Genomic Insights into the Pathogenic Evolution of *E. coli* Strain X during the 2011 Outbreak

M. Uzun[a,c], A. Fedorenko[b,c]

[a]*Research Center Biotechnology RAS, Moscow, Russia*
[b]*Skolkovo Institute of Science and Technology, Moscow, Russia*
[c]*Bioinformatics Institute, Saint Petersburg, Russia*

Abstract

Hemolytic Uremic Syndrome (HUS), characterized by bloody diarrhea, is often caused by pathogenic *Escherichia coli* producing Shiga toxins. This study investigates a 2011 *E. coli* outbreak in Germany, focusing on the novel pathogenic strain, *E. Coli* X, isolated from a Hamburg patient (TY2482). Through *de novo* genome assembly, we identified Shiga toxin and antibiotic resistance genes acquired by *E. Coli* through horizontal gene transfer facilitated by bacteriophages, leading to the pathogenic transformation of *E. coli* and rendering it resistant to antibiotic treatment.

Keywords: *Escherichia coli*, Hemolytic Uremic Syndrome, Shiga toxins, Genome assembly, Antibiotic Resistance

1. Introduction

Hemolytic Uremic Syndrome (HUS) - is a severe condition characterized by bloody diarrhea and can lead to kidney failure [1]. It is caused by pathogenic strains of *Escherichia coli* that produce Shiga toxins [2]. The genes encoding Shiga toxins and other pathogenic factors can be transmitted among *E. coli* strains through horizontal gene transfer (HGT) facilitated by bacteriophages [3]. This poses significant challenges for society, as non-pathogenic strains can acquire pathogenicity, leading to outbreaks, as witnessed, for instance, in Germany in 2011 [4].

The treatment of diseases caused by pathogenic strains may be complicated by the acquisition of new genes conferring resistance to beta-lactam antibiotics in their plasmids [5]. Plasmid genes with such characteristics might be missed if reads are aligned to a reference genome without plasmid information, leading to their omission and loss [6]. Therefore, in such cases, it is crucial to assemble genomes *de novo*.

In this study, we performed *de novo* assembly of the genome of a new strain of *E. coli* isolated from the TY2482 sample. Through genome analysis, we identified the presence of Shiga toxin genes and genes conferring resistance to beta-lactam antibiotics. We also determined that these genes were acquired through horizontal transfer, emphasizing the importance of molecular mechanisms contributing to the evolution and pathogenicity of these strains.

2. Materials and Methods

Samples of the sequencing data belong to the pathogenic isolate *E.Coli* X from the girl in Hamburg, called sample TY2482.

Basic statistics of raw reads was acquired using seqkit v.2.5.1 with the "stats" flag [7]. Raw reads were quality-checked with FastQC v.0.12.1 [8]. K-mer profile and genome size estimation were performed using jellyfish v.2.2.10 [9]. Histograms for kmers count were visualized in GenomeScope [10]. Raw mate-pair (SRR292862, SRR292770) and paired-end reads (SRR292678) were *de novo* assembled with SPAdes v3.15.2 [11]. The quality metrics were assessed using QUAST v5.0.2 [12]. The assembled genome was annotated using Prokka v.1.14.6 [13]. 16S rRNA genes were detected using Barrnap v.0.9 [14]. Reference and assembled genomes were aligned with ProggressiveMauve and visualized with mauve v.20150226 build 10 [15]. Genes responsible for antibiotic resistance were detected with ResFinder [16–18].

3. Results

The downloaded sequencing data comprised paired-end and mate-pair Illumina reads (Table 1). These reads were of good quality.

Table 1. Description of the reads in the studied samples and the number of reads

Sample SRR number	**Type of library**	**Insert size**	**Number of sequences (bp)**
SRR292678	paired-end	470 bp	5499346
SRR292862	mate-pair	2 kb	5102041
SRR292770	mate pair	6 kb	5102041

Two assemblies were generated for a single library of sequencing (paired-end) reads and for the three libraries (paired-ends and mate-pairs). The first one had 210 contigs with a total length of 5.29 Mb, N50 equal to 111860, and GC content of 50.56%. The second assembly had 105 contigs with a length of 5.35 Mb, N50 of 335515, and GC content of 50.59%. The second assembly was annotated for gene prediction. There

were found 80 tRNAs, 1 CRISPR region, 5064 CDS, and 2923 unique gene codes. rRNA genes were not found. Hence, Prokka annotation does not detect these genes by default; they were found using Barrnap. Detected 16S rRNA genes in the assembled *E.Coli* X strain are presented in Table 2.

Table 2. Detected 16S rRNAs in the assembled genome E.Coli X.

Position	Start	End	E-value	Strand	Length (bp)
PROKKA_01	326359	327896	0	-	1537
PROKKA_01	595966	597503	0	-	1537
PROKKA_01	2504403	2505940	0	-	1537
PROKKA_05	43835	45372	0	+	1537
PROKKA_05	85462	86999	0	+	1537
PROKKA_06	111955	113492	0	+	1537
PROKKA_71	314	719	9.80E-23	+	405

Blast of obtained 16S rRNA genes against RefSeq NCBI database allowed us to find the closest genome (NCBI Reference Sequence: NC_011748.1) set as a reference.

Assembled genome sequence was aligned to the reference genome. Two Shiga toxin genes, absent in the reference genome, were found in the insertion site of genome *E.Coli* X (Table 3).

Table 3. The Shiga toxin genes in the assembled genome *E.Coli* X.

Name	Length	Start position	End position
stxB *(Shiga toxin subunit B)*	269 bp	3483605	3483874
stxA *(Shiga toxin subunit A)*	959 bp	3483886	3484845

4. Discussion

In the assembled genome of *E. coli* X, Shiga toxin genes were identified, and an analysis of their genomic environment was conducted. It was observed that these genes were located within an insertion region of approximately 60 Mb in size. This region was flanked by an integrase on one side and a hypothetical protein of the *Escherichia* phage on the other. Moreover, it contained a substantial number of phage genes within its structure. The gene composition within this region strongly suggests its acquisition through HGT facilitated by a bacteriophage [3]. Consequently, this event led to the pathogenic transformation of *E. coli* X.

In addition, several new genes for certain antibiotic resistance were found in *E. coli* X plasmid, to which the reference strain is vulnerable. For example, genes of beta-lactamase CTX-M-1 (*bla_1*) and beta-lactamase TEM precursor (*bla_2*), which can hydrolyze drug molecules [19], were found in *E. coli* X plasmid. In the neighboring of these genes, phage integrases and mobile element genes were found, which means these genes were also acquired through HGT [20]. The suggested alternative treatment for affected patients is to use antibiotics to which this *E. coli* strain does not have resistance.

References

[1] George, J. N. & Nester, C. M. Syndromes of Thrombotic Microangiopathy. *N. Engl. J. Med.* **371**, 654–666 (2014).

[2] Joseph, A., Cointe, A., Kurkdjian, P. M., Rafat, C. & Hertig, A. Shiga Toxin-Associated Hemolytic Uremic Syndrome: A Narrative Review. *Toxins (Basel).* **12**, 1–46 (2020).

[3] Mir, R. A. & Kudva, I. T. Antibiotic-resistant Shiga toxin-producing *Escherichia coli*: An overview of prevalence and intervention strategies. *Zoonoses Public Health* **66**, 1–13 (2019).

[4] Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, Bernard H, F. A. & Prager R, Spode A, Wadl M, Zoufaly A, Jordan S, Kemper MJ, Follin P, Müller L, King LA, Rosner B, Buchholz U, Stark K, K. G. H. I. T. Epidemic Profile of Shiga-Toxin–Producing. *N. Engl. J. Med.* **365**, 1771–1780 (2011).

[5] Matinfar, S., Ahmadi, M., Sisakht, A. M., Sadeghi, J. & Javedansirat, S. Phylogenetic and antibiotics resistance in extended-spectrum B-lactamase (ESBL) Uropathogenic *Escherichia coli*: An update review. *Gene Reports* **23**, 101168 (2021).

[6] Liao, X. et al. Current challenges and solutions of de novo assembly. *Quant. Biol.* **7**, 90–109 (2019).

[7] Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**, 1–10 (2016).

[8] FastQC. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

[9] Jellyfish. *https://github.com/gmarcais/Jellyfish*.

[10] Bankevich, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

[11] Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

[12] Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

[13] Barrnap. *https://github.com/tseemann/barrnap*.

[14] Mauve. *https://darlinglab.org/mauve/download.html*.

[15] Bortolaia, V. et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**, 3491–3500 (2020).

[16] Zankari, E. et al. PointFinder: A novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J. Antimicrob. Chemother.* **72**, 2764–2768 (2017).

[17] Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9 (2009).

[18] Palzkill, T. Structural and mechanistic basis for extended-spectrum drug-resistance mutations in altering the specificity of TEM, CTX-M, and KPC -lactamases. *Front. Mol. Biosci.* **5**, 1–19 (2018).

[19] Bevan, E. R., Jones, A. M. & Hawkey, P. M. Global epidemiology of CTX-M -lactamases: Temporal and geographical shifts in genotype. *J. Antimicrob. Chemother.* **72**, 2145–2155 (2017).