

Exploring the Potential Association between PAI and CRISPR CAS Region in *E.coli* CFT073 – An Insilico Approach

Jayanthi S¹, Santhiya K², Nisha Nandhini J³, Prakrithi P⁴
{ jay.bio@psgtech.ac.in¹, santhiyakc17494@gmail.com²,
nishanandhini23@gmail.com³, prakrithi29@gmail.com⁴}

^{1,3,4} Department of Biotechnology, PSG College of Technology, Coimbatore (Tamilnadu), India
² Department of Biotechnology, Faculty of Engineering, Karpagam Academy of Higher Education, Coimbatore (Tamilnadu), India

Abstract. Antimicrobial resistance (AMR) continues to be a major problem that jeopardises public health and environment all around the world. New methods for controlling resistant bacterial infections are desperately needed. CRISPR Cas, the recent genome editing technology has potential applicability in fighting against AMR bacterial infections since it has an ability to target resistance genes with specificity. The current study aims to investigate the relationship among antibiotic resistance, virulence genes and insertion sequence along with CRISPRcas locus region in *E.coli* CFT073 strain through an insilico approach in an attempt to apply CRISPRcas technology to prevent the dissemination of pathogenic strains. Using the CARD database, the existence of eight resistance genes (mdtH, H-NS, baeR, evgA, emrR, gadW, cpxA, and EC-5) was discovered, and the ISs, ISEc10, and IS200c were shown to be prevalent in the *E.coli* CFT073 strain. Thirteen virulence genes from varied family were identified using virulence finder tool. The study found 19 spacer regions as well as the existence of cas genes using CRISPRcas ++ tool in *E.coli* genome and also identified the PAM positions along with their flanking IS regions using E CRISP. Two regions (mdtH-IS21 (ISEc10) and emrR-IS110 (ISEc20)) were identified as a promising PAI sites while comparing the locations of PAM, transposase, and antibiotic resistance genes where a knockout mechanism can be applied using CRISPRcas technology to disseminate the spread of AMR among microbial species.

Keywords: Antibiotic resistance, Genome editing, Insertion sequence, Pathogenicity island, Virulence genes, CRISPR Cas, *E coli* CFT073.

1 Introduction

Uropathogenic *Escherichia coli* (UPEC) are the primary causal agent in almost 90% of community acquired urinary tract infection (UTI) and 50% of nosocomial UTIs in humans. UPEC colonizes the urinary system primarily in the bladder, causing cystitis, and may aspire into the kidneys via the ureters, causing pyelonephritis. To colonize the urinary system, UPEC binds to urothelial cells via P fimbriae and begins to generate alpha and beta hemolysins, which can induce lysis of urinary tract cells [1]. And, of all the UPEC strains (such as CFT073, 536, and J96), CFT073 is the most prevalent and well-studied, having the whole genome annotated both manually and automatically with high accuracy [2]. The CFT073

genome differs mostly due to five distinct cryptic inserted prophage genomes that include a substantial amount of virulence/ virulence-associated genes, known as pathogenicity islands (PAIs). In-depth research indicates that 75.0% of the genes are conserved backbone genes, whereas the remaining (25.0%) are CFT073-specific islands that insert into the backbone areas in a hybrid fashion. Ever since the first publication, the annotation has not only given an overview of the pathogen's lifetime, but it has also served as a reference for experiment design [3].

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system is an immunological system present in bacterial system that has been modified/ used for genome editing purposes. The techniques like zinc finger nucleases (ZFNs) or transcription-activator-like effector nucleases (TALENs) relied on the usage of customised DNA-binding protein nucleases prior to CRISPR/Cas9 technology [4]. It requires researchers to engineer and produce a new nuclease-pair for each genomic target. CRISPR, on the other hand, has become one of the most popular methods for genome editing, owing to its versatility and ease of use. The bacterial CRISPR-Cas9 genome editing method has been used by many species, including bacteria and yeasts. A tracrRNA:crRNA duplex (or chimeric guide RNA (gRNA)) directs the Cas9 protein to cleave a target DNA sequence containing a protospacer motif (PAM) [5].

Horizontal gene transfer (HGT) is an important mechanism in microbial adaptability to diverse environment conditions because it allows for the acquisition of genetic material such as antibiotic resistance genes; virulence and insertion sequence [6]. Furthermore, microorganisms have developed an innate immune system, CRISPR and their associated Cas proteins (CRISPR-Cas), which aids in the limitation of genetic material procurement and defence against intrusive bacterial sp, phages and plasmids. The objective of this study is to identify PAM sequence region adjacent to the genes coding for PAI (antibiotic resistance, virulence factor and transposase enzymes) for knocking out using CRISPR-Cas technology for prevention and dissemination of bacterial virulence by HGT.

2. Methodology

2.1 Sequence retrieval:

The whole genome sequence of *E.coli*CFT073(5.23143Mb) was retrieved from NCBI (https://www.ncbi.nlm.nih.gov/genome/167?genome_assembly_id=299448) and used for analyses.

2.2 Identification ofPAIs:

Pathogenicity islands for the strain *E.coli* CFT073 was identified using the tools Comprehensive Antibiotic Resistance Database (CARD), Virulence Finder and ISSaga database. CARD (<https://card.mcmaster.ca>) is a web - based bioinformatics tool that provides the molecular mechanism of bacterial antibiotic resistance (AMR) for DNA and protein sequence [7]. The online OMICS tool, Virulence Finder was used to discover and extract *E. coli* virulence genes from whole genome sequencing (WGS) data, as well as their genomic positions. The insertion sequences or transposons flanking each of the virulence factors were also discovered and identified [8]. The mobile factors, such as Insertion Sequence, found in the *E.coli* CFT073 was examined using the ISSaga database (http://issaga.biotoul.fr/issaga_index.php), which operates on the ISFinder platform. It does computer analysis to yield good quality IS annotation [9].

2.3 Analysis of CRISPR loci and spacer sequence:

CRISPR Finder tool was used to locate CRISPR loci in raw genome sequences. The fasta sequence of *E.coli* genome was submitted to the CRISPR-Cas++ tool (<https://crisprcas.i2bc.paris-saclay.fr/>) [10]. The Crispr locus is characterized by spacers, leader sequence and cas genes. This tool was used to detect the presence of spacers and the casgene..

2.4 Location of the PAM positions in relation with ISs:

The protospacer adjacent motif (PAM) sequence located in the *E.coli* strain was identified using the tool, E-CRISP [11]. E-CRISP has indeed been enhanced by utilising quick and precise approaches to build CRISPR gRNA sequences that can hit any sequence of nucleotides from one exon to whole genomes. Usability in experimental applications has received special attention during the design phase. The PAM regions specific for Cas enzyme which are located either upstream or downstream of the ISs was found manually.

3 Result And Discussion

3.1 PAIs of *E.coli* CFT073:

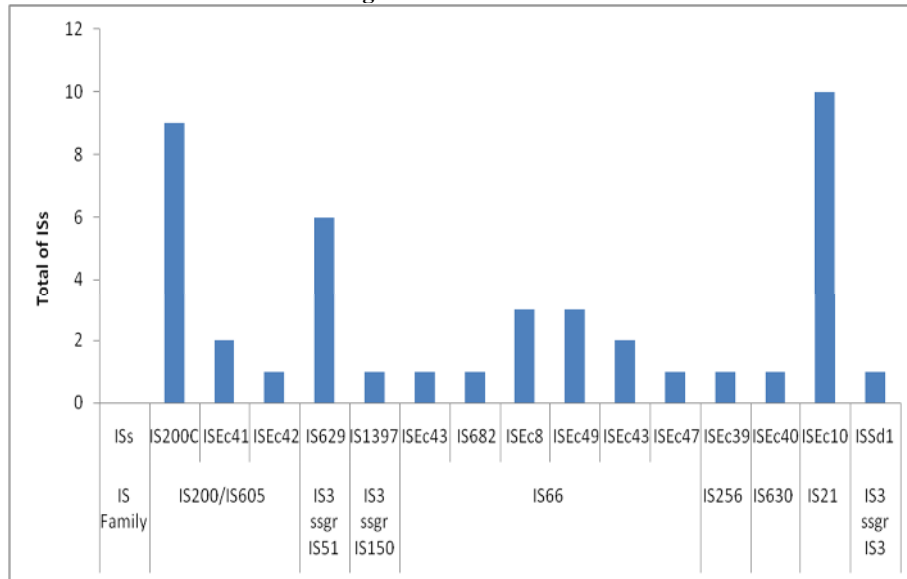
PAI comprises pathogen-specific genetic elements that encode mobile elements, virulence and resistance genes which are often transmitted by horizontal gene transfer to the recipient organism and plays a significant role in pathogenicity and evolution of microbes [12]. The antibiotic resistance genes of *E.coli* CFT073 strain was found using CARD database. Eight resistant genes (mdtH, H-NS, baeR, evgA, emrR, gadW, cpxA and EC-5) were found with 100% identity. Out of 8 antibiotic genes identified, 7 genes follow antibiotic efflux resistance mechanism whereas only one gene (EC-5) follows antibiotic inactivation mechanism. Antimicrobial resistance can be caused by the development of sophisticated bacterial machines capable of extruding a poisonous chemical from the cell is called as efflux mechanism. This resistance mechanism impacts a wide variety of antimicrobial classes except for polymyxin. Among five major classes of efflux pumps, *E.coli* CFT073 has two main classes: major facilitator superfamily (MFS) and resistance-nodulation-cell-division family (RND) [13]. Modification of antibiotic is a popular method for making an antibiotic ineffective; particularly it's applicable for aminoglycoside antibiotics, beta lactam and chloramphenicol. The bacteria will generate enzymes that permanently alter and inactivate the antibiotics; beta lactamases are one of the well-studied enzymes for this purpose [14].

Table 1: List of AMR present in *E.coli* CFT073

| Sl No | APC | AMR gene family | Drug class | Resistance Mechanism |
|-------|------|--|---|-------------------------|
| 1 | mdtH | major facilitator superfamily (MFS) antibiotic efflux pump | fluoroquinolone antibiotic | antibiotic efflux |
| 2 | H-NS | major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell-division (RND) antibiotic efflux pump | macrolide antibiotic, fusidic acid antibiotic, erythromycin, rifampicin, penam, tetracycline antibiotic | antibiotic efflux |
| 3 | baeR | resistance-nodulation-cell-division (RND) antibiotic efflux pump | aminoglycoside antibiotic, sulbactam antibiotic | antibiotic efflux |
| 4 | evgA | major facilitator superfamily (MFS) antibiotic efflux pump, resistance-nodulation-cell-division (RND) antibiotic efflux pump | macrolide antibiotic, fusidic acid antibiotic, penam, tetracycline antibiotic | antibiotic efflux |
| 5 | emrR | major facilitator superfamily (MFS) antibiotic efflux pump | fluoroquinolone antibiotic | antibiotic efflux |
| 6 | gadW | resistance-nodulation-cell-division (RND) antibiotic efflux pump | macrolide antibiotic, fusidic acid antibiotic, penam | antibiotic efflux |
| 7 | cpxA | resistance-nodulation-cell-division (RND) antibiotic efflux pump | aminoglycoside antibiotic, sulbactam antibiotic | antibiotic efflux |
| 8 | EC-5 | beta-lactamase | actinomycin | antibiotic inactivation |

A mobile element in the *E.coli* strain was identified using ISSaga tool. The ISs which shows 100% identity without duplication was chosen criteria for this analysis. As per the criteria, 15 distinct insertion sequences from 8 IS families, totaling 43 elements were considered based on ISSaga result. Among them, ISEc10 and IS200c were found to be predominant (high copy number) in the genome. It has been found that presence/ insertion of ISEc10 into the promoter or regions of the chromosomal ampC gene causes high expression of ampC and also responsible for cephalosporins resistance [15]. IS200/IS605 family members transpose utilising ssDNA intermediates. This family of ISs are different from traditional IS in terms of structure. They have subterminal palindromic structures at their terminals rather than inverted repetitions. Generally, IS200 group of ISs carries tnpA transposases whereas IS605 group members carriers both tnpA and tnpB. tnpA helps in mobility of ISs whereas tnpB has no such activity [16].

Figure 1: ISs of *E.coli* CFT073



Virulence genes found in the *E.coli* strain was found using virulence finder tool. There are about 13 virulence genes were identified along with their flanking ISs. Out of 13 virulence genes, four genes (mcmA, mchB, mchC and mchF) belongs microcin family. Microcins are peptides that are generated by enterobacterial strains contain urovirulence factor and most common in UPEC strains. Microcin are found on bacterial chromosomes and include genes for immunity, and secretion [17]. It also carries encoding serine protease autotransporters of Enterobacteriaceae (SPATE) toxin genes (vat, pic and sat). SPATE is a broad collection of autotransporter proteins which triggers virulence activity especially in *Shigella sp* and *E. coli*. Because of their proteolytic activity, this can enhance the breakdown of host cell substrates and provoke an inflammatory response, in *E.coli* intestinal pathogenic infection [18]. The virulence genes ireA and iron were belongs to iron uptake virulence factor category. Many viral illnesses are known to be exacerbated by iron excess, and iron deficiency is an essential defence mechanism for mammalian hosts. Because the fungus detects iron to regulate the

development of the polysaccharide capsule, which is the main virulence component during infection [19].

Table 2: Virulence genes in *E.coli* CFT073

| S. No | Virulence Factor | Protein Function | Location | Flanking ISs/ Transposons | |
|-------|------------------|--|-----------------|------------------------------------|------------------------------------|
| | | | | Left side | Right side |
| 1 | gal | Glucanase deacetylase | 1736318-1737718 | IS200(1574110-1574619) | IS200(2148324-2148833) |
| 2 | gal | Glucanase deacetylase | 4104924-4105324 | Retrontransposon (3970061-3970254) | Transposon (4291056-4292046) |
| 3 | ita | Adhesion protein | 3448776-3449286 | IS2610 (3445858-3447357) | IS66(3482090-3482628) |
| 4 | itaA | fibronectin receptor | 4935990-4936382 | IS2610 (4920904-4921303) | Retrontransposon (2161353-2162373) |
| 5 | itaN | Enterobacteraid receptor protein | 1198161-1200039 | IS2610(1196895-1197932) | IS200(1209626-1209145) |
| 6 | isa | Increased serum survival | 1422880-1423221 | IS2610(1403672-1403111) | IS200(1574110-1574619) |
| 7 | msbB | Microcin H47 part of colicin | 1176756-1177049 | Transposon (1175364 - 1175588) | IS2610(1196895 - 1197932) |
| 8 | msbC | MsbC protein | 1177321-1178871 | Transposon (1175364 - 1175588) | IS2610(1196895 - 1197932) |
| 9 | msbF | ABC Transceptor protein msbF | 1180751-1182085 | Transposon (1175364 - 1175588) | IS2610(1196895 - 1197932) |
| 10 | msmA | Microcin B part of colicin | 1183119-1183397 | Transposon (1175364 - 1175588) | IS2610(1196895 - 1197932) |
| 11 | wt | Vanadylphosphate spacer toxin | 371877-375007 | Transposon (365918-365896) | IS200 wp(342302-342311) |
| 12 | pic | Serine protease autotransporters of Enterobacteriaceae (SEPAE) | 324509-330324 | Transposon (303308-305696) | IS200 wp(342302-342311) |
| 13 | sat | Serine autotransporter toxin | 3454362-3460349 | IS2610(3445858-3447357) | IS66(3482090-3482628) |

3.2 CRISPR locations of *E.coli* CFT073:

CRISPR are unique structures discovered in many bacterial genomes that exhibit both sequential and interspaced repeat features. Repeats, spacers, a leader sequence, and the cas gene characterize a CRISPR locus. A CRISPR is a sequence of 23-47 bp repeats separated by unique sequences of comparable length is known as spacer region. The CRISPR locus is often

bordered on one flank with an AT-rich leader region of 100-350 bp that acts as a promoter for pre-crRNA production. CRISPR-associated genes are those that have been discovered to be tightly related to repetitive sequences. Using Crispr cas++ tool, we found that the particular strain of *E.coli* CFT073 has 19 spacer regions from the sequence position of 2871.857kb to 2898.840kb along with cas genes. The evolutionary causes for the multiplicity of different CRISPR-Cas system types often reside in almost the same genome, still mostly unknown. It is indeed unclear how certain CRISPR-Cas systems contain just one or a few spacers, whereas others have numerous. The number of spacers in an array is widely recognized to be the consequence of a trade-off between greater protections against numerous, varied, and rapidly changing pathogens [20].

Figure 2: Pictorial representation of Crispr regions in *E.coli* CFT073 (Crispr Cas ++)

| Element | CRISPR Id / Cas Type | Start | End | Spacer/ Gene | Repeat consensus/ cas genes | Direction | Evidence level |
|-------------|--------------------------|-----------|-----------|--------------|---|-----------|----------------|
| CRISPR | CP028309_3 | 2,871,857 | 2,872,801 | 15 | CGGTTTATCCCCGCTGGCGC GGGGAACAC | - | 4 |
| Cas cluster | CP028309.1 CAS-TypeI_E_1 | 2,873,191 | 2,881,474 | 7 | Cas1_0_IE_1, Cas3_0_I_7, Cas5_0_IE_3, Cas6_0_IE_2, Cas7_0_IE_4, Cse1_0_IE_6, Cse2_0_IE_5 | | |
| CRISPR | CP028309_4 | 2,898,568 | 2,898,840 | 4 | CGGTTTATCCCCGCTGGCGC GGGGAACAC | - | 4 |

Based on the topology of effector modules, the CRISPR–Cas system is classified into two primary groups. Class 1 system of CRISPRcas includes effector modules consisting of several cas proteins in which a few helps in forming crRNA-binding complex whereas class 2 system have a single multidomain crRNA binding protein that integrates all functions necessary for interference. *E.coli* CFT073 strain has Class 1 system of CRISPR cas with subtype Cas – Type1 E gene. There are several core cas genes in bacterial crisprcas system in which cas genes (1, 3, 5, 6, 7), cse1, and cse2 were identified in *E.coli* strain CFT073 using CRISPRcas ++ tool. All of the cas genes are linked to various Cas systems and serve unique functions. Cas1 aids in the spacers integration into direct repeats of CRISPR, Cas3 has helicase activity which separates DNA strands for integration purpose, Cas5 interacts with larger subunit and cas7 during interference and adaptation, and can also substitute for Cas6 if catalytically active whereas Cas6 is also a subunit of cascade system and Cas7 involved in cleavage of RNA when binds with crRNA and may be involved in RNA cleavage and finally the last two genes cse1 and cse2 small subunits of cascade complex [21].

Table 3: Spacer sequences in *E.coli*CFT073

| S.No | Spacer sequence | Strain(s) | Kingdom | Host's sequence(s) | CRISPR name(s) | CRISPR |
|------|-------------------------------|---------------------------------|----------|--------------------|----------------|------------|
| 1 | AAAGCTAAGCTTTCCTTAAGTTTATAGCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 2 | ACGCGGACGCGGATCGCGGACGCGGATCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 3 | GTGAAAGTACTGCGGATCGCGGATCGCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 4 | TGTTTTCGCGGATCGCGGATCGCGGATCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 5 | CGGATCGCGGATCGCGGATCGCGGATCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 6 | TGCGGATCGCGGATCGCGGATCGCGGAT | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 7 | CGGATCGCGGATCGCGGATCGCGGATCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 8 | GTGCGGATCGCGGATCGCGGATCGCGG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 9 | TGCGGATCGCGGATCGCGGATCGCGGAT | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 10 | CGGATCGCGGATCGCGGATCGCGGATCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 11 | TGCGGATCGCGGATCGCGGATCGCGGAT | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 12 | CGGATCGCGGATCGCGGATCGCGGATCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 13 | GTGCGGATCGCGGATCGCGGATCGCGG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 14 | ATGCGGATCGCGGATCGCGGATCGCGG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 15 | TGCGGATCGCGGATCGCGGATCGCGGAT | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 16 | CGGATCGCGGATCGCGGATCGCGGATCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 17 | GTGCGGATCGCGGATCGCGGATCGCGG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 18 | AAAGCTAAGCTTTCCTTAAGTTTATAGCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |
| 19 | AAAGCTAAGCTTTCCTTAAGTTTATAGCG | <i>Escherichia coli</i> O157:H7 | Bacteria | CP002111.1 | CP002111.1 | CP002111.1 |

3.3 PAM positions with respect to PAI:

PAM is very short sequence of length 2-6bp which is necessary for cas nuclease to cut the target region. Before unwinding the DNA, the cas nuclease will check for the perfect PAM site that matches with the guide RNA to make a cut in the sequence. The PAM positions in the strain were identified using E crisper tool and the flanking ISs regions around the PAM site was screened manually. About 12 different transposase family predicted to have PAM regions in it. While making comparison among the location of PAM, transposase and antibiotic resistant gene; two regions found to be appropriate PAI region where knock out mechanism can be applied through CRISPRcas technology to control the dissemination. The two PAI regions consist of mdhH-IS21(ISEc10) and emrR – IS110(ISEc20) along with the PAM region.

Table 4: Transposase and PAM sequence locations in *E.coli*CFT073

| #. Isr | Transposase Family | Location | PAM Sequence and Location |
|--------|--|-------------------|--|
| 1 | IS200 | 260-1000 | CGG (289-1000) |
| | | 67723-68202 | CGG (68203-67723) |
| 2 | IS21(IS21.10) | 1124405-1238115 | AGG (1238116-1244026) |
| | | 2343423-2343737 | CGG (2343738-2343715) |
| | | 3445265-3448174 | AGCG(3445267-3448272), CGG (3448150-3448152) |
| | | 1407285-1408114 | TGG (1408115-1408136) |
| | | 4932011-4932320 | AGG (4932321-4932317) |
| | | 235435-237747 | CGG (237723-237723) |
| | | 1234405 - 1238115 | TGG (1238117-1238138) |
| 3 | IS2643 | 1164475-1167426 | AGG (1167446-1167442) |
| | | 1167423-1167920 | CGG (1167920-1167920) |
| | | 23221-23247 | AGG(23222-23212) Y6999d |
| 4 | IS2640 | 1219181-1219447 | CGG(1219448-1219447) |
| | | 1192823-1197878 | CGG(1197874-1197874) |
| 5 | IS2641 | 5122285-5123501 | TGG(5123502-5123502) |
| 6 | IS2645 | 3423128-3427741 | CGG(3427746-3427746) |
| 7 | IS2638 (IS110) | 2023950 - 2024240 | CGG (2024240 - 2024240) |
| | | 2341285-2342645 | CGG(2342644-2342644) |
| 8 | IS3 | 113265 - 113282 | AGG (113283-113283) |
| | | 242223 -242213 | CGG (242214-242214) |
| | | 1174228-1174220 | CGG (1174229-1174229) |
| 9 | IS4 | 1132251-1132272 | AGG (1132273-1132273) |
| | | 9232771-9234120 | CGG (9234112-9234114) |
| 10 | IS5 | 1423222-1423226 | CGG (1423227-1423227) |
| 11 | IS2677 | 2643471-2643825 | AGG(2643821-2643821) |
| 12 | Kpn family accumulation-promoting endonuclease homologue | 137173 -137112 | AGG(137177-137175) |
| | | 1671261-1671654 | CGG(1671641-1671641) |
| | | 5161251-5162273 | CGG(5162261-5162261) |

4 Conclusion

Analyzing the presence of PAI in infectious strains has revealed interesting information regarding the virulence nature of pathogens. However, analysis of the factors of dissemination and a mechanism to combat dissemination was the focus of the current study. The results reveal a considerable association with respect to virulence genes of the PAI, presence of Insertion sequence (factors of dissemination of AMR genes through horizontal gene transfer) and the CRSPiR Cas/PAM sequences in proximity/within the PAI/IS sequences. Utilizing the CRISPR Cas technology to knock down the transposase genes of the IS sequences in these locations is proposed to be a potential means to control the dissemination of PAI's and hence be effective in the control of these infectious strains in the clinical scenarios and the environment where the surge of AMR genes is a challenge. The results present the details with respect to the uropathogenic strain of *E.coli* CFT073. The study when extended to other strains and genomes will enhance the understanding on this targeted approach of controlling the dissemination of PAIs utilizing CRISPR Cas.

References

- [1] Terlizzi, M. E., Gribaudo, G., & Maffei, M. E. (2017). UroPathogenic *Escherichia coli* (UPEC) Infections: Virulence Factors, Bladder Responses, Antibiotic, and Non-antibiotic Antimicrobial Strategies. *Frontiers in microbiology*, 8, 1566. <https://doi.org/10.3389/fmicb.2017.01566>.
- [2] Luo, C., Hu, G. Q., & Zhu, H. (2009). Genome reannotation of *Escherichia coli* CFT073 with new insights into virulence. *BMC genomics*, 10, 552. <https://doi.org/10.1186/1471-2164-10-552>
- [3] Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., Perna, N. T., Mobley, H. L., Donnenberg, M. S., & Blattner, F. R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26), 17020–17024. <https://doi.org/10.1073/pnas.252529799>
- [4] Karginov, F. V., & Hannon, G. J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. *Molecular cell*, 37(1), 7–19. <https://doi.org/10.1016/j.molcel.2009.12.033>
- [5] Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T., & Zhao, H. (2015). Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS synthetic biology*, 4(5), 585–594. <https://doi.org/10.1021/sb500255k>
- [6] Santhiya K, Jayanthi S, Ananthasubramanian M. (2021). Insilico Comparative Genome Analysis On Resistome Determinants And Mobile Elements Of Ndm Resistance. *Annals of the Romanian Society for Cell Biology*, 20886–20897. Retrieved from <https://annalsofscb.ro/index.php/journal/article/view/9813>
- [7] Alcock, B. P., Raphenya, A. R., Lau, T., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., Faltyn, M., ... McArthur, A. G. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research*, 48(D1), D517–D525. <https://doi.org/10.1093/nar/gkz935>.
- [8] Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic acids research*, 47(D1), D687–D692. <https://doi.org/10.1093/nar/gky1080>.
- [9] Varani, A.M., Siguier, P., Goubeyre, E. et al. ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* 12, R30 (2011). <https://doi.org/10.1186/gb-2011-12-3-r30>
- [10] Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalík, J., Néron, B., Rocha, E., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). CRISPRCasFinder, an update of CRISPRfinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic acids research*, 46(W1), W246–W251. <https://doi.org/10.1093/nar/gky425>
- [11] Heigwer, F., Kerr, G., & Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nature methods*, 11(2), 122–123. <https://doi.org/10.1038/nmeth.2812>
- [12] Ho Sui, S. J., Fedynak, A., Hsiao, W. W., Langille, M. G., & Brinkman, F. S. (2009). The association of virulence factors with genomic islands. *PloS one*, 4(12), e8094. <https://doi.org/10.1371/journal.pone.0008094>
- [13] McMurry, L., Petrucci, R. E., Jr, & Levy, S. B. (1980). Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 77(7), 3974–3977. <https://doi.org/10.1073/pnas.77.7.3974>
- [14] Jacoby, G. A., & Munoz-Price, L. S. (2005). The new beta-lactamases. *The New England journal of medicine*, 352(4), 380–391. <https://doi.org/10.1056/NEJMra041359>
- [15] Haldorsen, B., Aasnaes, B., Dahl, K. H., Hanssen, A. M., Simonsen, G. S., Walsh, T. R., Sundsfjord, A., & Lundblad, E. W. (2008). The AmpC phenotype in Norwegian clinical isolates of *Escherichia coli* is associated with an acquired ISEcp1-like ampC element or hyperproduction of the endogenous AmpC. *The Journal of antimicrobial chemotherapy*, 62(4), 694–702. <https://doi.org/10.1093/jac/dkn257>

- [16] Sadler, M., R Mormile, M., & L Frank, R. (2020). Characterization of the IS200/IS605 Insertion Sequence Family in *Halanaerobium Hydrogeniformans*. *Genes*, 11(5), 484. <https://doi.org/10.3390/genes11050484>
- [17] Azpiroz, M. F., Poey, M. E., & Laviña, M. (2009). Microcins and urovirulence in *Escherichia coli*. *Microbial pathogenesis*, 47(5), 274–280. <https://doi.org/10.1016/j.micpath.2009.09.003>
- [18] Habouria, H., Pokharel, P., Maris, S., Garénaux, A., Bessaiah, H., Houle, S., Veyrier, F. J., Guyomard-Rabenirina, S., Talarmin, A., & Dozois, C. M. (2019). Three new serine-protease autotransporters of Enterobacteriaceae (SPATEs) from extra-intestinal pathogenic *Escherichia coli* and combined role of SPATEs for cytotoxicity and colonization of the mouse kidney. *Virulence*, 10(1), 568–587. <https://doi.org/10.1080/21505594.2019.1624102>
- [19] Jung, W. H., Sham, A., White, R., & Kronstad, J. W. (2006). Iron regulation of the major virulence factors in the AIDS-associated pathogen *Cryptococcus neoformans*. *PLoS biology*, 4(12), e410. <https://doi.org/10.1371/journal.pbio.0040410>
- [20] Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A. Transcription Profile of *Thermus thermophilus* CRISPR Systems after Phage Infection. *Journal of Molecular Biology*. 2010;395(2):270–281. pmid:19891975
- [21] Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. *Methods in molecular biology* (Clifton, N.J.), 1311, 47–75. https://doi.org/10.1007/978-1-4939-2687-9_4