Mutations in ampicillin-resistant *Escherichia coli* reveal mechanisms of its resistance

Nikita Konstantinovsky, Ekaterina Kirillova, Bioinformatics Institute

Abstract

In this study, we investigated the sequencing data of the ampicillin-resistant *Escherichia coli* strain for mutations that cause antibiotic resistance. We performed alignment-based analysis, discovered mutations in five genes and suggest that a combination of three of them (in *ftsI, envZ* and *acrB)* cause ampicillin resistance. This review focuses on investigating the mechanisms behind the acquisition of *E. coli* resistance and the results obtained may help to find a new effective way to fight bacteria.

Introduction

E. coli is one of the most common causes of bacterial infections in humans and animals. The use of antibiotic agents against *E. coli* infections threatens to acquire resistance. Antibiotic resistance of bacterial pathogens is an acute problem and a serious threat to human health around the world [1]. The high genetic plasticity of bacteria allows them to respond to stress caused by antibiotics using several adaptation strategies: target site alteration, antibiotic inactivation or modification, alteration of metabolic pathway to compensate, reduction of drug amount in cell (by efflux pumps or decreasement of permeability).

For some mutations, determining the mechanism that mediates drug resistance is not difficult, for example, when a specific drug target is changed. But sometimes the relationship between mutation and drug resistance is complex and ambiguous, because bacteria often require multiple mutations to develop resistance to a drug [2]. A better understanding of the molecular mechanisms of *E. coli* resistance to *β*-lactam antibiotics is important for providing new methods for the prevention and control of multidrug-resistant bacteria and for selecting targets for new antibiotics.

Methods

In this study, we used raw Illumina pair end run sequencing reads from shotgun sequencing of ampicillin-resistant *E. coli* strain. We also used the reference sequence of the parental (unevolved, not resistant to antibiotics) *E. coli* K-12 strain substrain MG1655 [3].

First of all, we inspected the raw sequencing data with fastqc tool [4]. For the analysis, the default settings were used. To eliminate the detected issues, it was decided to filter the reads using Trimmomatic tool [5]. The following values were used as filtering parameters: *LEADING:20, TRAILING:20, SLIDINGWINDOW:10:20, MINLEN:20*.

After filtering, we used BWA package [6] to align the sequences to the reference. For this task, the bwa-mem tool [7] was used; the reference sequence and filtered reads were used as parameters. As a result, a SAM file was created, which we compressed, sorted and indexed using the samtools [8] to BAM file. To detect mutations we used samtools mpileup program [9] with default parameters to create a mpileup file. To find actual variants, we used VarScan tool [10] with mpileup2snp command. As parameters we have set: *--min-var-frequency 0.8, --variants flag, --output-vcf 1*. We used the IGV browser [11] to track mutations and visualize the results.

Results

Detailed results are presented in the laboratory notebook, links are provided in supplementary materials. Fastqc report showed the issues in the data in terms of "per base sequence quality" and "per

tile sequence quality". Figure 1 in supplementary materials shows quality distribution for forward run sequence.

After filtering, 439769 of the 455876 original reads remained, more than 96% high quality reads were selected. After that, we ran fastqc tool for filtered data, the issue "per base sequence quality" disappeared. Figure 2 and Figure 3 in supplementary materials shows quality distribution for filtered forward and reverse run sequence.

As a result of alignment, 99.87% of the reads were mapped. Next, we launched the VarScan tool, it found 5 SNPs, we examined them using an IGV browser, all detected substitutions in protein-coding genes are shown in Table 1.

Table 1. SNPs in protein-coding genes.

Gene id	**Position**	**Ref Codon**	**Ref AA**	**Reads Codon**	**Reads AA**	**Gene**
b0084	93043	GCC	A	GGC	G	*ftsI*
b0462	482698	CAG	Q	CTG	L	*acrB*
b3404	3535147	GTA	V	GGA	G	*envZ*
b4161	4390754	GCC	A	GCA	A	*rsgA*

We also detected SNP in the *rybA* gene (rna-b4416) with A→G substitution at position 852762. The product of this gene is small RNA RybA.

Discussion

We discovered a mutation in *ftsI* gene, which occurs Ala-to-Gly substitution in penicillin-binding catalytic domain of FtsI protein. FtsI is inhibited by *β*-lactams, but amino acid substitution near the site of the transpeptidase domain can decrease the affinity for ampicillin. However, there is the study showing that mutations in the *E. coli ftsI* gene by themselves did not increase antibiotic resistance, but the combination of *envZ* and *ftsI* mutations had a synergistic effect and significantly increased the minimum inhibitory concentration of antibiotics [12].

Mutation was also identified in the *envZ* gene. EnvZ protein is a member of the two-component regulatory system EnvZ/OmpR involved in osmoregulation (particularly of genes *ompF* and *ompC*). EnvZ is involved in regulation of these proteins and the *envZ* mutation can cause down-regulation of porins genes [12]. Mutants both lacking OmpC and OmpF proteins are resistant to *β*-lactams, because these proteins facilitate the penetration of *β*-lactams through the bacterial cell membrane [13].

Investigated *E. coli* strain also acquired mutation in *acrB* gene. This gene encodes multidrug efflux pump subunit AcrB, a member of AcrA-AcrB-AcrZ-TolC drug efflux protein complex. The effect of this mutation may be to overexpress the pump to reduce intracellular drug concentration.

In summary, this study suggests that under antibiotic stress, *E. coli* can employ a combination of several defence strategies, such as target alteration, decreasement of membrane permeability due to the lack of porin proteins, and increasement of efflux pump activity. An antibiotic with a different mode of action than *β*-lactams can be used as an alternative treatment for infection occurring by resistant *E. coli*. For example, ones that inhibit protein synthesis, interfere with the synthesis of nucleic acids, or that disrupt the cell membrane of bacteria can be used. It is also necessary to improve

new non-antibiotic approaches, such as natural and synthetic antimicrobial peptides and the competitive exclusion of pathogenic bacteria using probiotics.

Citations

1. "Antimicrobial resistance: global report on surveillance 2014," *World Health Organization*, 17-Aug-2016. [Online]. Available: https://www.who.int/drugresistance/documents/surveillancereport/en/. [Accessed: 05-Nov-2020].
2. J. M. Munita and C. A. Arias, "Mechanisms of Antibiotic Resistance," *Microbiology Spectrum*, vol. 4, no. 2, 2016.
3. "*Escherichia coli* (ID 167)," *National Center for Biotechnology Information*. [Online]. Available: https://www.ncbi.nlm.nih.gov/genome/167. [Accessed: 05-Nov-2020].
4. S. Andrews, "FastQC: A Quality Control Tool for High Throughput Sequence Data," [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. [Accessed: 05-Nov-2020].
5. A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for Illumina sequence data," *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, 2014.
6. *Burrows-Wheeler Aligner*. [Online]. Available: http://bio-bwa.sourceforge.net/. [Accessed: 05-Nov-2020].
7. H. Li, "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM," arXiv:1303.3997v1 [q-bio.GN], Mar. 2013.
8. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, "The Sequence Alignment/Map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.
9. H. Li, "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data," *Bioinformatics*, vol. 27, no. 21, pp. 2987–2993, 2011.
10. D. C. Koboldt, Q. Zhang, D. E. Larson, D. Shen, M. D. Mclellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, and R. K. Wilson, "VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing," *Genome Research*, vol. 22, no. 3, pp. 568–576, 2012.
11. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov, "Integrative genomics viewer," *Nature Biotechnology*, vol. 29, no. 1, pp. 24–26, 2011.
12. M. Adler, M. Anjum, D. I. Andersson, and L. Sandegren, "Combinations of mutations in *envZ*, *ftsI*, *mrdA*, *acrB* and *acrR* can cause high-level carbapenem resistance in *Escherichia coli*," *Journal of Antimicrobial Chemotherapy*, vol. 71, no. 5, pp. 1188–1198, 2016.
13. 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*, vol. 22, no. 6, pp. 942–948, 1982.

Supplementary materials

1. Day 1 laboratory notebook: https://github.com/MsKittin/What-causes-antibiotic-resistance-/blob/main/1_day_30_10_2020.md
2. Day 2 laboratory notebook: https://github.com/MsKittin/What-causes-antibiotic-resistance-/blob/main/2_day_02_11_2020.md