E.coli outbreak investigation

Authors: Anna Rybina, Daria Nikanorova

Abstract

In this study, we analyzed a novel *Escherichia coli* X strain that caused severe outbreak in Germany, 2011, characterized by hemolytic-uremic syndrome. We performed *de novo* assembly of *E.coli* X combining paired end and mate pair libraries to obtain greater accuracy of the final consensus sequence. The closest relative to *E. coli* X occured to be enteroaggregative strain *E. coli* 55989 as their 16S rRNA genes shared 100 % similarity. However genome-wide comparative analysis demonstrated that *E. coli* X, opposed to its closest relative, possessed Shiga-toxin related genes (*stxB* and *stxS*) specific to enterohemorrhagic *E. coli* (EHEC) strains and had broader set of antibiotic resistance genes, including β-lactamase gene *(bla)*, dihydropteroate synthase (*sul*) and dihydropteroate reductase (*aph*) genes. Investigating gene neighborhood, we came to the conclusion that *E. coli* X could be enteroaggregative strain which had gained Shiga-toxin genes and antibiotic resistance genes such as *bla* genes via horizontal gene transfer with bacteriophage and Tn3 transposon genetic element, respectively. Acquisition of these genes increased virulence and survival ability of the *E. coli* X strain. Applying methods of *de novo* assembly and comparative genomics might help to conquer bacterial diseases.

Introduction

In 2011, a novel strain *Escherichia coli* X triggered a severe outbreak of foodborne illness mainly located in Germany. Main symptoms of the disease included bloody diarrhea followed by hemolytic-uremic syndrome (HUS). HUS is a condition characterized by low red blood cells, acute kidney failure, and low platelets (Cody and Dixon 2019)
Increased pathogenicity of the *E. coli* X strain can be the result of acquiring a new toxin or a new gene responsible for antibiotic resistance. Such virulence genes are frequently obtained by bacteria from phages, other bacteria, or some mobile genetic elements via horizontal gene transfer (HGT) (Gyles and Boerlin 2014). For example, horizontal gene transfer via transduction, when Shiga toxin-encoding phage integrates into nontoxigenic genomes turning them pathogenic, plays a significant role in the evolution of *Shigella* (Strauch, Lurz, and Beutin 2001).
De novo assembly of the genome allows us to reveal important large-scale variations absent in the closely related organisms. Revealing the difference in gene content between candidate novel strain and highly similar one might shed light on increased virulence of the *E. coli X.*
In this work, we aimed to clarify the genetic basis of severe pathogenicity of a novel strain *E. coli* X. We performed de novo assembly using three sequencing libraries to achieve better accuracy and continuity and compared the draft genome with its closest relative of complete assembly to reveal new virulence genetic factors and suggest possible origin of these acquired genes.

Methods

Genome of *E. coli* strain X was sequenced and three libraries of reads were obtained: SRR292678 (pair-end reads), SRR292862 (mate-pair reads) and SRR292770 (mate-pair reads) (**Table 1**). Quality of reads in each library was checked with FastQC. *E. coli X* genome was assembled de novo from pair-end library and both pair-end and mate-pair libraries using SPAdes v3.13.1 (Bankevich et al. 2012). Assembly quality was checked with QUAST v5.1.0rc1 (**see Supplementary materials, section 3, Fig. S2**) (Gurevich et al. 2013). K-mer profile and genome size were estimated for initial reads and reads corrected during assembling by jellyfish (**see Supplementary materials, section 2**, **Fig. S1a**). As far as the quality of the second assembly (with mate-pair reads) was higher, we used it for further annotation. Genome was annotated using Prokka 1.12 (Seemann 2014). In order to find the closest relative to *E. coli* X we obtained a highly evolutionarily conserved sequence for 16S rRNA and ran BLAST 2.9.0+ (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (**see Supplementary materials, section 6, Fig. S3**). ResFinder was used to find genes responsible for antibiotic resistance (https://cge.cbs.dtu.dk/services/ResFinder/). In order to detect possible insertions of shiga-toxin and antibiotic resistance genes we used a software Mauve 2.3.1 (http://darlinglab.org/mauve/user-guide/introduction.html). We also inspected surrounding genes to uncover a possible source of these genes. Protein sequences encoded by neighboring genes were subjected to blastp search to identify their possible functions and origin.

Results

De novo assembly

In this study, we obtained two *de novo* assemblies of the studied *E. coli* strain X processing three sequencing libraries **(Table 1**).

Table 1. Description of libraries analyzed in the study

Sample	Orientation	Insert size	Read length	Total size
SRR292678	paired end	470 bp	90 bp	5499346 x 2
SRR292862	mate pair	2 kb	49 bp	5102041 x 2
SRR292770	mate pair	6 kb	49 bp	5102041 x 2

Compared to single library assembly (using SRR292678 sample), three libraries assembly resulted in a higher N50 parameter, lower total number of contigs and increased number of contigs with length >= 50 000 bp (**Fig. S2, Table 2**). As the three libraries assembly showed better statistics, we chose it for further analysis.

Table 2. Statistics for single library and three library assemblies obtained using QUAST.

Type of assembly	N50 (contigs)	N50 (scaffolds)	# contigs
single library	111860	111860	210
three libraries	335515	2815616	105

***Effect of read correction on K-mer distribution**

K-mer distribution of initial reads (without error correction) was calculated for pair-end library SRR292678. 46,189 reads occur 125 times each in all reads (peak on **Fig. 1a**). After read correction peak value decreases a bit (the highest frequency for the greatest part of reads (45,665) equals 124) (**Fig. 1b**). But the main difference lies in the first peak (**Fig. S1**). This first peak primarily due to rare sequencing errors in reads. Actually, the first peak is much higher than the second one, so we trimmed it on a plot to see clearly the second peak (it was done for **Fig. 1**). Comparing k-mer frequencies for uncorrected and corrected reads we can notice that there are fewer k-mers with low frequencies (< 10) in corrected reads (**Table 3**). For example, 13,894,353 of k-mers of length 31 occur only once in uncorrected reads, while in corrected reads the number of k-mers with this frequency equals 905,758 (it explains different heights of the first peak for corrected and uncorrected reads on **Fig. S1**). Consequently, the number of first rows in k-mer frequency data that should be skipped to trim the first peak differs for uncorrected and corrected reads (red rows in **Table 3**). It means that correction reduces a number of k-mers with errors by at least 13 times.

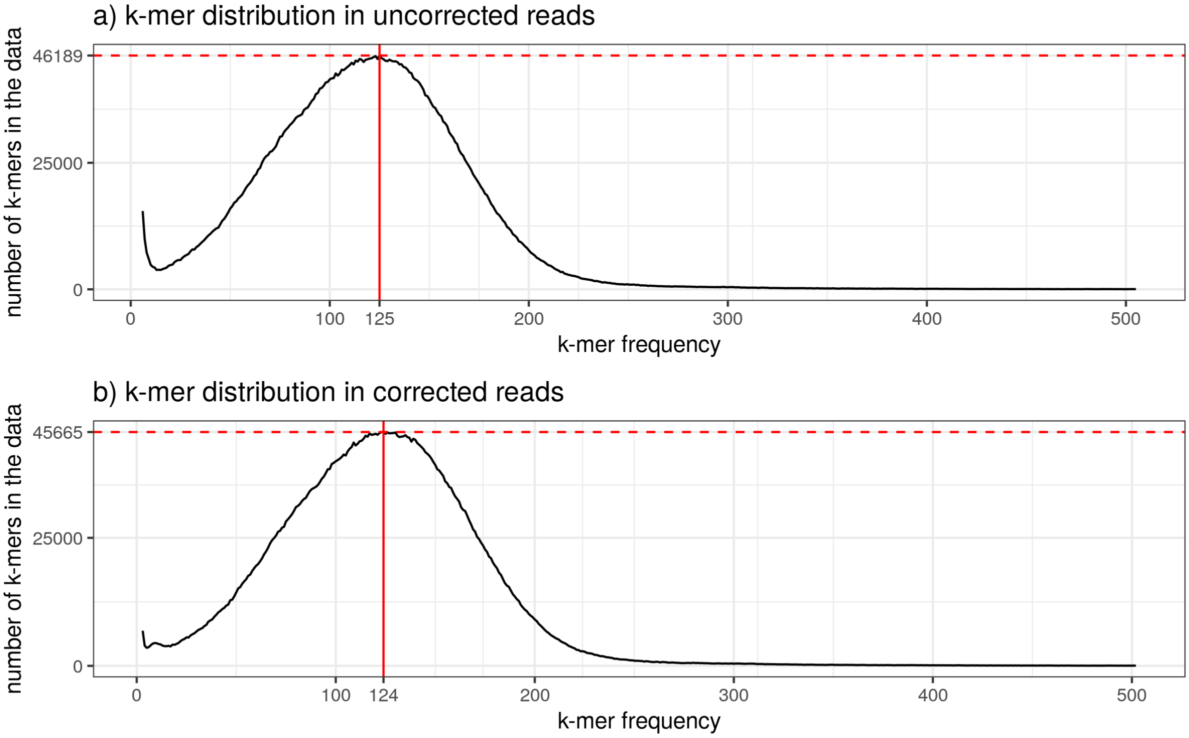

Figure 1. K-mer distribution for the pair-end library (SRR292678). *a - uncorrected reads: peak = 125, genome size = 5146818 ~ 5,15 Mb, number of k-mers = 643,352,262 ; b - reads corrected by SPAdes: peak = 124, genome size = 5252481 ~ 5,25 Mb, number of k-mers = 658,771,469*

Table 3. K-mer frequencies for uncorrected and corrected by SPAdes reads (SRR292678)

frequency	number of k-mers	
	uncorrected reads	corrected reads
1	13894353	905758
2	822668	26076
3	154910	6855
4	53078	3952
5	26464	3524
6	15473	3651
7	9911	4001
8	7249	4339
9	6073	4456
10	4854	4424

Annotation

We annotated the three libraries assembly using prokka. Among 342 contigs we discovered 5316 genes, 7 ribosomal RNAs (3 full and 1 partial 16S rRNAs, 2 of 5S rRNAs, and single 23S rRNA), 5032 CDS, 218 miscellaneous small RNAs, 58 transfer RNAs, and 1 transfer-messenger RNA (tmRNA).

Identification of the closest relative

We found the closest relative of the studied *E . coli* strain by the resemblance of their 16 S rRNAs. First, we predicted 16 S rRNA (**Table S1**) and then searched for its closest homologs running BLAST nucleotide search against RefSeq Genomes Database. According to the blastn result, we obtained *E. coli* 55989 (RefSeq accession: NC_011748.1, GenBank accession of identical sequence: CU928145.2) as the best hit with the identity of 100% and query coverage of 100% suggesting its genome is the closest relative to studied E. coli X and therefore can be used as a reference genome.

Toxin gene detection

To reveal the genetic basis of HUS, we compared the genome of the studied strain with the reference *E. coli* 55989. One of the common *E. coli* toxins that cause internal bleeding are Shiga toxins.

We identified two genes: *stxB* and *stxA* encoding Shiga toxin subunit B and A, respectively (**Table 4, Fig. 2**). Analysis of their gene neighborhood showed that *strxB* and *strxA* are located between phage-associated genes (**Supplementary materials, section 8, and Fig. S3**).

Table 4. Identified Shiga-toxin genes in the *Escherichia coli X.*

Shiga toxin gene	Product	Location (start, end)	Length
stxB	Shiga toxin subunit B	596, 865	270
stxA	Shiga toxin subunit A	877, 1836	960

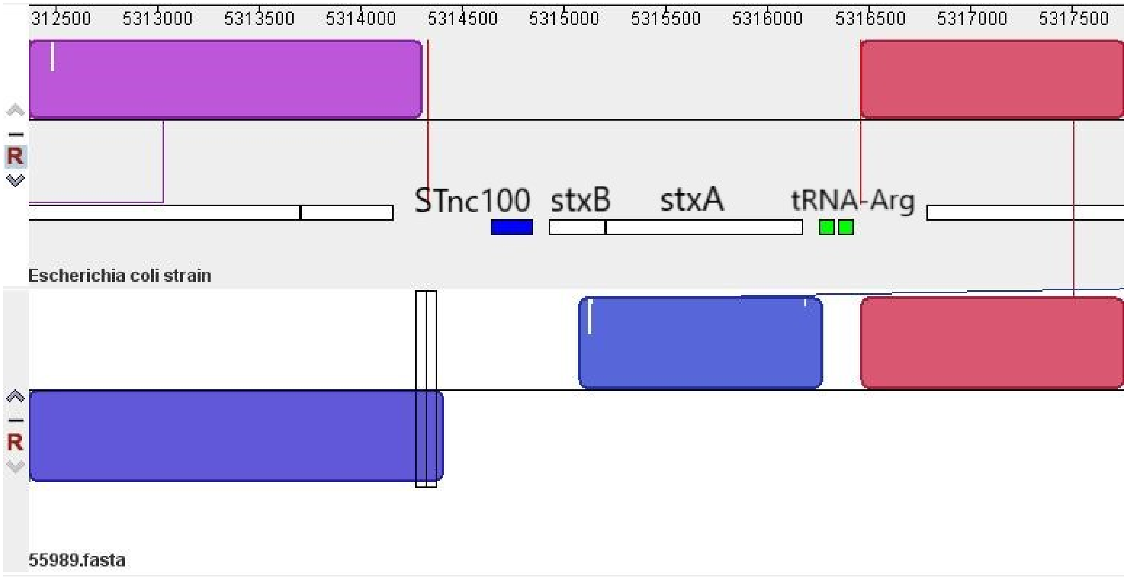

Figure 2. Shiga-toxin related genes of *E.coli X* on the genome-wide alignment with the reference *E. coli* 55989. Prokka annotation was used to obtain the alignment. Visualisation was made in Mauve (http://darlinglab.org/mauve/user-guide/introduction.html).

Antibiotic resistance detection

We identified genes responsible for antibiotic resistance in the studied *E. coli* and the reference genomes. According to ResFinder results, both strains are resistant to tetracycline due to common among *E. coli* strains *tetB* gene encoding tetracycline efflux MFS transporter. Another gene likely shared by both strains is *mdfA* that encodes multidrug transporter and could ensure insensitivity to a broad spectrum of antibiotics including macrolides. We found several genes providing *E. coli* X with resistance to different β-lactams, folate pathway antagonists and aminoglycoside antibiotics which the reference *E. coli* 55989 is vulnerable to (**Table 5**): *bla* (*blaCTX-M-15, blaTEM-1B), sul (sul1* and *sul2), dfrA7*, and *aph* (*aph(3'')-Ib* and *aph(6)-Id*) genes.

Table 5. Genes responsible for antibiotic resistance which are specific to *E. coli* X and absent in the *E. coli* 55989.

Antimicrobial agent	Class	WGS-predicted phenotype	Genetic background	Product
ampicillin	β-lactam	Resistant	*blaCTX-M-15, blaTEM-1B*	*β*-lactamase
cefotaxime	β-lactam	Resistant	*blaCTX-M-15*	*β*-lactamase
ceftazidime	β-lactam	Resistant	*blaCTX-M-15*	*β*-lactamase
cefepime	β-lactam	Resistant	*blaCTX-M-15*	*β*-lactamase
sulfamethoxazole	folate pathway antagonist	Resistant	*sul1, sul2*	Dihydropteroate synthase (type-1 and type-2, respectively)
trimethoprim	folate pathway antagonist	Resistant	*dfrA7*	Dihydrofolate reductase
streptomycin	aminoglyco side	Resistant	*aph(3'')-Ib* *aph(6)-Id*	Aminoglycoside 3'-phosphotransferase, Aminoglycoside O-phosphotransferase

To clarify how *E. coli* X could obtain *β*-lactamase genes, we investigated possible function and origin of hypothetical proteins whose genes are in close proximity to *bla* genes. Based on protein homologs search via blastp, we may assume that these neighbouring genes are related to mobile genetic elements (MGE): four of them shared about 100 % of similarity with some transposase genes (mainly, Tn3 family transposase) from other Enterobacteriaceae (**Supplementary materials, section 8**). Gene at position 84687..85244 according to prokka annotation might encode site-specific DNA recombinase and according to blastp result encode resolvase.

Discussion

In this work, we investigated *E. coli* strain X caused HUV outbreak in Germany, 2011, to reveal its pathogenetic features. First, we obtained *de novo* assembly with the use of three Illumina libraries including paired-end and mate-pair reads. Using mate pairs could reduce assembly fragmentation via joining multiple contigs into a larger one, increasing the assembly contiguity. Mate pair libraries of large insert size better cover highly repetitive regions and resolve misassemblies related to genomic rearrangements. Gaps between mate pairs can be filled by paired-end reads with short insert size. Combining paired end and mate pair libraries led to greater accuracy of the final consensus sequence.

We identified the *E. coli* strain 55989 as the closest relative of *E. coli* X based on the 16S rRNA gene similarity search and used as a reference genome in our study. Mossoro et al. (Mossoro et al. 2002) showed that *E. coli* 55989 belongs to enteroaggregative *E. coli*

(EAEC) strain possessing aggregative adherence fimbriae (AAF) genes located on the pAA plasmid. According to their work, *E. coli* strain 55989 did not result in blood disorders like HUS. Studied *E. coli X* could have gained the ability to cause HUS as opposed to the reference strain due to acquiring a new virulence factor or a new gene responsible for antibiotic resistance.
It is known that Shiga toxins, common among some pathogenic *E. coli* (for example, enterohemorrhagic *E. coli* strains), are associated with blood disorders. Taking it into account, we suggested that Shiga toxin genes should be present in *E. coli X* and absent in *E. coli* 55989 allowing the first to cause HUS unlike the last. Our results confirmed the presence of Shiga-toxin related genes *stxB* and *stxA* in the *E coli* X genome. Next to them, gene coding for STnc100 was found. STnc100 belongs to the Rfam group RF02076 consisting of bacterial small RNAs and, as some sRNA do, might regulate expression of virulence factors such as genes *sxtB* and *sxtA*. Moreover, both genes *stxB* and *stxA* are located between phage-associated ones (**Fig. S3**), implementing that *E. coli* X had acquired the ability to produce Shiga toxin from bacteriophage via HGT.

In terms of virulence, another important property specific to *E . coli* X is resistance to a broader set of antibiotics compared to *E. coli* 55989. For example, *E. coli* X possesses *bla* gene encoding *β*-lactamase (**Table 4**) that inactivates β-lactam drugs by hydrolyzing a specific site in the β-lactam ring (Reygaert 2018). Dihydropteroate synthase (encoded by *sul* genes) and dihydropteroate reductase (encoded by *dfrA7*) could bind and modify antimicrobial agents due to structural similarity of respective drugs to their natural substrates (Huovinen et al. 1995). Kinases encoded by *aph* genes phosphorylate aminoglycosides disrupting mechanism of drug action (Kotra, Haddad, and Mobashery 2000). Antibiotic resistance might have helped the bacteria to survive and persist in the environment.

According to our results, *β*-lactamase are surrounded by genes that share about 100 % of identity with genes related to transposon Tn3: either transposase or resolvase encoding. It was reported earlier that *bla* genes could be transferred with mobile genetic elements ((Partridge and Hall 2005), (Kapoor, Saigal, and Elongavan 2017) . These findings point towards the idea that *E. coli* acquired β-lactamase genes via HGT from the Tn3 transposon.

Overall, there are several mechanisms of antibiotics resistance covered by *E. coli* X: drug inactivation (*bla* genes), drug modification based on competitive binding to the enzyme (*sul* genes and *dfrA7*) or based on its phosphorylation (*aph* genes). Various spectrum of genes responsible for antibiotic resistance in *E coli* X made the treatment by antimicrobial agents more complicated and required more careful antibiotics administration than in case of *E. coli* 55989. Nevertheless, there are several antimicrobial drugs that could be used to fight against *E. coli* X infections as respective antibiotic resistance genes were not found in its genome. Among them are fosfomycin (class fosfomycin, inhibits bacterial cell wall synthesis by inactivating MurA enzyme (Brown et al. 1995)), ciprofloxacin (class fluoroquinolone, inhibits nucleic acid synthesis inactivating DNA gyrase (Andersson and MacGowan 2003)), and carbapenems (class β-lactams, binds to penicillin-binding proteins). The last one is commonly used to treat different multidrug-resistant bacterial infections (Vardakas et al. 2012).

References

Andersson, Monique I., and Alasdair P. MacGowan. 2003. “Development of the Quinolones.” *The Journal of Antimicrobial Chemotherapy* 51 Suppl 1 (May): 1–11.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 19 (5): 455–77.

Brown, E. D., E. I. Vivas, C. T. Walsh, and R. Kolter. 1995. “MurA (MurZ), the Enzyme That Catalyzes the First Committed Step in Peptidoglycan Biosynthesis, Is Essential in Escherichia Coli.” *Journal of Bacteriology* 177 (14): 4194–97.

Cody, Ellen M., and Bradley P. Dixon. 2019. “Hemolytic Uremic Syndrome.” *Pediatric Clinics of North America* 66 (1): 235–46.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. “QUAST: Quality Assessment Tool for Genome Assemblies.” *Bioinformatics* 29 (8): 1072–75.

Gyles, C., and P. Boerlin. 2014. “Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease.” *Veterinary Pathology* 51 (2): 328–40.

Huovinen, P., L. Sundström, G. Swedberg, and O. Sköld. 1995. “Trimethoprim and Sulfonamide Resistance.” *Antimicrobial Agents and Chemotherapy* 39 (2): 279–89.

Kapoor, Garima, Saurabh Saigal, and Ashok Elongavan. 2017. “Action and Resistance Mechanisms of Antibiotics: A Guide for Clinicians.” *Journal of Anaesthesiology, Clinical Pharmacology* 33 (3): 300–305.

Kotra, L. P., J. Haddad, and S. Mobashery. 2000. “Aminoglycosides: Perspectives on Mechanisms of Action and Resistance and Strategies to Counter Resistance.” *Antimicrobial Agents and Chemotherapy* 44 (12): 3249–56.

Mossoro, Christian, Philippe Glaziou, Simon Yassibanda, Nguyen Thi Phuong Lan, Claudine Bekondi, Pierre Minssart, Christine Bernier, Chantal Le Bouguénec, and Yves Germani. 2002. “Chronic Diarrhea, Hemorrhagic Colitis, and Hemolytic-Uremic Syndrome Associated with HEp-2 Adherent Escherichia Coli in Adults Infected with Human Immunodeficiency Virus in Bangui, Central African Republic.” *Journal of Clinical Microbiology* 40 (8): 3086–88.

Partridge, Sally R., and Ruth M. Hall. 2005. “Evolution of Transposons Containing blaTEM Genes.” *Antimicrobial Agents and Chemotherapy* 49 (3): 1267–68.

Reygaert, Wanda C. 2018. “An Overview of the Antimicrobial Resistance Mechanisms of Bacteria.” *AIMS Microbiology* 4 (3): 482–501.

Seemann, Torsten. 2014. “Prokka: Rapid Prokaryotic Genome Annotation.” *Bioinformatics* 30 (14): 2068–69.

Strauch, E., R. Lurz, and L. Beutin. 2001. “Characterization of a Shiga Toxin-Encoding Temperate Bacteriophage of Shigella Sonnei.” *Infection and Immunity* 69 (12): 7588–95.

Vardakas, Konstantinos Z., Giannoula S. Tansarli, Petros I. Rafailidis, and Matthew E. Falagas. 2012. “Carbapenems versus Alternative Antibiotics for the Treatment of Bacteraemia due to Enterobacteriaceae Producing Extended-Spectrum β-Lactamases: A Systematic Review and Meta-Analysis.” *The Journal of Antimicrobial Chemotherapy* 67 (12): 2793–2803.