

## Increase in bacteraemia cases in the East Midlands region of the UK due to MDR *Escherichia coli* ST73

Alhashash, Fahad; Wang, Xiaohui; Paszkiewicz, Konrad; Diggle, Mathew; Zong, Zhiyong; McNally, Alan

DOI:

[10.1093/jac/dkv365](https://doi.org/10.1093/jac/dkv365)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Alhashash, F, Wang, X, Paszkiewicz, K, Diggle, M, Zong, Z & McNally, A 2016, 'Increase in bacteraemia cases in the East Midlands region of the UK due to MDR *Escherichia coli* ST73: high levels of genomic and plasmid diversity in causative isolates', *Journal of Antimicrobial Chemotherapy*, vol. 71, no. 2, pp. 339-343.  
<https://doi.org/10.1093/jac/dkv365>

[Link to publication on Research at Birmingham portal](#)

### **Publisher Rights Statement:**

This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Journal of Antimicrobial Chemotherapy* following peer review. The version of record Fahad Alhashash, Xiaohui Wang, Konrad Paszkiewicz, Mathew Diggle, Zhiyong Zong, and Alan McNally 'Increase in bacteraemia cases in the East Midlands region of the UK due to MDR *Escherichia coli* ST73: high levels of genomic and plasmid diversity in causative isolates', *J. Antimicrob. Chemother.* (2016) 71 (2): 339-343 first published online October 30, 2015, doi:10.1093/jac/dkv365 is available online at 10.1093/jac/dkv365.

### **General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### **Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

1 **Increase in bacteraemia cases in the East Midlands region of the United**  
2 **Kingdom due to multi-drug resistant *Escherichia coli* ST73: High levels of**  
3 **genomic and plasmid diversity in causative isolates.**

4

5 Fahad Alhashash<sup>1</sup>, Xiaohui Wang <sup>2</sup>, Konrad Paszkiewicz<sup>3</sup>, Mathew Diggles<sup>4</sup>,  
6 Zhiyong Zong <sup>2</sup>, Alan McNally<sup>1\*</sup>

7

8 <sup>1</sup>Pathogen research group, Nottingham Trent University, Clifton Lane,  
9 Nottingham NG11 8NS, United Kingdom

10 <sup>2</sup>Department for Infection Control, West China Medical School, Chengdu, China

11 <sup>3</sup>Wellcome Trust Biomedical Informatics Hub, University of Exeter, Exeter, EX4

12 4QD, United Kingdom

13 <sup>4</sup>Nottingham University Hospitals NHS trust, QMC campus, Nottingham, NG7

14 2UH, United Kingdom

15

16 \*Corresponding author: Dr Alan McNally, Pathogen Research Group, Nottingham  
17 Trent University, Clifton Lane, Nottingham, NG11 8NS. Tel 0044 1158483324.

18 [alan.mcnally@ntu.ac.uk](mailto:alan.mcnally@ntu.ac.uk)

19

20 Running title: *E. coli* ST73 genome analysis

21 Keywords: *E. coli*; bacteraemia; genomics, ESBL

22 **Abstract**

23 **Objectives** To determine the population structure of *E. coli* ST73 isolated from  
24 human bacteraemia and urinary tract infections

25 **Methods** The genomes of 22 *E. coli* ST73 isolates were sequenced using the  
26 Illumina HiSeq platform. High resolution SNP typing was used to create a  
27 phylogenetic tree. Comparative genomics were also performed using a  
28 pangenome approach. *In silico* and S1-PFGE plasmid profiling was conducted,  
29 and isolates were checked for their ability to survive exposure to human serum

30 **Results** *E. coli* ST73 isolates circulating in clinically unrelated episodes show a  
31 high degree of diversity at a whole genome level, though exhibit conservation in  
32 gene content, particularly in virulence associated gene carriage. The isolates also  
33 contain a highly diverse plasmid pool that confers multi-drug resistance via  
34 carriage of CTX-M genes. All strains are highly serum resistant and uniformly  
35 carry genes shown to be essential for serum resistance.

36 **Conclusions** Our data shows that a rise in incidence of multi-drug resistant *E.*  
37 *coli* ST73 clinical isolates is not due to a circulating outbreak strain as in *E. coli*  
38 ST131. Rather the ST73 circulating strains are distantly related and carry a  
39 diverse set of resistance plasmids. This suggests that the evolutionary events  
40 behind emergence of drug resistant *E. coli* differ between lineages.

41

## 42 **Introduction**

43 Extra-intestinal pathogenic *Escherichia coli* (ExPEC) is the term used to describe  
44 strains of *E. coli* which can asymptotically colonise the intestinal tract of  
45 humans and animals, but cause disease in non-intestinal sites.<sup>1</sup> In humans ExPEC  
46 most commonly cause urinary tract infections, which is thought to affect as many  
47 as 70% of the global female population.<sup>1</sup> ExPEC are also capable of causing  
48 bacteraemia infections, where large numbers of bacterial cells gain entry to the  
49 bloodstream causing a potentially life-threatening infection. The incidence of  
50 bacteraemia caused by ExPEC has been increasing rapidly in the past 10 years,  
51 with ExPEC now the most common cause of bacteraemia in Europe, overtaking  
52 MRSA and *Clostridium difficile* bloodstream infections.<sup>2</sup>

53 The rise in cases of ExPEC bacteraemia is mirrored by a marked increase in the  
54 carriage of multi-drug resistance (MDR) plasmids in ExPEC. In particular ExPEC  
55 are associated with the sustained carriage and dissemination of genes encoding  
56 ESBL, and especially the CTX-M variant. In some countries as many as 50% of  
57 bacteraemia ExPEC isolates are ESBL positive isolates.<sup>2</sup> Numerous  
58 epidemiological studies have shown the *E. coli* ST131 clone to be the most  
59 commonly isolated MDR ExPEC strain type from human clinical cases.<sup>3,4</sup> ST73 is  
60 another phylogroup B2 strain type that is also frequently isolated from human  
61 clinical cases.<sup>4</sup> Unlike ST131, which has been extensively studied and  
62 characterised at a population and genomic level,<sup>5-7</sup> very little is known about  
63 ST73 beyond the reference ExPEC strain CFT073.<sup>8</sup>

64 We recently conducted a molecular epidemiological survey of bacteraemia  
65 ExPEC isolates from the East Midlands area of the United Kingdom.<sup>9</sup> Our study  
66 found that MDR ExPEC were significantly more abundant in bacteraemia

67 samples than clinical urine samples over a concomitant time frame. Perhaps  
68 more surprisingly our study also showed that ST73 prevalence had risen to  
69 become the most commonly isolated MDR ExPEC strain type from bacteraemia  
70 samples, and not ST131 as observed in a previous study in the same region.<sup>4</sup>  
71 Given that the rapid increase in clinical cases of MDR *E. coli* ST131 is attributable  
72 to rapid global dissemination of a successful clone,<sup>6,7</sup> we sought to determine if  
73 the high incidence of MDR ST73 clinical isolates from our bacteraemia study was  
74 also due to the emergence of a successful dominant clone.

75

## 76 **Methods**

77 **Bacterial isolates.** An epidemiological investigation of bacteraemia and urinary  
78 tract infection (UTI) *E. coli* isolates conducted by our group in 2013 identified an  
79 increase in the number of *E. coli* ST73 clinical isolates containing the CTX-M gene  
80 conferring multi-drug resistance.<sup>9</sup> Twenty-two isolates were selected for  
81 sequencing incorporating 10 ESBL positive blood isolates, 2 ESBL negative blood  
82 isolates, 3 ESBL positive UTI isolates, and 7 ESBL negative UTI isolates (table 1).  
83 These were selected to represent the diversity in ESBL phenotype in the samples  
84 population.

85 **Genome sequencing and analysis.** Isolates were sequenced on the Illumina  
86 HiSeq2500 platform using 2 x 250bp PE sequencing (Table 1). Genome  
87 assemblies were performed using Velvet and PAGIT,<sup>10</sup> which reordered contigs  
88 based on the CFT073 reference genome.<sup>8</sup> Assembled genomes were annotated  
89 using Prokka.<sup>11</sup> Progressive Mauve was used to create a whole genome  
90 alignment of the assembled genomes.<sup>12</sup> High-resolution SNP typing was  
91 performed by mapping fastQ files against the reference ST73 genome CFT073

92 using SMALT ([https://www.sanger.ac.uk/resources/software/smalt/#t 2](https://www.sanger.ac.uk/resources/software/smalt/#t2)) and  
93 Samtools. Resulting VCF files were filtered using vcftools<sup>13</sup> to retain only SNPs  
94 with a MinQ 30, MinDP 10, and MinAF 0.8. The filtered VCF files were used to  
95 produce a consensus sequence for each strain relative to CFT073. The sequences  
96 were aligned using Mugsy<sup>14</sup> from which a maximum likelihood phylogeny was  
97 created using RaxML implementing the GTR-Gamma model.<sup>15</sup> All raw sequence  
98 data has been deposited in the European Nucleotide Archive under project  
99 accession number PRJEB9931.

100 **Pangenome analysis.** A pangenome of the 22 sequenced strains and CFT073  
101 was made using Gegenees.<sup>16</sup> To determine if there were loci associated with  
102 bacteraemia in ST73, the genetic content of bacteraemia isolates was compared  
103 against UTI isolates using a cut-off of 80% identity across 80% of bacteraemia  
104 strains, and 80% identity across 20% of UTI strains. An identical analysis was  
105 conducted for ESBL positive against ESBL negative to attempt to identify loci  
106 associated with ESBL carriage. Presence of virulence-associated genes<sup>17</sup> was  
107 determined by BlastN analysis of gene sequences against the de novo assembled  
108 genome of each strain.

109 **Plasmid typing.** *In silico* plasmid typing was performed using a locally installed  
110 version of the PlasmidFinder database.<sup>18</sup> Assembled genomes were compared to  
111 the database using BlastN to identify plasmid types present in each genome.  
112 Plasmid profiling was also performed using the S1-PFGE method.<sup>19</sup>

113

## 114 **Results**

115 **The observed increase in MDR *E. coli* ST73 clinical isolates is due to a highly**  
116 **diverse group of strains.**

117 Sequence data for all 22 isolates was mapped against the CFT073 reference  
118 genome and a high-resolution SNP phylogenetic tree was constructed (Fig 1).  
119 The phylogenetic tree shows that bacteraemia and UTI isolates are intermixed  
120 throughout the phylogeny, as are ESBL positive and negative isolates. Pairwise  
121 SNP distance calculations between isolates showed that the minimum SNP  
122 distance between any two isolates was 416 SNPs, and the maximum distance  
123 6,026 SNPs (Fig S1.A).

124 **Comparative genomic analysis indicates diversity between ST73 genomes**  
125 **occurs at single base pair mutation level, and in plasmid repertoire.**

126 An alignment of all the ST73 genomes using progressiveMauve indicated genetic  
127 variation predominantly occurring in small contigs of the assemblies (Fig S2.A)  
128 suggestive that most gene-content variation occurs in plasmids and other mobile  
129 genetic elements (MGE). We created a pangenome of the ST73 genomes using  
130 Gegenees (Fig S2.B) showing a core genome of 3.81Mbp, and 1201 conserved  
131 CDS from a total of 10,696 CDS, consistent with analyses performed on the *E. coli*  
132 species and on *E. coli* ST131.<sup>20,21</sup> We performed *in silico* analysis to determine the  
133 presence of the major ExPEC virulence-associated genes in our data set (Fig  
134 S2.C). This shows some differences in carriage of virulence genes but a relatively  
135 fixed virulence gene profile. The comparison of UTI and bacteraemia isolates for  
136 virulence gene carriage also showed identical profiles between the two groups.  
137 We sought to identify the presence of any loci over-represented in the UTI or  
138 bacteraemia group of strains, or in the ESBL positive and ESBL negative group of  
139 strains using Gegenees. This analysis failed to identify any loci associated with a  
140 propensity towards bacteraemia or ESBL carriage.

141 **Highly diverse plasmid repertoire in circulating clinical *E. coli* ST73**  
142 **isolates.**

143 Given the observations of our pangenome analysis we sought to determine the  
144 extent of MGE diversity in our ST73 isolates, focussing primarily on plasmids.  
145 Using the PlasmidFinder database we performed *in silico* plasmid typing on our  
146 22 isolates (Table 1). Our analysis showed that FII, FIA and FIB plasmid types  
147 were predominant. To further investigate this we performed S1-PFGE plasmid  
148 profiling of every isolate. No plasmids were detected in the CTX-M negative  
149 isolates, but a large number of plasmid molecules were detected in the remaining  
150 isolates (Table 1). A 112Kbp plasmid was found in 6 isolates which showed the  
151 most similar accessory gene content in the pangenome analysis. Superimposing  
152 of the plasmid typing data on the phylogenetic tree showed that the 112Kbp  
153 plasmid is present in the 6 isolates that showed the lowest amount of core  
154 genome diversity (Fig 1). We compared the similarity of genomes at gene  
155 content level using the fragmented all-against-all comparison in Gegenees to  
156 show that the 6 strains sharing the 112kb plasmid also showed gene content  
157 similarity above 95% (Fig S1.B) suggesting that the plasmid pool in these 6  
158 strains is highly similar if not identical.

159 **Discussion**

160 Epidemiological studies in the East Midlands area of the UK have highlighted an  
161 increase in incidence of *E. coli* ST73 MDR isolates over the past 5 years.<sup>4,9</sup> In this  
162 study we present the genomic analysis of 22 ST73 isolates from human clinical  
163 bloodstream and UTI cases, all isolated within a 3-month period from the same  
164 region of the United Kingdom. Our analysis shows levels of diversity in the  
165 hundreds or thousands of SNPs between isolates. This is in stark contrast to



166 ST131, where isolates from the identical UK region over a 6 month period  
167 showed diversity of under 10 SNPs between strains isolated from unrelated  
168 clinical episodes, and a maximum diversity of dozens of SNPs.<sup>5</sup>

169 Analysis of our ST73 genomic data set identified the presence of a limited  
170 number of plasmid types based on *in silico rep* typing, however both genomic  
171 analysis and classical plasmid profiling show plasmid diversity in the small ST73  
172 population sampled here. The presence of a 112Kbp plasmid was inferred in 6  
173 isolates, which were also the 6 most closely related isolates phylogenetically and  
174 at gene content level. It is tempting to speculate there may be a circulating sub-  
175 clone of ST73 but such inference is hampered by our small and geographically  
176 restricted sample size.

177 The small population we have sequenced limits the inferences we can make from  
178 our data set. However there are several key points that our study highlights. The  
179 first is that the evolution and emergence of MDR lineages of ExPEC does not have  
180 a one-size-fits-all model. *E. coli* ST131 became a predominant clinical ExPEC  
181 isolate by clonal expansion and rapid global dissemination of an MDR clone of  
182 the wider ST131 lineage.<sup>7</sup> Our data of clinically unrelated ST73 isolates shows a  
183 highly diverse population of circulating ST73 strains, with a diverse plasmid pool  
184 driving multi-drug resistance in this lineage. In order to gain a more  
185 comprehensive understanding of the emergence and population structure of this  
186 important lineage of pathogenic *E. coli* it is vitally important that larger global  
187 isolate collections are analysed. Equally as important is that these collections  
188 include non-human reservoir isolates. By doing this we will acquire a far greater  
189 understanding of the ways in which ExPEC lineages can emerge as dominant  
190 MDR clinical isolates, and move our focus beyond just *E. coli* ST131.

191 **Acknowledgements**

192 All sequencing was performed at the Exeter Sequencing service at University of  
193 Exeter.

194 **Funding**

195 This study was funded by an EMDA iNET award to AM and MD, a Kuwaiti  
196 government studentship award to FA, and Royal Society grant award IE121459  
197 to ZZ and AM.

198 **Transparency declaration**

199 The authors declare there are no competing interests in the research conducted  
200 or in the reporting of this research.

201

202 **References**

203 1. Foxman B. The epidemiology of urinary tract infection. *Nat Rev Urol* 2010; **7**: 653–  
204 60.

205 2. De Kraker MEA, Jarlier V, Monen JCM, *et al.* The changing epidemiology of  
206 bacteraemias in Europe: trends from the European Antimicrobial Resistance  
207 Surveillance System. *Clin Microbiol Infect* 2013; **19**: 860–8.

208 3. Johnson JR, Johnston B, Clabots C, *et al.* *Escherichia coli* Type ST131 as the Major  
209 Cause of Serious Multidrug-Resistant *E. coli* Infections in the United States. *Clin Infect*  
210 *Dis* 2010; **51**: 286–94.

211 4. Croxall G, Hale J, Weston V, *et al.* Molecular epidemiology of extraintestinal  
212 pathogenic *Escherichia coli* isolates from a regional cohort of elderly patients  
213 highlights the prevalence of ST131 strains with increased antimicrobial resistance in  
214 both community and hospital care settings. *J Antimicrob Chemother* 2011; **66**: 2501–  
215 8.

216 5. Clark G, Paszkiewicz K, Hale J, *et al.* Genomic and molecular epidemiology analysis  
217 of clinical *Escherichia coli* ST131 isolates suggests circulation of a genetically  
218 monomorphic but phenotypically heterogeneous ExPEC clone. *J Antimicrob*  
219 *Chemother* 2012; **67**: 868–77.

- 220 6. Price LB, Johnson JR, Aziz M, *et al.* The epidemic of extended-spectrum- $\beta$ -  
221 lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic  
222 subclone, H30-Rx. *MBio* 2013; **4**: e00377–13.
- 223 7. Petty NK, Zakour NL Ben, Stanton-Cook M, *et al.* Global dissemination of a  
224 multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci U S A* 2014; **111**:5694-9.
- 225 8. Welch RA, Burland V, Plunkett G, *et al.* Extensive mosaic structure revealed by the  
226 complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci*  
227 *USA*. 2002; **99**: 17020–4.
- 228 9. Alhashash F, Weston V, Diggle M, *et al.* Multidrug-Resistant *Escherichia coli*  
229 Bacteremia. *EmergInfectDis* 2013; **19**: 1699–701.
- 230 10. Swain MT, Tsai IJ, Assefa SA, *et al.* A post-assembly genome-improvement toolkit  
231 (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc* 2012; **7**: 1260–84.
- 232 11. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;  
233 **30**: 2068–9.
- 234 12. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment  
235 with gene gain, loss and rearrangement. *PLoS One* 2010; **5**: e11147.
- 236 13. Danecek P, Auton A, Abecasis G, *et al.* The variant call format and VCFtools.  
237 *Bioinformatics* 2011; **27**: 2156–8.
- 238 14. Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole  
239 genomes. *Bioinformatics* 2011; **27**: 334–42.
- 240 15. Stamatakis A, Ludwig T, Maier H. RAxML-III: a fast program for maximum  
241 likelihood-based inference of large phylogenetic trees. *Bioinformatics* 2005; **21**: 456.
- 242 16. Agren J, Sundstrom A, Hafstrom T, *et al.* Gegenees: fragmented alignment of  
243 multiple genomes for determining phylogenomic distances and genetic signatures  
244 unique for specified target groups. *PLoS One* 2012; **7**: e39107.
- 245 17. Johnson JR, Stell AL. Extended virulence genotypes of *Escherichia coli* strains  
246 from patients with urosepsis in relation to phylogeny and host compromise. *J Infect*  
247 *Dis* 2000; **181**: 261–72.
- 248 18. Carattoli A, Zankari E, Garcia-Fernandez A, *et al.* In silico detection and typing of  
249 plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob*  
250 *Agents Chemother* 2014; **58**: 3895–903.
- 251 19. Barton BM, Harding GP, Zuccarelli AJ. A general method for detecting and sizing  
252 large plasmids. *Anal Biochem* 1995; **226**: 235–40.

253 20. Alqasim A, Scheutz F, Zong Z, *et al.* Comparative genome analysis identifies few  
254 traits unique to the *Escherichia coli* ST131 H30Rx clade and extensive mosaicism at  
255 the capsule locus. *BMC Genomics* 2014; **15**: 830.

256 21. McNally A, Cheng L, Harris SR, *et al.* The evolutionary path to extra intestinal  
257 pathogenic, drug resistant *Escherichia coli* is marked by drastic reduction in  
258 detectable recombination within the core genome. *Genome BiolEvol* 2013; **5**: 699–  
259 710.

260

261 **Table 1. List of isolates and genome assembly statistics used in this study**

Isolate	PCR ESBL type	Genome size (bps)	N Contigs	N50 contig size	% mapped reads	S1-PFGE plasmid profile	<i>In silico</i> Inc typing
B10	CTX-M-15	5173276	106	108731	94.5	112Kbp	FIB(AP001918), FII, Col156
B14	Negative	5099552	158	113745	90.21	Negative	
B18	CTX-M-15	5120683	125	122417	91.93	33.5Kbp, 48.5Kbp	Non-typable
B29	CTX-M-15	5261474	168	101820	93.7	112Kbp	FIB(AP001918), FII Col156
B36	CTX-M-15	5191523	152	125321	92.26	145Kbp	FIB(pB171), FII, Col156
B40	CTX-M-15	5257611	165	103459	91.43	140Kbp	FIA, FIB(AP001918)
B72	CTX-M-15	5158804	110	134654	84.53	33.5Kbp, 82 Kbp	FII(pRSB107)
B73	CTX-M-15	5150717	156	121329	94.38	112Kbp	FIB(AP001918), FII, Col156
B84	CTX-M-15	5182704	137	134972	93.42	112Kbp	FIB(AP001918), FII, Col156
B91	CTX-M-15	5155911	197	79515	90.23	120Kbp	FIB(S), FII, Col156
B102	Negative	5075956	160	87164	93.51	Negative	
B134	OXA-1 CTX-M-15	5230535	154	116039	93.61	82Kbp	FIB(AP001918), FII, FIA
U1	Negative	5243352	151	123112	86.52	Negative	
U7	Negative	5176031	145	126228	93.16	Negative	
U21	Negative	5145668	162	113459	91.81	Negative	
U24	Negative	5120446	147	110560	89.83	Negative	
U30	Negative	5287542	160	139416	87.12	Negative	
U36	Negative	5162072	138	114804	91.04	Negative	
U42	CTX-M-15	5188710	155	106920	93.92	112Kbp	FIB(AP001918), Col156, Col8282, Col(MG828)
U48	Negative	5080928	112	113440	87.44	Negative	
U50	CTX-M-15	5256879	145	117621	94.03	48.5Kbp	FII
U76	CTX-M-15	5179037	140	133761	94.11	112Kbp	FIB(AP001918), FII, Col156

262 Isolates with the prefix B were isolated from bacteraemia cases, those with prefix U from UTI. % reads  
 263 mapped equates to reads mapped against the CFT073 genome  
 264 N50 is a weighted median statistic such that 50% of the entire assembly is contained in contigs or  
 265 scaffolds equal to or larger than this value  
 266  
 267

268

269 **Figure 1.** Maximum likelihood phylogenetic tree of clinical ST73 isolates, with S1-  
270 PFGE and *in silico* plasmid profiling superimposed. Plasmid sizes as determined by  
271 S1-PFGE, and inc-types as determined by *in silico* analysis are indicated in the legend  
272 to the right. This figure appears in colour in the online version of *JAC* and in black  
273 and white in the printed version of *JAC*

274

275 **Figure S1.** (A) Pairwise distance matrix of the number of SNPs difference between  
276 any two isolates on the phylogenetic tree. Numbers of SNPs are relative to those  
277 obtained from mapping against the CFT073 reference genome for each isolate. (B)  
278 Pairwise comparison of percentage similarity between each genome at gene content  
279 level, as determined by fragmented-all-against-all comparison in Gegenees. This  
280 figure appears in colour in the online version of *JAC* and in black and white in the  
281 printed version of *JAC*

282

283 **Figure S2.** Comparative genomics of ST73 isolates. (A) Mauve alignment of all 22  
284 isolates alongside CFT073. Co-coloured blocks indicate genome segments containing  
285 syntenic genetic loci. Regions to the 3' end of the alignment indicate low levels of  
286 synteny. (B) Pangenome analysis of the 22 ST73 isolates alongside CFT073. Levels of  
287 nucleotide identity between genomic regions are indicated as heatmap colours.  
288 Green regions indicate genomic segments with levels of identity above 80% at  
289 nucleotide level, down to red regions that indicate levels of identity below 20%. (C)  
290 Heatmap representation of carriage of common ExPEC virulence associated genes in  
291 bacteraemia and UTI isolates of ST73. This figure appears in colour in the online  
292 version of *JAC* and in black and white in the printed version of *JAC*