The identification of single nucleotide polymorphisms in the ampicillin-resistant strain of *E. coli.*

Michael Greenberbg, Polina Vaganova
Bioinformatics Institute, Saint Petersburg, Russia

Abstract. The rise of antibiotic-resistant strains among pathogenic bacteria is a pressing worldwide issue, demanding swift adaptation of our defenses against them. Modern sequencing technologies allow for a thorough investigation into the underlying factors and specific mechanisms driving the emergence of antibiotic resistance. In this study, we conducted a variant calling analysis on an *E. coli* strain resistant to ampicillin. Our findings revealed single nucleotide polymorphisms within the *ftsI*, *acrB*, and *envZ* genes, potentially influencing antibiotic resistance of this strain.

Keywords: *E. coli*, antibiotic resistance, ampicillin, SNP, *ftsI*, *acrB*, *envZ* .

1 Introduction

Escherichia coli, often referred to as *E. coli*, is a type of Gram-negative bacterium. While many strains are typically part of the natural gut flora, specific pathogenic variations have the potential to lead to a wide range of infections, including urinary tract infections and severe gastrointestinal diseases.[1] The overuse of commonly prescribed antibiotics, like ampicillin — a beta-lactam antibiotic that targets bacterial cell wall synthesis — can lead to the emergence of antibiotic resistance. This significantly complicates the treatment of such infections. Mutations within crucial genes is a substantial contributor to the development of antibiotic resistance in bacteria.[2–4] Identifying these mutations in antibiotic-resistant strains offers valuable insights into the underlying mechanisms driving resistance emergence, enabling a more precise tailoring of therapeutic strategies. In this study, we focus on identifying specific single nucleotide polymorphisms (SNPs) that may be responsible for antibiotic resistance in an ampicillin-resistant strain of *E. coli*.

2 Materials and methods

For this study raw Illumina sequencing reads from shotgun sequencing of an *E. coli* strain that is resistant to the antibiotic ampicillin were used. Read quality was assessed using FastQC[5] tool (version 0.12.1). Trimming of of low quality bases was performed using Trimmomatic[6] tool (version 0.36) in paired end mode using phred33 quality scale (-phred33) and following options (LEADING:20 TRAILING:20 SLIDINGWINDOW:10:20 MINLEN:20).

Alignment and indexation of reads was performed using BWA (version 0.7.17).[7] Sorting and indexation of alignment file was performed using SAMtools[8] (version 1.3.1). Variants were identified using Varscan[9] (version 2.4.6) with a minimum variant frequency parameter set to 20% (–min-var-frequency=0.2). Automated SNP annotation was performed using SnpEff[10] (version 5.2). Results were analyzed using IGV[11] (version 2.16.2).

3 Results

Genome sequence of the ampicillin-resistant *E. coli* strain included a total of 455,876 pairs of forward and reverse reads. Each read covered 101 bases, with an average GC content of 50%. The initial FastaQC analysis identified quality issues in the forward sequences, particularly in terms of quality per base and per tile. Similar quality concerns were noted for the reverse sequences, but they were limited only to quality per base. Subsequent trimming with Trimmomatic resolved the problem for per-base quality problem for both forward and reverse reads. However, the per-tile quality issue persisted in the forward reads. Following trimming, 446,259 high-quality paired forward and reverse sequences, ranging from 20 to 100 bases in length, remained for further analysis (See Supplementary materials). Filtered reads were then indexed and aligned to the reference genome of the K-12 substrain MG1655. The resulting coverage was 99.8%. Aligned reads were examined for variant calling with Varscan with minimum variant frequency parameter set to 20% as recommended by authors of the software tool.[9] Examination revealed five variant positions, all of which were identified as SNPs. VCF file, containing all these SNPs was then automatically annotated. Annotation results are shown in Table 1.

SNP position in genome	SNP Type	Gene	Base change	Amino Acid change
93043	missense variant	*ftsI*	G –>C	Ala544 –>Gly
482698	missense variant	*acrB*	T –>A	Gln569 –>Leu
852762	upstream/downstream gene variant	intragenic region	A –>G	–
3535147	missense variant	*envZ*	A –>C	Val241 –>Gly
4390754	synonymous variant	*rsgA*	G –>T	Ala252 –>Ala

Table 1: *Results of automatic annotation by SnpEff.*

4 Discussion

There are four main mechanisms, that confer antibiotic resistance to bacteria: limiting the intake of a drug, modification of the drug's target, inactivation of a drug, and active drug expel from the cell.[12] Variant calling analysis revealed three SNPs that directly alter the products of the *ftsI*, *acrB*, and *envZ* genes. Notably, all these genes are related to antibiotic resistance, suggesting that mutations in them may lead to ampicillin resistance.

The product of the *ftsI* gene, a transpeptidase, plays a crucial role in bacterial cell division. FtsI together with glycosyltransferase FtsW constructs the cell wall during septal formation.[13] It is worth mentioning that FtsI is the target of ampicillin, and it proposed that a mutation in *ftsI* gene may alter the binding of ampicillin to FtsI.[14] The product of the *acrB* gene, together with AcrA and TolC proteins, forms an RND efflux pump, which is part of

the multidrug resistance (MDR) efflux pumps in bacteria. These pumps actively move drugs including antibiotics out of the cell.[15] The product of the *envZ* gene, in conjunction with OmpR, forms the EnvZ/OmpR signaling system, which plays a pivotal role in osmoregulation and stress responses by controlling the expression of outer membrane porins (OMPs)[16,17] . Studies have shown that in response to β-lactams EnvZ/OmR activation leads to reduction in outer membrane permeability through changes in the expression of specific OMPs, which serve as a major route for the entry of small hydrophilic antibiotics.[17,18]

In conclusion, our data suggests that some of the identified SNPs in the genome of the studied *E. coli* strain may potentially contribute to its resistance to ampicillin. However, this requires proper experimental validation. An effective approach may involve generating strains harboring only one mutation and examining their response to the antibiotic. This would help identify which mutations play a pivotal role in ampicillin-resistance in this strain.

References

1 J. B. Kaper, J. P. Nataro, and H. L. T. Mobley, "Pathogenic escherichia coli," *Nature Reviews Microbiology* **2**, 123–140 (2004).

2 N. Woodford and M. Ellington, "The emergence of antibiotic resistance by mutation," *Clinical Microbiology and Infection* **13**, 5–18 (2007).

3 B. Ramanathan, H. M. Jindal, C. F. Le, *et al.*, "Next generation sequencing reveals the antibiotic resistant variants in the genome of pseudomonas aeruginosa," *PLOS ONE* **12**, e0182524 (2017).

4 T. B. Johansen, L. Scheffer, V. K. Jensen, *et al.*, "Whole-genome sequencing and antimicrobial resistance in brucella melitensis from a norwegian perspective," *Scientific Reports* **8** (2018).

5 S. Andrews, "Fastqc. a quality control tool for high throughput sequence data," (2010).

6 A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for illumina sequence data," *Bioinformatics* **30**, 2114–2120 (2014).

7 H. Li and R. Durbin, "Fast and accurate short read alignment with burrows–wheeler transform," *Bioinformatics* **25**, 1754–1760 (2009).

8 P. Danecek, J. K. Bonfield, J. Liddle, *et al.*, "Twelve years of SAMtools and BCFtools," *GigaScience* **10** (2021).

9 D. C. Koboldt, Q. Zhang, D. E. Larson, *et al.*, "VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing," *Genome Research* **22**, 568–576 (2012).

10 P. Cingolani, A. Platts, L. L. Wang, *et al.*, "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff," *Fly* **6**, 80–92 (2012).

11 J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, *et al.*, "Integrative genomics viewer," *Nature Biotechnology* **29**, 24–26 (2011).

12 W. C. R. and, "An overview of the antimicrobial resistance mechanisms of bacteria," *AIMS Microbiology* **4**(3), 482–501 (2018).

13 M. Attaibi and T. den Blaauwen, "An updated model of the divisome: Regulation of the septal peptidoglycan synthesis machinery by the divisome," *International Journal of Molecular Sciences* **23**, 3537 (2022).

14 S. Tristram, M. R. Jacobs, and P. C. Appelbaum, "Antimicrobial resistance in haemophilus influenzae," *Clinical Microbiology Reviews* **20**, 368–389 (2007).

15 D. J. Eaves, V. Ricci, and L. J. V. Piddock, "Expression of acrb, acrf, acrd, mara, and soxs in salmonella enterica serovar typhimurium: role in multiple antibiotic resistance," *Antimicrobial Agents and Chemotherapy* **48**, 1145–1150 (2004).

16 S. J. Cai and M. Inouye, "EnvZ-OmpR interaction and osmoregulation in escherichia coli," *Journal of Biological Chemistry* **277**, 24155–24161 (2002).

17 D. Ko and S. H. Choi, "Mechanistic understanding of antibiotic resistance mediated by envz/ompr two-component system in salmonella enterica serovar enteritidis," *Journal of Antimicrobial Chemotherapy* **77**, 2419–2428 (2022).

18 A. Jaffe, Y. A. Chabbert, and O. Semonin, "Role of porin proteins OmpF and OmpC in the permeation of beta-lactams," *Antimicrobial Agents and Chemotherapy* **22**, 942–948 (1982).

5 Supplementary materials

1. Reads from ampicillin-resistant *E. coli* strain: https://doi.org/10.6084/m9.figshare.10006541.v3

2. FastaQC analysis results: google disk