the relatedness of a large collection of temporally and
99 spatially diverse ST131 isolates. This investigation revealed the emergence of a global *H*27 lineage
100 of EAEC linked to a single acquisition of the pAA plasmid that likely improved the strain's ability to
101 persistently colonize its human host.

102

103 **Results**

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

104 **Identification of historic Danish EAEC ST131**. Our previous PFGE-based analysis of ST131 O25
105 isolates within the WHO *E. coli* collection identified an apparently unrecognized outbreak of UTI,
106 and possibly also diarrhea, in Denmark in 1998-2000 that was caused by a group of seemingly
107 similar EAEC isolates (*10*). To better estimate these isolates' relatedness and to correct for
108 recombination (*6*), all 128 ST131 isolates from the WHO collection were subjected to whole-genome
109 sequencing. The *in silico* analyses regarding the genotypic features of the isolates revealed that the
110 most frequent *fimH* alleles were *H*30 (51%, n=65), *H*22 (34%, n=44) and *H*27 (7%, n=9). Most of
111 the isolates harbored multiple antibiotic resistance genes. The most common ESBL gene was bla_{CTX-}
112 $_{M-15}$ (37%, n=47), which occurred almost exclusively within the *H*30 subgroup (98%, 46/47). See
113 **Suppl. Table 1.**

114 To ensure that only high-quality genome data sets were included in our SNP analysis,
115 we excluded the genome data for seven (5%) of the 128 ST131 isolates due to low sequencing depth.
116 In addition, we included genome sequences from a published collection of human clinical ST131 *E.*
117 *coli* (n = 93) collected in the USA and Germany between 2010 and 2012 (*6*). Across these genomes,
118 11,529 SNPs were identified within ~38.4% of the JJ1886 genome that was conserved across all
119 isolates. After purging of recombinant regions, 4,241 SNPs were used to infer relationships between
120 the isolates.

121 The SNP-based phylogeny showed distinct overall clustering of isolates in accordance
122 to their *fimH* alleles, with all 116 isolates carrying the *fimH*30 allele clustering as a monophyletic
123 clade (**Suppl. Fig. 1**). Likewise, all 10 isolates carrying the *fimH*27 allele clustered within a single
124 clade which – like the *H*30 clade – appeared to descend from an ancestral *H*22 group. Of these
125 *fimH*27-carrying isolates, eight – all from the WHO collection – qualified molecularly as EAEC
126 based on presence of ≥1 of the EAEC-associated putative virulence genes *aggR*, *aatA*, and *aaiC (19)*.
127 On the basis of previous PFGE data (*10*), seven of these eight EAEC isolates had formed a cluster,
128 whereas the remaining isolate (C796-00) appeared quite distinct from the others. By contrast, here all

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

129 eight isolates clustered together, with C796-00 being most closely related to three of the other seven
130 EAEC isolates (**Suppl. Fig. 1**). Included in this cluster of *fimH*27-carrying EAEC isolates were C86-
131 04, a fecal isolate from Vietnam (2004), and seven Danish isolates, including two urine isolates, one
132 lower respiratory tract isolate, and four fecal isolates from patients with diarrhea, suggesting that this
133 lineage was able to cause both extraintestinal infections and diarrhea.

134 Unlike the PFGE analysis, the SNP-based analysis additionally revealed a distinct
135 cluster of isolates located within the *H*22 group consisting of the remaining four EAEC isolates from
136 the WHO collection (**Suppl. Fig. 1**). Three of the isolates carried *fimH*22*,* whereas one isolate
137 carried *fimH*298, which differs from *fimH*22 by only a single nucleotide. All four isolates were from
138 urine samples collected within a 10-month period in 1998, from elderly patients admitted to four
139 different departments of the same hospital in the Capital Region in Denmark, suggesting an
140 unrecognized nosocomial UTI outbreak. With the exception of Vietnamese isolate C86-04, all 11
141 EAEC isolates were recovered from Danish patients. The Vietnamese isolate carried $bla_{CTX\text{-}M\text{-}27}$,
142 whereas no ESBL genes were present in the other 11 EAEC isolates. (**Suppl. Table 1**).

143

144 **EAEC virulence genes in historic Danish EAEC ST131.** The four clustered EAEC urine isolates
145 that carried *fimH*22 or *fimH*298 all harbored the *aggDCBA* gene cluster encoding the AAF/I variant,
146 whereas seven of the eight clustered *fimH27*-carrying EAEC isolates of diverse sources all harbored
147 *agg5DCBA* encoding the recently described AAF/V variant (*20*). The remaining *H*27 isolate, C1883-
148 99, lacked the *agg5A* gene (**Suppl. Table 2**). All but two of the 12 EAEC isolates contained genes
149 encoding AggR, Aar, dispersin (*aap*), and the AatPABCD transporter system, plus two additional
150 AggR-regulated open reading frames (ORFs), ORF3 and ORF4, that are assumed to play a role in
151 isoprenoid biosynthesis (*21*). The remaining two isolates, C1883-99 (*H*27) and C167-00 (*H*22), both
152 lacked most of these EAEC-specific genes, implying partial deletion of the pAA plasmid in isolates
153 from both these *fimH*-associated clades.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

154

155 **Invasive EAEC ST131 isolates with *bla*$_{\text{CTX-M-101}}$ obtained from Danish patients**. To determine
156 whether EAEC ST131 strains were present among contemporary Danish patients, we searched for
157 EAEC-specific virulence genes within a recently published collection of 552 whole-genome
158 sequenced ESBL-producing *E. coli* (ESBL-*Ec*) isolates obtained from patients with BSIs in Denmark
159 between 2014 and 2015 (*22*). ST131 accounted for 50% (n = 258) of the isolates (*22*). Among these,
160 25 isolates harbored the genes encoding AAF/V, AggR, Aar, Aap, AatPABCD, and ORF3/4 (**Suppl.**
161 **Table 2**). All 25 carried *fimH*27 and *bla*$_{\text{CTX-M-101}}$. In total, 27 isolates in the collection were *bla*$_{\text{CTX-M-}}$
162 $_{\text{101}}$-positive (*10*), of which thus 93% qualified molecularly as EAEC. The EAEC pathotype did not
163 occur in conjunction with any other ESBL-encoding gene; hence, the association between *bla*$_{\text{CTX-M-}}$
164 $_{\text{101}}$ and the EAEC pathotype was highly significant (χ^2, p<0.0001), suggesting that the resistance
165 gene may be co-localized with EAEC-specific virulence genes on the pAA plasmid. As shown by
166 Roer *et al.*, the 27 ST131 strains with *bla*$_{\text{CTX-M-101}}$ formed a distinct cluster with nine or fewer SNP
167 differences, supporting a recent emergence (*22*).

168 To assess whether the same ESBL-EAEC ST131 strain could be detected in the urine
169 of the patients from whom the positive blood samples were collected, urine isolates from eight of the
170 258 patients with an ESBL-EAEC ST131 BSI were available and sequenced. Notably, all of these
171 patients had presented with UTI either prior to, concurrent with, or subsequent to the time of blood
172 sampling. Thus, some patients provided more than one urine sample. Specifically, one patient
173 presented 19 days after an initial bacteremia episode with both recurrent bacteremia and UTI, all
174 involving the ESBL-EAEC ST131 strain, and four other patients had recurrent same-strain UTI
175 diagnosed from three weeks to eight months after their initial UTI episode. SNP analysis
176 demonstrated that the corresponding blood and urine isolates were nearly identical (i.e., differed by
177 ≤6 SNPs), whereas different pairs were separated by zero to 11 SNPs. Of the eight pairs, five
178 clustered distinctively in the phylogeny thus strongly implying their relatedness. For the remaining

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

179 three pairs of blood and urine isolates, the data were inconclusive (**Suppl. Fig. 2**). These findings
180 suggest strongly that the ESBL-EAEC ST131 strains were capable of long-term persistence in these
181 hosts and/or their immediate environment and of causing repeated episodes of UTI and bacteremia.
182
183 **pAA plasmid characterization in ST131 isolates.** To characterize in detail the pAA plasmids
184 present in the ESBL-EAEC ST131 strains, we applied MinION sequencing (Oxford Nanopore
185 Technologies) to obtain the complete plasmid sequence for one such plasmid, which we designated
186 pAA-ST131, from a single representative Danish *E. coli* isolate (ESBL20150001, Sequence Read
187 Archive (SRA) ID: DF215RRW). The complete plasmid sequence was 142,646 bp in length, with an
188 average G+C content of 48.7% (**Fig. 1,** GenBank accession no. KY706108). A total of 197 open
189 reading frames (ORFs) were predicted and annotated, of which 142 were functionally assigned. Two
190 replicons were identified, RepFII and RepFIB, with the multireplicon F plasmid FAB formula of
191 F1:A-:B33.
192 pAA-ST131 contained several genes associated with plasmid stability, including *parA*,
193 *parM*, and *stbB*, plus three toxin-antitoxin (TA)-based addiction systems: *ccdAB*, *vagCD*, and *relBE*
194 (**Fig. 1**). Furthermore, it harbored a 33 kb complete *tra* region encoding transfer components (24 *tra*
195 genes, 8 *trb* genes, and *finO*), implying that the plasmid may be conjugally transferable. It also
196 contained genes encoding putative transmembrane proteins and proteins involved in catabolism and
197 metabolism, plus several integrated mobile elements. It contained all the EAEC virulence factor
198 genes already identified by Illumina sequencing (as expected), but no additional known virulence
199 factor genes or antibiotic resistance genes including $bla_{\text{CTX-M-101}}$. Because isolate ESBL20150001
200 contains no additional plasmids (**Suppl. Fig. 3**), we conclude that $bla_{\text{CTX-M-101}}$ is chromosomally
201 located.
202 A BLAST comparison of pAA-ST131 with five complete pAA plasmids (Genbank
203 accession IDs NC_018666 (O104:H4), NC_011752 (55989), FN554767 (042) and NC_008460

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

204 (DIJ1) and SRA ID SRA055981 (226)) showed that all these plasmids shared common features (**Fig.**
205 **1**), mostly corresponding to EAEC virulence genes. The entire genomic sequences were not available
206 for all corresponding EAEC isolates but they belonged to at least three STs (ST40, ST414, and
207 ST678). However, they also varied substantially for genetic content, as described previously for *E.*
208 *coli* virulence plasmids (*23*). Specifically, regions containing putative metabolic, catabolic, and
209 transmembrane proteins were unique to pAA-ST131 (**Fig. 1**). Additionally, although (like pAA-
210 ST131) four of the five reference pAA plasmids also contained the RepFIB and/or the RepFIIA
211 replicon, only three contained both. Moreover, only two contained the *tra* region.
212
213 **Global emergence of EAEC ST131 isolates carrying *fimH*27.** To assess the global extent of
214 EAEC ST131 isolates, we screened for *aggR* (the global regulator of EAEC virulence) among all
215 >3,500 *E. coli* ST131 genomes (as of November 15th 2017) available in EnteroBase
216 (http://enterobase.warwick.ac.uk), a database with >100,000 genomic assemblies of enteric bacteria,
217 including *E. coli*. Disregarding the *aggR*-positive, $bla_{CTX-M-101}$-containing Danish EAEC isolates, we
218 identified another 25 international *aggR*-positive isolates, which carried the following *fimH* alleles
219 (no. of isolates, % of 25): *fimH*27 (18, 66%), *fimH*22 (one, 4%), *fimH*298 (two, 8%), *fimH*30 (one,
220 4%), *fim*H5 (two, 8%), and *fim*H54-like (one, 4%) (**Suppl. Table 2**). The *aggR* gene occurred in 18
221 (38%) of the 47 other *fimH*27-carrying ST131 isolates in EnteroBase.
222 To determine whether the 12 EAEC ST131 isolates from the WHO collection were
223 clonally related to other available EAEC ST131 isolates, a SNP-based phylogeny was constructed
224 using 1) all 121 ST131 isolates from the WHO collection (*10*); 2) 93 German/US ST131 isolates
225 from 2010-2012 (*6*); 3) the 27 Danish $bla_{CTX-M-101}$-containing ST131 isolates from 2014-2015 (*22*);
226 and 4) 46 international *fimH*27-carrying ST131 isolates from Enterobase. Focusing first on isolates
227 within the *H*22 clade (**Fig 2A**), WHO collection EAEC isolate C180-00, carrying *fimH298*, clustered
228 together with two *fimH*298 EAEC EnteroBase isolates (SRA IDs SRR2970775 and SRR2970774)

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

229 from Cambodia (2009-2010). In contrast, the three WHO collection EAEC isolates carrying *fimH*22
230 were not closely related to the only *fimH22* ST131 isolate from EnteroBase.

231 In the rooted phylogeny, most EAEC isolates nested within the *H*27 clade (**Fig. 2A**).
232 To improve resolution, this clade's 79 isolates (76 with *fimH27*, two with *fimH5*, one with *fimH54*)
233 were analyzed separately. The 27 $bla_{\text{CTX-M-101}}$-containing Danish EAEC isolates formed a distinct
234 cluster with very short branches indicating a recent emergence (**Fig. 2B**). By contrast, none of the
235 *fimH*27-carrying international (i.e., non-Danish) isolates clustered either with the Danish isolates or
236 with one another, suggesting that the Danish outbreak was confined to Denmark, and that no other
237 focal outbreaks were captured. The WHO collection's *H*27 clade EAEC isolates were intermingled
238 with isolates from the UK, Thailand, and Canada (**Fig. 2B**), suggesting global spread of a common-
239 ancestry strain.

240 To determine the mosaicism of the EAEC-specific virulence genes among *fimH*27-
241 carrying ST131 isolates, sequence reads from all 79 *H*27 clade isolates were mapped against pAA-
242 ST131 from ESBL20150001. Intriguingly, pAA-ST131 was highly conserved among the *H*27 EAEC
243 isolates (**Fig. 2B**). To validate the *in silico* results, plasmid gel profiling was done for representative
244 *fimH*27-carrying $bla_{\text{CTX-M-101}}$-containing EAEC isolates and EAEC isolates from the WHO
245 collection. This confirmed that all but three of the ten tested isolates contained a single plasmid of
246 conserved size (~140 kb). When mapping the sequence reads to pAA-ST131 these all showed 99-
247 100% coverage. The three exceptions included WHO isolate C1883-99, which harbored a slightly
248 smaller plasmid, and isolates ESBL20150196 and ESBL20150300 (both obtained from Danish
249 patients in 2015), each of which harbored a single 30-35-kb plasmid that, based on *in silico* mapping,
250 was the result of a single major deletion that left only the RepFIB replicon (**Suppl. Fig. 3**).

251 Next, we performed a SNP-based analysis of pAA-ST131 across all 57 *H*27 isolates
252 with more than 70% coverage of the plasmid, with the four *H*22 clade EAEC isolates from the WHO
253 collection used as an outgroup. This identified 197 SNPs within the ~25% of pAA-ST131 that was

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

254 conserved across all isolates. Phylogenetic analysis based on these SNPs showed that pAA-ST131 is
255 highly conserved among the *H*27 isolates (including the three with a truncated version of the
256 plasmid), but differs considerably between the *H*27 and non-*H*27 isolates (**Suppl. Fig. 4**), strongly
257 suggesting a single ancestral acquisition of the plasmid within the *H*27 clade, with subsequent partial
258 deletions (to give the observed smaller variants).

259 Finally, we performed a BLAST analysis to screen the *H*27 isolates for classical
260 ExPEC virulence genes (**Suppl. Table 3**). All *H*27 isolates carried the *sfa*, *iutA*, and *kpsM* II genes,
261 which qualified them molecularly as ExPEC (*24*). In addition, they all contained the *chuA*, *fyuA*, and
262 *yfcV* genes, which moreover qualified them molecularly as uropathogenic *E. coli* (UPEC) (*25*).

263

264 **Discussion**

265 *E. coli* ST131 is a notorious MDR ExPEC lineage associated with both UTI and BSI (*1*). Little is
266 known about which genes (whether promoting virulence or other phenotypes) make this clonal
267 lineage so successful. ST131 strains have been shown to exhibit high levels of genomic plasticity
268 including frequent recombination and plasmid flux, particularly involving IncF-type plasmids,
269 facilitating spread of antibiotic resistance and virulence genes (*4, 26*). Indeed, ST131 lineages exhibit
270 extensive variation with regard to acquired virulence genes (*2, 27*). Many typical ExPEC-associated
271 virulence factors – including P fimbriae, hemolysins, and factors conferring increased serum survival
272 and iron uptake – have been identified in ST131 isolates (*9, 10*). In contrast, virulence traits
273 associated with diarrheagenic *E. coli* (DEC) pathotypes have thus far largely been absent from
274 ST131 isolates *(10)*.

275 In developed countries, EAEC is known mostly as a cause of self-limiting diarrhea of
276 mild to moderate severity. Indeed, long-term carriage of EAEC has been suggested to lead to
277 colonization rather than infection (*11*). By contrast, in developing countries EAEC is a leading cause
278 of childhood diarrhea (*11, 28*). The pathogenic potential of EAEC is underscored by its ability to

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

279 cause major foodborne outbreaks of diarrheal disease (*19, 29, 30*). Intriguingly, recent studies have
280 associated EAEC with UTI, suggesting that what classically has been regarded as a DEC pathotype
281 may also qualify as ExPEC and cause both diarrhea and UTI (*31-33*).

282 Like other DEC pathotypes, EAEC has been shown to encompass diverse genetic
283 lineages, reflecting a high level of phylogenetic heterogeneity (*16, 34*). Although until recently no
284 extraintestinal EAEC ST131 had been documented, recent reports have described the occurrence in
285 certain ST131 subclonal lineages of EAEC-specific traits in geographically distinct areas (*10, 13,*
286 *16*). ESBL-producing ST38 and other ST types of various serotypes have been indicated as emerging
287 hybrid strains of UPEC and EAEC involved primarily in UTIs from Germany, the Netherlands and
288 the United Kingdom (*35*). In a study of fecal and urine isolates from Danish patients in 1998-2000,
289 we made the novel observation of EAEC ST131 of serotype O25 (*10*). Subsequently, another study
290 documented the emergence of a $bla_{CTX\text{-}M\text{-}14}$-containing EAEC ST131 O25:H4 strain in stool samples
291 of diarrheic patients in Japan from 2003 onwards (*13*).

292 Our previous PFGE analysis of Danish ST131 O25 *E. coli* isolates from the WHO
293 collection demonstrated that seven of the 12 identified EAEC strains were highly similar and
294 appeared to have been part of an unrecognized UTI and diarrhea outbreak in Denmark in 1998 to
295 2000 (*10*). Although the genomic approach we used here to analyze the same isolates confirmed the
296 suspected outbreak, it also yielded certain substantially different conclusions from the PFGE
297 analysis, including identification of a second distinct cluster of EAEC urine isolates. These
298 discrepancies between WGS SNP-based analysis and PFGE analysis among ST131 isolates
299 correspond with previous findings (6).

300 Here we also document the current presence of a novel ESBL-EAEC ST131 sublineage
301 as a cause of bacteremia in patients admitted to Danish hospitals. The isolates, as collected across
302 Denmark over a 16-month period in 2014-2015, all carried $bla_{CTX\text{-}M\text{-}101}$. Notably, to date the only
303 other reports of CTX-M-101-positive *E. coli* are from China (*36-38*). Roer *et al*. used a genomic

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

304 approach to establish that the Danish ESBL EAEC isolates were highly clonal, strongly suggesting a
305 recent common source (*22*). Here, by analyzing sequential urine isolates from eight of the patients
306 with the Danish outbreak-associated ESBL-EAEC ST131 bacteremia, we found that five patients had
307 recurrent UTI caused by the outbreak strain. The time of paired urine sampling ranged from one
308 week before to eight months following the initial blood sampling. The Danish ESBL-EAEC ST131
309 outbreak strain thus seems capable of persistently colonizing patients, resulting in occasional clinical
310 manifestations. Less likely, the patients may have been continuously exposed from an external
311 source.

312 Using the EnteroBase collection of ST131 genome sequences, we identified the global
313 presence of *fimH*27-carrying EAEC ST131 isolates. These isolates originated from Africa, Asia,
314 Europe and North America, and together with the Danish EAEC ST131 outbreak isolates spanned
315 more than two decades. Strikingly, all the EAEC ST131 strains studied here proved to be clonal, i.e.,
316 to share a common ancestor. Moreover, they share a common virulence plasmid, designated pAA-
317 ST131, which appears to have been acquired once in the *H*27 clade and transmitted vertically,
318 although it has undergone recombination and deletion events in some sublineages. MinION-based
319 sequencing of this plasmid in a representative ESBL-EAEC strain identified an array of classical
320 plasmid-encoded EAEC-defining virulence genes, including those encoding AAFs and the global
321 regulator of virulence AggR.

322 The acquisition of pAA plasmids by *E. coli* ST131 is interesting, considering that the
323 ST131 lineage is thought of as a host generalist (*39*), whereas EAEC appears to be highly adapted to
324 humans, without a natural animal reservoir (*40, 41*). However, the EAEC ST131 lineage described
325 here is reminiscent of the UTI outbreak in Copenhagen in 1991 caused by an *E. coli* O78:H10 clonal
326 group in that both fulfilled the molecular criteria for EAEC and contained multiple ExPEC virulence
327 genes (*32*). That the O78:H10 outbreak strains' EAEC-associated virulence factors were found to
328 increase uropathogenicity (*33*) suggests that this may be true also for EAEC ST131 strains. The

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

329 O78:H10 EAEC outbreak strain, which never was found outside the Capital Region of Denmark and
330 was not known to have caused BSI. In contrast, ST131 is a highly successful pandemic clonal group,
331 which makes the potential of future outbreaks of UTI and bacteremia caused by this novel strain, or
332 other EAEC ST131 lineages, a cause for serious concern.

333 Despite the genetic evidence that all 25 *bla*$_{CTX-M-101}$-containing Danish EAEC isolates
334 were derived from a common source, based on the limited available epidemiological data we have
335 been unable to establish a patient link or explain the spread of the lineage across regions. UTI is
336 often caused by enteric *E. coli* that enter the urinary tract via the fecal-perineal-urethral route, and in
337 some instances may have as their proximate external source food products or animals (*42*).
338 Interestingly, the gender ratio among the 25 present cases (68% females, 32% males), differs
339 significantly (p = 0.01) from that across the entire collection of 552 ESBL-producing ST131
340 bloodstream isolates from Danish hospitals in 2014-2015 (42% females, 58% males) (*22*). A similar
341 overrepresentation of women was observed among cases during the 2011 multinational European
342 outbreak caused by a novel multi-pathotype, Shiga-toxin-producing EAEC O104:H4 strain (*43*).
343 Bean sprouts were found the most likely vehicle of infection, and the high proportion of female cases
344 in that study was thought to be driven by the tendency of women to be more health conscious and
345 perhaps suggestive of a food-related source for the Danish ESBL EAEC ST131 outbreak. Further
346 screening, e.g. of fecal and urine samples from past and current patients who present with diarrhea or
347 UTI, are warranted to determine the true clinical impact of the ESBL EAEC ST131 strain and to
348 provide the demographic and epidemiological data needed to identify potential sources.

349 In conclusion, we hypothesize that acquisition of the pAA plasmid made this *H*27
350 ESBL EAEC ST131 lineage highly successful at persistently colonizing patients, thereby allowing it
351 to occasionally cause UTIs and diarrhea. The presence of multiple ExPEC virulence factors –
352 including P fimbriae, α-hemolysin, and Sat – in turn facilitates dissemination from the urinary tract
353 to the bloodstream. Furthermore, we have demonstrated the ability of a non-*H*30 ST131 lineage to

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

354 acquire EAEC-specific pAA virulence plasmids and to disseminate across multiple continents over
355 the past two decades. An *H*27 EAEC ST131 strain that has also acquired the $bla_{CTX\text{-}M\text{-}101}$ gene was
356 causing bacteremia outbreaks in several geographic regions in Denmark, seemingly associated with
357 recurrent infections. These findings emphasize the potential for different pathogens to evolve, thus
358 potentially generating important new pathotypes that require continuous vigilance. They also
359 illustrate the power of whole-genome sequencing to elucidate the historical and current molecular
360 epidemiology and evolution of emerging pathogens of high public health importance.

361

362 Materials and methods

363 *Genome data and sequencing*

364 WGS was performed on a previously described international collection of 128 ST131 *E. coli* human isolates of
365 serotype O25:H4 or O25:H- (1968 to 2011) from the WHO Collaborating Centre for Reference and Research
366 on *Escherichia* and *Klebsiella* (www.ssi.dk) (*10*). DNA samples were prepared for multiplexed, paired-end
367 sequencing using a combination of Illumina MiSeq and HiSeq (*6, 10*). We also included a published *E. coli*
368 ST131 sequence dataset comprising 93 US and German isolates of human and animal origin (1967 to 2011)
369 (*6*), and sequences from 27 ESBL-producing ST131 bloodstream isolates (carrying $bla_{CTX\text{-}M\text{-}101}$) collected from
370 Danish patients within the national surveillance program for antimicrobial resistance (DANMAP) for ESBL-
371 producing *E. coli* (2014 – 2015) (*22*). Furthermore, all >3,500 *E. coli* ST131 genome data available at
372 EnteroBase (http://enterobase.warwick.ac.uk, accessed November 15th 2017) were analyzed for EAEC
373 characteristics (presence for the *aggR* gene) and positive isolates were included. Finally, we sequenced and
374 included 13 *E. coli* isolates obtained from urine from the source patients for eight of the ESBL-producing
375 blood isolates. The sequences were analyzed using the Bacterial Analysis Platform (BAP) from Center for
376 Genomic Epidemiology (*44*).

377

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

378 *Plasmid sequencing and analysis*

379 Plasmid DNA was sequenced on both a MiSeq instrument (Illumina) and a MinION flow cell
380 (Oxford Nanopore Technologies). The MiSeq library was made using the Nextera XT kit (Illumina)
381 and sequencing was performed as a paired-end 250 bp run, yielding 372,720 reads with an average
382 length of 237 bp. The MinION library was prepared using the Genomic Sequencing SQK-MAP006
383 kit and was sequenced on a FLO-MAP003 Early Access flow cell according to the manufacturer's
384 instruction. Fast5 read files were subjected to base calling via a two-direction workflow using
385 Metrichor software (ONT), yielding 4,856 passed 2D read files. Mixed assembly was performed by
386 combining MiSeq and MinION reads using the SPAdes assembler (v3.9.0). Finally, CLCbio
387 Genomics Workbench (v9.5.2) was used for end trimming of the assembled plasmid and for final
388 error correction by mapping trimmed MiSeq reads against the plasmid contig obtained after the
389 mixed SPAdes assembly.

390 pAA-ST131 was annotated using RAST (*45*), with putative hypothetical genes curated manually
391 using NCBI BLASTn and BLASTp searches. A BLASTn atlas of pAA-ST131 and other pAA
392 virulence plasmids was constructed using BLAST Ring Image Generator v0.95 (BRIG) (*46*).

393

394 *Virulence genotyping*

395 Isolates qualified molecularly as EAEC if positive for ≥1 of EAEC-associated putative virulence
396 genes *aggR*, *aatA*, and *aaiC (10).* Isolates were regarded as ExPEC if positive for ≥2 of *papA* and/or
397 *papC* (P fimbriae), *sfa* and *foc* (S and F1C fimbriae), *afa* and *dra* (Dr-binding adhesins), *kspM* II
398 (group 2 capsule), and *iutA* (aerobactin siderophore system) (*24*). Isolates were considered UPEC if
399 positive for ≥2 of *chuA* (heme uptake), *fyuA* (yersiniabactin siderophore system), *vat* (vacuolating
400 toxin), and *yfcV* (adhesin) (*25*).

401

402 *Identification of SNPs*

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

403 SNPs in the core chromosome or plasmid, depending on analysis, were identified using the Northern
404 Arizona SNP Pipeline (NASP) (*47*). Briefly, duplicate regions of the reference chromosome JJ1886
405 or pAA-ST131 plasmid (GenBank accession no. CP006784 and KY706108, respectively), were
406 identified by aligning the reference against itself with NUCmer (*48*), followed by mapping of
407 Illumina raw reads against the reference using the Burrows–Wheeler Aligner (BWA) (*49*) with
408 identification of SNPs using GATK (*50*). Purging of recombinant region in the chromosome-based
409 SNPs was performed using Gubbins v.2.2 (*51*).

410

411 *Phylogenetic analysis*

412 Relatedness between isolates according to core genome SNPs was inferred using RAxML v8.2.10 using the
413 GTRCAT model (*52*). Relatedness of the plasmids was inferred using using PHyML with Smart Model
414 Selection (*53*) with tree searching using SPR and 100 bootstrap replicates.

415

416 *Plasmid profiling*

417 Plasmids were purified as described by Kado and Liu (*54*) and visualized by separation on a 0.8%
418 agarose gel electrophoresis and stained with GelRed (Biotium, Hayward, Ca, USA). *E. coli* strain
419 39R861, which contains four plasmids of sizes 147, 63, 36, and 7 kb (*55, 56*), served as a size
420 marker.

421

422 *Statistical Analysis*

423 Comparisons of categorical variables were tested using Pearson's χ^2 test with Yates' continuity
424 correction, using R v3.3.2 statistical software (https://www.r-project.org). The criterion for statistical
425 significance was $p<0.05$.

426

427 *Accession of Sequence data*

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

428 The accession numbers for the Illumina sequences generated from the 134 *E. coli* ST131 isolates
429 presented in this study are available in the European Nucleotide Archive (ENA;
430 https://www.ebi.ac.uk/ena) under the following accession numbers: PRJEB27194. Sequences can
431 also be located in the ENA using the following study summary: "Emergence of enteroaggregative
432 *Escherichia coli* within the ST131 lineage as a cause of extraintestinal infections".
433 The sequence of the pAA-ST131 plasmid has been deposited in GenBank under accession number
434 KY706108.

435

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

436 **Acknowledgements**

437 We thank Michala T. Sørensen at Statens Serum Institut for performing the plasmid gel
438 electrophoresis. We also thank Karin Sixhøj Pedersen at Statens Serum Institut for her excellent
439 WGS work.

440

441 **Funding**

442 This work was supported financially by the Danish Council for Research, grant DFF-1331-00161 to
443 C. Struve, and the Danish Ministry of Health, as part of The Integrated Surveillance of
444 ESBL/AmpC-producing *E. coli* and Carbapenemase-Producing Bacteria. This work was also
445 supported in part by Office of Research and Development, (United States) Department of Veterans
446 Affairs, grant #2I01CX000920-04 to J. Johnson.

447

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

448 **References**

449 1. Rogers BA, Sidjabat HE, Paterson DL. 2011. *Escherichia coli* O25b-ST131: a pandemic,
450 multiresistant, community-associated strain. J Antimicrob Chemother 66: 1-14.

451 2. Johnson JR, Johnston B, Clabots C, Kuskowski MA, Castanheira M. 2010. *Escherichia coli* sequence
452 type ST131 as the major cause of serious multidrug-resistant *E. coli* infections in the United States.
453 Clin Infect Dis 51: 286-294.

454 3. Banerjee R, Johnson JR. 2014. A new clone sweeps clean: the enigmatic emergence of *Escherichia*
455 *coli* sequence type 131. Antimicrob Agents Chemother 58: 4997-5004.

456 4. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW,
457 Kasarskis A, Batty EM, Kos V, Wilson DJ, Phetsouvanh R, Wyllie D, Sokurenko E, Manges AR,
458 Johnson TJ, Price LB, Petro TE, Johnson JR, Didelot X, Walker AS, Crook DW, Modernizing
459 Medical Microbiology Informatics Group (MMMIG). 2016. Evolutionary History of the Global
460 Emergence of the E*scherichia coli* Epidemic Clone ST131. MBio 7: e02162-15.

461 5. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook
462 M, Schembri MA, Beatson SA. 2016. Sequential Acquisition of Virulence and Fluoroquinolone
463 Resistance Has Shaped the Evolution of *Escherichia coli* ST131. MBio 7: e00347-00316.

464 6. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M,
465 Chattopadhyay S, Stegger M, Andersen PS, Pearson T, Riddell K, Rogers P, Scholes D, Kahl B, Keim
466 P, Sokurenko EV. 2013. The epidemic of extended-spectrum-β-lactamase-producing *Escherichia coli*
467 ST131 is driven by a single highly pathogenic subclone, H30-Rx. MBio 4: e00377-00313.

468 7. Matsumura Y, Johnson JR, Yamamoto M, Nagao M, Tanaka M, Takakura S, Ichiyama S, Kyoto-
469 Shiga Clinical Microbiology Study Group. 2015. CTX-M-27- and CTX-M-14-producing,
470 ciprofloxacin-resistant *Escherichia coli* of the H30 subclonal group within ST131 drive a Japanese
471 regional ESBL epidemic. J Antimicrob Chemother 70: 1639-1649.

472 8. Abe CM, Salvador FA, Falsetti IN, Vieira MA, Blanco J, Blanco JE, Blanco M, Machado AM, Elias
473 WP, Hernandes RT, Gomes TA. 2008. Uropathogenic *Escherichia coli* (UPEC) strains may carry
474 virulence properties of diarrhoeagenic *E. coli*. FEMS Immunol Med Microbiol 52: 397-406.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

475 9. Blanco J, Mora A, Mamani R, Lopez C, Blanco M, Dahbi G, Herrera A, Marzoa J, Fernandez V, de la
476 Cruz F, Martinez-Martinez L, Alonso MP, Nicolas-Chanoine MH, Johnson JR, Johnston B, Lopez-
477 Cerero L, Pascual A, Rodriguez-Bano J, Spanish Group for Nosocomial Infections (GEIH). 2013.
478 Four main virotypes among extended-spectrum-β-lactamase-producing isolates of *Escherichia coli*
479 O25b:H4-B2-ST131: bacterial, epidemiological, and clinical characteristics. J Clin Microbiol 51:
480 3358-3367.
481 10. Olesen B, Frimodt-Møller J, Leihof RF, Struve C, Johnston B, Hansen DS, Scheutz F, Krogfelt KA,
482 Kuskowski MA, Clabots C, Johnson JR. 2014. Temporal trends in antimicrobial resistance and
483 virulence-associated traits within the *Escherichia coli* sequence type 131 clonal group and its H30 and
484 H30-Rx subclones, 1968 to 2012. Antimicrob Agents Chemother 58: 6886-6895.
485 11. Hebbelstrup Jensen B, Olsen KE, Struve C, Krogfelt KA, Petersen AM. 2014. Epidemiology and
486 clinical manifestations of enteroaggregative *Escherichia coli*. Clin Microbiol Rev 27: 614-630.
487 12. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS,
488 Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank
489 D, Redman JC, Steyert SR, Frimodt-Møller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP,
490 Schadt EE, Waldo MK. 2011. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic
491 syndrome in Germany. N Engl J Med 365: 709-717.
492 13. Imuta N, Ooka T, Seto K, Kawahara R, Koriyama T, Kojyo T, Iguchi A, Tokuda K, Kawamura H,
493 Yoshiie K, Ogura Y, Hayashi T, Nishi J. 2016. Phylogenetic Analysis of Enteroaggregative
494 *Escherichia coli* (EAEC) Isolates from Japan Reveals Emergence of CTX-M-14-Producing EAEC
495 O25:H4 Clones Related to Sequence Type 131. J Clin Microbiol 54: 2128-2134.
496 14. Guiral E, Mendez-Arancibia E, Soto SM, Salvador P, Fabrega A, Gascon J, Vila J. 2011. CTX-M-15-
497 producing enteroaggregative *Escherichia coli* as cause of travelers' diarrhea. Emerg Infect Dis 17:
498 1950-1953.
499 15. Amaya E, Reyes D, Vilchez S, Paniagua M, Möllby R, Nord CE, Weintraub A. 2011. Antibiotic
500 resistance patterns of intestinal *Escherichia coli* isolates from Nicaraguan children. J Med Microbiol
501 60: 216-222.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

502 16. Zhang R, Gu DX, Huang YL, Chan EW, Chen GX, Chen S. 2016. Comparative genetic
503 characterization of Enteroaggregative *Escherichia coli* strains recovered from clinical and non-clinical
504 settings. Sci Rep 6: 24321.
505 17. Nataro JP. 2005. Enteroaggregative *Escherichia coli* pathogenesis. Curr Opin Gastroenterol 21: 4-8.
506 18. Santiago AE, Ruiz-Perez F, Jo NY, Vijayakumar V, Gong MQ, Nataro JP. 2014. A large family of
507 antivirulence regulators modulates the effects of transcriptional activators in Gram-negative
508 pathogenic bacteria. PLoS Pathog 10: e1004153.
509 19. Boisen N, Scheutz F, Rasko DA, Redman JC, Persson S, Simon J, Kotloff KL, Levine MM, Sow S,
510 Tamboura B, Toure A, Malle A, Malle D, Panchalingam S, Krogfelt KA, Nataro JP. 2012. Genomic
511 characterization of enteroaggregative *Escherichia coli* from children in Mali. J Infect Dis 205: 431-
512 444.
513 20. Jønsson R, Struve C, Boll EJ, Boisen N, Joensen KG, Sørensen CA, Jensen BH, Scheutz F, Jenssen
514 H, Krogfelt KA. 2015. Novel aggregative adherence fimbria variant of enteroaggregative *Escherichia*
515 *coli*. Infect Immun 83: 1396-1405.
516 21. Morin N, Santiago AE, Ernst RK, Guillot SJ, Nataro JP. 2013. Characterization of the AggR Regulon
517 in Enteroaggregative *Escherichia coli*. Infect Immun 81: 122-132.
518 22. Roer L, Hansen F, Thomsen MCF, Knudsen JD, Hansen DS, Wang M, Samulioniene J, Justesen US,
519 Røder BL, Schumacher H, Østergaard C, Andersen LP, Dzajic E, Søndergaard TS, Stegger M,
520 Hammerum AM, Hasman H. 2017. WGS-based surveillance of third-generation cephalosporin-
521 resistant *Escherichia coli* from bloodstream infections in Denmark. J Antimicrob Chemother 72:
522 1922-1929.
523 23. Johnson TJ, Nolan LK. 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. Microbiol
524 Mol Biol Rev 73: 750-774.
525 24. Johnson JR, Murray AC, Gajewski A, Sullivan M, Snippes P, Kuskowski MA, Smith KE. 2003.
526 Isolation and molecular characterization of nalidixic acid-resistant extraintestinal pathogenic
527 *Escherichia coli* from retail chicken products. Antimicrob Agents Chemother 47: 2161-2168.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

528 25. Spurbeck RR, Dinh PC Jr, Walk ST, Stapleton AE, Hooton TM, Nolan LK, Kim KS, Johnson JR,
529 Mobley HL. 2012. *Escherichia coli* isolates that carry vat, fyuA, chuA, and yfcV efficiently colonize
530 the urinary tract. Infect Immun 80: 4115-4122.
531 26. Lanza VF, de Toro M, Garcillan-Barcia MP, Mora A, Blanco J, Coque TM, de la Cruz F. 2014.
532 Plasmid flux in *Escherichia coli* ST131 sublineages, analyzed by plasmid constellation network
533 (PLACNET), a new method for plasmid reconstruction from whole genome sequences. PLoS Genet
534 10: e1004766.
535 27. Coelho A, Mora A, Mamani R, Lopez C, Gonzalez-Lopez JJ, Larrosa MN, Quintero-Zarate JN, Dahbi
536 G, Herrera A, Blanco JE, Blanco M, Alonso MP, Prats G, Blanco J. 2011. Spread of *Escherichia coli*
537 O25b:H4-B2-ST131 producing CTX-M-15 and SHV-12 with high virulence gene content in
538 Barcelona (Spain). J Antimicrob Chemother 66: 517-526.
539 28. Okhuysen PC, Dupont HL. 2010. Enteroaggregative *Escherichia coli* (EAEC): a cause of acute and
540 persistent diarrhea of worldwide importance. J Infect Dis 202: 503-505.
541 29. Itoh Y, Nagano I, Kunishima M, Ezaki T. 1997. Laboratory investigation of enteroaggregative
542 *Escherichia coli* O untypeable:H10 associated with a massive outbreak of gastrointestinal illness. J
543 Clin Microbiol 35: 2546-2550.
544 30. Dallman TJ, Chattaway MA, Cowley LA, Doumith M, Tewolde R, Wooldridge DJ, Underwood A,
545 Ready D, Wain J, Foster K, Grant KA, Jenkins C. 2014. An investigation of the diversity of strains of
546 enteroaggregative *Escherichia coli* isolated from cases associated with a large multi-pathogen
547 foodborne outbreak in the UK. PLoS One 9: e98103.
548 31. Park HK, Jung YJ, Chae HC, Shin YJ, Woo SY, Park HS, Lee SJ. 2009. Comparison of *Escherichia*
549 *coli* uropathogenic genes (*kps*, *usp* and *ireA*) and enteroaggregative genes (*aggR* and *aap*) via
550 multiplex polymerase chain reaction from suprapubic urine specimens of young children with fever.
551 Scand J Urol Nephrol 43: 51-57.
552 32. Olesen B, Scheutz F, Andersen RL, Menard M, Boisen N, Johnston B, Hansen DS, Krogfelt KA,
553 Nataro JP, Johnson JR. 2012. Enteroaggregative *Escherichia coli* O78:H10 -- the Cause of an
554 Outbreak of Urinary Tract Infection. J Clin Microbiol 50: 3703-11.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

555 33. Boll EJ, Struve C, Boisen N, Olesen B, Stahlhut SG, Krogfelt KA. 2013. Role of enteroaggregative
556 *Escherichia coli* virulence factors in uropathogenesis. Infect Immun 81: 1164-1171.
557 34. Okeke IN, Wallace-Gadsden F, Simons HR, Matthews N, Labar AS, Hwang J, Wain J. 2010. Multi-
558 locus sequence typing of enteroaggregative *Escherichia coli* isolates from Nigerian children uncovers
559 multiple lineages. PLoS One 5: e14093.
560 35. Chattaway MA, Jenkins C, Ciesielczuk H, Day M, DoNascimento V, Day M, Rodiguez I, van Essen-
561 Zandbergen A, Schink AK, Wu G, Threlfall J, Woodward MJ, Coldham N, Kadlec K, Schwarz S,
562 Dierikx C, Guerra B, Helmuth R, Mevius D, Woodford N, Wain J. 2014. Evidence of evolving
563 extraintestinal enteroaggregative *Escherichia coli* ST38 clone. Emerg Infect Dis 20: 1935-1937.
564 36. Zhang J, Zheng B, Zhao L, Wei Z, Ji J, Li L, Xiao Y. 2014. Nationwide high prevalence of CTX-M
565 and an increase of CTX-M-55 in *Escherichia coli* isolated from patients with community-onset
566 infections in Chinese county hospitals. BMC Infect Dis 14: 659.
567 37. Xia S, Fan X, Huang Z, Xia L, Xiao M, Chen R, Xu Y, Zhuo C. 2014. Dominance of CTX-M-type
568 extended-spectrum β-lactamase (ESBL)-producing *Escherichia coli* isolated from patients with
569 community-onset and hospital-onset infection in China. PLoS One 9: e100707.
570 38. Tong P, Sun Y, Ji X, Du X, Guo X, Liu J, Zhu L, Zhou B, Zhou W, Liu G, Feng S. 2015.
571 Characterization of antimicrobial resistance and extended-spectrum β-lactamase genes in *Escherichia*
572 *coli* isolated from chickens. Foodborne Pathog Dis 12: 345-352.
573 39. McNally A, Oren Y, Kelly D, Pascoe B, Dunn S, Sreecharan T, Vehkala M, Välimäki N, Prentice
574 MB, Ashour A, Avram O, Pupko T, Dobrindt U, Literak I, Guenther S, Schaufler K, Wieler LH,
575 Zhiyong Z, Sheppard SK, McInerney JO, Corander J. 2016. Combined Analysis of Variation in Core,
576 Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of
577 Bacterial Populations. PLoS Genet 12: e1006280.
578 40. Trung NV, Nhung HN, Carrique-Mas JJ, Mai HH, Tuyen HT, Campbell J, Nhung NT, Van Minh P,
579 Wagenaar JA, Mai NT, Hieu TQ, Schultsz C, Hoa NT. 2016. Colonization of Enteroaggregative
580 *Escherichia coli* and Shiga toxin-producing *Escherichia coli* in chickens and humans in southern
581 Vietnam. BMC Microbiol 16: 208.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

582 41. Puño-Sarmiento J, Medeiros L, Chiconi C, Martins F, Pelayo J, Rocha S, Blanco J, Blanco M,
583 Zanutto M, Kobayashi R, Nakazato G. 2013. Detection of diarrheagenic *Escherichia* coli strains
584 isolated from dogs and cats in Brazil. Vet Microbiol 166: 676-680.
585 42. Mitsumori K, Terai A, Yamamoto S, Yoshida O. 1997. Virulence characteristics and DNA
586 fingerprints of *Escherichia coli* isolated from women with acute uncomplicated pyelonephritis. J Urol
587 158: 2329-2332.
588 43. Frank F, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, Bernard H, Fruth A, Prager R,
589 Spode A, Wadl M, Zoufaly A, Jordan S, Kemper MJ, Follin P, Müller L, King LA, Rosner B,
590 Buchholz U, Stark K, Krause G, HUS Investigation Team. 2011. Epidemic profile of Shiga-toxin-
591 producing *Escherichia coli* O104:H4 outbreak in Germany. N Engl J Med 365: 1771-1780.
592 44. Thomsen MC, Ahrenfeldt J, Cisneros JL, Jurtz V, Larsen MV, Hasman H, Aarestrup FM, Lund O.
593 2016. A Bacterial Analysis Platform: An Integrated System for Analysing Bacterial Whole Genome
594 Sequencing Data for Clinical Diagnostics and Surveillance. PLoS One 11: e0157718.
595 45. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM,
596 Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D,
597 Parczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko
598 O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9: 75.
599 46. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG):
600 simple prokaryote genome comparisons. BMC Genomics 12: 402.
601 47. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, Driebe EM, Drees KP, Hicks ND,
602 Williamson CHD, Hepp CM, Smith DE, Roe C, Engelthaler DM, Wagner DM, Keim P. 2016. NASP:
603 an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input
604 and output formats. Microb Genom 2: e000074.
605 48. Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome
606 alignment and comparison. Nucleic Acids Res 30: 2478-2483.
607 49. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform.
608 Bioinformatics 25: 1754-1760.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

609 50. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D,
610 Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for
611 analyzing next-generation DNA sequencing data. Genome Res 20: 1297-1303.
612 51. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015.
613 Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using
614 Gubbins. Nucleic Acids Res 43: e15.
615 52. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
616 phylogenies. Bioinformatics 30: 1312-1313.
617 53. Lefort V, Longueville JE, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. Mol Biol Evol
618 34: 2422-2424.
619 54. Kado CI, Liu ST. 1981. Rapid procedure for detection and isolation of large and small plasmids. J
620 Bacteriol 145: 1365-1373.
621 55. Threlfall EJ, Rowe B, Ferguson JL, Ward LR. 1986. Characterization of plasmids conferring
622 resistance to gentamicin and apramycin in strains of *Salmonella typhimurium* phage type 204c
623 isolated in Britain. J Hyg (Lond) 97: 419-426.
624 56. Schjørring S, Struve C, Krogfelt KA. 2008. Transfer of antimicrobial resistance plasmids from
625 *Klebsiella pneumoniae* to *Escherichia coli* in the mouse intestine. J Antimicrob Chemother 62: 1086-
626 1093.

627

628

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

629 **Figure legends**

630 **Fig. 1.** Circular map of plasmid pAA-ST131 compared to other publicly available pAA plasmids.
631 The outer ring shows predicted ORFs of pAA-ST131. Colors represent different putative functions:
632 gray, hypothetical proteins; red, EAEC-specific virulence factors; blue, plasmid replication and
633 maintenance; maroon, catabolism and metabolism; orange, membrane and transporter proteins;
634 green, conjugational transfer proteins (*tra* and *trb* genes); light blue, regulatory genes; purple,
635 miscellaneous; and black, mobile elements. Within the circles representing pAA plasmids from other
636 EAEC strains (labeled one to five), the darkest color indicates >90% nucleotide identity, the lightest
637 color >80% identity.
638

639 **Fig. 2.** A) Unrooted phylogenetic tree of ST131 genomes from the WHO collection, the US/German
640 collection from 2010-2012, the Danish *bla*$_{\text{CTX-M-101}}$-containing isolates from 2014-2015, and the
641 EnteroBase database of international isolates carrying *fimH*27 (n=287). The distant *H*41 cluster is not
642 shown. EAEC isolates highlighted in red. B) Rooted phylogenetic tree with all isolates within the
643 *H*27 clade (n=79). ESBL-enzymes: CARB-2 (grey), CTX-M-101 (purple), CTX-M15 (pink), CTX-
644 M-15+OXA-10 (dark blue), CTX-M-27 (yellow) and SHV-12 (light blue). EAEC-positive isolates
645 are marked by an asterisk. Scale bar represents substitution rate in the conserved core genome.
646

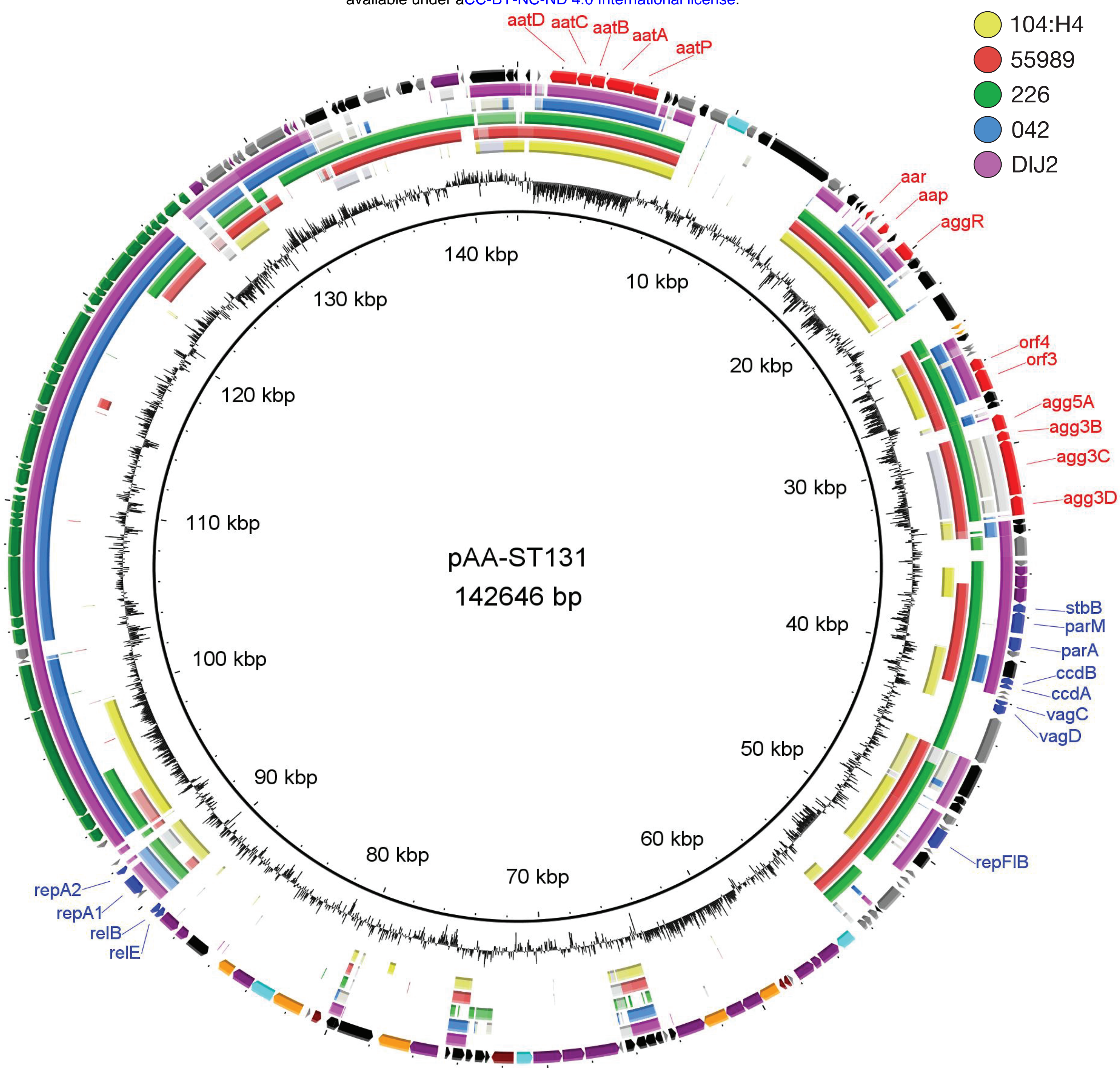
647 **Suppl. Fig. 1.** Rooted phylogenetic tree of ST131 *E. coli* strains from the WHO collection (n=121)
648 and the US/German collection (n=93). Isolates fulfilling the molecular criteria for EAEC are marked
649 with an asterisk.
650

651 **Suppl. Fig. 2.** Midpoint-rooted phylogeny based on 58 SNPs of paired blood (red) and urine
652 (yellow) isolates from eight Danish cases as well as all remaining Danish ST131 EAEC *H*27 *bla*$_{\text{CTX-}}$
653 $_{\text{M-101}}$-positive isolates. The analysis depicts the distinct relatedness of five corresponding blood and

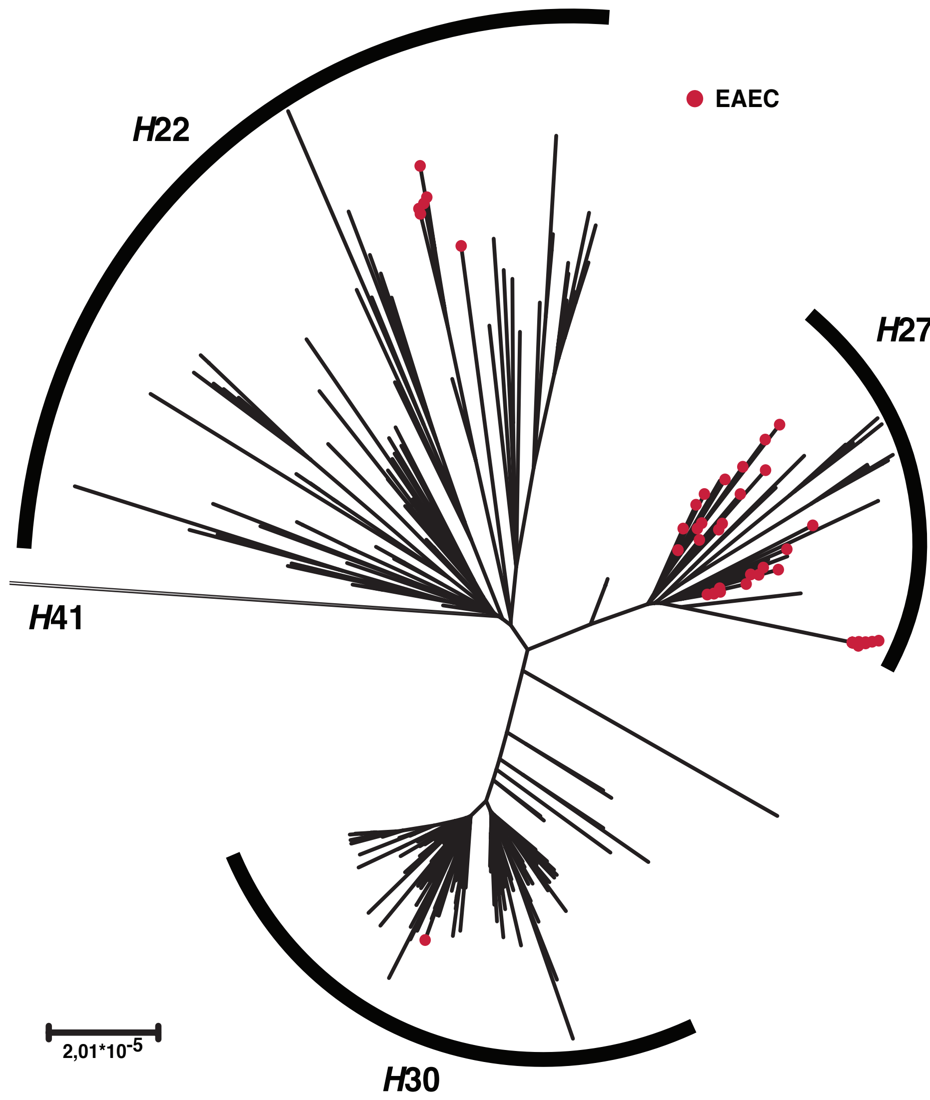
bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

654 urine isolates (Cases 1, 4, 5, 7, and 8), whereas the data is inconclusive for the remaining isolates
655 (Cases 2, 3 and 6).
656
657 **Suppl. Fig. 3**. Plasmid profiles of representative *fimH*27-carrying ST131 isolates either containing
658 $bla_{\text{CTX-M-101}}$ or from the WHO collection. Plasmid profiles with *E. coli* 39R861 as a marker (147, 63,
659 36 and 7kb) are included in lane 1. All ten *fimH*27 strains harbored only a single a plasmid. In seven
660 of the strains, this plasmid was 140-145 kb in size, whereas a slightly smaller size plasmid was
661 present in WHO isolate C1883-99. ESBL20150196 and ESBL20150300 both harbored a
662 significantly smaller plasmid of approximately 30 kb.
663
664 **Suppl. Fig. 4.** Phylogenetic tree of pAA from ST131 *H*27 isolates with more than 70% coverage of
665 pAA-ST131 from representative Danish isolate ESBL20150001. Three of the *H*22/*H*298 EAEC
666 isolates from the WHO collection (C156-00, C168-00 and C180-00) were included as an outgroup.
667 A total of 197 SNPs was identified equivalent to approximately 25% of pAA-ST131.

bioRxiv preprint doi: https://doi.org/10.1101/435941; this version posted October 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



A



B