

What causes ampicillin resistance of *E.coli* strain?

Abstract

Antibiotic resistance is a big problem in modern science and medicine. Understanding its mechanism is necessary to solve this problem. In this work, we tried to find out which mutations in the genome of *E.coli* led to the emergence of resistance to ampicillin.

Introduction

Antibiotics are truly miracles. They made a real breakthrough in medicine, saving thousands of lives every day. Humanity has known antibiotics for less than 100 years, but we cannot imagine life without them. We are used to the fact that people do not die from colds or even more serious infections, operations have become commonplace, women can give birth to children without fear of complications from infection (stop me when you want). It's all thanks to antibiotics.

But we understand that winning the battle does not mean winning the war. Over the years of antibiotic oppression, the bacteria have become angry, have grown teeth and are ready for a new battle. Resistance to antibiotics is a huge problem in modern medicine. There are plenty of ways for bacteria to avoid death from antibiotics. The studying of these mechanisms is the first step to the development of new, better drugs, or approaches to treat a patient.

In this project, we worked with real sequencing data from a strain of *E. coli* resistant to the antibiotic ampicillin. We tried to analyze that data to locate the mutations responsible for giving *E. coli* its antibiotic resistance property.

Methods

For our study we use raw Illumina sequencing reads from shotgun sequencing of an *E. coli* strain that is resistant to the antibiotic ampicillin [1]. For reference we used the genome of *E.coli* strain K-12 substrain MG1655 [2].

The quality of raw reads was assessed using the program FastQC (version 0.10.1) [3]. To remove low quality reads was used program Trimmomatic (version 0.39) [4] with options CROP:90 SLIDINGWINDOW:10:20 MINLEN:20 (cut bases off the end of a read if quality below 20, the quality of starts of reads was fine).

To align reads to the reference sequences was used aligner BWA-MEM (version 0.7.5a-r405) [5]. To compress, index and simply statistic of SAM-file with alignments was used Samtools *view*, *index*, *flagstat*, *sort* (version 0.1.19-44428cd) [6].

Samtools *mpileup* was used to search for mutations. To annotate SNP in our data was used program VarScan (version 2.3.9.), command *mpileup2snp* [7]

To visualize results was used IGV Browser (version 2.8.11) [8].

Results

First of all, we check the quality of read using FastQC. We observed a degradation in the quality of reading by the end of the read. So, next we cut bases off the end of a read if quality below 20 (using Trimmomatic). After that we run FastQC again. On the Fig.1 you can see per base sequence quality before and after trimming.

Table 1 shows the quantity of reads before and after trimming and the quantity of aligned reads.

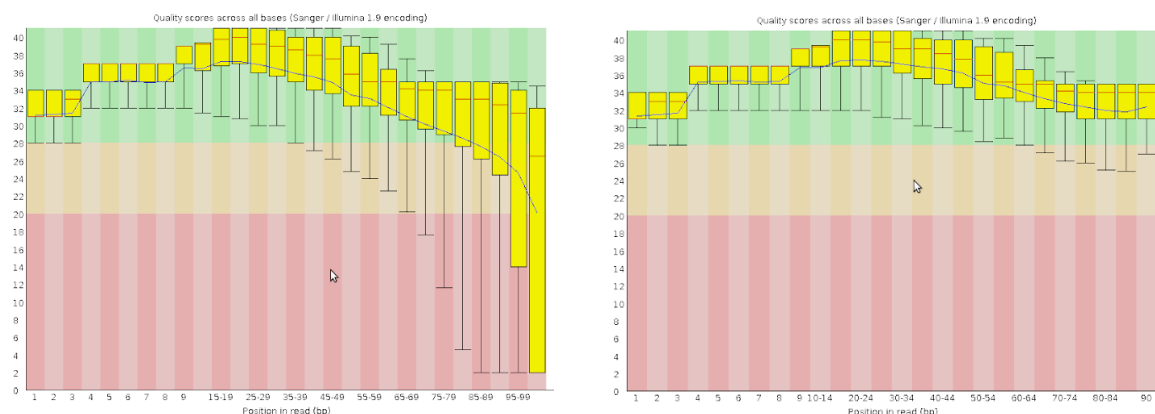


Fig. 1. Per base sequence quality of reads before (left) and after (right) trimming.

Before trimming	455876 x 2
After trimming	445888 x 2
Aligned	892011

Table 1. The quantity of reads before and after trimming and the quantity of aligned reads

Next step reads were mapped to the reference and scanned to identify positions that likely contained mutations. We found seven SNPs (not using options **--min-var-frequency**). Of these, one has 28.57% of non-reference bases, and this position is not in the coding region of the genome. Six SNPs have 100% of non-reference bases. One of these nucleotide replacements doesn't lead to the replacement of amino acid. So, only five mutations remain interesting for us.

Discussion

SNP in gene *ftsI* causes mutation A544G in protein called DD-transpeptidase, also known as penicillin-binding protein (PBD). It is involved in bacterial cell wall biosynthesis, namely, the transpeptidation that crosslinks the peptide side chains of peptidoglycan strands. β -lactam antibiotic, being a structural analogue of the terminal dipeptide of peptidoglycan D-alanyl-D-alanine, binds covalently to the active site of this protein. This binding inhibits the transpeptidation reaction and stops peptidoglycan synthesis, eventually collapsing the bacterial cell [10]. We can suggest that replacement amino acid substitution can change the binding site of PBD, the antibiotic cannot bind the enzyme, resulting in resistance.

Mutation in gene *envZ* also can cause antibiotic resistance. Histidine kinase EnvZ is a member of the two-component regulatory system EnvZ/OmpR involved in osmoregulation, particularly if regulates transcription of genes *ompF* and *ompC*. In article [11] were shown, that a decreased level of porin OmpF can lead to decreased susceptibility to a number of hydrophilic antibiotics including ampicillin. We can

assume that the mutation M241G led to changes in the regulation of porin OmpF expression, that caused resistance to ampicillin in our strain.

One way to develop resistance is to pump the antibiotic out of the cell. There are membrane proteins, called efflux pumps, that are supposed to pump out different toxic substances. One of these is coding by genes *acrA* and *acrB*. It has been repeatedly shown that the AcrAB kit plays an important role in antibiotic resistance [12, 13]. We can suggest that mutations Q569L make this protein more effective in combating antibiotics.

Nº	Position	Gene	Strain	Ref. Codon	Ref. amino acid	Alt. Codon	Alt. amino acid	Mutations
1	93043	ftsI	+	GCC	Alanine (A)	GGC	Glycine (G)	A544G
2	482698	acrB	-	CAG	Glutamine (Q)	CTG	Leucine (L)	Q569L
3	852762	rybA	-	TTT	Phenylalanine (F)	TCT	Serine (S)	?
4	1905761	mntP	+	GGT	Glycine (G)	GAT	Aspartic acid (D)	G25D
5	3535147	envZ	-	GTA	Valine (V)	GGA	Glycine (G)	V241G
6	4296060	----						
7	4390754	<i>rsgA</i>	-	GCC	Alanine (A)	GCA	Alanine (A)	

Amino acid biochemical properties:

	Non-polar	Polar	Acidic
--	-----------	-------	--------

Table 2. Found SNPs: the position of nucleotide in the genome, mutated gene, direction of the DNA chain, codon and amino acid in the reference sequence, codon and amino acid in the read sequence, mutations. Colors show amino-acid biochemical properties [9]. Mutations Nº 1 - 5 lead to amino acid replacement. Mutation Nº 6 is not in the coding region of the genome, and mutations Nº 7 don't change the amino acid sequence.

Citations

1. Illumina sequencing data: <http://public.dobzhanskycenter.ru/mrayko/>
2. Reference sequence and gene annotation: ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Escherichia_coli/all_assembly_versions/GCA_000005845.2_ASM584v2/
3. FastQC: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
4. Trimmomatic: <http://www.usadellab.org/cms/?page=trimmomatic>
5. BWA: <http://bio-bwa.sourceforge.net/bwa.shtml>

6. Samtools: <https://www.htslib.org/>
7. VarScan: <http://dkoboldt.github.io/varscan/>
8. IGV Browser: <https://igv.org/>
9. https://en.wikipedia.org/wiki/DNA_codon_table
10. Karen Bush, Patricia A. Bradford. (2016). β -Lactams and β -Lactamase Inhibitors: An Overview. *Cold Spring Harb Perspect Med.* **6**, a025247
11. Duval V., Nicoloff H., Levy S.B. (2009). Combined inactivation of lon and ycgE decreases multidrug susceptibility by reducing the amount of OmpF porin in *Escherichia coli*. *Antimicrob. Agents Chemother.* 53:4944-4948
12. Sun J, Deng Z, Yan A (October 2014). "Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations". *Biochemical and Biophysical Research Communications.* 453 (2): 254-67
13. Okusu H, Ma D, Nikaido H (January 1996). "AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants". *Journal of Bacteriology.* 178 (1): 306-8. doi:10.1128/jb.178.1.306-308.1996. PMC 177656. PMID 8550435
14. UniProt: <https://www.uniprot.org/uniprot>
15. EcoCyc: <https://ecocyc.org/>