The presence of Type I-F CRISPR/Cas systems is associated with antimicrobial susceptibility in *Escherichia coli*

Seyid AYDIN[1], Yoann PERSONNE[1], Enas NEWIRE[2], Rebecca LAVERICK[1], Oliver RUSSELL[1], Adam P. ROBERTS[2,3] and Virve I. ENNE[1*]

[1]UCL, Division of Infection and Immunity, Royal Free Hospital Campus, Rowland Hill Street, London, NW3 2PF, United Kingdom;

[2]Department of Microbial Diseases, UCL Eastman Dental Institute, 256 Gray's Inn Road, London, WC1X 8LD, United Kingdom.

[3] Department of Parasitology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, United Kingdom.

* Corresponding author: Tel: +44 (0)207 794 0500 Ext 3353, E-mail: v.enne@ucl.ac.uk

Running title: Type I-F CRISPR/Cas and antimicrobial susceptibility in *E. coli*

Synopsis

Background Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated *cas* genes are sequence specific DNA nuclease systems found in bacteria and archaea. CRISPR/Cas systems use RNA transcripts of previously acquired DNA (spacers) to target invading genetic elements with the same sequence, including plasmids. In this research we studied the relationship between CRISPR/Cas systems and multi-drug resistance in *Escherichia coli*.

Methods The presence of Type I-E and Type I-F CRISPR systems were investigated among 82 antimicrobial susceptible and 96 MDR clinical *E. coli* isolates by PCR and DNA sequencing. Phylogrouping and MLST were performed to determine relatedness of isolates. RT-PCR was performed to ascertain the expression of associated *cas* genes.

Results Type I-F CRISPR was associated with the B2 phylogroup and was significantly overrepresented in the susceptible group (22.0%) compared to the MDR group (2.1%). The majority of CRISPR I-F containing isolates had spacer sequences that matched IncF and IncI plasmids. RT-PCR demonstrated that Type I-F *cas* genes were expressed and therefore potentially functional.

Conclusion The CRISPR I-F system is more likely to be found in antimicrobial susceptible *E. coli*. Given that the Type I-F system is expressed in wild-type isolates, we suggest that this difference could be due to the

CRISPR system potentially interfering with the acquisition of antimicrobial resistance plasmids, maintaining susceptibility in these isolates.

Introduction

Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) loci were first observed over 20 years ago and have since been found in the genomes of many bacteria and archaea.[1] Along with their associated Cas proteins, the system collectively known as CRISPR/Cas, has been described as providing adaptive immunity for bacteria, targeting potentially deleterious or costly invading DNA such as phages or plasmids. CRISPR loci consist of short 21 to 47 base pair (bp) repeats separated by similarly sized non-repeating sequences called spacers.[2] The repeat arrays are often, but not always, associated with *cas* genes which encode the proteins involved in the function of the CRISPR/Cas system. The CRISPR/Cas system leads to the enzymatic cleavage of double stranded DNA in precise sites determined by the sequence of the spacer.[3] The process can be divided into two stages; acquisition and interference. In the acquisition stage, Cas1 and Cas2 proteins scan invading DNA for a short 3-6 bp motif (called the Protospacer Adjacent Motif or PAM). Sequences immediately next to the PAM are processed and integrated into the CRISPR array; these spacers are then transcribed and processed into CRISPR RNA (crRNA), and this RNA is used in the interference stage to guide the Cas nuclease complex to cleave complementary DNA.[1]

With recent attention focused on the genetic engineering potential of CRISPR/Cas, its natural role has received less attention. The existence of an adaptive immune system that rids bacteria of mobile genetic elements (MGEs) is paradoxical in terms of survival. Indeed, the ubiquitous distribution of mobile genetic elements among bacterial species suggests that CRISPR

systems are not always functional, or that they may have other roles such as the regulation of gene expression[4] and/or as yet undiscovered roles. In some environments, host bacteria clearly benefit from plasmid-encoded traits such as antimicrobial resistance and possession of CRISPR systems to rid the cell of such as plasmids is likely to be rapidly selected against. The assumption that CRISPR functions as an immune system has been called into question in *E. coli.*[5] We set out to explore this paradox.

Two subtypes of CRISPR are known in *E. coli*, Type I-E and Type I-F.[6] In both types, the genes are clustered and closely flanked by two repeat arrays each; CRISPR1 and CRISPR2 for Type I-E, and CRISPR3 and CRISPR4 for I-F. Both systems are similar but the Type I-E system has 8 genes whereas Type I-F has 6.[7] The functionality of the Type I-E system has been brought in to question due to the finding that Type I-E *cas* genes are repressed by the global regulator H-NS under laboratory conditions.[8] Conversely, Type I-F *cas* genes have been shown to be constitutively expressed.[9] Due to the spacer content of the two systems, it has previously been hypothesized that the Type I-E system may be specialized in targeting bacteriophages whereas the Type I-F system is more associated with plasmids.[6]

In this study, we examine the relationship between CRISPR and antimicrobial resistance plasmids in *E. coli* by comparing the prevalence of CRISPR Type I-E and I-F systems in antimicrobial susceptible and resistant isolates. Additionally, by investigating expression of Type I-F *cas* genes, we aim to gain an insight into the activity of these systems and potential interference against natural antimicrobial resistance plasmids.

Methods

E. coli isolates and antimicrobial susceptibility testing

A total of 178 clinical *E. coli* isolates, derived from three sources, were available for the study. Isolates were split into two groups; MDR[10] comprising 96 isolates and fully susceptible, comprising 82 isolates, based on known susceptibility testing results to 10 or more antimicrobials. The three sources comprised; 90 (33 susceptible and 57 MDR) recent urine isolates from the Royal Free Hospital (RFH), London collected between 2014 and 2015; 39 MDR isolates from Jaroden Hospital and Alexandria University in Egypt between 2009 and 2011, and 49 susceptible community urine isolates from South West of England collected between 2005-2006. Egyptian isolates were chosen on the basis of resistance to 3rd generation cephalosporins and carbapenems. RFH isolates were picked at random from available fully susceptible or MDR *E.* coli isolates from the urine bench. The isolates from South West England represented the first 50 fully susceptible isolates in a larger collection forming part of another study. All isolates were subjected to additional susceptibility testing to antimicrobials commonly associated with plasmid-acquired genes (ampicillin, gentamicin, tetracycline, ciprofloxacin, chloramphenicol and sulfamethoxazole) using the EUCAST disc diffusion method. All isolates are listed in Table S1.

PCR and DNA sequencing

PCR was used to screen for four known CRISPR arrays with primers from Touchon *et al.*[5] (listed in Table 1). PCR reactions were prepared using HotStar Taq Mastermix (Qiagen) according to manufacturer's instructions

(12.5 µl MasterMix, 0.2-1 mM of each primer and 20-100 ng of DNA up to 25 µl total volume). PCR products were visualized in agarose/ethidium bromide gels under UV light. The presence of CRISPR3 and CRISPR4 arrays were confirmed with Sanger DNA sequencing (Beckman Coulter Genomics) followed by CRISPR identification using CRISPRfinder.[2]

Analysis of *E. coli* by phylogrouping, MLST and plasmid replicon typing

Phylogenetic groups were determined using multiplex PCR according to the revised method of Clermont *et al.* 2013[11] and isolates that were unclassified according to the method were re-confirmed to be *E. coli* by MALDI-TOF. MLST was also performed on Type I-F CRISPR containing isolates using the 7 gene Achtman method.[12] O25b-ST131 clones were detected using PCR.[13] PCR-based replicon typing was used to screen CRISPR I-F containing *E. coli* for the presence of IncF and IncI group plasmids.[14]

Spacer analysis

CRISPRfinder[2] was used to determine the number and sequences of the spacers within CRISPR3 and CRISPR4 repeat arrays. Nucleotide BLAST and CRISPRTarget[15] were used to search for matching sequences for Type I-F spacers and a subset of Type I-E spacers. An identity score of 29 was used as a lower threshold for plasmid matches of interest, excluding matches to CRISPR regions from other isolates.

RT-PCR for expression analysis

RT-PCR for Type I-F *csy1* and *cas1* was performed using One Step RT-PCR kit (Qiagen) using previously described primers (Table 1). The housekeeping gene *rpsL* was used as a control. RNA was extracted from bacteria in the

logarithmic phase using an RNeasy minikit (Qiagen) and treated with DNase using Turbo DNA (Ambion) according to manufacturers' instructions. Extracts were confirmed to be devoid of detectable DNA with PCR using the HotStar Taq kit (Qiagen).

Statistics

Results were analysed using GraphPad Prism 7. A significance level of α: 0.05 was used for all statistics. Fisher's exact test was used for comparisons of CRISPR presence between the susceptible and resistant isolates.

Results and Discussion

All 178 *E. coli* isolates were screened for four CRISPR arrays (CRISPR 1-4) found in this species. Overall, over half of the *E. coli* isolates had at least one of the screened CRISPR arrays (53.9%) and Type I-E repeat arrays (CRISPR1 and/or 2) were more common (39.9%) than Type I-F repeats (CRISPR3 and/or 4) (15.7%). The overall distribution of CRISPR array types differed significantly between susceptible and MDR groups ($P < 0.0001$); CRISPR1 and 2 arrays were overwhelmingly the most prevalent amongst resistant isolates, whereas in susceptible isolates approximately equal numbers of both array types (CRISPR1/2 or CRISPR3/4) were found (Figure 1). Type I-E and Type I-F repeats were largely mutually exclusive among the isolates; only 4 out of the 178 isolates studied had repeats associated with both CRISPR types, in line with previous findings.[5] None of the isolates had all four repeat arrays.

In addition to screening for the individual repeat arrays, isolates that were shown to have Type I-F repeats (CRISPR3 and CRISPR4) were also screened for Type I-F *cas* genes. Out of the 82 susceptible isolates, 18 (22.0%) had Type I-F systems, defined here as having CRISPR3 and CRISPR4 as well as the associated genes, and an additional 8 (9.8%) had only CRISPR3 repeat arrays but without the *cas* genes. This differs significantly from the resistant isolates ($P < 0.0001$) where only two isolates out of 97 (2.1%) had Type I-F systems and none had CRISPR3 on their own. Type I-F overrepresentation in susceptible isolates was also demonstrated in only the Royal Free Hospital subset of isolates collected from the same hospital and over the same time period ($p = 0.0108$). 21.2% of the 33 susceptible RFH isolates had CRISPR I-F whereas only 3.5% of the 57 resistant RFH isolates had the system. None of the highly resistant Egyptian isolates had CRISPR3 or 4.

There were a total of 65 distinct CRISPR3 spacers and 39 distinct CRISPR4 spacers with no overlap between the two arrays in terms of spacer content. Some spacers were common and appeared in multiple non-clonal isolates, including both susceptible and resistant isolates (Table 2, Figure 2). Type I-E repeats from 49 susceptible isolates were also sequenced for comparison. Interestingly, there were 152 and 117 distinct spacers for CRISPR1 and CRISPR2 respectively, which is greater than the number of distinct spacers for CRISPR3 and CRISPR4 for the entire group of 178 isolates. However, none of the CRISPR1 and CRISPR2 associated spacers corresponded to known plasmids and only one corresponded to a known phage. Most spacers were cryptic with no homology to any known genes. This

188 is in contrast to the work of Diez-Villasenor *et al*. who reported a much larger
189 proportion of spacers with a known origin.[16] On the other hand, Nucleotide
190 BLAST for the Type I-F spacers revealed that five of the spacers matched
191 conserved regions within IncFII, IncFIB and IncI1 type plasmids with a
192 minimum of 97% homology (31/32 nucleotides) (Table 2). One spacer
193 corresponding to *klcA,* encoding a putative anti-restriction protein, appeared in
194 20 isolates in total. The *klcA* gene is conserved among IncI1 and IncFII
195 plasmid scaffolds, including those associated with the epidemic *E. coli* ST131;
196 typified by the CTX-M-15 encoding plasmids pEK516 and pEK499.[17, 18]
197 Interestingly, three of the spacers identified (2, 3 and 4 in Table 2) are found
198 in the same, largely cryptic, region which is shared between IncI1 and IncFII
199 plasmids. PCR-based screening for plasmid replicons confirmed the absence
200 of plasmids corresponding to the spacer content of susceptible *E. coli* isolates
201 containing Type I-F CRISPR loci. The contrast between the spacer content
202 between Type I-E and I-F systems supports the hypothesis that the systems
203 have different functions within *E. coli* with the Type I-F seemingly being
204 associated more with plasmids.

205 Phylogrouping was performed for all 178 isolates and their phylogroup
206 composition breaks down as follows: A 10.7%, B1 5.6%, B2 55.1%, C 2.3%,
207 D 14.6%, E 2.8%, F 7.9%, unclassified 1.1%. The results confirmed a
208 previously reported[16] association between CRISPR subtype and phylogenetic
209 groups. CRISPR I-F systems were only found in B2 group isolates whereas I-
210 E systems never appeared in B2 isolates. However, lone CRISPR3 arrays
211 without *cas* genes or CRISPR4 were detected in phylogroups A, D, E and F
212 (Table S1). Our results show that B2 isolates are more common in the

susceptible group (65.4%) than in the resistant group (49.0%). Since Type I-F systems only appear in the B2 group, we analysed Type I-F presence within the B2 group only and found a significant difference between susceptible and MDR resistant isolates within B2 ($p = 0.0001$) (Figure 3). Further, MLST analysis was performed on all isolates with Type I-F systems as well as 10 randomly selected susceptible isolates without Type I-F CRISPR. All isolates underwent PCR screening for the O25b-ST131 epidemic clone.[13] The MLST types of Type I-F positive isolates suggested clustering within particular STs, with nine of the 20 isolates with Type I-F systems belonging to the ST95 clonal complex, including one of the resistant isolates with Type I-F systems. The second most common sequence type among Type I-F isolates was ST141 (3/20). Plasmid-corresponding spacers were largely limited to these two sequence types (Figure 2). The MLST types of isolates without I-F CRISPR systems were representative of urinary *E. coli* found in other studies (Table S1).[19]

A previous study found that unlike the Type I-E system, Type I-F genes can be expressed under laboratory conditions.[9] We therefore used RT-PCR to investigate the expression of the *csy1* and *cas1* genes at the log phase of growth in 7 Type I-F isolates (6 from the susceptible group and 1 from the MDR group). These two genes are the first in the two putative transcriptional units of the Type I-F *cas* genes.[9] In all seven of the Type I-F strains tested, both transcriptional units were expressed in the log phase of growth.

In this work we show that the presence of the Type I-F CRISPR systems is strongly associated with antimicrobial susceptibility in *E. coli*. Reinforcing previous research,[5, 16] we also demonstrate that Type I-F systems

238 are only typically associated with the B2 phylogenetic group. However, our
239 results are in contrast to a previous study, which did not show an association
240 between the distribution of CRISPR and antimicrobial resistance plasmids in
241 *E. coli.*[20] When only considering the B2 group, which is the only group that
242 can contain Type I-F genes, the presence of Type I-F system is still strongly
243 associated with antimicrobial susceptibility. We also show that Type I-F genes
244 are expressed in a number of clinical isolates of *E. coli* and therefore
245 theoretically capable of interfering with antimicrobial resistance plasmids.
246 Indeed, a study by Almendros *et al.* demonstrated that an isolate with an
247 expressed Type I-F system was also capable of interfering with plasmid
248 constructs containing matching spacers.[9]

249 Previous work has shown that some B2 lineages such as ST131 are
250 associated with antimicrobial resistance.[21, 22] None of the ST131 isolates
251 included within our study contained Type I-F systems. What was more striking
252 was the finding that CRISPR3 and CRISPR4 arrays incorporated spacer
253 sequences derived from IncFII and IncI1 plasmid scaffolds commonly linked
254 to resistant *E. coli* clones such as ST131. ST95 strains are often
255 underrepresented in resistant groups of *E. coli*[23, 24] and our data suggest that
256 CRISPR may be a contributing factor, given that 9/20 of the Type I-F positive
257 isolates reported here belonged to the ST95 clonal complex. We suggest that
258 B2 strains with active Type I-F CRISPR systems may be interfering with the
259 uptake or survival of antimicrobial resistance plasmids within the isolate,
260 hence helping to keep them susceptible to antimicrobials.

261 The observation that some of these spacers still persist in multiple
262 isolates and different sequence types may be an indication that they are

advantageous, particularly since the spacers can correspond to more than one plasmid. In environments where antimicrobials are scarce or absent, plasmids may confer a fitness cost,[25] and in these conditions, B2 strains with Type I-F systems may have an advantage. While we did find two MDR isolates with Type I-F spacers that correspond to antimicrobial resistant plasmids, this could be explained by the fact that CRISPR systems have been shown to have leakage and are not functionally perfect even with exact spacer matches and optimal PAMs.[9, 26] There is also the possibility that the system has been deactivated as reported in *Staphylococcus epidermis.*[25]

Taken together our findings suggest a role for Type I-F CRISPR in the distribution of antimicrobial resistance among *E. coli* B2 lineages.

Acknowledgements

We would like to thank staff at the Royal Free Hospital Microbiology laboratory for their assistance with collecting recent clinical isolates. We would also like to thank Prof Barry Cookson for initial discussions.

Also, we are grateful to the US. Naval Medical Research Unit-3 (US. NAMRU-3), and the Centers for Disease Control and Prevention (CDC) partnership Global Disease Detection and Response Program (GDDRP) for supplying the Egyptian *E. coli* isolates (as part of Enas Newire's PhD), and partly supporting in their basic phenotypic testing.

Funding

285 Enas Newire is supported by the Schlumberger Foundation's Faculty for the
286 Future Program. No specific funding was received for this work.

287 Transparency declarations

288 None to declare.

289

290 References

291 1. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed
292 adaptive immunity in bacteria and archaea. *Nat Rev Genet* 2010; **11**: 181-90.
293 2. Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools
294 to display CRISPRs and to generate dictionaries of spacers and repeats.
295 *BMC Bioinform* 2007; **8**: 172.
296 3. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria
297 and archaea. *Science* 2010; **327**: 167-70.
298 4. Louwen R, Staals RH, Endtz HP *et al.* The role of CRISPR-Cas
299 systems in virulence of pathogenic bacteria. *Microbiol Mol Biol Rev* 2014; **78**:
300 74-88.
301 5. Touchon M, Charpentier S, Clermont O *et al.* CRISPR distribution
302 within the *Escherichia coli* species is not suggestive of immunity-associated
303 diversifying selection. *J Bacteriol* 2011; **193**: 2460-7.
304 6. Touchon M, Rocha EP. The small, slow and specialized CRISPR and
305 anti-CRISPR of *Escherichia* and *Salmonella*. *PloS One* 2010; **5**: e11126.
306 7. Makarova KS, Wolf YI, Alkhnbashi OS *et al.* An updated evolutionary
307 classification of CRISPR-Cas systems. *Nat Rev Microbiol* 2015; **13**: 722-36.

8. Westra ER, Pul U, Heidrich N *et al.* H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol Microbiol* 2010; **77**: 1380-93.

9. Almendros C, Guzman NM, Diez-Villasenor C *et al.* Target motifs affecting natural immunity by a constitutive CRISPR-Cas system in *Escherichia coli*. *PloS One* 2012; **7**: e50797.

10. Magiorakos AP, Srinivasan A, Carey RB et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for intern standard definitions for acquired resistance. Clinical Microbiol Infect 2012; 18: 268-81

11. Clermont O, Christenson JK, Denamur E *et al.* The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 2013; **5**: 58-65.

12. Wirth T, Falush D, Lan R *et al.* Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 2006; **60**: 1136-51.

13. Clermont O, Dhanji H, Upton M *et al.* Rapid detection of the O25b-ST131 clone of *Escherichia coli* encompassing the CTX-M-15-producing strains. *J Antimicrob Chemother* 2009; **64**: 274-7.

14. Carattoli A, Bertini A, Villa L *et al.* Identification of plasmids by PCR-based replicon typing. *J Microbiol Meth* 2005; **63**: 219-28.

15. Biswas A, Gagnon JN, Brouns SJ et al. CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets. RNA Biol 2013; 10: 817-827

16. D Díez-Villaseñor C, Almendros C, García-Martínez J et al. Diversity of CRISPR loci in Escherichia coli. Microbiol 2010; 156: 1351-1361

17. Enne VI, Delsol AA, Roe JM *et al.* Evidence of antibiotic resistance gene silencing in *Escherichia coli*. *Antimicrob Agents Chemother* 2006; **50**: 3003-10.

18. Woodford N, Carattoli A, Karisik E *et al.* Complete nucleotide sequences of plasmids pEK204, pEK499, and pEK516, encoding CTX-M enzymes in three major *Escherichia coli* lineages from the United Kingdom, all belonging to the international O25:H4-ST131 clone. *Antimicrob Agents Chemother* 2009; **53**: 4472-82.

19. Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin Microbiol Infect* 2014; **20**: 380-390.

20. Touchon M, Charpentier S, Pognard D *et al*. Antibiotic resistance plasmids spread among natural isolated *Escherichia coli* in spite of CRISPR elements. *Microbiology* 2012; **158**: 2997-3004.

21. Mathers AJ, Peirano G, Pitout JD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. *Clin Microbiol Rev* 2015; **28**: 565-91.

22. Adams-Sapper S, Diep BA, Perdreau-Remington F *et al.* Clonal composition and community clustering of drug-susceptible and -resistant *Escherichia coli* isolates from bloodstream infections. *Antimicrob Agents Chemother* 2013; **57**: 490-7.

23. Bengtsson S, Naseer U, Sundsfjord A *et al.* Sequence types and plasmid carriage of uropathogenic *Escherichia coli* devoid of phenotypically detectable resistance. *J Antimicrob Chemother* 2012; **67**: 69-73.

24. Golding GR, Persaud N, Levett PN *et al.* Characterization of *Escherichia coli* urinary tract infection isolates in remote northern

357 Saskatchewan communities: the Northern Antibiotic Resistance Partnership.
358 *Diagn Microbiol Infect Dis* 2012; **74**: 242-7.
359 25. Vogwill T, MacLean RC. The genetic basis of the fitness costs of
360 antimicrobial resistance: a meta-analysis approach. *Evol Appl* 2015; **8**: 284-
361 95.
362 26. Jiang W, Maniv I, Arain F *et al.* Dealing with the evolutionary downside
363 of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet* 2013; **9**:
364 e1003844.
365

366 **Tables**

367

Name	Sequence (5' to 3')	Purpose	Source
C1 fw	GTTATGCGGATAATGCTACC	CRISPR screening	[5]
C1 rev	CGTAYYCCGGTRGATTTGAA		
C2 fw	AAATCGTATGAAGTGATGCAT	CRISPR screening	[5]
C2 g rev	TCGATAATTGTGAACYTMTC		
C3 fw	GCGCTGGATAAAGAGAAAAAT	CRISPR screening	[5]
C3 rev	GCCCACCATTCACCTGTA		
C4 fw	CTGAACAGCGGACTGATTTA	CRISPR screening	[5]
C4 rev	GTACGACCTGAGCAAAG		
Csy1 fw	TCAGTCATGGTGATTCT	*cas* gene screening & RT-PCR	[9]
Csy1 rev	GCAACAGGGAAATAGA		
Cas1 fw	CGGGGTGATGGTAGGCTTTT	*cas* gene screening & RT-PCR	This study
Cas1 rev	TGGTTTTCTGCCGCGTCTAT		
RPSL fw	CTCGCAAAGTTGCGAAAAGC	RT-PCR control	[17]
RPSL rev	TTCACGCCATACTTGGAACG		

368

369 **Table 1.** Oligonucleotides used for CRISPR screening and gene expression
370 studies. Oligonucleotides used for phylogenetic grouping, MLST and plasmid
371 replicon typing are primers not included.

372

Name of spacer	CRISPR array	Sequence (5' to 3')	No. of isolates	Protospacer match
Spacer1	CRISPR3	AGCATCTGCATGGTGC CCGTGGTCTTAACAAG	1	IncFII/FIB plasmids
Spacer2	CRISPR3	TGATGGCGCAGCAGTC CTCCCTCCTGCCGCCA	13	Non-coding region of IncI1 and IncFII plasmids
Spacer3	CRISPR3	CTGAACGTTGAAGAGT GCGACCGTCTCTCCTT	20	Putative anti-restriction protein KlcA on IncI1 and IncFII plasmids
Spacer4	CRISPR3	GGAAGAGACGGATGTT GACCAGCGAAATCCGA	1	Hypothetical protein found on IncFII and IncI1 plasmids
Spacer5	CRISPR4	TGTGGCGCTGATGCGT CTGGGCGTCTTTGTAC	8	*repA* gene of IncFIB plasmids

Table 2. Spacer sequences matching antimicrobial resistance plasmids. Five spacers that correspond to plasmid sequences were found using nucleotide BLAST in CRISPR3 and CRISPR4 containing strains, including isolates which had repeat arrays but no I-F genes.

Figures

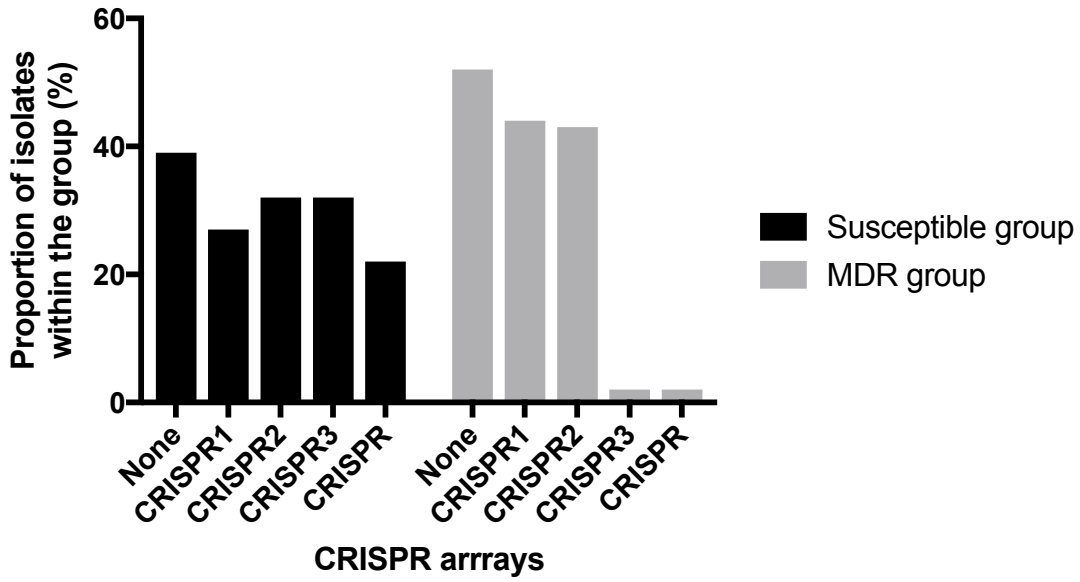
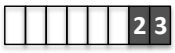
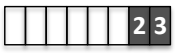
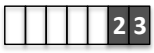
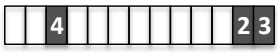


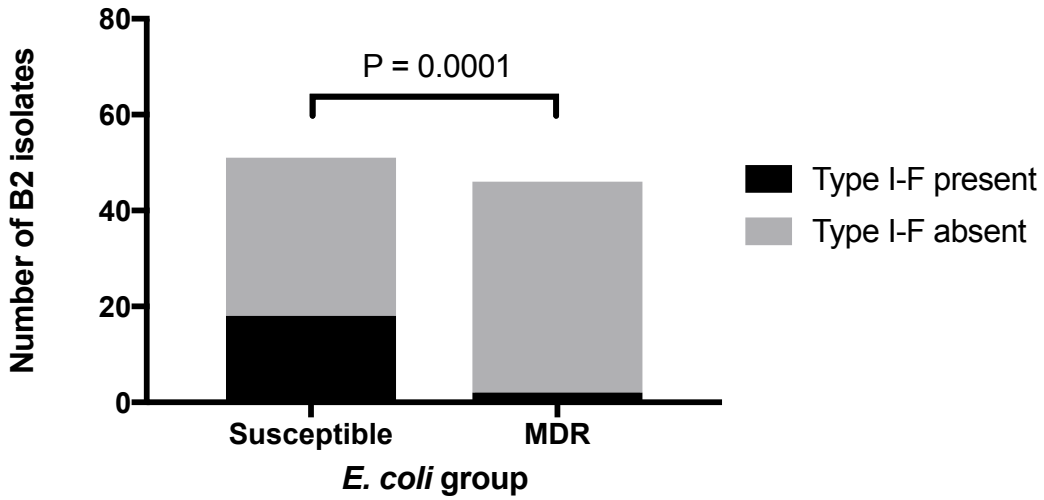
Figure 1. Proportion of *E. coli* isolates with CRISPR 1, 2, 3 and 4 repeat arrays within the susceptible and MDR groups. Overall the two groups had a significantly different distribution of CRISPR arrays ($P < 0.0001$). CRISPR 1 & 2 and CRISPR 3 & 4 are often, but not always, found in pairs.

CRISPR3 arrays	CRISPR4 arrays	No. of isolates	Sequence types
		3	ST80, ST5351
		1	ST141
		1	ST141
		1	ST141
		3	ST95, ST5168
		2	ST95
		1	ST95
		1	ST95
		1	ST95
		1	ST421 (ST95 complex)
		1	ST2015
		1	ST420
		1	ST2582
		1	ST4075
		1	ST4075

388

389 **Figure 2.** CRISPR3 and CRISPR4 array profiles found in isolates containing
390 Type I-F genes. Each box represents a spacer sequence. Isolates of the
391 same sequence type have similar spacer profiles but often with missing or
392 additional spacers. Shaded spacers correspond to known antimicrobial
393 resistance plasmids listed in Table 2. MLST types of all isolates with complete
394 Type I-F systems are listed

395



396

397 **Figure 3.** Presence of Type I-F systems in susceptible and resistant B2

398 isolates. 'Type I-F systems' are defined here as presence of both Type I-F

399 repeat arrays and *cas* genes.

400

401