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

Fahad Alhashash[1], Xiaohui Wang [2], Konrad Paszkiewicz[3], Mathew Diggle[4], Zhiyong Zong [2], Alan McNally[1*]

[1]Pathogen research group, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom

[2]Department for Infection Control, West China Medical School, Chengdu, China

[3]Wellcome Trust Biomedical Informatics Hub, University of Exeter, Exeter, EX4 4QD, United Kingdom

[4]Nottingham University Hospitals NHS trust, QMC campus, Nottingham, NG7 2UH, United Kingdom

[*]Corresponding author: Dr Alan McNally, Pathogen Research Group, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS. Tel 0044 1158483324. alan.mcnally@ntu.ac.uk

Running title: *E. coli* ST73 genome analysis

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

Abstract

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

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

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

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

Introduction

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

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

We recently conducted a molecular epidemiological survey of bacteraemia ExPEC isolates from the East Midlands area of the United Kingdom.[9] Our study found that MDR ExPEC were significantly more abundant in bacteraemia

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

Methods

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

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

using SMALT (https://www.sanger.ac.uk/resources/software/smalt/#t_2) and Samtools. Resulting VCF files were filtered using vcftools[13] to retain only SNPs with a MinQ 30, MinDP 10, and MinAF 0.8. The filtered VCF files were used to produce a consensus sequence for each strain relative to CFT073. The sequences were aligned using Mugsy[14] from which a maximum likelihood phylogeny was created using RaxML implementing the GTR-Gamma model.[15] All raw sequence data has been deposited in the European Nucleotide Archive under project accession number PRJEB9931.

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

Plasmid typing. *In silico* plasmid typing was performed using a locally installed version of the PlasmidFinder database.[18] Assembled genomes were compared to the database using BlastN to identify plasmid types present in each genome. Plasmid profiling was also performed using the S1-PFGE method.[19]

Results

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

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

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

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

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

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

Discussion

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

ST131, where isolates from the identical UK region over a 6 month period showed diversity of under 10 SNPs between strains isolated from unrelated clinical episodes, and a maximum diversity of dozens of SNPs.[5]

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

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

Acknowledgements

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

Funding

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

Transparency declaration

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

References

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

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

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

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

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

6. Price LB, Johnson JR, Aziz M, *et al.* The epidemic of extended-spectrum-β-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio* 2013; **4**: e00377–13.

7. Petty NK, Zakour NL Ben, Stanton-Cook M, *et al.* Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci U S A* 2014; **111:**.5694-9.

8. Welch RA, Burland V, Plunkett G, *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA.* 2002; **99**: 17020–4.

9. Alhashash F, Weston V, Diggle M, *et al*. Multidrug-Resistant *Escherichia coli* Bacteremia. *EmergInfectDis* 2013; **19**: 1699–701.

10. Swain MT, Tsai IJ, Assefa SA, *et al*. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc* 2012; **7**: 1260–84.

11. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–9.

12. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010; **5**: e11147.

13. Danecek P, Auton A, Abecasis G, *et al.* The variant call format and VCFtools. *Bioinformatics* 2011; **27**: 2156–8.

14. Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 2011; **27**: 334–42.

15. Stamatakis A, Ludwig T, Maier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 2005; **21**: 456.

16. Agren J, Sundstrom A, Hafstrom T, *et al*. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One* 2012; **7**: e39107.

17. Johnson JR, Stell AL. Extended virulence genotypes of Escherichia coli strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis* 2000; **181**: 261–72.

18. Carattoli A, Zankari E, Garcia-Fernandez A, *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014; **58**: 3895–903.

19. Barton BM, Harding GP, Zuccarelli AJ. A general method for detecting and sizing large plasmids. *Anal Biochem* 1995; **226**: 235–40.

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

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

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	*In silico* 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

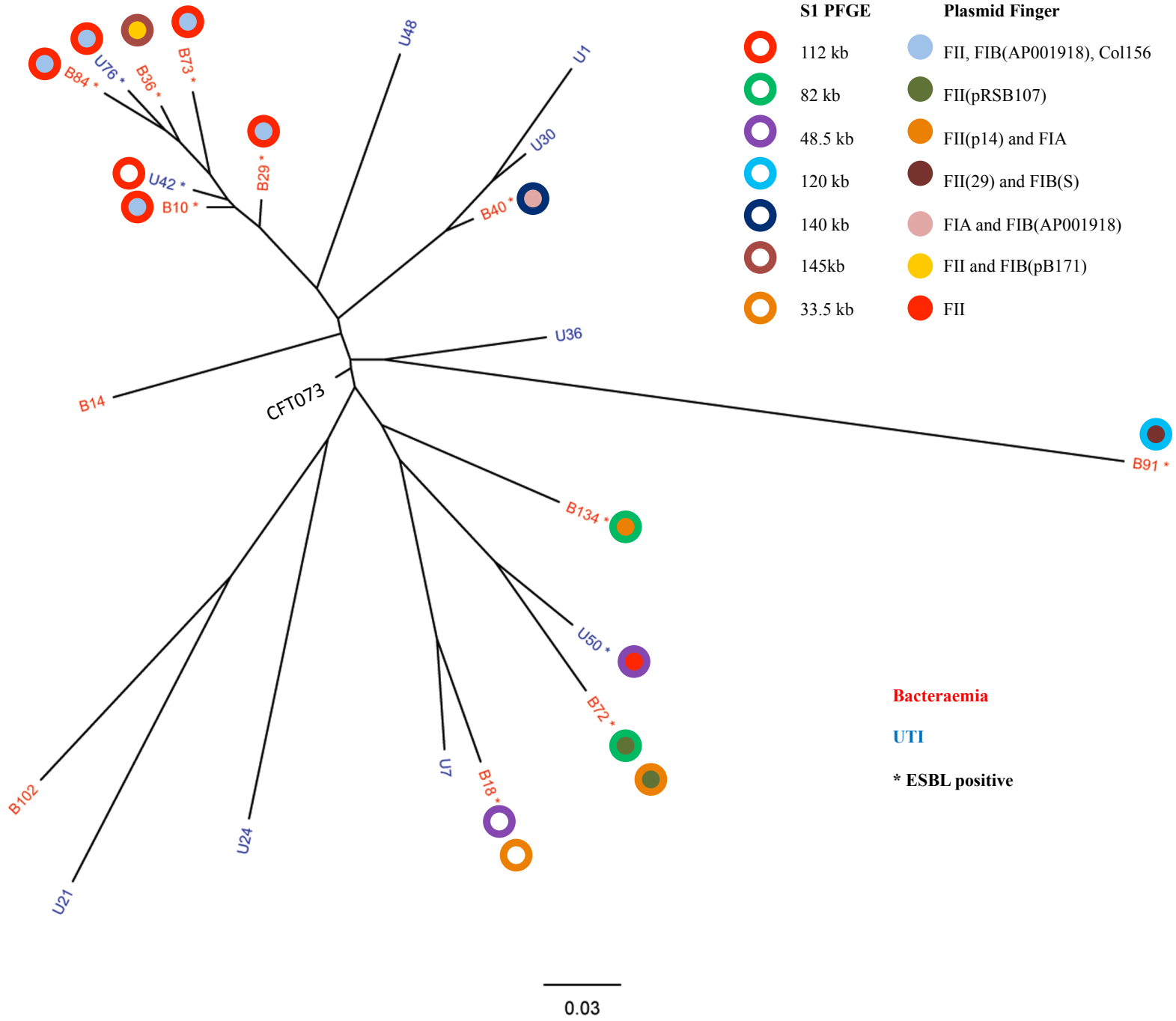
Isolates with the prefix B were isolated from bacteraemia cases, those with prefix U from UTI. % reads mapped equates to reads mapped against the CFT073 genome

N50 is a weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value

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

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

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

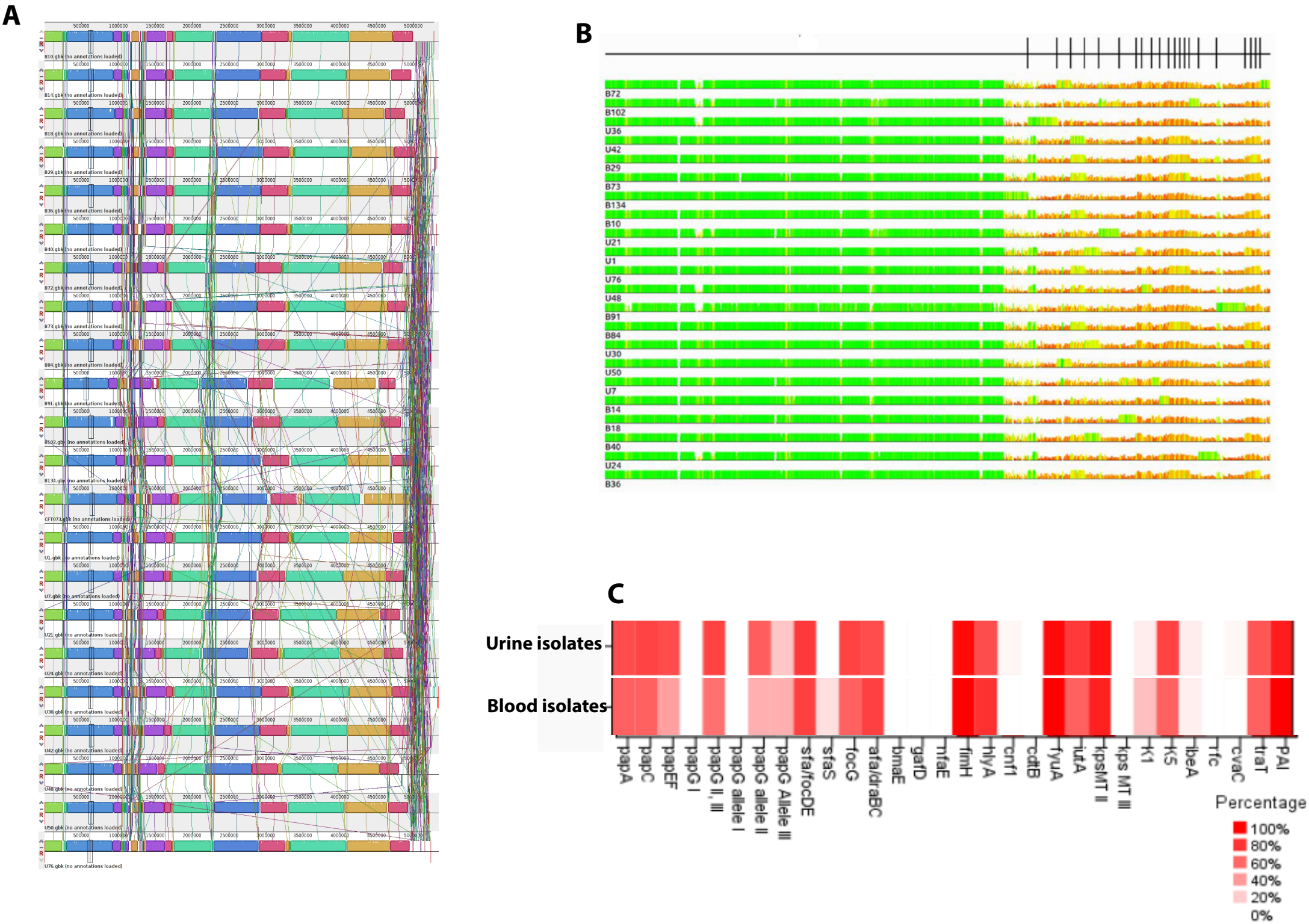

S1 PFGE
Plasmid Finger
112 kb
FII, FIB(AP001918), Col156
82 kb
FII(pRSB107)
48.5 kb
FII(p14) and FIA
120 kb
FII(29) and FIB(S)
140 kb
FIA and FIB(AP001918)
145kb
FII and FIB(pB171)
33.5 kb
FII
B84 *
U76 *
B36 *
B73 *
U48
U1
U30
B40 *
B29 *
U42 *
B10 *
U36
B14
CFT073
B91 *
B134 *
U50 *
B72 *
U7
B18 *
B102
U24
U21
Bacteraemia
UTI
* ESBL positive
0.03

A

	Ref	U48	U50	B84	B29	U42	B14	B40	B73	U24	U30	U1	U76	B18	U36	B72	U7	B10	B36	B91	B102	U21	B134
Ref																							
U48	1887																						
U50	1874	3109																					
B84	1800	2475	2901																				
B29	1079	1966	2411	1079																			
U42	1223	2046	2523	989	564																		
B14	1589	2596	2836	2658	2074	2210																	
B40	1232	2278	2443	2337	1663	1779	2206																
B73	1551	2011	2772	1133	790	836	2350	1896															
U24	2471	3679	3743	3510	3061	3187	3514	3196	3427														
U30	1598	2458	2696	2421	1932	2020	2412	816	2017	3340													
U1	2117	2916	3077	2898	2417	2513	2925	1276	2525	3792	1039												
U76	1724	2336	2870	728	979	841	2539	2209	1037	3468	2309	2776											
B18	2235	3497	2840	3230	2727	2869	3294	2868	3085	3831	3047	3520	3198										
U36	1252	2359	2712	2093	1874	1825	2395	2117	2216	3355	2424	2936	2068	3188									
B72	2182	3262	1453	3173	2664	2777	3145	2669	2997	3714	2854	3234	3119	3119	2936								
U7	2168	3444	2781	3153	2656	2795	3224	2814	3042	3676	3022	3432	3136	1364	3107	3008							
B10	1207	2032	2487	944	499	416	2175	1736	811	3166	2027	2479	840	2861	1833	2772	2788						
B36	1664	2282	2814	687	901	779	2501	2155	941	3377	2258	2732	626	3130	2023	3061	3051	773					
B91	3854	4848	5082	5025	4545	4617	4502	4531	4771	5711	4752	5201	4983	5483	4361	5221	5387	4611	4932				
B102	2828	3942	3505	3697	3192	3338	3786	3484	3545	3923	3616	4075	3695	3565	3727	3413	3471	3339	3618	5963			
U21	3028	4142	4013	4139	3587	3705	4134	3611	3949	4392	3936	4288	4123	4248	3913	3995	4141	3686	4090	6026	3206		
B134	1506	2948	2282	2697	2103	2217	2616	2206	2503	3427	2427	2853	2606	2690	2433	2634	2435	2221	2545	4757	3443	3798	

B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1: B10	100	96	96	96	96	96	95	92	92	92	91	91	91	92	91	90	89	90	94	94	94	85
2: B29	96	100	96	96	96	96	95	92	92	92	91	91	90	92	91	90	88	89	94	94	94	85
3: U76	97	97	100	97	97	96	96	93	93	93	92	92	91	93	92	90	89	90	95	94	94	86
4: B36	97	97	97	100	97	96	96	93	93	93	92	92	91	93	92	90	89	90	95	95	95	86
5: B84	96	97	97	97	100	96	96	93	93	93	92	92	92	93	92	90	89	91	95	94	94	86
6: U42	96	97	97	97	97	100	96	93	93	93	92	92	92	93	92	91	90	91	95	94	94	86
7: B73	96	97	97	97	97	96	100	93	93	93	91	92	91	93	92	90	89	90	95	95	94	86
8: B14	94	95	94	95	95	94	94	100	93	93	92	92	92	93	93	92	90	90	95	94	95	86
9: U48	95	95	95	95	95	95	94	93	100	93	93	92	91	93	93	91	90	92	94	94	94	88
10: U50	93	94	93	94	94	94	93	92	91	100	96	94	93	95	91	91	90	91	94	94	94	87
11: B72	92	92	92	92	93	92	91	91	91	95	100	93	92	93	91	91	91	91	93	92	92	87
12: U7	92	93	93	93	93	92	92	91	91	94	93	100	95	93	90	90	90	91	93	93	93	85
13: B18	92	92	92	92	93	92	92	92	90	93	93	96	100	94	91	91	91	90	93	93	93	85
14: B134	92	92	92	92	93	92	91	91	90	93	92	92	92	100	91	90	90	90	93	93	93	85
15: U36	92	92	92	92	92	92	91	91	91	91	91	90	90	92	100	90	89	89	92	92	92	85
16: U24	91	92	92	92	92	92	91	91	91	91	91	91	91	92	91	100	91	91	92	92	92	85
17: B102	91	92	92	92	92	92	91	91	91	92	93	92	93	93	91	92	100	94	92	92	92	86
18: U21	91	92	91	91	92	91	91	90	91	92	92	92	91	92	90	91	93	100	92	91	91	86
19: U30	93	93	93	93	94	93	93	92	91	92	91	91	90	92	90	89	88	89	100	96	96	84
20: B40	93	93	93	93	93	93	93	92	91	92	91	91	91	93	90	89	88	89	97	100	96	84
21: U1	93	93	93	93	94	93	93	92	91	92	91	91	91	93	91	90	89	89	97	97	100	84
22: B91	86	87	87	87	87	86	86	86	87	87	87	86	85	87	86	84	84	86	87	86	86	100


A
B
B72
B102
U36
U42
B29
B73
B134
B10
U21
U1
U76
U48
B91
B84
U30
U50
U7
B14
B18
B40
U24
B36
C
Urine isolates
Blood isolates
papA
papC
papEF
papG I
papG II, III
papG allele I
papG allele II
papG Allele III
sfa/focDE
sfaS
focG
afa/draBC
bmaE
gafD
nfaE
fimH
hlyA
cnf1
cdtB
fyuA
iutA
kpsMT II
kps MT III
K1
K5
ibeA
tfc
cvaC
traT
PAI
Percentage
100%
80%
60%
40%
20%
0%