Lab Report

Dmitrii Traktirov & Valentin Mikhalchuk

Investigation of the genome E.coli X strain: properties, evolutionary origins and pathogenic potential

Abstract. A new strain of E. coli X causes a fatal hemolytic uremic syndrome. Researching a new strain can help to choose the right treatment. To understand its pathogenicity, properties, evolutionary origin, antibiotic resistance, this study was carried out using bioinformatic analysis of the sequencing data of the TY2482 isolate. As a result, it turned out that strain X is most similar to *Escherichia coli 55989* strain. A few genes which are responsible for pathogenicity were detected in *E.Coli* X strain, such as *stxA* and *stxB* genes, that are responsible for shiga toxin expression. Also, it was found that *E.Coli* X is resistant to a wide range of antibiotics, such as tetracycline, antibiotics of beta-lactam family and some others.

Introduction. The pathogenicity of E.coli, like any other bacteria, is determined by virulence factors. For example, it could be the shiga toxin that causes hemolytic uremic syndrome (HUS). Virulence factors can be encoded on the bacterial chromosome or plasmid. Other pathogenic agents are phages. These are viruses that cannot reproduce and therefore infect bacteria. They pass their own genome into the bacterial genome. Phages are an example of horizontal gene transfer (HGT). For example, phages can transform a non-virulent strain into a virulent one. [1]

It is important in some cases to assemble the genome de novo instead of aligning reads to a reference, because in this case it is possible to find transcribed regions, the sequences of which are absent in the genomic assembly. Having a de novo assembly and a reference genome, it is possible to detect transcripts of exogenous origin.[10]

Methods. Three libraries with different insert sizes were available for analysis. To confirm the quality of the received reads and count them, we used the FastQC program [2]. Jellyfish count tool with parameters -m 31 -C -s 6M was used to access the k-mer distribution, with further visualization and genome size estimation using python 3.8 [3]. SPAdes assembler with one SRR292678 library and all three libraries was used to assemble *E.Coli* X genome and compare qualities of both assemblies, and such quality was compared with QUAST-5.0.2 [4, 5]. After assembly, Prokka --compliant --prefix scaffolds --force was used to annotate the genome and predict genes [6]. Then, 16S rRNA gene sequence was found in *E.Coli* X with barrnap, and the closest relative strain was determined using the obtained sequence and NCBI BLAST [7]. Finally, *E.Coli* X assembly was examined for differences in the presence of genes using Mauve, and ResFinder 4.1 was used to search for genes responsible for antibiotic resistance [8, 9].

Results. Three libraries from the TY2482 sample, which were generated at Beijing Genome Institute, with different insert sizes were used: SRR292678 - paired end, insert size 470 bp; SRR292862 – mate pair, insert size 2 kb; SRR292770 – mate pair, insert size 6 kb. Based on file with forward reads from SRR292678 library we successfully estimated genome size:

$$N = \frac{(M*L)}{(L-K+1)} = \frac{62*90}{90-31+1} = 93$$

$$Genome\_size = \frac{T}{N} = \frac{5499346}{93} \simeq 5.9 \text{ Mb.}$$

,

where N - Depth of coverage, M - Kmer peak, K - Kmer-size, L - avg read length, T- Total bases).

Two assemblies were performed using only SRR292678 (paired end, insert size 470 bp) library and using all three available libraries (table 1).

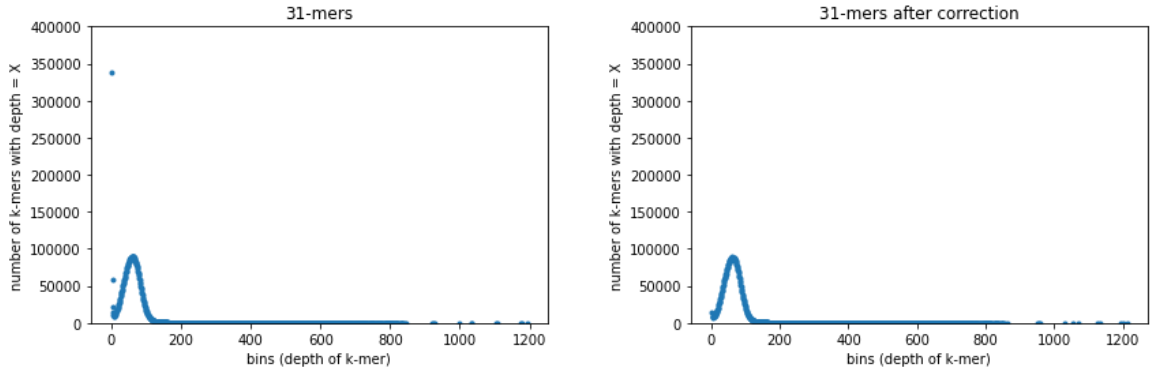
Table 1. Assemble quality for one- SRR292678 and three-library assemblies

	one-library assemble	three-library assemble
contigs (>= 500bp)	206	105
N50	105346	335515

As can be seen from table 1, N50 statistics is higher in "three-library" case, and average number of contigs is lower in this case, due to their greater length, which means higher quality of three-library assemble.

The results of plotting k-mer (k = 31) profile are shown on picture 1. It seems like correction error step of SPAdes corrects mainly first bins, eliminating outliers

A. **B.**



Picture 1. k-mer profile before (A) and after (B) SPAde's correction error step

During barrnap step, six similar to 16S rRNA sequences were found in *E.Coli* X strain with length 1537 each. Based on these sequences, the closest relative strain was found using the BLAST algorithm -- it turned out to be a Escherichia coli 55989 (NC_011748.1). Reference genome belongs to an enteroaggregative *E.Coli* strain, harboring the pAA plasmid, which contains aggregative adherence fimbria (AAF) genes allowing bacteria to stick to cells in the intestine (pic. 2).

Picture 2. Fimbria on E.Coli.

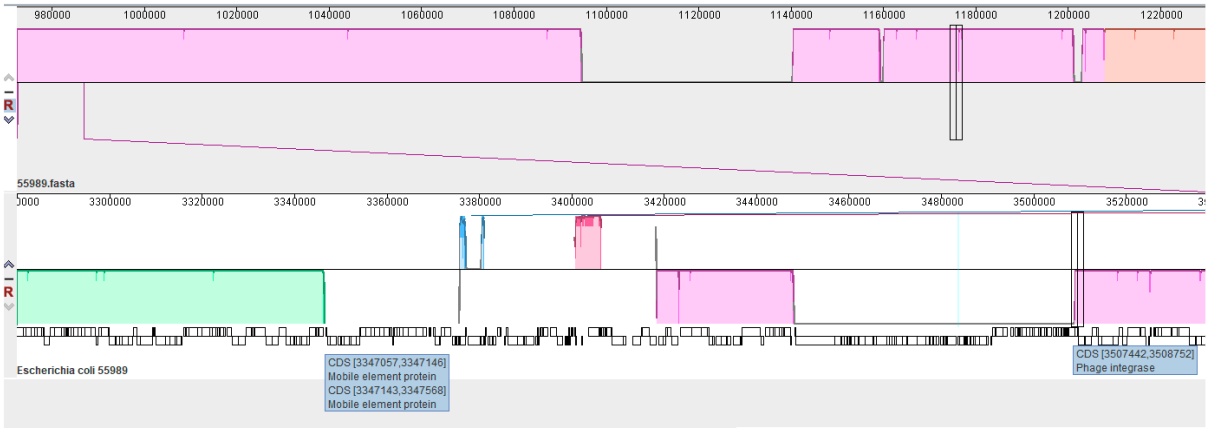
However, reference *E.Coli* does not contain shiga toxin genes whereas *E.Coli* X does - these are *stxA* and *stxB* genes (table 2).

Table 2. New genes obtained by E.Coli X

gene	gene product	position in *E.Coli* X	gene length
stxA	Shiga toxin subunit A precursor	3483886−3484845	959
stxB	Shiga toxin subunit B precursor	3483605−3483874	269
bla	class A beta lactamase (TEM family)	5199263 - 5200123	860
	class A beta lactamase (CTX-M family)	5195566 - 5196441	875
	class C beta lactamase (BlaEC family)	4758802 - 4759935	1133

It also turned out that *E.Coli* X is resistant to some antibiotics, such as cefepime, ampicillin, cefotaxime, ceftazidime from beta-lactam family, antibiotic tetracycline, and folate pathway antagonist antibiotics trimethoprim and sulfamethoxazole, from which reference *E.Coli* does not have protection.

Discussion. It is widely known that bacteria are able to gain new features through HGT. Examination of whole *E.Coli* X sequence segment, containing horizontally acquired genes has also revealed the content of “mobile” genes, such as phage integrase — enzymes that mediate unidirectional site-specific recombination between two DNA recognition sequences, the phage attachment site, attP, and the bacterial attachment site, attB (pic 2.).



Picture 3. Mauve alignment. Newly acquired unaligned region shown as wide white block. *bla* gene is shown as a faint blue line

These mobile genetic elements can be transferred from one species to another together with virulence factors and antibiotic resistance genes on them, and that is how *E.Coli* X acquired new genes: *stxA*, *stxB* and *bla*. Latter allowed *E.Coli* X to become resistant to beta-lactam antibiotics, as *bla* gene encodes beta-lactamase enzyme that is capable of breaking the β-lactam ring open, deactivating the molecule's antibacterial properties.

Since antibiotic therapy could not be used on patients with resistant *E.Coli* strain, there are also alternative ways to get rid of the pathogenic bacillus, ex bacteriophages are able to recognize and effectively destroy colonies of E. coli and antibiotic resistance does not reduce the efficiency of this process.

References

1. Adrien Joseph,Aurélie Cointe,Patricia Mariani Kurkdjian,Cédric Rafat, and Alexandre Hertig, Shiga Toxin-Associated Hemolytic Uremic Syndrome: A Narrative Review. Toxins (Basel). 2020 Feb; 12(2): 67.
2. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
3. Guillaume Marcais and Carl Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics (2011) 27(6): 764-770; doi: <https://doi.org/10.1093/bioinformatics/btr011>
4. Prjibelski, Andrey; Antipov, Dmitry; Meleshko, Dmitry; Lapidus, Alla; Korobeynikov, Anton (2020). Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics, 70(1), –. doi:10.1002/cpbi.102
5. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Gloeckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Opens external link in new window Nucl. Acids Res. 41 (D1): D590-D596.
6. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068–2069. doi:10.1093/bioinformatics/btu153

7. Seemann, T. (2013). Barrnap: BAsic Rapid Ribosomal RNA Predictor. Available online at: https://github.com/tseemann/barrnap
8. N. Conrad, A Whole Genome Alignment Visualization Tool for the Web, (2019), GitHub repository, https://github.com/nconrad/mauve-viewer
9. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. (2020) PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. Journal of Antimicrobial Chemotherapy 72(10) 2764-2768.
10. Adrien Joseph,Aurélie Cointe,Patricia Mariani Kurkdjian,Cédric Rafat, and Alexandre Hertig, Shiga Toxin-Associated Hemolytic Uremic Syndrome: A Narrative Review. Toxins (Basel). 2020 Feb; 12(2): 67.