Polymorphisms in E. coli genes causing antibiotic resistance

Chernyavskaya K., Kupaeva D.

Abstract

Antibiotic resistance is becoming one of the most important problems of modern medicine. Exploring the process and metabolic pathways in which this resistance occurs can help find ways of avoiding it. We investigated the sequence of the ampicillin resistant E. coli strain in order to search for SNPs affecting the development of antibiotic resistance. In this study we identified three most significant polymorphisms in ftsI, acrB and envZ genes of E. coli.

Introduction

Increasing microorganisms antibiotic resistance is the predictable outcome of their uncontrolled use by mankind since antibiotics were invented. The possibility of treating infectious diseases will decrease with the growth of antibiotic resistance and we risk entering the post-antibiotic era in the near future. Such a transition will be characterized by a huge number of infectious diseases for which there will be no treatment. “If we want to change the future state, and have long-term availability of effective antimicrobial therapy for infections, we need to think disruptively and challenge long-standing and sometimes cherished assumptions.” [1]

Ampicillin is one of the most common semi-synthetic antibiotics acting by the mechanism of irreversible blocking of transpeptidase involved in the synthesis of peptidoglycan of the cell wall which causes bacteriolysis. Ampicillin is effective against a large number of gram negative and gram positive bacteria, which is why it’s widely used. However, it is ineffective against a number of microorganisms. Some of them form penicillinase (an enzyme that breaks down beta-lactam antibiotics), some show spontaneous insensitivity to the antibiotic. More important for us now is antibiotic resistance, which is not associated with the synthesis of penicillinase by a microorganism, since such resistance can arise due to mutations in the genome of one bacterium and be transmitted to the entire colony. As a result, the strain stops responding to the antibiotic and such infections therapy becomes ineffective.

There are a few general mechanisms of building antibiotic resistance such as decreasing uptake of antibiotic, modification or degradation of enzymes, altered target (PBP for example), efflux pumps and antibiotic inactivation.[2] The occurrence of each such mechanism corresponds to a change in the microbial genome. Therefore, we examined the genome of the ampicillin resistant e coli strain in order to search for different variants. Studying the causes of antibiotic resistance can help to figure out how to deal with or avoid it.

Methods

We used raw Illumina sequencing reads (paired end run) from shotgun sequencing of a E. coli strain resistant to the antibiotic ampicillin and reference sequence (FASTA and annotation) of the parental (not resistant to ampicillin) E.coli strain K-12 substrain MG1655[1]. Firstly, the quality of the reads was analyzed using fastqc [3]. Fastqc reported about next problems: per base sequence quality for both forward and reverse strands, and per tile sequence quality for forward strand reads. To improve the overall quality of sequencing reads we used Trimmomatic [4] with the following parameters: LEADING:20 TRAILING:20 MINLEN:20 SLIDINGWINDOW:10:20.

After that we aligned sequences to reference. We used BWA-MEM [5] for indexing genome and aligning reads (with default options of bwa index and bwa-mem commands) and got some statistics using compressed .sam file (.bam) generated by samtools (default options as well) [6]. We conclude that 99.88% of all reads were mapped. This is quite a good result for further analysis. So we sorted the .bam file for faster variants searching using samtools sort and indexed it with samtools index. To distinguish actual mutations from the sequencing errors we generate a mpileup file using samtools mpileup -f (as thought our reference in FASTA format). To call actual variants we created a .vcf file using VarScan [7]. We have established the minimum % of non-reference bases at a position required to call it a single nucleotide polymorphism in the sample as 0.8 (parameter --min-var-freq 0.8 for mpileup2snp option) and got 5 SNPs in the generated .vcf file. For getting .vcf file with indels we used a similar command (parameter --min-var-freq 0.8 for mpileup2indels option).

Finally we visualised our results with IGVbrowser [8,9] to find out current SNPs and aminacid's replacements.

For more information on the settings for using the software for this study, see the Supplemental Resources.

Results

Before mapping reads to reference we cut low quality bases. After trimming, 97.77% of the reads survived, but some of them were significantly shortened. Basic statistics of raw reads shown in Table 1, distribution of read quality before and after trimming shown in Supplementary figure 1.

	forward, before trimming	reverse, before trimming	forward, after trimming	reverse, after trimming
Total Sequences	455876	455876	445689	445689
Sequence length	101	101	20-101	20-101
%GC	50	50	50	50

Table 1. Basic statistic of raw reads before and after using Trimmomatic

Reads were mapped to the reference and scanned to identify positions that likely contained mutations. According flagstat samtools statistics, we successfully mapped 99.88% of all raw reads. With help of IGV genome browser in the base of samtools mpileup and VarScan result, we visualized 5 SNP and 3 indels, all of which are presented in table 2.

	Position	**Reference nucleotide**	**Alternate nucleotide**	**Reference amino acid**	**Alternate amino acid**	**Gene**
1.	**93043**	**C**	**G**	**A**	**G**	**ftsI**
2.	**482698**	**T**	**A**	**Q**	**L**	**arcB**
3.	852762	A	G	-	-	rybA
4.	**3535147**	**A**	**C**	**V**	**G**	**envZ**
5.	4390754	G	T	A	A	rsgA
6.	2173360-2173362	CC	-		Frameshift mutation	gatC, pseudogene

7.	3560455	-	G		Frameshift mutation	glpR, pseudogene
8.	4296380	-	CG			-

Table 2. SNPs and indels in E. coli strain resistant to the antibiotic ampicillin.

Of the 8 mutations found, only 3 lead to amino acid substitutions in protein-coding genes. They affect genes ftsI, arcB, envZ (lines 1, 2 and 4 of Table 1). The other 5 mutations have no effect on protein products: SNP on line 3 is in ribosomal RNA, line 5 is a synonymous mutation. All indels are in pseudogenes or non-coding DNA.

For all sense SNP in protein-coding genes we created a table 3 with annotation product in reference genome and function according databases.

Gene	Product according annotation	Function
ftsI	peptidoglycan DD-transpeptidase FtsI;	“Essential cell division protein that catalyzes cross-linking of the peptidoglycan cell wall at the division septum” (according https://www.uniprot.org/uniprot/P0AD68)
arcB	multidrug efflux pump RND permease AcrB; Aerobic respiration control sensor protein ArcB	"Member of the two-component regulatory system ArcB/ArcA. Sensor-regulator protein for anaerobic repression of the arc modulon. Activates ArcA via a four-step phosphorelay. ArcB can also dephosphorylate ArcA by a reverse phosphorelay involving His-717 and Asp-576." (according https://www.uniprot.org/uniprot/P0AEC3)
envZ	sensory histidine kinase EnvZ	“Member of the two-component regulatory system EnvZ/OmpR involved in osmoregulation (particularly of genes ompF and ompC) as well as other genes. EnvZ functions as a membrane-associated protein kinase that phosphorylates OmpR in response to environmental signals; at low osmolarity OmpR activates ompF transcription, while at high osmolarity it represses ompF and activates ompC transcription.” (according https://www.uniprot.org/uniprot/P0AEC3)

Table 3. Short summary of genes with sense mutation according to reference and Uniprot databases.

Discussion

In the result of our study, we found 3 SNP, which affect the amino acid sequence of their product. Since this strain of bacteria is resistant to antibiotics, we hypothesize that the mutations we found could affect the bacteria's ability to ignore the drug.

ftsI

The most obvious candidate for a "resistance gene" is ftsI gene, whose product is protein essential during cell division. Transpeptidase is the main protein, which is associated with penicillin and other beta-lactam antibiotics [10]. Actually, the existence of mutation on this gene after is not unusual. We suggest that this mutation reduces affinity drag to transpeptidase, and it has an effect on resistance on bacterias to antibiotics.And now known many cases, when mutation in this gene was associated with resistance to beta-lactam antibiotics. [11] According to Adler’s research [12] meropenem exhibits antimicrobial activity for strains with mutations in the ftsI genes, therefore it can be used as an alternative antibiotic therapy.

acrB

Gene arcB also may be associated with antibiotic resistance. Protein product of this gene included in third parts pomp with AcrA / AcrB / TolC, in charge of the efflux of cytotoxic substances and protons. The Resistance Nodulation Division (RND) ArcB includes 3 monomers, and they have 3 different conformations, which protein takes during work. RND locates in the inner membrane component, and responsibility for substrate-specificity. Mutation in RND associated with multidrug resistance. Possibly mutations in this gene lead to an increase expression by a feedback mechanism with acrR gene (the local repressor of AcrAB) [13] Mutations in acrB "...seem to increase (beta-lactam) carbapenem-binding affinity and improve export of the drug, but at the expense of reduced binding to other antibiotic classes"based on this fact, it can be assumed that antibiotics of a different series (not beta-lactam) may still affect the microbes of this strain. [12]

envZ

The last of our SNP located in envZ. EnvZ is a member of the two-component regulatory system EnvZ/OmpR, and they regulate the expression level of porin proteins ompF and ompC. Both provide transfer of substrates to the cell, but they have different sizes of chanel. OmpF have larger chanell and are associated with a high level of access for nutrients and drags together. OmpC have smaller chanell, and are less available for nutrients and drugs. In the normal condition ratio on this porin is consistent values, but in condition of permanent antibiotic stress may be beneficial to completely suppress the expression of OmpF. Replacing all OmpF channels to OmpC may increase resistance of bacteria to antibiotics. By this reason mutation in the porin regulation system, in envZ, may be helpful for bacteria in the condition of antibiotic stress, but may be less competitive in normal conditions[14] . Also envZ mutations can have effects on resistance independent of reducing ompCF expression, possibly through altered regulation of other genes [13].

In the article about mechanisms of ampicillin resistance authors work with strain, including ftsl, arcB, envZ mutations. They report about resistants to their strain to AMP, piperacillin, cefuroxime, cefazolin, cefoxitin, AMP/sulbactam, amoxicillin/clavulanic acid, piperacillin/tazobactam, and aztreonam. Respectively, they provide a list of drugs, which are effective on treatment of this strain: gentamicin, tobramycin, cefepime, ceftazidime, cefoperazone, tetracycline, peracillin/tazobartan, meropenem, imipenem, aztreonam, sulfamethoxazole, evofloxacin, ciprofloxacin. In the base of this data, we can recommend using this antibiotic against strain from our report.[15]

In addition, according to Madler Adler's research, increasing the antibiotic resistance of the bacterial strain may be caused by a combination of three genes (ftsI, acrB and envZ). When mutations were combined, resistance to

Bacteriophages can be used as an alternative therapy, since they affect only one type of even antibiotic-resistant bacterias. Now exists reports about successful therapy of multiresistance bacterial infection by combination of antibiotic and phags[16]. However, there are many problems associated with their use, the main of them is the difficulty of their synthesis and poor knowledge of side effects.

Citations

1. Spellberg, B. & Gilbert, D. N. The future of antibiotics and resistance: A tribute to a career of leadership by John Bartlett. *Clin. Infect. Dis.* **59**, S71–S75 (2014).

2. Kapoor, G., Saigal, S. & Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of Anaesthesiology Clinical Pharmacology* **33**, 300–305 (2017).

3. Andrews, S. FastQC A Quality Control Tool for High Throughput Sequence Data. *Babraham Bioinformatics* (2010). Available at: www.bioinformatics.babraham.ac.uk/projects/fastqc/.

4. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

5. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

6. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

7. Koboldt, D. C. *et al.* VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283–2285 (2009).

8. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

9. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24–26 (2011).

10. Houba-Hérin, N., Hara, H., Inouye, M. & Hirota, Y. Binding of penicillin to thiol-penicillin-binding protein 3 of Escherichia coli: identification of its active site. *MGG Mol. Gen. Genet.* **201**, 499–504 (1985).

11. Deghmane, A. E. *et al.* High diversity of invasive Haemophilus influenzae isolates in France and the emergence of resistance to third generation cephalosporins by alteration of ftsI gene. *J. Infect.* **79**, 7–14 (2019).

12. Adler, M., Anjum, M., Andersson, D. I. & Sandegren, L. Combinations of mutations in envZ, ftsI, mrdA, acrB and acrR can cause high-level carbapenem resistance in Escherichia coli. *J. Antimicrob. Chemother.* **71**, 1188–1198 (2016).

13. Seeger, M. *et al.* The AcrB Efflux Pump: Conformational Cycling and Peristalsis Lead to Multidrug Resistance. *Curr. Drug Targets* **9**, 729–749 (2008).

14. Ferenci, T. & Phan, K. How porin heterogeneity and trade-offs affect the antibiotic susceptibility of gram-negative bacteria. *Genes (Basel).* **6**, 1113–1124 (2015).

15. Li, M. *et al.* The resistance mechanism of escherichia coli induced by ampicillin in laboratory. *Infect. Drug Resist.* **12**, 2853–2863 (2019).

16. Letrado, P. *et al.* Bactericidal synergism between antibiotics and phage endolysin Cpl-711 to kill multidrug-resistant pneumococcus. *Future Microbiol.* **13**, 1215–1223 (2018).

Supplemental resources

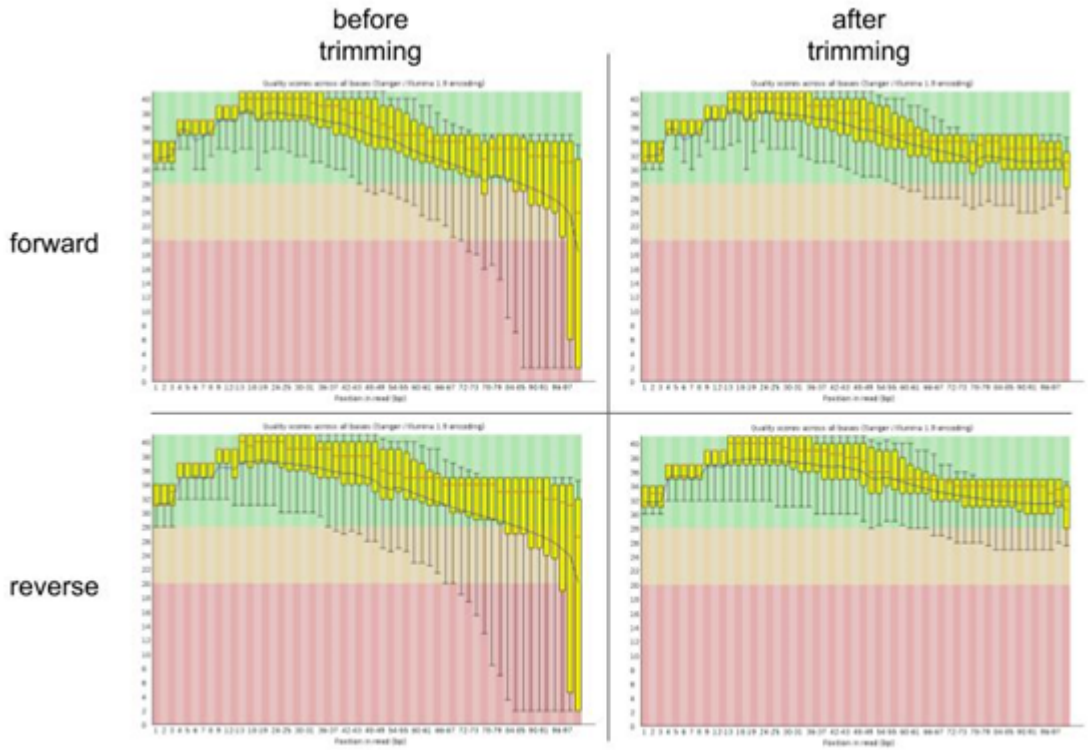

Figure 1. Distribution of bases quality before and after trimming.

Materials 2. Lab Notebook
https://docs.google.com/document/d/1mQLg14bLx7L3lyllnDo-m3UrrLXxCKQehe66pheavdI/edit