Project #1. “What causes antibiotic resistance?”

Alignment to reference, variant calling.

Anna Chechenina, Dana Melnikova
October 2022, Bioinformatics Institute

Introduction

Antibiotic resistance plays a significant role in the treatment of bacterial infections and can cause a large number of problems in clinics up to the need to replace or chemically modify the antibiotic. Antibiotic-resistant strains arise in the population due to mutational processes and later could gain the advantage over other strains due to evolutionary pressure. Depending on the type of antibiotic, there might be different mutations under selection. However, there are some general mechanisms of resistance, those are the most popular ways to bacteria to escape from the damaging effects of antibiotics. First of all, the bacteria can block the entry of the antibiotic into the cell or attachment to the receptor molecules if that is the case. This could be done by accelerating of synthesis of the shielding molecules like polysaccharides or peptidoglycans. Another effective way to fight antibiotics is to bind them and pump them out of the cell using protein pumps or transporters. Besides, there are numerous examples of direct antibiotic inhibition by, for example, irreversible binding or destroying the antibiotic structure.

There are many different types of antibiotics that could be classified by target bacterial group, mechanism of action or chemical structure. The main mechanisms include destroying the cell wall and cell membrane or interrupting vital processes like replication, transcription, and translation by targeting the key enzymes. Antibiotics could work either as bacteriostatic agents restricting cell growth and reproduction or as bactericidal agents causing cell death.

The main groups are beta-lactams, sulfonamides, aminoglycosides, tetracyclines, chloramphenicols, macrolides, quinoles, and others. Beta-lactams were the first group to discover at the beginning of the 20th century, and they include such broad groups of antibiotics as penicillins and cephalosporins. They work by interfering with the synthesis of peptidoglycan and are mostly used against gram-positive bacteria. Sulfonamides have a bacteriostatic effect and work by inhibiting the bacterial synthesis of the B vitamin folate, thus preventing the growth and reproduction of the bacteria. Antibiotics from the group of aminoglycosides, for example, streptomycin and kanamycin, are only effective against certain gram-negative bacteria and are widely used in clinics and research. They inhibit protein synthesis by binding to the 30S ribosomal subunit and impairing the proofreading that can result in the production of faulty proteins. Tetracyclines act against both gram-positive and gram-negative bacteria. Their mode of action is similar to the aminoglycosides – they also bind to the 30S ribosomal subunit but in another site, preventing the binding of the tRNA and inhibiting protein synthesis. Chloramphenicols also act by inhibiting protein synthesis but bind to the 50S ribosome subunit preventing the formation of the peptide bond.

In our research, we focused on the resistance of the *Escherichia coli* strain K-12 substrain MG1655 to ampicillin. Ampicillin is an antibiotic of the group of beta-lactams. Ampicillin binds to and

inactivates penicillin-binding proteins (PBPs) located on the inner membrane of the bacterial cell wall. The inactivation of PBPs interferes with the cross-linkage of peptidoglycan chains necessary for bacterial cell wall strength. This interrupts bacterial cell wall synthesis, results in its weakening, and causes cell lysis. The most common mechanism of beta-lactam resistance is based on the beta-lactamase enzyme activity (Sanders and Eugene Sanders 1992), which performs hydrolysis to break the beta-lactam ring of the antibiotic. That is why in clinical practice this type of antibiotic is usually used in complex with beta-lactamase inhibitors (for instance, ampicillin-sulbactam) (Bajaj, Singh, and Virdi 2016). However, in *E. coli* the majority of genes of known beta-lactamases are located on the plasmids, for example gene $bla_{\text{TEM-1}}$ (Cantón et al. 2008).

In multiple experimental articles (Boman, Nordström, and Normark 1974; Monner, Jonsson, and Boman 1971; PREHM et al. 1976) it was shown that ampicillin resistance in *E. coli K12* usually goes along with the loss of some carbohydrate parts of lipopolysaccharides (LPS). The changes include a loss of significant parts of the rhamnose, galactose, and glucose from the LPS (Monner et al. 1971), thus suggesting potential mutations in the glycosyltransferases. Since the complete oligosaccharides are phosphorylated, mutations in the specific kinases could be also a case (PREHM et al. 1976). The genes described to be connected with ampicillin resistance are *lpsABC* (responsible for glucose, galactose, and rhamnose synthesis), *ampC*, *ampA*, *envA*, *pgm*, *tolD*, *tolE*, *galU* (PREHM et al. 1976).

In the more recent study the *E.coli* strains 15743-32 and 15743-256 were explored to find some ampicillin-resistance mutations (Li et al. 2019). Multiple SNPs were found in the following genes: *ftsI* (penicillin-binding protein 3), *acrB* (multi-drug efflux pump gene), *OmpD* (porin outer membrane protein), *marR* (multi-drug resistance protein involved in transcriptional regulation), *frdD* (fumarate reductase subunit D), *vrgrG* (valine glycine repeat), *envZ* (osmotic pressure sensor histidine kinase). FRD enzyme was shown to catalyze the conversion between fumarate reductase and succinate dehydrogenase (Spencer and Guest 1973). This result highly intercept with another study where such mutations were identified for the focal *E. coli* strains (Baumgartner et al. 2020). According to the study there were some SNPs and the deletion mutations found in genes membrane-related genes (*ompR, ftsI, opgB*), stress response gene (*relA*), and transcription (*rpoC, rpoD*). Apart from this, they found mutations in other genes not described before too have relation to beta-lactam resistance. Those include genes *insN, gtrS, rpoD, opgB,* and *yaiO* (Baumgartner et al. 2020).

In our study, we focused on the K-12 *Escherichia coli strain.* In order to find and identify genes that could be important for the ampicillin resistance in the *Escherichia coli K-12* we performed an analysis of raw Illumina data provided by experimental work [1]. Reads were filtered depending on length and quality and then mapped to the reference genome. Then the most frequent mutations were detected and explored using annotation of the corresponding genome.

Methods

Genome and raw sequencing data

In our research, we used the full genome of the Escherichia coli strain K-12 substrain MG1655 (NCBI ID:167) as the reference sequence. We derived the raw Illumina sequencing data from the publication (Raiko 2019).

Raw data preparation

Raw reads were first analyzed by the "FastQC" program (Andrews 2010). Then we performed the trimming of the raw reads to filter out ones with the low quality and to cut off adaptors that might left on the reads ends. For this purpose we used "trimmomatic" package (Bolger, Lohse, and Usadel 2014) version 0.36+dfsg-3. Since our data contains forward and reverse reads we used the "paired ends" option, which is designed to eliminate a pair-read of the already mapped read from the further mapping. We performed the run with the quality threshold for the end and the start of the read equal to 20, sliding window of size 10 with average quality 20, and a length cut-off of 20 base-pairs. Trimming step significantly improved the quality of our reads (Supplementary, Fig. 1).

Aligning sequences to the reference

To index the reference genome and to map the filtered reads we used BWA-MEM tool followed by samtool to sort and index the output BAM file (Li et al. 2009). According to the samtools statistics 891649 reads (99.87% of the total number of reads) were successfully aligned to the reference genome.

The BAM file was then sorted and indexed in order to access positions faster in the next steps. Then we computed mpileup file with statistics on mutations on all positions.

Variant calling

For mutations analysis we used VarScan tool (Koboldt et al. 2012) version VarScan.v2.4.0. with minimal variant frequency parameter of 10% or 50%. The found mutations did not depend on a frequency level, because the only mutations present were one with the frequency equal to 100%. Then we annotated found mutations with the SnpEff tool (Cingolani et al. 2012) using default options.

For more details on tools and parameters consult the laboratory journal (you can find it in at in the corresponding repository https://github.com/checheanya/BI_git/tree/main/HW1).

Results and discussion

Raw data contained 455876 reads in total, to eliminate low quality reads we performed trimming which resulted in the 446259 final reads that we used for further analysis (for the quality comparison see Supplementary Fig. 1). Mean quality after trimming increased from minimum =2 to minimum =24.

Regardless of minimal variant frequency threshold, we yielded only five mutations, all with frequency 100%. Mutations located in the genes *ftsI, acrB, glnH, envZ, rsgA* (Table 1). In genes *ftsI, acrB* and *envZ* mutations turned out to be missense and resulted in the change of amino acids.

Table 1. Genes carrying SNP in the ampicillin resistant *E. coli*.

	Gene	*Position*	*Before*	*After*	*Freq*	*Effect*	*Impact*	*GeneID*
0	*ftsI*	93043	C	G	100	missense_variant	MODERATE	b0084
1	*acrB*	482698	T	A	100	missense_variant	MODERATE	b0462
2	*glnH*	852762	A	G	100	upstream_gene_variant	MODIFIER	b0811
3	*envZ*	3535147	A	C	100	missense_variant	MODERATE	b3404
4	*rsgA*	4390754	G	T	100	synonymous_variant	LOW	b4161

FtsI is peptidoglycan DD-transpeptidase, which operates in complex with FtsW protein controlling FtsZ treadmilling (Yang et al. 2021) and peptidoglycan synthesis. Especially FtsWI complex is essential for the correct performance of all peptidoglycan synthases within the divisome during cell division (Marmont and Bernhardt 2020). Both types of peptidoglycan synthases the *E. coli* has, are thought to participate in septal peptidoglycan biogenesis forming a complex with the FtsWI, SEDS-bPBP synthase of the division machinery (Adler et al. 2016). It was shown that the overproduction of FtsI along with FtsQ, FtsL, FtsB, and FtsW promotes the spontaneous formation of an activated PG synthase complex (Marmont and Bernhardt 2020). Moreover, the inactivation of FtsI by binding of beta-lactam antibiotics like ampicillin or mutagenesis, induces the SOS response via the DpiBA two-component signal transduction system. The resulting cell division arrest may enable survival of the cells despite exposure to otherwise lethal antibiotic (Miller et al. 2004). Thus taking these functional details into account we can suggest that the mutation in *ftsI* gene could induce the resistant state by changing the cell peptidoglycan state of by inducing the SOS response in the presence of antibiotic.

AcrB is a multidrug efflux pump RND permease, an inner membrane-associated part of the AcrAB-TolC multidrug efflux complex (Adler et al. 2016; Li et al. 2019). AcrB mediated proton transport proceeds by an alternating access mechanism – proton binding sites within the transmembrane domain become alternatively accessible to the periplasm or cytoplasm. AcrB adopts a novel antiport mechanism which involves the coupling of two remote alternating access mechanisms - one for drugs and one for protons - within each AcrB protomer (Eicher et al. 2014). Substrate molecules are captured by AcrB from the periplasm or the periplasm/inner membrane interface. So the suggested mechanism of AcrB involvement into b-lactam resistance is that the complex containing this protein could bind and pump out the molecules of antibiotic.

EnvZ is a dimeric sensor protein, kinase, in two-component system EnvZ-OmpR regulating production of outer membrane proteins OmpCF. OmpC and OmpF are trimeric outer membrane porins that are regulated differently depending on the environment conditions (Adler et al. 2016; Garrett, Taylor, and Silhavy 1983). Moreover, it was shown that the mutations in this gene increase the expression levels of some other genes, for example gene *ompR* (Adler et al. 2016). Based on this knowledge we can assume that EnvZ changes the membrane permability of the cell that could lead to changing of the cell membrane state thus activating some membrane complexes. Other than that OmpCF could be directly involved in the antibiotic removal. Mutations in this gene can also induce expression of some other genes that could be possibly involved in the antibiotics resistance (Li et al. 2019). In another study is was shown that AMP-resistant bacteria underwent protective surface modifications that led to a reduction in binding/entry upon peptide exposure (Georgieva et al. 2022).

GlnH codes the periplasmic binding protein of an L-glutamine ABC transport system. It is required for growth on L-glutamine as sole carbon source (Nohno, Saito, and Hong 1986). However, there is not much information on this protein in the antibiotic-resistance context as well as on *rsgA* gene. *rsgA* codes the ribosome small subunit-dependent GTPase A. It plays a role in the late stages of maturation of the 30S subunit of the ribosome (Campbell and Brown 2008), so we can assume that the mutation in such an essentian gene could lead to the wrong ribosome assembly or wrong operation of the small subunit.

Overall, we can say that all mutations we found were previously described as related to the ampicillin resistance (Li et al. 2019) and seems to have serious impact on the *E. coli* phenotype.

Literature

Adler, Marlen, Mehreen Anjum, Dan I. Andersson, and Linus Sandegren. 2016. "Combinations of Mutations in EnvZ, FtsI, MrdA, AcrB and AcrR Can Cause High-Level Carbapenem Resistance in Escherichia Coli." *Journal of Antimicrobial Chemotherapy* 71(5):1188–98.

Andrews, S. 2010. "FASTQC. A Quality Control Tool for High Throughput Sequence Data."

Bajaj, Priyanka, Nambram S. Singh, and Jugsharan S. Virdi. 2016. "Escherichia Coli β-Lactamases: What Really Matters." *Frontiers in Microbiology* 7(MAR):1–14.

Baumgartner, Michael, Florian Bayer, Katia R. Pfrunder-Cardozo, Angus Buckling, and Alex R. Hall. 2020. "Resident Microbial Communities Inhibit Growth and Antibiotic-Resistance Evolution of Escherichia Coli in Human Gut Microbiome Samples." *PLoS Biology* 18(4):1–30.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30(15):2114–20.

Boman, H. G., K. Nordström, and S. Normark. 1974. "Penicillin Resistance in Escherichia Coli K12: Synergism Between Penicillinases and a Barrier in the Outer Part of the Envelope." *Annals of the New York Academy of Sciences* 235(1):569–86.

Campbell, Tracey L., and Eric D. Brown. 2008. "Genetic Interaction Screens with Ordered Overexpression and Deletion Clone Sets Implicate the Escherichia Coli GTPase YjeQ in Late Ribosome Biogenesis." *Journal of Bacteriology* 190(7):2537–45.

Cantón, R., M. I. Morosini, O. Martin, S. De La Maza, and E. Gomez G. De La Pedrosa. 2008. "IRT and CMT β-Lactamases and Inhibitor Resistance." *Clinical Microbiology and Infection* 14(SUPPL. 1):53–62.

Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. "A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff." *Fly* 6(2):80–92.

Eicher, Thomas, Markus A. Seeger, Claudio Anselmi, Wenchang Zhou, Lorenz Brandstätter, François Verrey, Kay Diederichs, José D. Faraldo-Gómez, and Klaas M. Pos. 2014. "Coupling of Remote Alternating-Access Transport Mechanisms for Protons and Substrates in the Multidrug Efflux Pump AcrB." *ELife* 3:1–26.

Garrett, S., R. K. Taylor, and T. J. Silhavy. 1983. "Isolation and Characterization of Chain-Terminating Nonsense Mutations in a Porin Regulator Gene, EnvZ." *Journal of Bacteriology* 156(1):62–69.

Georgieva, Maria, Tytti Heinonen, Simone Hargraves, Trestan Pillonel, Christian Widmann, and Nicolas Jacquier. 2022. "The EnvZ/OmpR Two-Component System Regulates the Antimicrobial Activity of TAT-RasGAP$_{317-326}$ and the Collateral Sensitivity to Other Antibacterial Agents." *Microbiology Spectrum* 10(3):e02009-21.

Koboldt, Daniel C., Qunyuan Zhang, David E. Larson, Dong Shen, Michael D. McLellan, Ling Lin, Christopher A. Miller, Elaine R. Mardis, Li Ding, and Richard K. Wilson. 2012. "VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing." *Genome Research* 22(3):568–76.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25(16):2078–79.

Li, Mengchen, Qiaoli Liu, Yanli Teng, Liuyang Ou, Yuanlin Xi, Shuaiyin Chen, and Guangcai Duan. 2019. "The Resistance Mechanism of Escherichia Coli Induced by Ampicillin in Laboratory."

Infection and Drug Resistance 12:2853–63.
Marmont, Lindsey S., and Thomas G. Bernhardt. 2020. "A Conserved Subcomplex within the Bacterial Cytokinetic Ring Activates Cell Wall Synthesis by the FtsW-FtsI Synthase." *Proceedings of the National Academy of Sciences of the United States of America* 117(38):23879–85.
Miller, Christine, Line Elnif Thomsen, Carina Gaggero, Ronen Mosseri, Hanne Ingmer, and Stanley N. Cohen. 2004. "SOS Response Induction by β-Lactams and Bacterial Defense against Antibiotic Lethality." *Science* 305(5690):1629–31.
Monner, D. A., S. Jonsson, and H. G. Boman. 1971. "Ampicillin-Resistant Mutants of Escherichia Coli K-12 with Lipopolysaccharide Alterations Affecting Mating Ability and Susceptibility to Sex-Specific Bacteriophages." *Journal of Bacteriology* 107(2):420–32.
Nohno, T., T. Saito, and J. S. Hong. 1986. "Cloning and Complete Nucleotide Sequence of the Escherichia Coli Glutamine Permease Operon (GlnHPQ)." *Molecular & General Genetics : MGG* 205(2):260–69.
PREHM, Peter, Stephan STIRM, Barbara JANN, Klaus JANN, and Hans G. BOMAN. 1976. "Cell-Wall Lipopolysaccharides of Ampicillin-Resistant Mutants of Escherichia Coli K-12." *European Journal of Biochemistry* 66(2):369–77.
Raiko, Mike. 2019. "Amp_res_Ecoli_data." *Figshare. Dataset.* Retrieved (https://figshare.com/articles/dataset/amp_res_2_fastq_zip/10006541/3).
Sanders, Christine C., and W. Eugene Sanders. 1992. "SS-Lactam Resistance in Gram-Negative Bacteria: Global Trends and Clinical Impact." *Clinical Infectious Diseases* 15(5):824–39.
Spencer, M. E., and J. R. Guest. 1973. "Isolation and Properties of Fumarate Reductase Mutants of Escherichia Coli." *Journal of Bacteriology* 114(2):563–70.
Yang, Xinxing, Ryan McQuillen, Zhixin Lyu, Polly Phillips-Mason, Ana De La Cruz, Joshua W. McCausland, Hai Liang, Kristen E. DeMeester, Cintia C. Santiago, Catherine L. Grimes, Piet de Boer, and Jie Xiao. 2021. "A Two-Track Model for the Spatiotemporal Coordination of Bacterial Septal Cell Wall Synthesis Revealed by Single-Molecule Imaging of FtsW." *Nature Microbiology* 6(5):584–93.

Supplementary

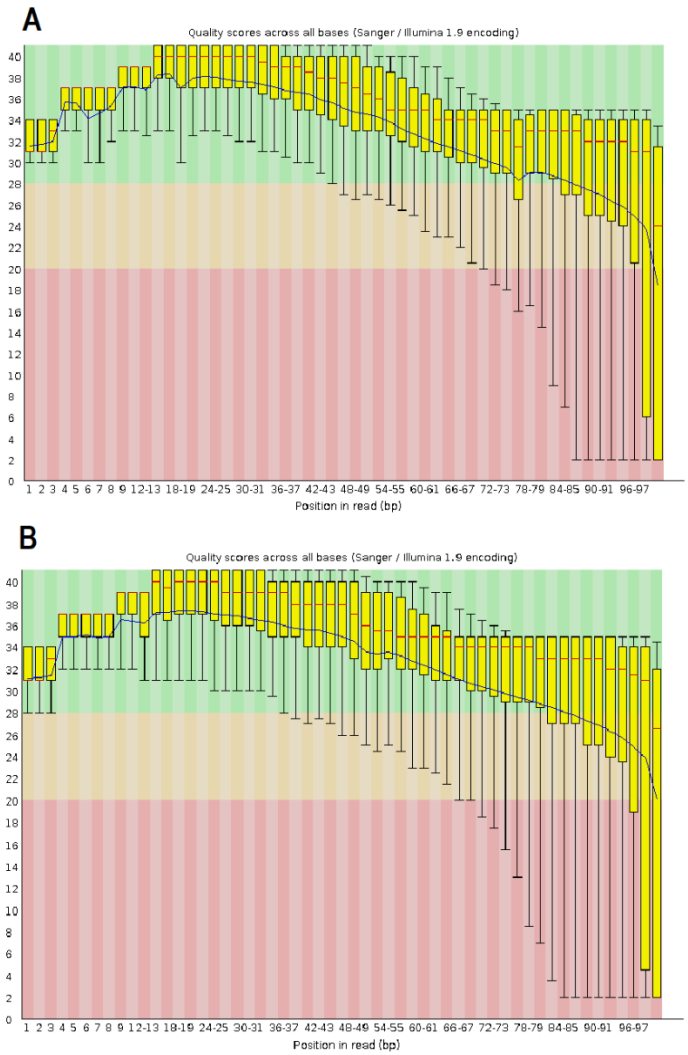

Figure 1a. Reads trimming using trimmomatic tool significantly improves the quality. A – forward reads quality before trimming; B – reverse reads quality before trimming.

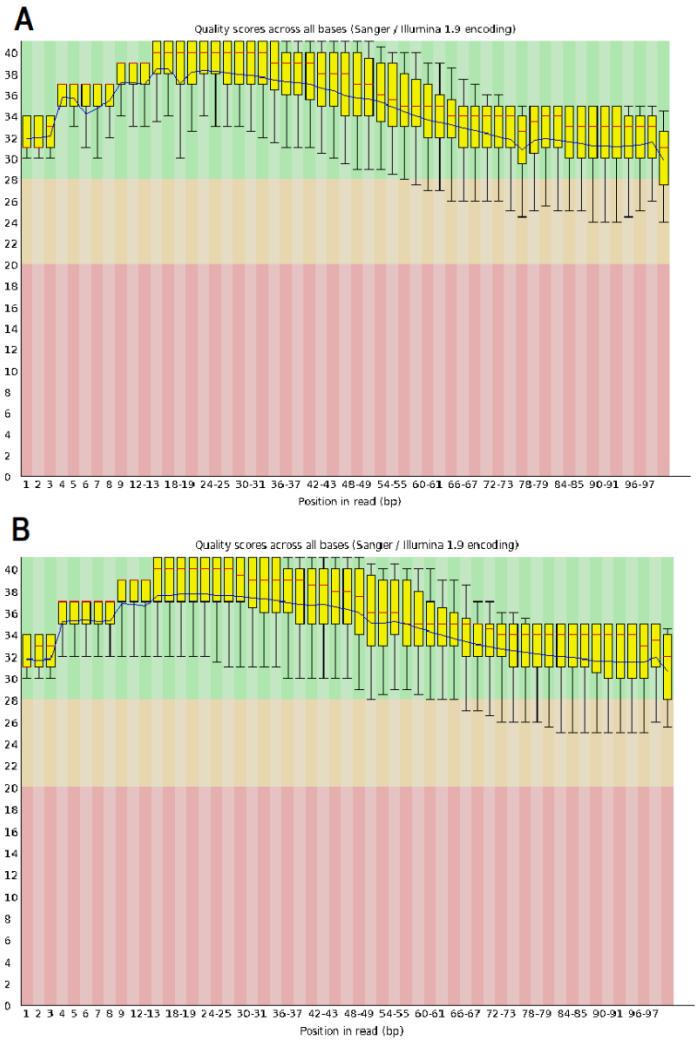

Figure 1b. Reads trimming using trimmomatic tool significantly improves the quality. A – forward reads quality after trimming; B – reverse reads quality after trimming.