BIOINFORMATICS INSTITUTE

Detective story: SNP-footprints reveal the causes of *Escherichia coli* ampicillin resistance

Bukreeva Anastasia[1, 2 *] and Nikita Vaulin[1, 3, * †]

[1]Bioinformatics Institute, Saint-Petersburg, Russia
[2]Faculty of Biology, Moscow State University, Moscow, Russia
[3]Skolkovo Institute of Science and Technology, Moscow, Russia

*Contributed equally.
†vaulin@ro.ru

Abstract

The report presents the results of the analysis of K12 sequencing data from the ampicillin-resistant *E.coli* strain. The analysis detected five valid mutations in the following genes: *ftsI*, *acrB*, *rybA*, *envZ*, and *rsgA*. The detected substitutions were analysed for significance. We identified mutations in *acrB* and *ftsI* genes as the most likely candidates for the role of genetic determinants of resistance.

Key words: *Escherichia coli*; antibiotics; ampicilin, NGS, SNP

Introduction

Antibiotic resistance is a form of natural adaptation of the organism to the effects of chemicals that inhibit the reproduction and development of bacteria, fungi, viruses and cells. When antibiotic resistance occurs, the bacterium is able to tolerate significantly higher concentrations of the substance than other microorganisms of a similar strain [1]. Resistant strains of microorganisms appear when the bacterial cell genome is altered by a spontaneous mutation or as a result of the acquisition by the bacterium of an extrachromosomal transposon element already carrying the gene or genes of resistance to a particular substance. Most strains of *E.coli* are intestinal commensals that migrate easily both within and between animal and human populations. This fact ensures that the majority of *E. coli* strains are resistant to a large number of antibacterial compounds. Thus, *E.coli* strains persisting in the body and having R-plasmids are potentially capable of transferring R-plasmids by conjugation to other bacteria, including pathogenic ones, which leads to ineffectiveness of possible future antibiotic therapy[2]. Resistance of microorganisms to antimicrobial products in the case of both plasmid and chromosomal localization of resistance determinants may be due to different mechanisms, depending on both the type of bacterium and the substance itself. For example, the predominant mechanism of resistance to β-lactams in Gram-negative bacteria is the production of β-lactamases, whereas resistance to these compounds in Gram-positive organisms is achieved mainly through modification of their target site, the penicillin-binding proteins (PBP) aedo2019stationary. To summarize, the mechanisms of antibiotic resistance production can be classified as follows: i) modification of the antibacterial substance, ii) prevention of reaching the antimicrobial target (blocking the entry into the cell or active removal of the substance from the cell) iii) change or circumvention of target sites, iv) resistance due to the global adaptive processes of the cell. Each of the described mechanisms involves specific biochemical pathways and has certain genetic determinants[3].

Materials and methods

Dataset

In this study we use sequencing data from ampicillin resistant *Escherichia coli* K-12 strain [4]. The data was obtained with Illumina shotgun pair-end sequencing on HiSeq platform. *E. coli* K-12 MG1655 strain (RefSeq GCF_000005845.2 Primary Assembly) is used as a reference.

Processing

Raw reads were analysed with FastQC v.0.11.9 [5]. Both forward and reverse reads fastq-files contained 455876 reads. All FastQC reports are presented in project repository (see Supplemental materials).

Reads were trimmed and filtered with Trimmomatic v.0.39 [6]. The quality threshold was set to 20 (at the beginning of the read, at the end of the read, and on average with a sliding window of 10 nucleotide; parameters LEADING, TRAILING and SLIDINGWINDOW). Reads shorter than 20 nucleotides were removed from the analysis (parameter MINLEN). For trimming, standard TruSeq3-PE.fa adapters were used (see Supplemental materials).

Compiled on: November 8, 2022.
Draft manuscript prepared by the author.

Reads alignment

Reads alignment to the reference genome was conducted with BWA MEM tools v. 0.7.17-r1188 [7]. Reference genome index was made prior to alignment with *bwa index*, Sequence alignment was transformed and stored as a binary file with *samtools view -Sb* [8]. Here and below *samtools* v. 1.16.1 is used. Alignment statistics was obtained with *samtools flagstat*, alignment was sorted with *samtools sort* and alignment index was build with *samtools index*.

Variant calling

In order to obtain single nucleotide polymerisms, it was first obtained pileup file with tabulated number of bases at each position (for this we used *samtools mpileup* tool).

Single nucleotide polymorphisms from this file were obtained using the *mpileup2snp* function from VarScan tool v. 2.4.4 with the following parameters: "*–min-var-freq 0.5 –variants –output-vcf 1*" (minimum SNP frequency 50%) [9]. VarScan returned the result of 6 SNPs one of which was filtered by the strand-filter so there are 5 SNPs left to analyze. The results of variat calling were visualised with IGV v. 2.14.0 [10].

An automatic annotation of the SNPs was also carried out using SnpEff v. 5.1 *ann* function with *k12* database name (database was created with *builg -genbank -v k12* function) [11]. File snpEff.config used to set up SnpEff is located in snpEff folder in GitHub (see Supplemental materials).

Results

Sequencing data processing

The initially received data contained 455876 101-nucleotide long both forward and reverse reads (results 911752 reads in total, see Supplemental materials for GitHub repository with fastqc_output folder). Most of the reads had average per read phred 33 quality over 34. Reads trimming and filtering with Trimmomatic resulted 445524 both forward and reverse reads (97.73% of reads remained). There were also several unpaired reads (9951 for forward and 271 for reverse) which were further excluded from the analysis.

There were several forward reads with a degenerate N nucleotide at position 6, however, they remained after filtering. This issue is also depicted at quality per tile plot from which we can conclude that this is due to some temporary problem in the sequencing, most likely an air bubble.

Reads alignment

Trimmed and filtered reads was aligned using BWA-MEM method from the BWA package. 892776 reads were mapped in total (446388 for both forward and reverse). 892776 of them were paired properly 99.53% of mapped reads).

Detected SNPs

SNP detection was carried out with VarScan and SnpEff tools as described in Materials and methods. Only nucleotide mutations with 50% or more frequency considered as SNPs. All detected SNPs are presented in the table 1.

One of the polymorphisms is a silent mutation in the non-coding region of the *rybA* gene. One substitution in the *rsgA* gene is synonymous and does not result in an amino acid substitution. Two mutations - in genes *ftsI* and *envZ* are neutral and they lead to the substitution of an amino acid with a similar one. Another mutation, in the *acrB* gene, is a missense and leads to an amino acid substitution with a severe change in properties. These amino acid substitutions and their possible impact on antibiotic resistance will be discussed in more detail below.

Discussion

The sequencing analysis of the ampicillin-resistant Escherichia coli strain K12 revealed the following candidates for the role of genetic determinants of resistance.

FtsI gene

The Fts (filamenting temperature-sensitive) family of proteins plays an important role in the maintenance of the bacterial cell wall and cell division [12]. FtsI (penicillin-binding protein 3) is a transpeptidase necessary for the final step of peptidoglycan synthesis in septa [13]. In addition, recent data suggest the presence of a number of non-coding RNA transcripts, rlmD int and mglC int, in the ORF (open reading frame) of FtsI, which function as RNA sponges. The authors of the work suggest that rnc-transcript FtsI functions as a sponge for RybB and CpxQ, which are induced by the misfolding of outer and inner membrane proteins, respectively, and downregulate the corresponding classes of proteins [14].

Thus, mutations in this region of the bacterial chromosome may be directly associated with changes in bacterial cell membrane function, leading, for example, to more efficient removal of antibiotic from the cell and consequently to bacterial resistance.

rybA

rybA is a small noncoding RNA with poorly understood functions. It derives from the region with the gene encoding the manganese-dependent transcription factor MntR in *E. coli*, which controls the transcription of proteins associated with manganese biogenesis in the cell. Transcription factor MntR is the main sensor and transducer of excess manganese. After binding to manganese, MntR binds promoter DNA to repress or activate transcription of its target genes. Due to the specificity of the work of transcriptional regulators in bacteria, it has been suggested that *rybA* is a part of the MntR regulon [15]. At the same time, manganese being cofactor for enzymes with a variety of functions, so it is quite difficult to assume, how exactly *rybA* malfunction can lead to the emergence of antibiotic resistance in bacteria.

envZ

EnvZ is a protein kinase that is involved in the regulation of osmosis with OmpR. It has also been shown to have some effect on antibiotic resistance in bacteria [16].

acrB

AcrB (Acriflavine resistance protein B) is an key element of the ArAB-TolC system [17, 18]. This system, in turn, is a huge intermembrane complex, which carries out the specific pumping of a wide variety of substances outside the cell. The TolC complex is located in the outer membrane, while the AcrAB complex is located in the inner one. The Q569 residue is located in the C-terminal porter subdomain 1 (PC1) and plays a role in the specificity of molecular recognition and transfer.

This protein forms a channel out of itself that can also pump chemicals through itself due to conformational changes similar to those that ATP synthase undergoes during its cycle. resistance. The Q569L substitution changes both the size and charge of the amino acid, which has a dramatic effect on microenvironment.

Table 1. Single nucleotide polymorphisms found in sequencing data

Genome Position, bp [1]	SNP [2]	DNA strand	AA replacement[3]	Gene	Type of SNP
93 043	C/G	+	A544G	*ftsI*	Neutral
482 698	T/A	-	Q569L	*acrB*	Missense
852 762	A/G	+	—	*rybA*	Silent
3 535 147	A/C	-	V241G	*envZ*	Neutral
4 390 754	G/T	-	A252A	*rsgA*	Synonymic

Thus, a missense mutation in this protein is an obvious candidate for the cause of antibiotic.

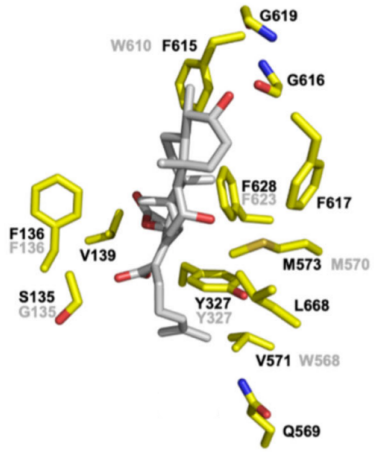

Figure 1. Crystal structure data of the acrB protein when bound to the fucidic acid molecule, which shows that the residue Q569 is also involved in the interaction with the substrate (PDB: 5NC5)[18]

References

1. Bryskier A, et al. Antimicrobial agents: antibacterials and antifungals. ASM press; 2005.
2. Gilmore MS, Clewell DB, Ike Y, Shankar N. Enterococci: From Commensals to Leading Causes of Drug Resistant Infection [Internet] 2014;.
3. Miller WR, Munita JM, Arias CA. Mechanisms of antibiotic resistance in enterococci. Expert review of anti-infective therapy 2014;12(10):1221–1236.
4. Raiko M, Ampres *E.coli* data. Figshare; 2020. https://figshare.com/articles/amp_res_2_fastq_zip/10006541/3, accessed: 2022-10-27.
5. FastQC tool;. Accessed: 2022-10-27. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
6. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014 Apr;30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25:1754 – 1760.
8. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience 2021 02;10(2). https://doi.org/10.1093/gigascience/giab008, giab008.
9. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research 2012;22 3:568–76.
10. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology 2011 Jan;29(1):24–26. https://doi.org/10.1038/nbt.1754.
11. Cingolani P, Platts A, Coon M, Nguyen T, Wang L, Land SJ, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 2012;6(2):80–92.
12. Mobley HL, Mendz GL, Hazell SL. Helicobacter pylori: physiology and genetics 2001;.
13. Daniel RA, Harry EJ, Errington J. Role of penicillin-binding protein PBP 2B in assembly and functioning of the division machinery of Bacillus subtilis. Molecular microbiology 2000;35(2):299–311.
14. Adams PP, Baniulyte G, Esnault C, Chegireddy K, Singh N, Monge M, et al. Regulatory roles of Escherichia coli 5'UTR and ORF-internal RNAs detected by 3'end mapping. Elife 2021;10:e62438.
15. Waters LS, Sandoval M, Storz G. The Escherichia coli MntR miniregulon includes genes encoding a small protein and an efflux pump required for manganese homeostasis. Journal of bacteriology 2011;193(21):5887–5897.
16. Li M, Liu Q, Teng Y, Ou L, Xi Y, Chen S, et al. &ltp&gtThe resistance mechanism of &ltem&gtEscherichia coli&lt/em&gt induced by ampicillin in laboratory&lt/p&gt. Infection and Drug Resistance 2019 Sep;Volume 12:2853–2863. https://doi.org/10.2147/idr.s221212.
17. Kobylka J, Kuth MS, Müller RT, Geertsma ER, Pos KM. AcrB: a mean, keen, drug efflux machine. Annals of the New York Academy of Sciences 2019;1459.
18. Ornik-Cha A, Wilhelm J, Kobylka J, Sjuts H, Vargiu AV, Malloci G, et al. Structural and functional analysis of the promiscuous AcrB and AdeB efflux pumps suggests different drug binding mechanisms. Nature Communications 2021;12.

Supplemental materials

GitHub repository of the project: GitHub