Genomic characterization of ampicillin resistant Echerichia coli

Filippov M.

Abstract
Resistant to antibiotics bacteria strands are emerging treat to human health, as antibiotics usage increases worldwide. In this project, open-source data from shotgun sequencing of ampicillin resistant E. coli was analyzed using common bioinformatic tools. Comparsion of sequenced strand genetic data with reference E. coli genome (strain K-12 substrain MG1655) shown 6 mutations, 4 of which are missense. Some of the mutated genes' products are proteins with known role in antibiotic resistance. Probable mechanisms of ampicillin resistance were suggested.

Introduction
E. coli is an important part of normal human microbiome, the most popular procariote model object in biology and a crucial producent in biotechnology. In these three large fields of human interest there are place for interaction of E. coli with antibiotics.First, this bacteria under certain circumstances has pathogenic effects, which include intestinal and urinary tract infections, sepsis and even meningitis in infants. Treatment for E. coli induced deseases is antibiotic therapy [1]. Second, as a model object, E. coli is a important „military training ground" in attempt to develop antibiotic treatment for other, much more severe, bacterial infections. Resistant strains wont allow proper experiments. Third, common practice in biotechnology is to transform E. coli cells with plasmids containing antibiotic resistance genes to control transformation itself. In this case resistant E. coli is also a huge problem.
There are several ways for bacteria to become resistant to antibiotics:
1. Mutation in antibiotic binding site
2. Removing the antibiotic from the cell with efflux pumps
3. Preventing the entry of an antibiotic into the cell
4. Inactivation or destruction of the antibiotic
5. Compensating damaged pathway with alternative one
Since all these methods are possible only when the enzymes or proteins of the cell are altered in some way, genetic changes must be present. And if there are genetic changes present, they can be detected in the genomic data. Therefore, in most cases, it is possible to determine the exact cause of antibiotic resistance.
In this study we examine ampicillin resistant E. coli strain. Ampicillin is a penicillin-like antibiotic that has a beta-lactam ring in its structure. It is an irreversible inhibitor of transpeptidase — bacterial enzyme contributing cell wall building. In presence of ampicillin bacteria can't maintain its cell wall integrity and undergoes lysis — destruction of membrane and loss of content [2].

Methods
Ampicillin resistant E. coli paired end shotgun sequencing data from Illumina in .fastq format was obtained [3]. Reference genome is strain K-12 substrain MG1655 E. coli genome in .fasta, .gff and .gbff formats, available in genebank [4]. Primary inspection of sequencing data was performed with FastQC 0.11.9 [5]. Reads was trimmed with Trimmomatic 0.39 [6] with following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10, LEADING:20, TRAILING:20, SLIDINGWINDOW:10:20, MINLEN:20.
Aligment was done with BWA 0.7.17-r1188 [7]. Reference genome was indexed with «bwa index», then reads from sequencing data was aligned with «bwa mem».
Samtools 1.13 [8] was used to compress .sam file («samtools view -b [data] > [data].bam»), show basic statisctics («samtools flagstat»), sort («samtools sor»t) and index («sam index») alignment.
For variant calling VarScan 2.4.4 [9] was used. .mpileup file was created with «samtools mpileup -f», VasrScan was executed from java file («java -jar Varscan.v2.4.4.jar mpileup2snp») with following parameters: --min-var-freq 0.5, --variants, – output-vcf 1.

Automatic annotation of revealed SNPs was performed with SnpEff [10]. .config file was created with command «echo k12.genome : ecoli_K12 > snpEff.config», then reference genome in .gbff format was renamed as genes.gbk and database was created with command «java -jar snpEff.jar build -genbank -v k12». Annotation was done by command «java -jar snpEff.jar ann k12 [data].vcf > [annotated data].vcf».
Inpection of obtained data was performed with Integrative Gemomic Viewer (IGV) [11]

Results

Basic inspection of sequencing data in FastQC shown that there are 455876 reads. Basic properties shown in *table 1*. Forward and reverse sequencing have inequal quality per tile: one in forward sequence is marked as unusual by the program (*picture 1*). Per base quality in both reverse and forward files is marked as unusual too (*pictures 2 and 3*). Other parameters are marked as normal or slightly abnormal.
Trimming with Trimmomatic was performed, and Per base quality in both filed is now marked as normal (*pictures 4 and 5*). Number of reads after trimming is 445524, which is 97,7% of that in untrimmed files (*table 2*).
Basic statistic after alignment with bwa mem performed with samtools flagstat is shown in *picture 6*. 890189 reads, which is 99,87% of total numer of reads, are properly mapped.
With VarScan setted to at least 50% reads support 6 SNPs was identified. Manual inspection in IGV showed that mutated genes are fstI, acrB, rybA, mntP, rsgA and envZ. It was possible to determine types of mutations, and results of such analyzis demonstrated in *table 3*. Automatic annotation using SnpEff confirmed manually obtained results and made it possible to know exact changes in mutated genes' products' aminoacid sequence. There changes are also displayed in *table 3* .

Discussion

Three of missense-mutated genes — acrB, envZ and ftsI, has clear role in providing ampicillin resistance in E. coli. acrB is a multidrug efflux pump, which, as mentioned in the introduction, can pump number of drugs out of the cell . Note that it uses proton force to execute its function. It can be assumed that in wild type E. coli this pump is not able to excrete ampicillin, but in our case it mutated to get that ability [12]. envZ, on the other hand, is the gene that codes protein EnvZ, which is the part of EnvZ/OmpR osmoregulation system. Interesting that it is suggested that particulary EnvZ protein is sensitive to cytoplasm acidification. Maintain normal cytoplasmic pH is crucial when abnormal use of protonic effllux pump takes place: as drug pumped out, protons pumped in, that necessarily will cause pH lowening in cytoplasm [13]. Finally, ftsI is the gene coding transpeptidase FtsI, which is the direct substrate for ampicillin. Presumably, detected mutation changes its structure in such a way that ampicillin can't bind anymore [14].
Significance of two other missense-mutated genes is unclear. mntP and rybA are part of manganese homeostasis system. mntP is probable manganese efflux pump and rybA codes small RNA, that can be translated into mntS small protein, which function is not fully studied, but strongly associated with low levels of manganese ions in the cell [15]. Analyzing probable functional role of all these mutation, it can by assumed, that changes in envZ work can disrupt manganese or osmotic homeostasis in general, and this disruption is compensated by changes in mntP/rybA system.
Another possible explanation is that
Further studies can be devoted to clarifying the role of the found changes in primary structure of mutated proteins and their exact role in promoting ampicillin resistance. Other probable direction of scientific work is to research possibilities of changing mutated bacteria's genome in vivo and suppress emerged resistance.
Recomendation for treatment of patients which has been diagnosed with E. coli infection caused by strain with similar mutations is to prescribe non-beta-lactam antibiotic with no transpeptidase inhibition activity. Such antibiotics with shown their efficience on E. coli are doxycycline and levofloxacin.

Substitutes

Measure	Value
Filename	amp_res_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	455876
Sequences flagged as poor quality	0
Sequence length	101
%GC	50

Table 1: Basic properties of raw sequencing data (FastQC)

Measure	Value
Filename	amp_res_1_trimmed.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	445524
Sequences flagged as poor quality	0
Sequence length	20-101
%GC	50

Table 2: Basic properties of trimmed sequencing data (FastQC)

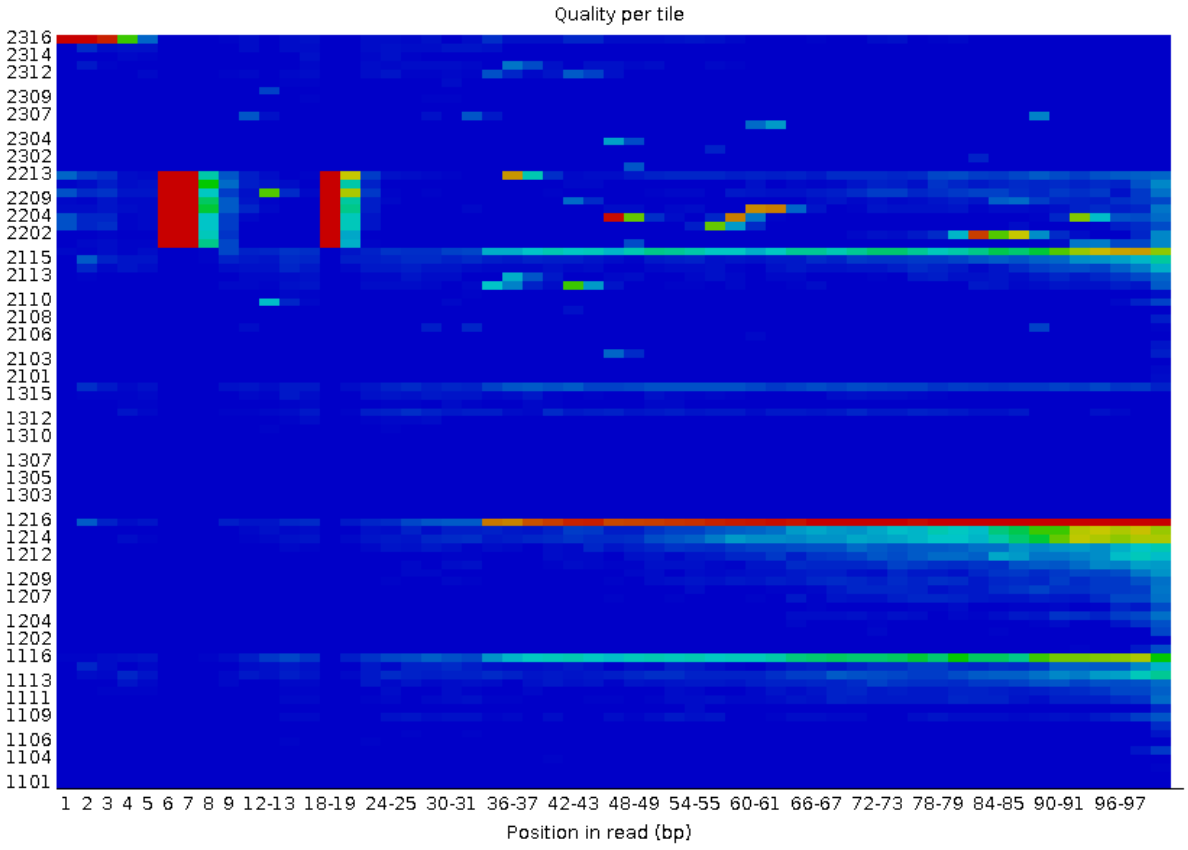

Figure 1: Per tile quality of forward sequence (FastQC)

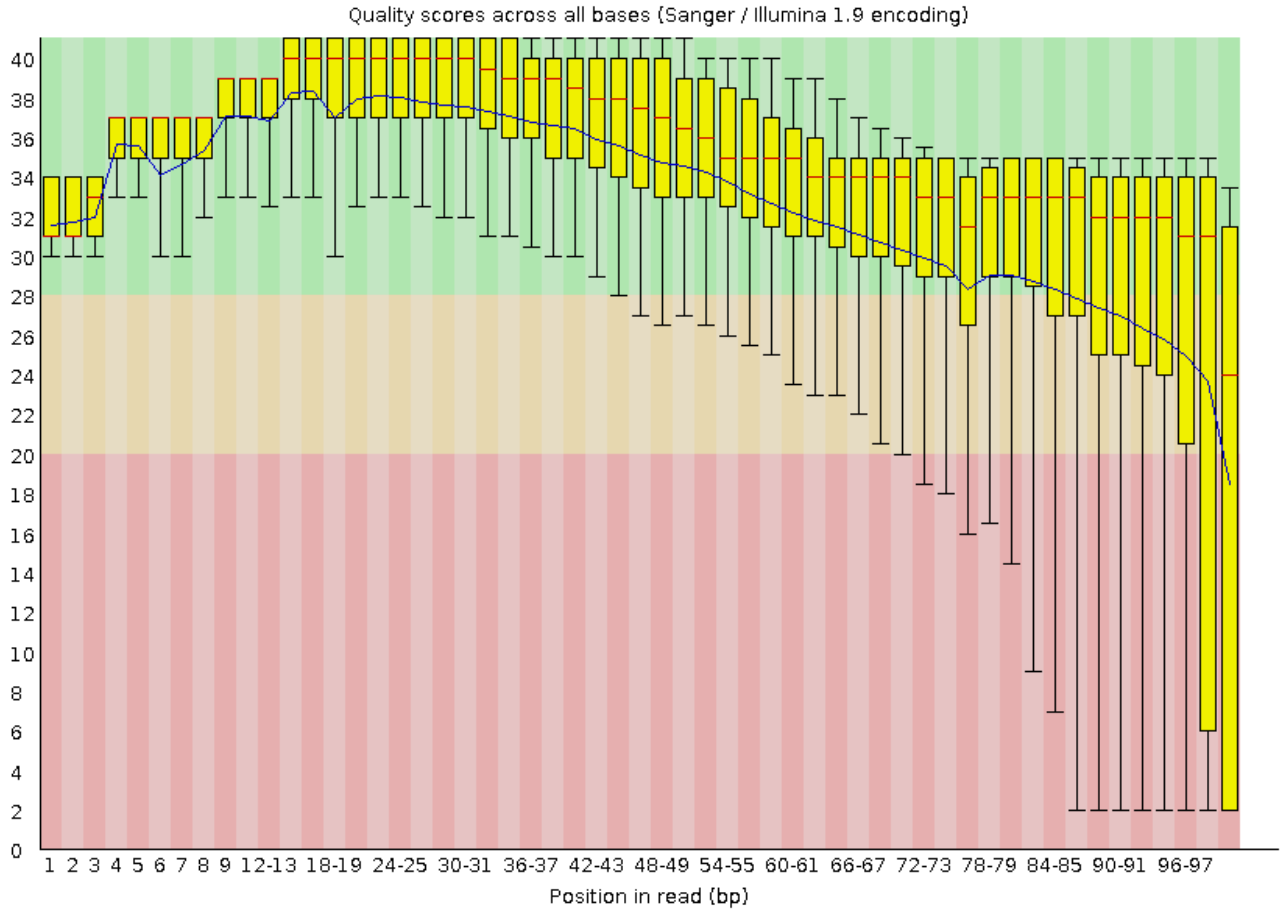

Figure 2: Per base quality of forward sequence (FastQC)

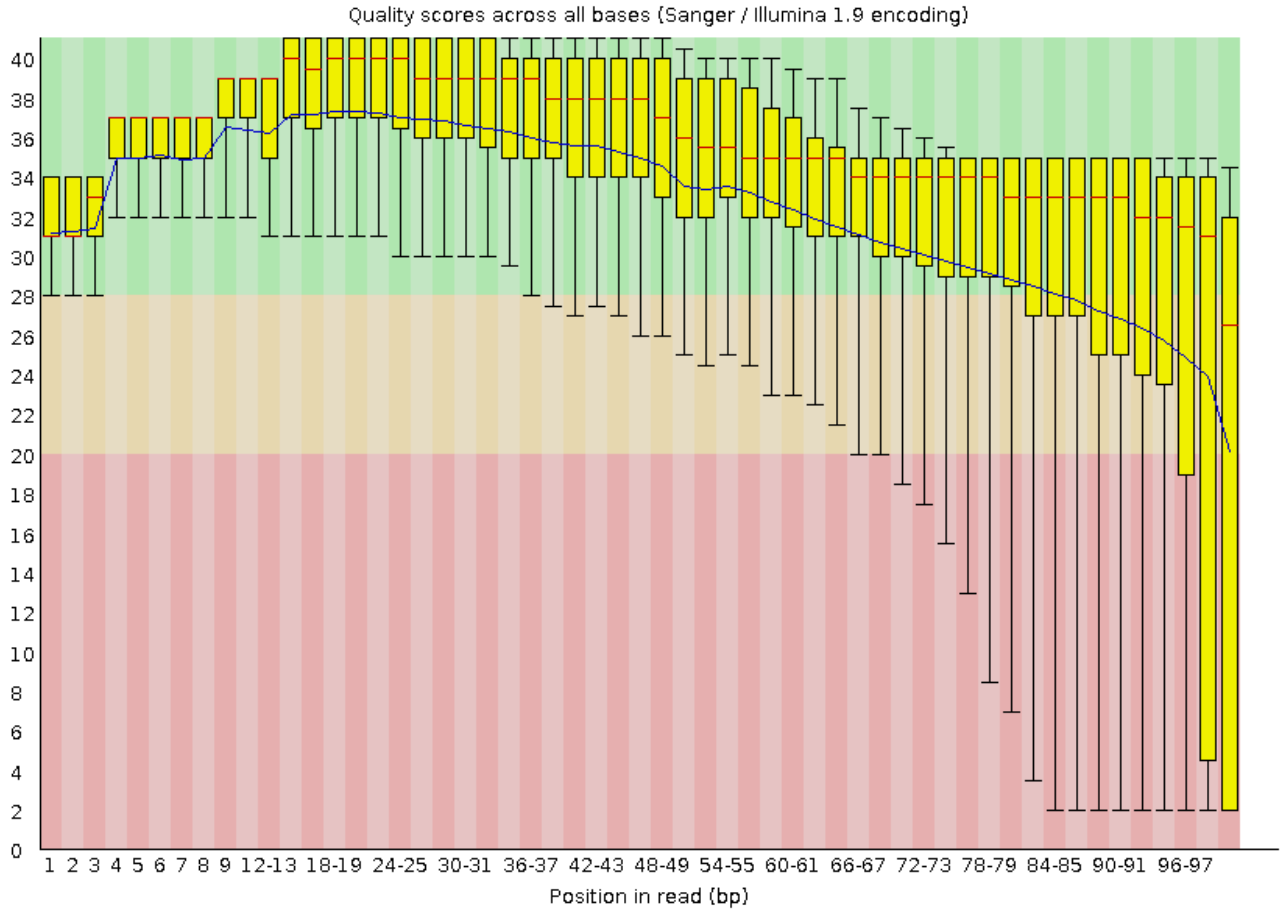

Figure 3: Per base quality of reverse sequence (FastQC)

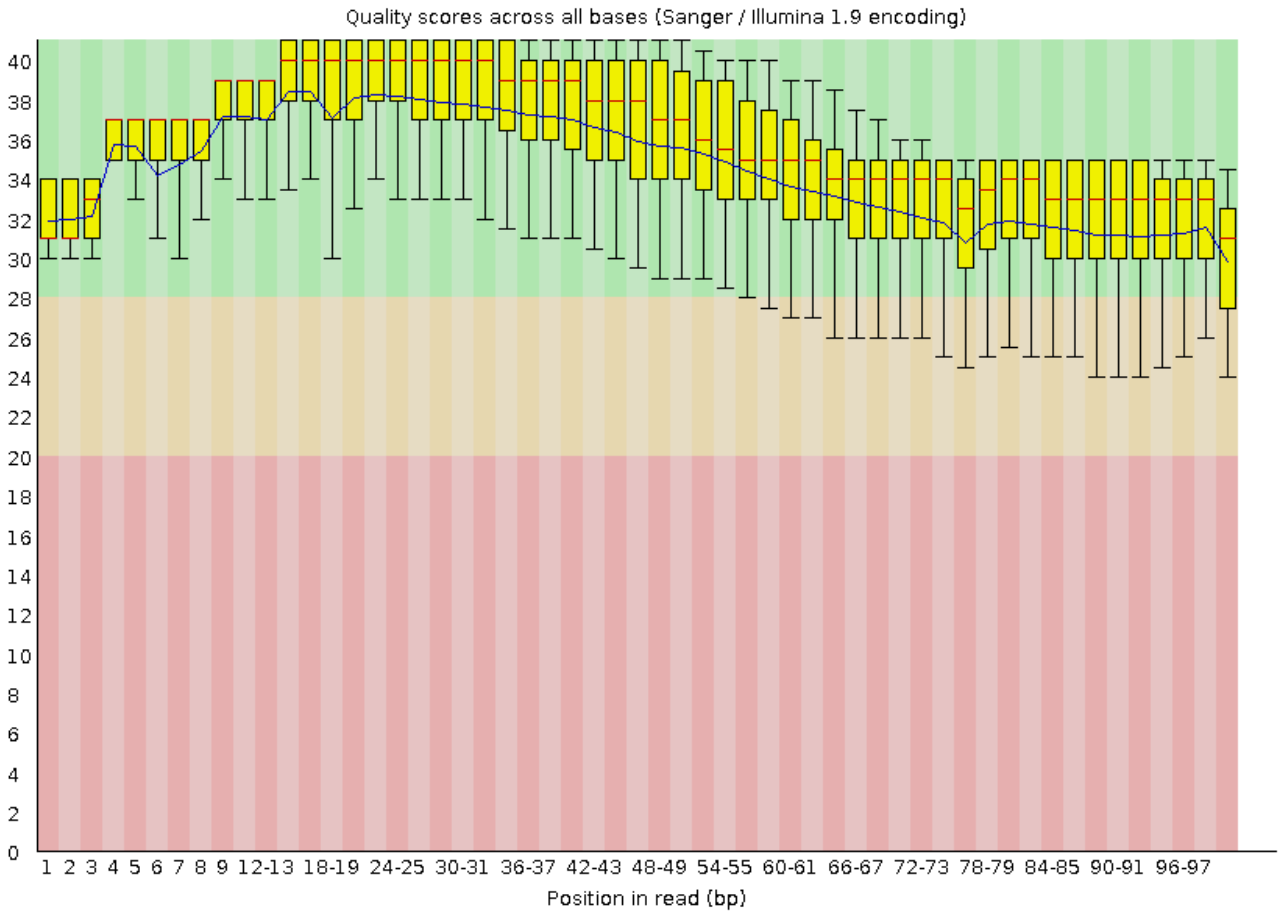

Figure 4: Per base quality of forward sequence after trimming (FastQC)

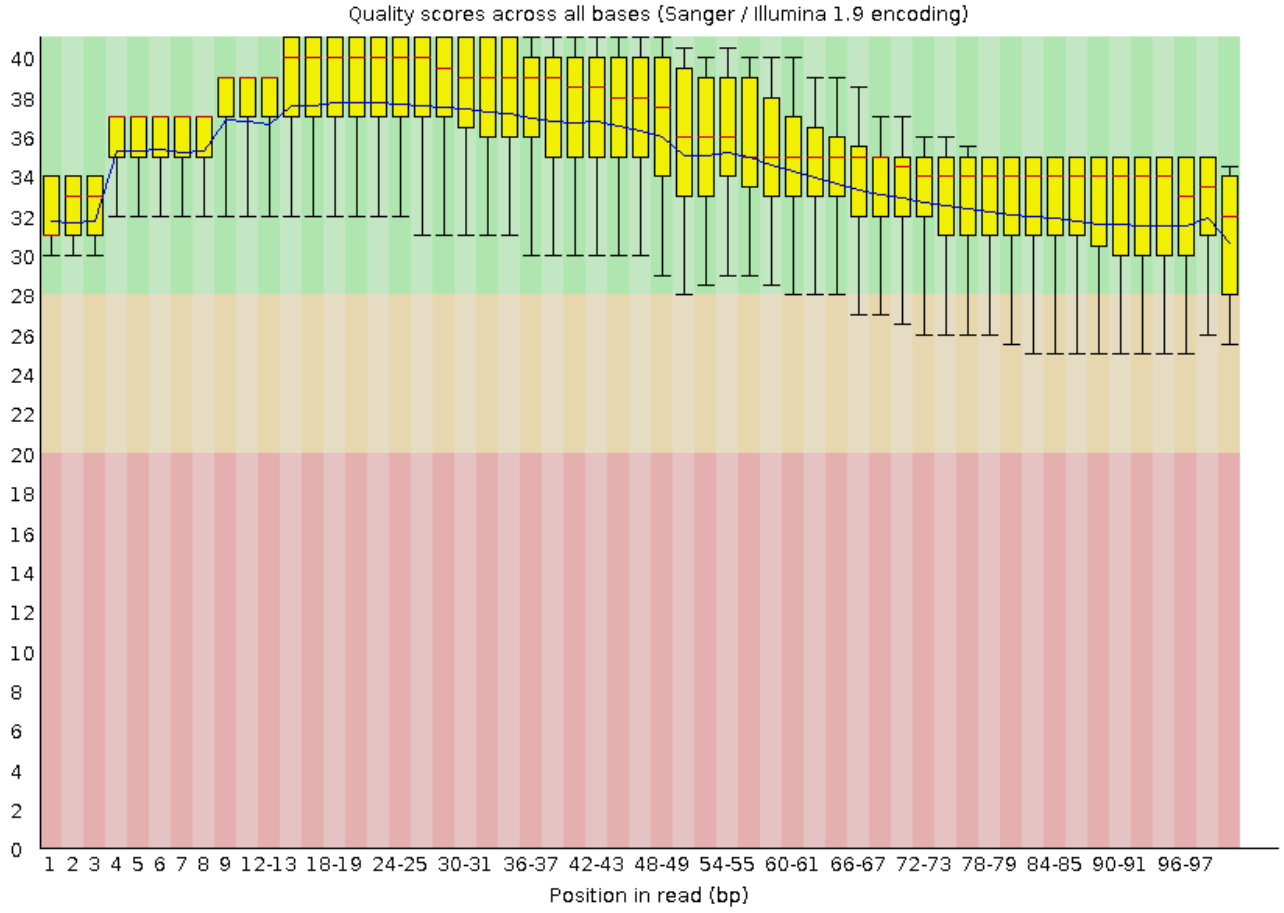

Figure 5: Per base quality of reverse sequence after trimming (FastQC)

```
891306 + 0 in total (QC-passed reads + QC-failed reads)
891048 + 0 primary
0 + 0 secondary
258 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
890189 + 0 mapped (99.87% : N/A)
889931 + 0 primary mapped (99.87% : N/A)
891048 + 0 paired in sequencing
445524 + 0 read1
445524 + 0 read2
887108 + 0 properly paired (99.56% : N/A)
888960 + 0 with itself and mate mapped
971 + 0 singletons (0.11% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Figure 6: Samtools flagstat result

Gene name	Mutation type	Aminoacid change
ftsI	missense	Ala544gly
acrB	missense	Gln569Leu
envZ	missense	Val241Gly
rybA	-	-
mtnP	missense	Gly25Asp
rsgA	synonimous	-

Table 3: Mutations found with SnpEff

References

1. Lim JY, Yoon J, Hovde CJ. A brief overview of Escherichia coli O157:H7 and its plasmid O157. J Microbiol Biotechnol. 2010 Jan;20(1):5-14. PMID: 20134227; PMCID: PMC3645889

2. Petri (2011). "Penicillins, Cephalosporins, and Other Beta-Lactam Antibiotics". In Brunton L, Chabner B, Knollman B (eds.). Goodman and Gilman's the Pharmacological Basis of Therapeutics, Twelfth Edition (12th ed.). New York: McGraw Hill Professional. pp. 1477–1504.

3. https://figshare.com/articles/dataset/amp_res_2_fastq_zip/10006541/3

4. https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF_000005845.2_ASM584v2/

5. Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

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

7. Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]

8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup, The Sequence alignment/map (SAM) format and SAMtools, Bioinformatics (2009) 25(16) 2078-9 [19505943]

9. 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
URL**:** http://varscan.sourceforge.net

10. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012 Apr-Jun;6(2):80-92. doi: 10.4161/fly.19695. PMID: 22728672; PMCID: PMC3679285

11. 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)

12. Hobbs EC, Yin X, Paul BJ, Astarita JL, Storz G. Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. Proc Natl Acad Sci U S A. 2012 Oct 9;109(41):16696-701. doi: 10.1073/pnas.1210093109. Epub 2012 Sep 24. PMID: 23010927; PMCID: PMC3478662.

13. Chakraborty S, Winardhi RS, Morgan LK, Yan J, Kenney LJ. Non-canonical activation of OmpR drives acid and osmotic stress responses in single bacterial cells. Nat Commun. 2017 Nov 14;8(1):1587. doi: 10.1038/s41467-017-02030-0. PMID: 29138484; PMCID: PMC5686162

14. Spratt BG. Distinct penicillin binding proteins involved in the division, elongation, and shape of Escherichia coli K12. Proc Natl Acad Sci U S A. 1975 Aug;72(8):2999-3003. doi: 10.1073/pnas.72.8.2999. PMID: 1103132; PMCID: PMC432906

15. Martin JE, Waters LS, Storz G, Imlay JA. The Escherichia coli small protein MntS and exporter MntP optimize the intracellular concentration of manganese. PLoS Genet. 2015 Mar 16;11(3):e1004977. doi: 10.1371/journal.pgen.1004977. Erratum in: PLoS Genet. 2015 Jun;11(6):e1005322. PMID: 25774656; PMCID: PMC4361602