Search for mutations responsible for antibiotic resistance in *E. coli* genome

Abstract

Antibiotic resistance of some bacterial strains is one of the most important problems in modern medicine. The widespread use of antibiotics has led to the emergence of a wide range of mutations that allow bacteria to escape the action of drugs. Our investigation is aimed at identifying mutations in the *E.coli* genome leading to antibiotic resistance. In our research we have found 5 SNPs between *E.coli* strain K-12 substrain MG1655, widely used in laboratories, and *E. coli* strain that is resistant to ampicillin. Identification of mutations that lead to antibiotic resistance can become the basis for the development of new synthetic antibiotics.

Introduction

Antibiotics were discovered in 1928 and for a long time they were a reliable shield against bacterial infection for human beings. But uncontrolled use of these drugs has led humanity to a difficult fight against antibiotic-resistant bacteria. But how exactly did this happen? The action of an antibiotic is aimed at recognizing and disrupting the work of individual molecules that are encountered only in the bacterial cell. A bacterial cell can avoid antibiotic exposure in several ways: modifying the target of the drug during mutagenesis, inactivation or modification of antibiotic itself, alter metabolic pathway to compensate for the effect of the antibiotic, and reduce the concentration of antibiotic in cytoplasm or medium [1].

In this article, we have identified several ways to acquire antibiotic resistance, belonging to the groups of "modification of the antibiotic target" and “reduce of concentration of antibiotic”. The high division speed of bacterial cells leads to a high probability of polymerase error during DNA replication. A very small percentage of these mutations can lead to antibiotic resistance. We compared the genome of *E. coli* resistant to ampicillin with the reference genome and found out which mutations most likely led to antibiotic resistance. Information on how exactly the antibiotic target has changed can be used to create new generations of antibiotics that get round bacterial resistance.

Methods

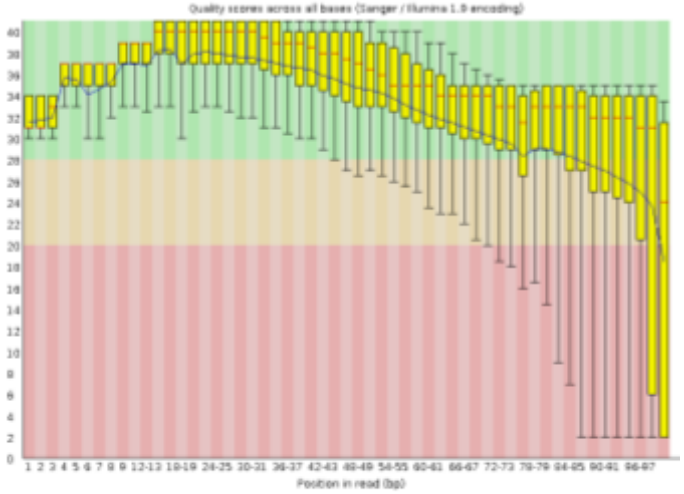
In this work, the genome of *E. coli* strain K-12 substrain MG1655 was used as reference: GCA_000005845.2_ASM584v2 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000005845.2/). We analyzed raw Illumina sequencing reads from shotgun sequencing of an *E. coli* strain that is resistant to the antibiotic ampicillin (https://figshare.com/articles/dataset/amp_res_2_fastq_zip/10006541). To confirm the quality of the received reads, we used the FastQC program [2]. Trimmomatic program with the following arguments was used to improve the quality of reads: `PE -phred33 LEADING:20 TRAILING:20 SLIDINGWINDOW:10:20 MINLEN:20` [3]. To

create indices based on the reference genome, we used the program `bwa index` [4]. The next step was to align the reads to the reference genome using the BWA-MEM algorithm (`bwa mem`). To convert the extension .sam to .bam we used the program Samtools (`samtools view -S -b`) [5]. On the next step we sorted and indexed .bam file with samtools with the following commands respectively: `samtools sort` and `samtools index`. `Samtools mpileup -f` was used to create an intermediate file with information about how many bases in each read match (or don't match) the reference genome. Then we called variants with `mpileup2snp` from VarScan.v2.3.9 using following parameters: `--min-var-freq 0.8 --variants --output-vcf 1` [6]. To determine how many reads have nucleotides that differ from the reference genome, we used online program to visualize the alignment results (IGV browser) and the console program (`bedtools intersect -loj`) [7, 8].

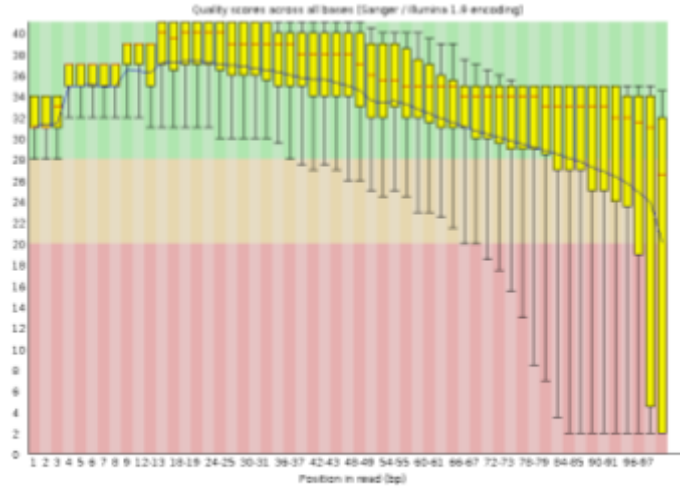
Results

Inspect and cleanup sequencing data

At the beginning of the study, we conducted raw data and appreciated its quality with the FastQC program (Pic. 1-2).

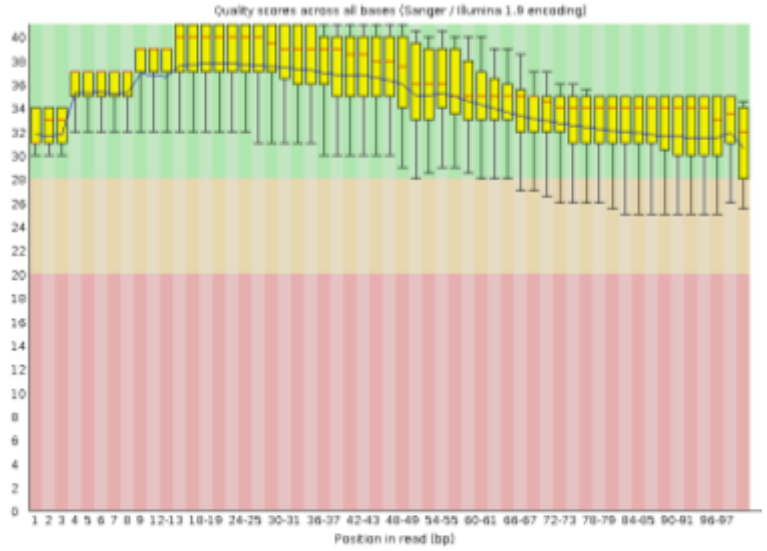


Pic. 1. Quality scores of raw sequence reads (reverse) across all bases

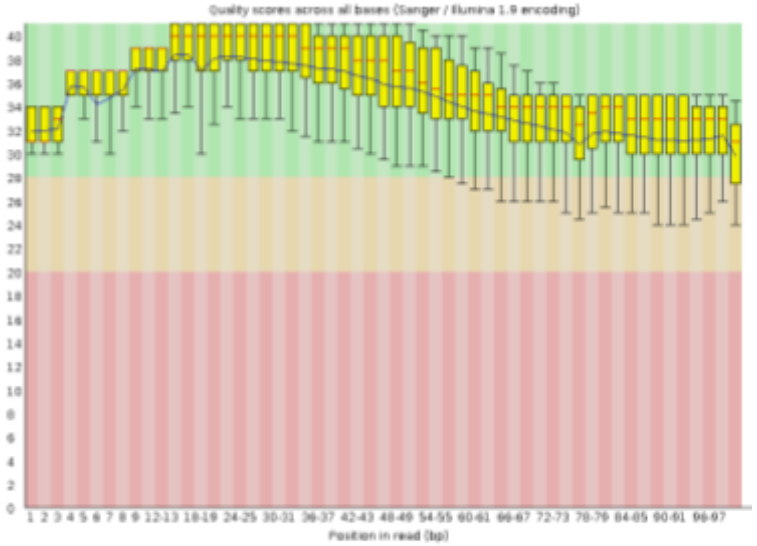


Pic. 2. Quality scores of raw sequence reads (forward) across all bases

The quality of the input data was low, so we used the Trimmomatic program to get rid of the low quality reads (and low quality parts of reads) (table 1). It can be seen that overall quality has increased (pic. 3-4.). In numbers, it increased from 32.9±4.4 to 34.5±2.6, and we got rid of big errors at the ends of reads.



Pic. 3. Quality scores of reads (forward) across all bases after trimming



Pic. 4. Quality scores of reads (reverse) across all bases after trimming

Table 1. Result of cleanup sequencing data with Trimmomatic

Read pairs	Both Surviving	Forward Only Surviving	Reverse Only Surviving	Dropped
455876	446259 (97.89%)	9216 (2.02%)	273 (0.06%)	128 (0.03%)

Aligning sequences to reference

First, an index file for the reference genome was created using `bwa index`. Index file makes it faster to search. Then, reads were mapped to the reference using `bwa mem` command. As a result, 99.87% of the reads left from the Trimmomatic step were correctly mapped to the reference (table 2).

Alignment file (.sam) was compressed with `samtools view -S -b`. The result, .bam file with compressed alignment, was sorted by sequence coordinate on reference (using `samtools sort`) and indexed (using `samtools index`) for faster search.

Table 2. Overall change in number of reads after cleanup and aligning

Read pairs in the beginning	Both surviving after Trimmomatic	Properly paired and mapped with bwa-mem
455876	446259 (97.89%)	444277 (97.45%)

Variant calling

Intermediate file with information about how many bases in each read match (or don't match) the reference genome was made using `samtools mpileup -f` command. Using this file we called variants with `mpileup2snp` command and `--min-var-freq` set to `0.8`, which means that the same SNP in at least 80% reads is needed to call it mutation. As a result, we got 5 variant positions reported (table 2).

Table 3. Mutations in resistant to ampicillin *E. coli* strain

Position	Reference base -> Alternative base	Ref aa -> Alt aa	Gene	Gene type	Mutation type
93043	C -> G	Ala -> Gly	*ftsI* (forward)	Protein coding	Missense
482698	T -> A	Gln -> Leu	*acrB* (reverse)	Protein coding	Missense
852762	A -> G	-	*rybA* (reverse) (possibly mntS)	ncRNA	-
3535147	A -> C	Val -> Gly	*envZ* (reverse)	Protein coding	Missense
4390754	G -> T	Ala -> Ala	*rsgA* (reverse)	Protein coding	Synonymous

Mutation in the ftsI gene is missens, and it leads to change in the amino acid alanine to glycine, differing only by the -CH3 group. The *ftsI* gene encodes a Peptidoglycan D,D-transpeptidase protein, that catalyzes cross-linking of the peptidoglycan cell wall at the division septum.

acrB gene encodes multidrug efflux pump subunit of a large AcrA-AcrB-AcrZ-TolC protein complex, which works as a proton-substrate antiporter. The mutation leads to an amino acid substitution (Gln -> Leu) with a change in hydrophobic properties [9].

rybA is a gene that encodes small RNA. It was found that the expression of the sRNA is strongly upregulated under cold shock and peroxide stress, which means that *rybA* might play a role during cold shock and/or oxidative stress [10]. It seems like *rybA* regulates expression of genes for aromatic amino acid biosynthesis - it can negatively regulate *aroL* and *aroF* expression during peroxide stress, and positively regulate *cusR* regulon at the same time [11]. Also, it was found that *rybA* encodes small protein mntS that can function as a chaperone that makes manganese more available [12].

SNP in *envZ* gene leads to missense mutation. Gene encodes sensor histidine kinase EnvZ, a part of EnvZ/OmpR protein system. It can respond to environmental signals such as low or high osmolarity, and activate/repress membrane porins F and C transcription through OmpR phosphorylation [13].

rsgA gene encodes ribosome small subunit-dependent GTPase A. Mutation in this gene is synonymous.

Discussion

The mutation in the *rsgA* gene is synonymous and the functions of the *rybA* gene are different - its work is mainly associated with the regulation of magnesium and silver ions, and weakly correlates with the emergence of resistance. Thus, it is the mutations in the *ftsI*, *acrB*, and *envZ* genes that, in our opinion, could lead to antibiotic resistance in *E. coli*.

The *ftsI* gene encodes the enzyme transpeptidase, which is needed by bacteria to make the cell wall. Ampicillin acts as an irreversible inhibitor of transpeptidase, which leads to the death of bacteria. A mutation in this gene probably changes the conformation of the protein, which leads to the impossibility of attachment of ampicillin and the correct operation of transpeptidase.

It is known that the *acrB* gene is responsible for resistance to some antibiotics [14]. It is possible that a mutation in the acrB gene can increase pump efficiency, which will lead to improved elimination of ampicillin from the cell. This mutation can lead to antibiotic resistance.

The third mutated gene is *envZ*. This gene is responsible for activating the transcription of porin genes in response to changes in osmotic molarity. The loss of porins OmpF and OmpC is known to be associated with antibiotic resistance in several reports [15]. Thus, a mutation in a gene can lead to a change in the level of porin transcription and the emergence of resistance.

In most cases, you can use an alternative antibiotic that acts on other targets of the bacterial cell. It can be chloramphenicol [16], rifampicin [17] and antibiotics of the tetracycline series [18] (their principle of operation is based on inhibition of RNA binding to ribosome), as well as nalidixic acid [19] (inhibits the replication of DNA of bacterium). But there is a possibility that these types of antibiotics can be not effective because of mutations in acrB and envZ genes. There are also alternative ways to get rid of the pathogenic bacillus. Bacteriophages are able to recognize and effectively destroy colonies of *E. coli* and antibiotic resistance does not reduce the efficiency of this process.

References

[1] Martinez JL. General principles of antibiotic resistance in bacteria. Drug Discov Today Technol. 2014 Mar;11:33-9. doi: 10.1016/j.ddtec.2014.02.001. PMID: 24847651.

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

[3] Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

[4] Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60. [PMID: 19451168]

[5] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H, Twelve years of SAMtools and BCFtools, GigaScience (2021) 10(2) giab008 [33590861]

[6] Koboldt, D., Zhang, Q., Larson, D., Shen, D., McLellan, M., Lin, L., Miller, C., Mardis, E., Ding, L., & Wilson, R. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing Genome Research DOI: 10.1101/gr.129684.111

[7] James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 (2011).

[8] Quinlan AR and Hall IM, 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 6, pp. 841–842.

[9] Du D, Wang Z, James NR, et al. Structure of the AcrAB-TolC multidrug efflux pump. Nature. 2014;509(7501):512-515. doi:10.1038/nature13205

[10] Gerstle, Kirstin; Klätschke, Kristin; Hahn, Ulrich; Piganeau, Nicolas (2012). The small RNA RybA regulates key-genes in the biosynthesis of aromatic amino acids under peroxide stress in E. coli. RNA Biology, 9(4), 458–468. doi:10.4161/rna.19065

[11] Gerstle K, Klätschke K, Hahn U, Piganeau N. The small RNA RybA regulates key-genes in the biosynthesis of aromatic amino acids under peroxide stress in E. coli. RNA Biol. 2012;9(4):458-468. doi:10.4161/rna.19065

[12] https://www.uniprot.org/uniprot/P0DKB3

[13] Forst S, Delgado J, Inouye M. Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the ompF and ompC genes in Escherichia coli. Proc Natl Acad Sci U S A. 1989;86(16):6052-6056. doi:10.1073/pnas.86.16.6052

[14] Okusu H, et al. 1996. J Bacteriol 178(1): 306-308. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of Escherichia coli multiple-antibiotic-resistance (Mar) mutants. (PMID 8550435)

[15] Ziervogel BK, Roux B. The binding of antibiotics in OmpF porin. Structure. 2013;21(1):76-87. doi:10.1016/j.str.2012.10.014

[16] Bartlett JG. Chloramphenicol. Med Clin North Am. 1982 Jan;66(1):91-102. doi: 10.1016/s0025-7125(16)31444-4. PMID: 7038343.

[17] Binda G, Domenichini E, Gottardi A, Orlandi B, Ortelli E, Pacini B, Fowst G. Rifampicin, a general review. Arzneimittelforschung. 1971 Dec;21(12):1907-77. PMID: 4945790.

[18] Weinberg MA, Bral M. Tetracycline and its analogues: a therapeutic paradigm in periodontal diseases. Crit Rev Oral Biol Med. 1998;9(3):322-32. doi: 10.1177/10454411980090030501. PMID: 9715369.

[19] Petersen U. Die Evolution der Chinolone: Von der Nalidixinsäure zu den Chinolonen der dritten Generation [The evolution of quinolone: from nalidixic acid to the quinolones of the third generation]. Pharm Unserer Zeit. 2001;30(5):376-81. German.