Genome assembly as a way finding differences associated with antibiotic resistance and virulence in bacterial strains

Vinogradova Sofiya, Zherko Nikita
Moscow Institute of Physics and Technology

November 25, 2023

Abstract

A 2011 bacterial outbreak in Germany forced people to look for solutions to combat a new virulent strain. In this work, we tried to find changes in the *E. coli* X genome that led to virulence and acquisition of resistance to other types of antibiotics. By assembling the *E. coli X* genome, we were able to find regions missing from the reference genome, in which we found toxicity and antibiotic resistance genes of interest.

1 Introduction

The phenomena of horizontal gene transfer is characteristic of bacteria and plays an important role in the process of their evolution. This transfer of genes from a host organism to a recipient organism can occur both between bacteria alone and between bacteria and viruses. Horizontal gene transfer contributes to the acquisition of new traits, including those related to antibiotic resistance and virulence. It is also worth noting that this type of genetic information exchange can make non-virulent bacterial strains virulent due to the incorporation of the virus genome into their genome [1].

The German case is an example of horizontal gene transfer with subsequent consequences. Specifically, a non-virulent bacterium has acquired pathogenic properties that have caused HUS (hemolytic uremic syndrome).

In our work, we have resorted to genome assembly rather than alignment to a reference for the reason that We can use one of the two given techniques depending on the experiment. In our case, the first option has an advantage because get information about both the depth of sequencing and the base quality scores.

2 Materials and Methods

We worked with three libraries from the TY2482 sample: paired-end library SRR292678 (insert size 470 bp) and two mate pair libraries SRR292862 (insert size 2 kb), SRR292770 (insert size 6 kb). Forward and reverse reads from each library were checked with FastQC [2].

For counting k-mers we used Jellyfish [3] and using the written R script we estimated the genome size.

The genome was assembled using Spades v.3.15.4 (Bankevich et al., 2012). After genome assembly, QUAST [4] should be used to obtain basic assembly statistics. These operations are performed for both the single-library and the three-library.

To search for a reference, take a ready-made archive PROKKA.zip [5]. In order to find the best matching genome with the *E. coli X* genome we use the conserved 16S rRNA genes that we were able to obtain by rRNA genes prediction tool Barrnap [6]. After make nucleotide blast [7] of this file and select the Reference Genome Database (refseq_genomes) in the Database field, and Escherichia coli in the Organism field.

The search for toxin genes and genes for antibiotic resistance is performed through a program Mauve [8] that visualises the genome-wide alignment of *E. coli X* and the reference genome

To determine which antibiotics are resistant to *E. coli X* and reference, we're going to use ResFinder [9].

3 Results

We have three libraries and all the files in these libraries were analyzed using the FastQC tool. You can show all results in "fastqc_results" directory on github.

For further studies, we estimated the genome size by analysing the paired-end library (SRR292678). We ended up estimation with a genome size of 5168217.

After assembly we get a whole genome in 2 variants (after single-library assembly and three-library assembly). Below in Table 1 are the basic statistics for each of the two assembly. You can find more information in the "quast_results" directory on github.

Also annotation statistics: assembly length = 5390599, contigs = 247, N50 = 335515. Results are in the "prokka" directory on github.

To find a reference genome, we used a conserved gene 16S ribosomal RNA. *E. coli X* genome include 7 repeats of this gene: six of similar length (1538 bp) and one is differ (406 bp). Next, using Blast, we were able to find the most similar genome to *E. coli X* and took it as a reference (Escherichia coli 55989, complete sequence).

Assembly	Total length	Number of contigs	N50
single-library	5295721	210	111860
three-library	5350156	105	335515

Table 1: Basic statistics of assemblies

Shiga toxin genes have been discovered in the *E. coli X* genome and in order to speculate on their pathway into the genome we need to analyse the flanking genes.

stxA, stxB stxA, stxB are flanked on the 5'-end by Phage DNA adenine methylase, Phage antitermination protein Q, Phage protein NinH, Phage recombination protein NinG, Gifsy-2 prophage protein, DNA primase, phage associated, Putative ATP-dependent helicase.

stxA, stxB are flanked on the 3'-end by ORF B78, Phage head completion protein, Phage holin/antiholin component S, Phage lysozyme R, Phage antirepressor protein, Phage endopeptidaze Rz.

We also decided to investigate the antibiotic resistance of strain *E. coli X* and using ResFinder we found these antibiotics for strain *E. coli X* and the reference strain.

E. coli X resistant to streptomycin, amoxicillin, ampicillin, cefepime, cefotaxime, ceftazidime, piperacillin, aztreonam, ticarcillin, ceftriaxone, cephalothin, sulfamethoxazole, trimethoprim, tetracycline, doxycycline

Reference resistant to tetracycline, doxycycline, minocycline.

In order to understand the mechanism of antibiotic resistance, we will also analyse the flanking genes that encode beta-lactamases of different classes.

bla_1, bla_2 are flanked on the 5'-end by Mobile element protein. bla_1, bla_2 are flanked on the 3'-end by Tryptophan syntase, Mobile element protein

Class C beta-lactamase gene is flanked on the 5'-end by Small multidrug resistance (SMR) efflux transporter Class C beta-lactamase gene is flanked on the 3'-end by Fumarate-reductase group genes

Class A beta-lactamase gene is flanked on the 5'-end by error-prone genes Class A beta-lactamase gene is flanked on the 3'-end by mobile element protein

4 Discussion

The causes of virulence in *E. coli* may be different and encode on chromosomal, plasmid, transposon, or temperate bacteriophage DNA [10]. Analyses of the annotated genome showed that the Shiga toxin genes appeared in the bacterial genome in a region unaligned with the reference genome. The analysis of flanking genes gives us a complete understanding that this toxin is viral in nature and has been transferred into the genome by virus infection.

Antibiotic resistance occurs by the mechanism of enzyme (beta-lactamase) inactivation. Having analysed the three regions where beta-lactamases of different types are located, it can be concluded that resistance has emerged through mobile elements such plasmids, transposons, because genes are located in aligned with the reference regions.

Supportive therapy remains the mainstay of treatment of HUS, managing fluid balance, electrolyte abnormalities and hypertension if present. Up to 80% of patients will require transfusion with blood or platelets during their illness. There is no optimal antibiotic treatment, because when bacterial cells die, a toxin is released. however, there are cases where antibiotic treatment has helped: azithromycin, fosfomycin. [**treatment**]

5 References

[1] Brian J. Arnold, I-Ting Huang, and William P. Hanage. "Horizontal gene transfer and adaptive evolution in bacteria". In: *Nature Reviews Microbiology* 20 (4 Apr. 2022), pp. 206–218. ISSN: 1740-1526. DOI: 10.1038/s41579-021-00650-4.

[2] *FastQC*. June 2015. URL: https://qubeshub.org/resources/fastqc.

[3] Guillaume Marçais and Carl Kingsford. "A fast, lock-free approach for efficient parallel counting of occurrences of k-mers". In: *Bioinformatics* 27.6 (Jan. 2011), pp. 764–770. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr011. eprint: https://academic.oup.com/bioinformatics/article-pdf/27/6/764/48866141/bioinformatics_27_6_764.pdf. URL: https://doi.org/10.1093/bioinformatics/btr011.

[4] Alexey Gurevich et al. "QUAST: quality assessment tool for genome assemblies." In: *Bioinformatics (Oxford, England)* 29 (8 Apr. 2013), pp. 1072–5. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btt086.

[5] URL: https://disk.yandex.ru/d/4ZzBdc2IxzZUb.

[6] URL: https://github.com/tseemann/barrnap.

[7] M. Johnson et al. "NCBI BLAST: a better web interface". In: *Nucleic Acids Research* 36 (Web Server May 2008), W5–W9. ISSN: 0305-1048. DOI: 10.1093/nar/gkn201.

[8] Aaron C E Darling et al. "Mauve: multiple alignment of conserved genomic sequence with rearrangements." In: *Genome research* 14 (7 July 2004), pp. 1394–403. ISSN: 1088-9051. DOI: 10.1101/gr.2289704.

[9] Alfred Ferrer Florensa et al. "ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes." In: *Microbial genomics* 8 (1 Jan. 2022). ISSN: 2057-5858. DOI: 10.1099/mgen.0.000748.

[10] Johnny W. Peterson. *Medical Microbiology. Chapter 7. Bacterial Pathogenesis.* 4th ed. 1996. ISBN: 0963117211.