Genomic Insights into Antibiotic Resistance in E. coli: Unraveling SNPs and Treatment Strategies

Zherko Nikita, Vinogradova Sofiya
Moscow Institute of Physics and Technology

October 28, 2023

Abstract

Antibiotic resistance is currently a major problem in the world. It is though of extreme importance to understand the mechanisms and reasons for resistance to cope with that and to save patients. In this work, we used bioinformatical tools to work with sequencing reads to identify single-nucleotide polymorphisms (SNPs) responsible for ampicillin resistance in *Escherichia coli*. We managed to identify 6 SNPs, 5 of which correspond to genes that could potentially help bacteria escape from the antibiotic. We also give recommendations for how to treat patients with such resistance.

1 Introduction

Escherichia coli is one of the major pathogen causing diarrhea, sepsis, and other symptoms in human and animals [1]. Ampicillin, a widely used semi-synthetic β-lactam antibiotic, has been effective against *E. coli* infections. However, there is a growing concern about the increasing resistance to Ampicillin. Understanding the mechanisms that underlie this resistance is essential for developing effective therapeutic strategies and minimizing its impact.

Ampicillin resistance can occur through mechanisms such as encoding β-lactamase enzymes, altering cell wall target proteins, reducing outer membrane permeability, and boosting drug efflux pump expression [2]. In order to combat resistance it is necessary to clearly understand which genetic mutations cause the cell to go down this pathway.

In this work we use Next Generation Sequencing (NGS) data from ampicillin resistant *E. coli* and various informatics tools to identify genetic mutations that cause resistance and provide therapeutic recommendations to treat patients with such resistance.

2 Materials and Methods

2.1 Raw data

The reference *E. coli* genome was downloaded from NCBI FTP [3]. Illumina sequencing reads from the ampicilline resitant *E. coli* strain was downloaded from [4].

2.2 Manual data inspection

NGS files were inspected using the following commands.

```
$ head -20
$ cat
$ wc -l
```

2.3 FastQC data inspection

NGS files were inspected with FastQC [5] using the following command.

```
$ fastqc -o . [filename.fastq.gz]
```

2.4 Filtering reads

The reads were filtered using Trimmomatic tool [6] with the following parameters: cut bases off the start of a read if quality below 20, cut bases off the end of a read if quality below 20, trim reads using a sliding window approach, with window size 10 and average quality within the window 20, drop the read if it is below length 20.

```
$ trimmomatic PE [filenames]
$ ILLUMINACLIP:TruSeq3-PE.fa:2:30:10
$ LEADING:20 TRAILING:20 SLIDINGWINDOW:10:20 MINLEN:20
```

2.5 Alignment to genome

The filtered reads were aligned to the reference genome using BWA MEM [7], compressed, sorted and index using samtools (version=1.18) [8]. The resulting files were visualized using IGV tool [9].

2.6 Variants calling

Possible SNPs were scanned for by VarScan.v2.4.0.jar [10] with N=20.

```
$ java -jar VarScan.v2.4.0.jar mpileup2snp [filename]
$ --min-var-freq 0.2 --variants --output-vcf 1 >
$ [results].vcf
```

2.7 SNP annotation

SNP annotation was performed both munaully using IGV and automatically using SnpEff [11].

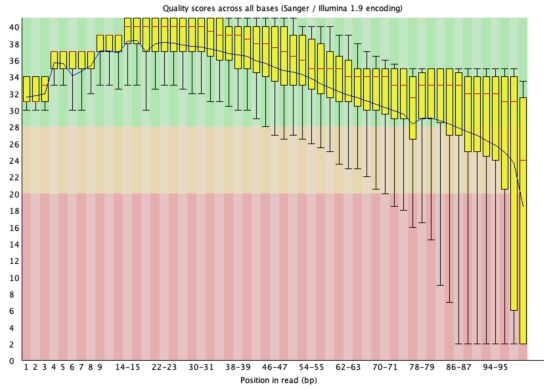
3 Results

3.1 Initial data overview

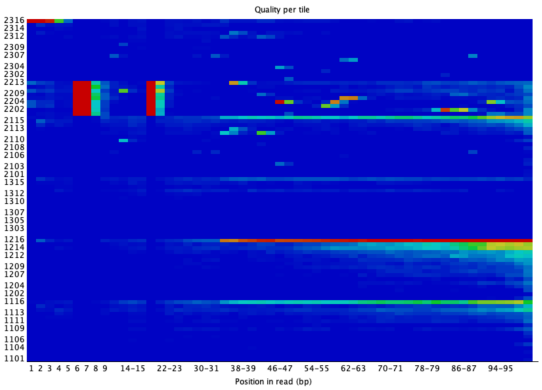
We first worked with files that correspond to ampicillin-resistant *Escherichia coli*. For each file with forward or reverse reads, the number of reads was calculated. These files were then analyzed using the FastQC tool (Table 1). The analysis revealed a high average quality of reads. Very unusual results were achieved in per base quality section and per tile quality section (Figure 1).

Sample	forward reads	reverse reads
Total Sequences	455876	455876
Sequences flagged as poor quality	0	0
Sequence length	101	101
GC-content, %	50	50

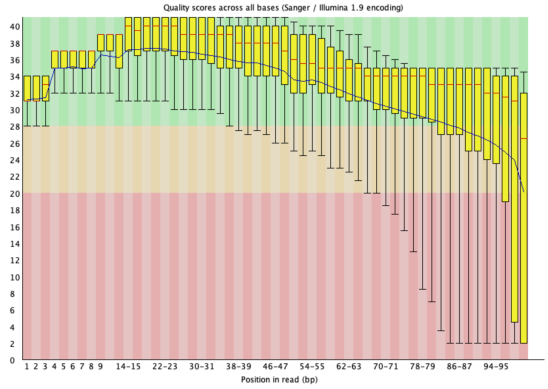
Table 1: FastQC basic statistic for forward and reverse reads



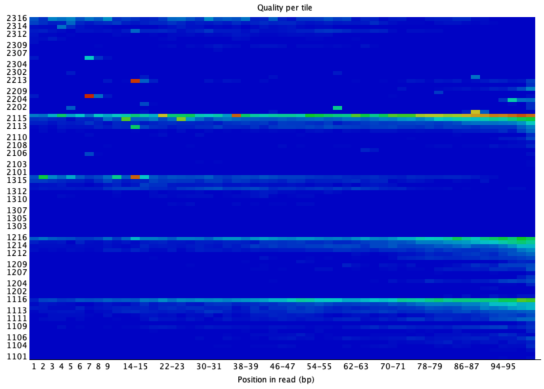
(a) Per base quality, forward reads



(b) Per tile quality, forward reads



(c) Per base quality, reverse reads



(d) Per tile quality, reverse reads

Figure 1: FastQC results for forwards and reverse reads

3.2 Reads filtering

Forwards and reverse reads were filtered using Trimmomatic tool, the results of the FastQC run are shown in the Supplementary (Figure 2, Table 4).

3.3 Alignment and SNP searching

Filtered reads were aligned to the non-resistance E.coli genome which resulted in 99.87 % coverage. To call the variants for each position in the genome the VarScan program was used. The threshold for calling a non-reference base a mutation was set to 20 %. SNPs were revealed both manually in the IGV browser and automatically. All in all, we managed to identify 6 SNPs (Table 2).

3.4 Number of reads at different stages of analysis

The numbers of reads at the beginning of the research, after trimming and the number of reads that aligned are shown below in table (Table 3).

Position in genome	Mutation type	Gene name	Gene function
93043	missense A ->G	*ftsI*	synthesis of peptidoglycan during cell division
482698	missense Q ->L	*acrB*	drug efflux
852762	no genes		
1905761	missense G ->D	*mntP*	manganese exporter
3535147	missense V ->A	*envZ*	responds to osmolarity changes
4390754	missense A ->S	*rsgA*	small ribosomal subunit biogenesis GTPase

Table 2: SNPs found in the ampicillin-resistant strain

Stage	Number of reads
Start	911752
After trimming	892518
After alignment	891649

Table 3: Number of reads during analysis

4 Discussion

In this work, we tried to identify SNPs responsible for ampicillin resistance in *E. coli*. To do so, we downloaded sequencing reads. The quality of data was rather high: the quality of bases was low only in the last nucleotides of each read which is rather common for Illumina sequencing. However, we considered filtered sequencing reads. The threshold for filtering was set as a common one, and since the quality of reads after filtering was satisfactory for us we did not change these conditions.
After alignment to the reference non-resistant *E. coli* genome we calculated the number of SNPs. The threshold for distinguishing Illumina error and real SNPs was set as 20 % since the coverage is rather high (about 20 reads per site) so it should be only 4 non-reference bases to consider them a mutation.
All in all, we found 6 SNPs, 5 of which correspond to a missense mutation in genes, *ftsI*, *acrB*, *mntP*, *envZ*, and *rsgA*.

4.1 Gene function overview and their role in resistance

AcrB is an inner membrane protein that plays a important role in antibiotic resistance in *E. coli*. It is a component of the AcrAB-TolC efflux pump system, which is responsible for pumping antibiotics out of the bacterial cell [12]. Mutations in AcrB can lead to the increased function of this protein, thus, actively reducing the

amount of Ampicillin in cells and providing them with resistance.

FtsI is a transpeptidase enzyme found in bacterial cells, and it plays a crucial role in the final stages of cell division. It is responsible for cross-linking the peptidoglycan strands in the cell wall during the synthesis of the division septum [13]. The inactivation of these protein by mutation may stall cell division and let cells survive the exposure to antibiotics since antibiotics mainly affect actively living and dividing cells [14].

RsgA is a GTPase involved in the biogenesis of the small ribosomal subunit in bacteria. It is responsible for the assembly and maturation of the 30S ribosomal subunit [15]. Mutations in RsgA, as with FtsI, alters protein's structure that results in the slower rates of ribosom assembly and overall cell growth which leads to Ampicillin resistance [16].

MntP is a manganese exporter protein and plays a role in maintaining manganese homeostasis in the cell [17].

EnvZ is a bacterial membrane protein that functions as a histidine kinase and phosphatase and plays a critical role in responding to changes in the osmolarity of the external environment [18].

4.2 Recommendations

In case of resistance to Ampicillin we provide several recommendations to treat patients with such resistance.

Doctors may use antibiotics that disrupt components of the AcrAB-TolC system (except tetracycline hydrochloride and ciprofloxacin) to avoid enhancing *acrAB* operon expression and, thus, *E. coli* resistance.

Another possible way to cope with resistance is to use another group of antibiotics, except penicillin group. It could antibiotics that do not affect cell growth but simply destroy the cell even it is not actively dividing. In this case lipopetides could be used.

5 Conclusion

In this work using informatics tools we managed to identify several SNPs in ampicillin-resistant strain of *E. coli*. We made several predictions of how these mutations could lead to resistance. We also provided recommendations for doctors to treat patients with ampicillin resistance. All in all, this work provides more genetic insights into the antibiotic resistance which is important for the development of new antibacterial treatments and therapeutics.

6 References

[1] Hidron AI, Edwards JR, Patel J, et al. Antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006-2007. *Infect Control Hosp Epidemiol.* 2008;29(11):996–1011. doi: 10.1086/591861

[2] Sifaoui, F.; Arthur, M.; Rice, L.; Gutmann, L.; Proulx, M.-E.; Désormeaux, A.; Marquis, J.-F.; Olivier, M.; Bergeron, M.G. Role of penicillin-binding protein 5 in expression of ampicillin resistance and peptidoglycan structure in *Enterococcus faecium. Antimicrob. Agents Chemother.* 2001, 45, 2594–2597.

[3] `https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF_000005845.2_ASM584v2/`

[4] `https://figshare.com/articles/dataset/amp_res_2_fastq_zip/10006541/3`

[5] `https://www.bioinformatics.babraham.ac.uk/projects/fastqc/`

[6] `http://www.usadellab.org/cms/?page=trimmomatic`

[7] `https://github.com/lh3/bwa`

[8] `https://www.htslib.org`

[9] `https://igv.org`

[10] `https://github.com/dkoboldt/varscan/blob/master/VarScan.v2.4.0.description.txt`

[11] `http://pcingola.github.io/SnpEff/`

[12] Pourahmad Jaktaji R, Jazayeri N. Expression of acrA and acrB Genes in Esherichia coli Mutants with or without marR or acrR Mutations. *Iran J Basic Med Sci.* 2013 Dec; 16(12): 1254-8. PMID: 24570831; PMCID: PMC3933802.

[13] Wissel MC, Weiss DS. Genetic analysis of the cell division protein FtsI (PBP3): amino acid substitutions that impair septal localization of FtsI and recruitment of FtsN. *J Bacteriol.* 2004 Jan;186(2):490-502. doi: 10.1128/JB.186.2.490-502.2004. PMID: 14702319; PMCID: PMC305773.

[14] Miller C, Thomsen LE, Gaggero C, Mosseri R, Ingmer H, Cohen SN. SOS response induction by beta-lactams and bacterial defense against antibiotic lethality. *Science.* 2004 Sep 10;305(5690):1629-31. doi: 10.1126/science.1101630. Epub 2004 Aug 12. PMID: 15308764.

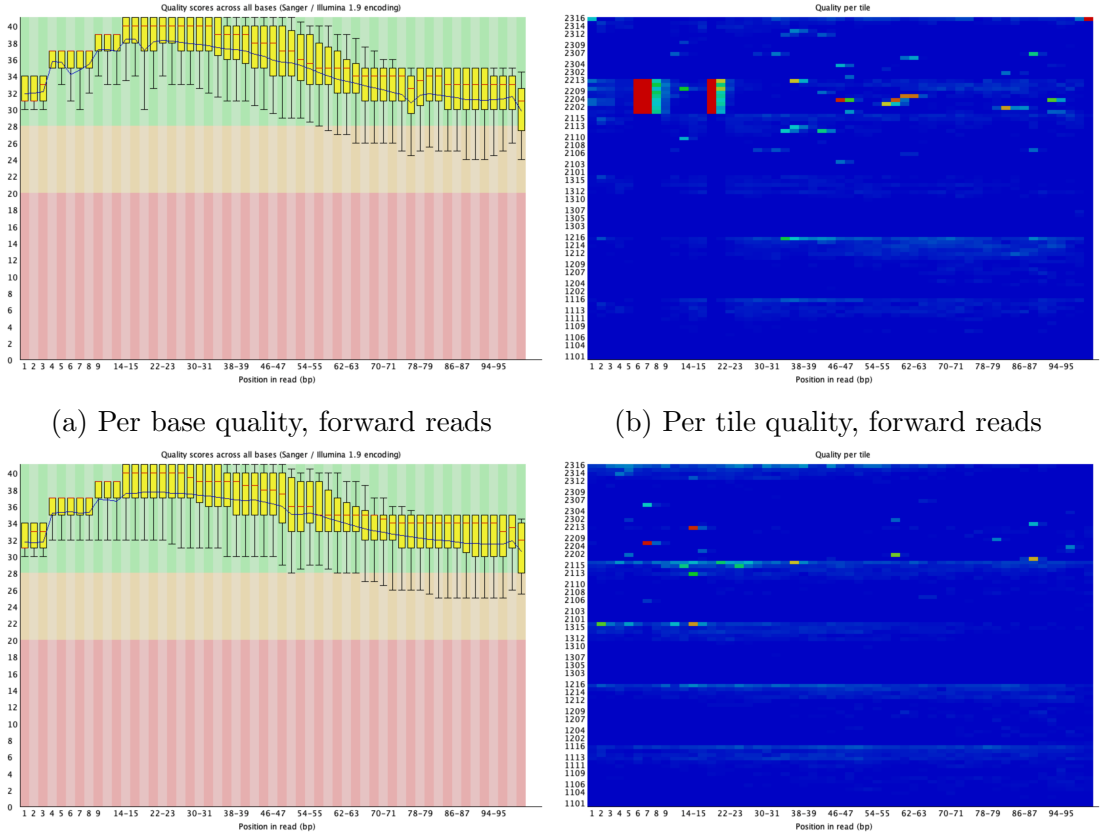
[15] `https://www.uniprot.org/uniprotkb/P39286/entry`

[16] Dörr T. Understanding tolerance to cell wall-active antibiotics. *Ann N Y Acad Sci.* 2021 Jul;1496(1):35-58. doi: 10.1111/nyas.14541. Epub 2020 Dec 3. PMID: 33274447; PMCID: PMC8359209.

[17] Martin JE, Waters LS, Storz G, Imlay JA. The *Escherichia coli* small protein MntS and exporter MntP optimize the intracellular concentration of manganese. *PLoS Genet.* 2015 Mar 16;11(3):e1004977. doi: 10.1371/journal.pgen.1004977. Erratum in: PLoS Genet. 2015 Jun;11(6):e1005322. PMID: 25774656; PMCID: PMC4361602.

[18] Cai SJ, Inouye M. EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*. *J Biol Chem.* 2002 Jul 5;277(27):24155-61. doi: 10.1074/jbc.M110715200. Epub 2002 Apr 24. PMID: 11973328.

7 Supplementary materials



(a) Per base quality, forward reads

(b) Per tile quality, forward reads

(c) Per base quality, reverse reads

(d) Per tile quality, reverse reads

Figure 2: FastQC results for forward and reverse filtered reads

Sample	forward reads	reverse reads
Total Sequences	446259	446259
Sequences flagged as poor quality	0	0
Sequence length	20-101	20-101
GC-content, %	50	50

Table 4: FastQC basic statistic for forward and reverse filtered reads