What causes antibiotic resistance?

Yulia Solomennikova, Yulia Barablina , Darya Panchenko

September 22, 2022

Abstract

Antibiotic-resistant bacterial infections pose a global threat. To treat infections in both the clinic and the community, accurate identification of antimicrobial resistance is essential for treatment decisions. Therefore, in this study, we performed SNP variant calling on real sequencing data from an ampicillin-resistant E. coli strain to predict antibiotic resistance in order to better understand antimicrobial susceptibility profiles.

1 Introduction

Antibiotic resistance of microorganisms is now widespread and poses a serious clinical threat. For example, after the introduction of penicillins into clinical practice, this problem became evident (Premlatha, 2019). The mechanism of resistance refers to the insensitivity of bacteria to the antimicrobial actions of a given antibiotic, such as changes in the target site, modification of the antibiotic itself, changes in the metabolic pathway, efflux pumps, reduced permeability, and changes in the cell membrane (Kumar Schweizer, 2005). It is very important to know the mechanisms of resistance in order to make a proper medical recommendation and understand the evolution of bacteria. Therefore, in this report, we conduct a study on the classification of resistance mechanisms, in particular on the recognition of mutations, their location and type, based on data from real sequencing of an E. coli strain resistant to the antibiotic ampicillin.

2 Methods

2.1 Whole-genome sequencing for antibiotic susceptibility testing

DNA is extracted from bacteria in clinical samples or more commonly, from cultured bacterial colonies. Sequencing technologies fragment the DNA and then randomly sequence to produce a library of reads (stored in FASTQ files). The reads are assembled into genomic scaffolds in silico. Sequencing is performed either using short-read second-generation technology, which tends to produce fragmented whole-genome assemblies of high accuracy, or long-read third-generation technologies that have higher error rates but more complete assemblies. WGS-AST algorithms operate on the raw reads and/or assembled contigs (Su et al., 2019).

2.2 Obtaining data

First, we downloaded the reference data from NCBI. The Reference genome and the genome annotation are publicly available: https://www.ncbi.nlm.nih.gov/genome/167. It is E.coli strain K-12 substrain MG1655, widely used bacteria in laboratories. It consist of 58022 reads (command [wc -l]). Direct link to RefSeq entry: https://www.ncbi.nlm.nih.gov/nuccore/$NC_000913.3$. Then we obtained the raw Illumina sequencing reads from shotgun sequencing of an E. coli strain that is resistant to the antibiotic ampicillin: here. Then we inspected the raw data at a glance by using command [zless file — head -20] in fasta format.

2.3 Quality control

Filtering the reads and quality control check by using tool fastqc. For that we used used FastQC: https://www.bioinformatics.babraham.ac.uk/projects/fastqc.If there are some problems with the raw sequence, the report in HTML, tables or graphs will be reported. It is recommended to replace the sample in case if the quality of data is too low (Wingett Andrews, 2018)

2.4 Removing low quality base

The trimming was performed with certain quality threshold by TRIMMOMATIC http://www.usadellab.org/cms/?page=trimmomatic. The given parameters were used: EADING:30 TRAILING:30 SLIDINGWINDOW:10:30 MINLEN:30 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 and compared with more strict options (LEADING:20 TRAILING:20 SLIDINGWINDOW:10:20 MINLEN:20).

2.5 Aligning sequences to the reference

The resulting reads were mapped to Escherichia coli str. K-12 substr. MG1655, complete reference genome using Burrows-Wheeler Aligner (BWA) version 0.7.17-r1188 http://bio-bwa.sourceforge.net/bwa.shtml (Sarmento et al., 2021) with BWA-MEM algorithm. By command [bwa index ref] , we run bwa index on the reference sequence with the default options. As a result, many files were produced.

2.6 Compressing SAM file

The .sam files were converted into .bam files by using samtools 1.15.1 (using htslib 1.15.1) https://github.com/samtools/htslib/tree/1.15.1 by command [samtools view -S -b].

2.7 Sorting and index BAM file

Then the .bam files were sorted and indexed using samtools sort command and then indexed. To visualize these data, IGV browser https://software.broadinstitute.org/software/igv/ was used with reference genome and aligned and indexed .bam file.

2.8 Variant calling

Variants for each sample were obtained using germline variant calling tool VarScan2 (variant scanner) http://dkoboldt.github.io/varscan/. Firstly, we used an option [samtools mpileup -f] to "piles up" the reads. We used VarScan with option:

1. –min-var-frequency = 0.5
2. –variants
3. –output-vcf (creates VCF format)

2.9 Variant effect prediction

Have the mutations actually changed any proteins in the host? We upload in IGV browser reference genome, annotation, .bam files and VCF and saw where the mutations have occurred, whether they are missense, nonsense or synonymous with these steps:

1. ref1.fna and ref2.fna.fai → in Genomes
2. ref.gff, .bam, bam.bai, .vcf files → in Tracks

3 Results

The raw given data consist of 58022 reads. It is worth to be mentioned that the raw data have a low quality. Some of the main base calls are in the red zone and orange. Which, in turn, can have a bad effect on long-term results when working with this data, which is why we used trimmomatic. (see supplementary 1.)

For default options (LEADING:20 TRAILING:20 SLIDINGWINDOW:10:20 MINLEN:20) the length was only 1784060 bp for both paired-end reads (see supplementary 2.)

After using trimmomatic, the quality of basic calls improved. The substandard ones were cut off.The data is in the green zone. We continue their further analysis. For required parameters: LEADING:30 TRAILING:30 SLIDINGWINDOW:10:30 MINLEN:30 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 the length of paired reads are 1440012 bp (see supplementary 3). The resulted length of sequence are presented on the Table 1.

We obtained some statistic after alignment: 720144 + 0 in total (QC-passed reads + QC-failed reads), 720006 + 0 primary, 0 + 0 secondary, 138 + 0 supplementary, 0 + 0 duplicates, 0 + 0 primary duplicates, 720097 + 0 mapped (99.99 percent : N/A), 719959 + 0 primary mapped (99.99 percent : N/A), 720006 + 0 paired in sequencing, 360003 + 0 read1, 360003 + 0 read2, 718476 + 0 properly paired (99.79 percent : N/A), 719940 + 0 with itself and mate mapped.

Our sequence were mapped by 99.79 percent. Then we obtained data from variant calling: Min coverage: 8, Min reads2: 2, Min var freq: 0.5, Min avg qual: 15, P-value thresh: 0.01, 4640871 bases in pileup file, 8 variant positions (6 SNP, 2 indel), 0 were failed by the strand-filter, 6 variant positions reported (6 SNP, 0 indel).

Then we annotated .vcf file using snpEFF http://pcingola.github.io/SnpEff and obtained new html file which contains visualization and information about our founded

in the beginning	after trimmimg	alligned	alligned+sorted
221056	1823504	1440012	272273

Table 1: Lenght of sequence while using trimmomatic

Ref(f)	Ref(r)	Alt	Name	Function
C	G	G	ftsI	cross-linking of the cell wall
T	A	A	acrB	Efflux pump membrane transporter
A	T	G	rybA	as a chaperone when Mn is limited
G	C	C	mntP	manganese efflux pump
A	T	T	envZ	membrane-associated protein kinase
A	C	G	rsgA	enabling RNA binding

Table 2: SNPs detected in the analysis

SNPs. Total number of SNPs in studied strain was six, four of them were missense and one silent (Table 2, supplementary 4).

4 Discussion

The SNPs we found allow us to conclude that E. coli is resistant to various antibiotics. The ftsl gene is responsible for the most important cell division protein that catalyzes the crosslinking of peptidoglycans of the cell wall in the division septum. (Nicholas et al., 1985). With this mutation, it is possible to disrupt the action of the antibiotic. The antibiotic acts by disrupting the synthesis of the cell wall by inhibiting the synthesis of peptidoglycan. Most likely, the antibiotic will not be able to bind to the catalytic serine. E. coli will continue to develop.

The mntP gene functions as pumps for manganese outflow(Waters et al., 2011). The action of the antibiotic leads to disruption of the functioning of the membrane as a result of the formation of new ion channels or the cessation of the outflow of manganese. When mutated, alter metabolic pathway to compensate. This in turn neutralizes the properties of the antibiotic. In order to continue treatment, it is necessary to choose an antibiotic with a different mechanism of action. An antibiotic with a biological action mechanism that disrupts protein synthesis. Inhibition of amino acid activation and transfer, ribosome function.

The envZ gene performs a variety of functions: member of the two-component regulatory system EnvZ/Epr involved in osmoregulation (particularly of genes OmpF and OmpC), EnvZ functions as a membrane-associated protein kinase that phosphorylates OmpR in response to environmental signals. (Comeau et al., 1985)(Premlatha, 2019). If the antibiotic acts through a change in the membrane and a mutation occurs. Therefore, this antibiotic will not be able to do its job. Since the mutation of this gene can lead to a decrease in permeability. This, in turn, will not allow the antibiotic to get into the bacterium. As an option for further treatment of the patient, we can take an antibiotic whose mechanism of action is based on a violation of protein synthesis. The optimal solution for choosing an antibiotic for the treatment of Escherichia coli of this strain will be the choice of an antibiotic with a biological mechanism of action of a violation of protein synthesis, as it was written earlier. Since the mutations we

found do not affect this mechanism of action of the antibiotic. Therefore, prediction of antibiotic resistance based on genome sequences is an important part for understanding the antimicrobial susceptibility profiles and as a result for design and development of new therapeutic strategies.

5 References

1. Sarmento, C., Guimarães, S., Kılınç, G. M., Götherström, A., Pires, A. E., Ginja, C., Fonseca, N. A. (2021). A Study on Burrows-Wheeler Aligner's Performance Optimization for Ancient DNA Mapping. International Conference on Practical Applications of Computational Biology Bioinformatics, 105–114.

2. Comeau, D. E., Ikenaka, K., Tsung, K. L., Inouye, M. (1985). Primary characterization of the protein products of the Escherichia coli ompB locus: structure and regulation of synthesis of the OmpR and EnvZ proteins. Journal of Bacteriology, 164(2), 578–584.

3. Nicholas, R. A., Strominger, J. L., Suzuki, H., Hirota, Y. (1985). Identification of the active site in penicillin-binding protein 3 of Escherichia coli. Journal of Bacteriology, 164(1), 456–460. https://doi.org/10.1128/jb.164.1.456-460.1985

4. Premlatha, M. (2019). Microbial resistance to antibiotics. Bacterial Adaptation to Co-Resistance, 4(December), 61–80. https://doi.org/10.1007/978-981-13-8503

5. Waters, L. S., Sandoval, M., Storz, G. (2011). The Escherichia coli MntR miniregulon includes genes encoding a small protein and an efflux pump required for manganese homeostasis. Journal of Bacteriology, 193(21), 5887–5897.

6 Suppmementary materials:

1. Supplement1:https://docs.google.com/document/d/1rCcVzsLylNvr-nNIKpV3k8VV1JItMmpsj8hspA32cpw/edit?usp=sharing

2. Supplement2: https://docs.google.com/document/d/1KxQk2tRsK5YRVsD0pyvm704eitQOKLL1
HO78/edit?usp=sharing

3. Supplement3: https://docs.google.com/document/d/1oQtkJFs9WU0QCQ7YYkXQiRMtpEqYq2

4. Supplement4: https://drive.google.com/file/d/17XOEEazzq8Cjw70b5fwOP7LT9BdEKMR0/view