Virulence factors and antibiotic resistance genes of STEC, caused Germany 2011 HUS outbreak

Elizaveta Demina[1,2*†] and Bogdan Sotnikov[2,3*†]

[1] Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russian Federation.
[2] Bioinformatics institute, Kantemirovskaya street 2A, Saint Petersburg, 197342, Russia.
[3] Medical faculty, Kyrgyz-Russian Slavic university, Kievskaya street 44, Bishkek, 720000, Kyrgyzstan.

*Corresponding author(s). E-mail(s): demina.eiu@gmail.com; bogdan.sotnikov.1999@mail.ru;
†These authors contributed equally to this work.

Abstract

In this investigation we tried to elaborate, which virulence factors and antibiotic resistance genes caused hemolytic urinary syndrome outbreak in Germany in 2011 year. The aetiological agent of outbreak was new strain of *E. coli*. We assembled genome of this strain and found, that genes, producing Shiga toxin are answerable for most severe symptoms of this disease. Also we found, this strain of *E. coli* has resistance to β-lactams, tetrecyclines and trimethoprim. The main ways, this bacteria get pathological properties are horisontal transfer, including plasmids conjugation, transposons transfer and bacteriophage transduction. We recommended to use carbapenems to treat patients, compromised by this strain of *E. coli*

Keywords: hemolytic uremic syndrome, Shiga toxin-producing E. coli, horizontal transfer, Germany 2011 outbreak

Table 1 Data about libraries, used for *E.coli* genome assembling

Library name	Library type	Number of reads	Insertion size
SRR292678	Paired-end	5499346 (for each - forward and reverse)	470 bp
SRR292678	Mate pair	5102041 (for each - forward and reverse)	2 kb
SRR292770	Mate pair	5102041 (for each - forward and reverse)	6 bp

1 Introduction

Escherichia coli (*E. coli*) is a gram-negative bacteria, living in the humans intestine. It is usually opportunistic: typical strains don't cause disease. But some of the strains have some virulence properties, which may lead to different gastrointestinal, urinary, central nervous and generalized diseases. There are usual classified five types of pathogenic *E. coli*: enteropathogenic *E. coli* (EPEC), Shiga toxin-producing *E. coli* (STEC) (e.g., enterohemorrhagic *E. coli* [EHEC]), Shigella/enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC), and enterotoxigenic *E. coli* (ETEC) [1][2]. All of them may cause severe diseases, sometimes fatal.

One of the ways, innocent bacteria can change their genome is horizontal transfer [3]. There are three ways to realise this process: conjugation, transduction and transformation. The conjunction is a plasmid exchange via special bacterial organelles - piles. The transduction is the result of bacteriophage-bacteria interaction. The transformation demands of naked DNA in the extracellular space [4]. Some plasmids or phags may contain virulence and antibiotic resistance genes.

In cases of such genome changes, we can't use simple alignment to reference, because elements from plasmids aren't situated within chromosome sequence. Such cases need another approach, based on the *de novo* genome assembly.

In our case, we need to find out the cause of the *E. coli* virulence, which led to an outbreak of hemolytic uremic syndrome (HUS). We designate this strain below as *E. coli X*. Also, we tried to find, how *E. coli X* acquired its virulence factors, and which treatment strategy may be most effective and safety.

2 Methods

In our investigation, we used three libraries from the *E. coli X* for assembling: one with paired-ends and two mate pairs. Reads were generated at Beijing Genome Institute and deposited into the Short Read Archive. The brief information about three libraries from the TY2482 sample presented in table 1.

We used fastqc (v0.11.9) for checking the quality of the reads.

To assemble the genome of *E. coli X* we used SPAdes assembler (v3.15.5) [5] twice: the first way was with only paired-end reads and the second one with all three libraries. We used isolated mode for both assembles. For checking the

Table 2 Information about aseemblies quality

Assembly	One-library assemble (contigs)	One-library assemble (scaffolds)	Three-library assemble (contigs)	Three-library assemble (scaffolds)
contigs (>= 0 bp)	885	868	1154	1070
contigs (>= 1000 bp)	127	123	126	56
contigs (>= 5000 bp)	86	83	66	13
contigs (>= 10000 bp)	73	70	54	12
contigs (>= 25000 bp)	54	53	44	11
contigs (>= 50000 bp)	32	32	28	11
Total length (>= 0 bp)	5363438	5365096	5402541	5531597
Total length (>= 1000 bp)	5227528	5228807	5221609	5351356
Total length (>= 5000 bp)	5112460	5115861	5090134	5266301
Total length (>= 10000 bp)	5024868	5026097	5011459	5259820
Total length (>= 25000 bp)	4704658	4751300	4846400	5248736
Total length (>= 50000 bp)	3898474	3989113	4280675	5248736
contigs	162	168	169	108
Largest contig	300784	300784	391919	1046981
Total length	5252199	5260240	5251556	5386995
GC (%)	50.54	50.52	50.54	50.52
N50	114227	114589	157905	741159
N90	24915	25454	29147	426441
auN	126907.1	130855.3	177788.0	719107.4
L50	15	14	11	3
L90	55	53	40	7
N's per 100 kbp	0.00	32.47	0.00	2395.70

quality of the assembly we used QUAST (v. 2.31) [6] on assembles scaffolds and contigs. Findings from QUAST are defined in the results section.

To annotate the assembled genome we used Prokka (v. 1.14.6) [7]. We annotated only a three-libraries genome because it had better assembling statistics.

For extracting 16S rRNA clasters' sequences we used the barrnap tool (v. 0.9) [8]. We also used barrnap only for the second assembly. The bash command was the next:

```
barrnap -o rrna.fa < scaffolds.fasta > rrna.gff
```

Founded sequences were elaborated in BLASTn [9] with params "Entrez Query":1900/01/01:2011/01/01[PDAT]. The most relative sequence of *E. coli* was used in the next stages of investigation as a reference.

We elaborated difference between the reference strain and *E. coli X* using Mauve (snapshot_2015-02-13) [10]. For finding genes of antibiotic resistance we used ResFinder with scaffolds.fasta file from three-libraries assemble.

3 Results

We had completely normal results of fastqc quality checking with the absent of any abnormalities. Results of fastqc output presented in supplementary.

We assembled the genome of *E. coli X* from one library (paired-end reads) and all three libraries together. The results of the assembly quality analysis are presented in table 3.

As we can see, the biggest N50 and the smallest L50 are both in the three-libraries assembly scaffolds. It means the quality of the assembly was increased after adding two mate-pair libraries. The cause of this event is solving duplications in the genome, which is larger than 470 base pairs (mean insert size of a library with paired-end reads).

Table 3 Attributes of assembled genome annotation

organism	*E. coli* X
contigs	1070
bases	5531597
CDS	5012
rRNA	10
repeat_region	1
tRNA	65
tmRNA	1

Table 4 Brief data about 16s rRNA of *E. coli X*

	Coordinates	Strand
NODE_3	353640-355178	+
NODE_2	526836-528374	-
NODE_76	1-407	-

Table 5 Genes and their products, which present in *E. coli X* but not in reference genome

Gene	Coordinates	Product	Uniprot code
stxA	5308246-5309205	Shiga toxin	Q9FBI2
stxB	5307965-5308234	Shiga toxin	Q7BQ98
hoKe_3	5441483-5441635	Toxic protein HokE	P77091
tetA	5281235-5282434	Tetracycline resistance protein, class C	P02981
bla_1	5108214-5109083	TEM	P62593
bla_2	5111970-5112845	CTX-M1	P28585
dhfrI	4957362-5049029	dihydrofolate reductase type 1	P00382

For further stages of investigation, we annotated the three-libraries assembled genome, using the tool prokka. A brief information about annotation presented in table 4.

We searched the genome for 16S rRNA for further using blast to find more relative organisms. Data about founded 16S rRNA is in table 5.

After using BLAST for finding resembled sequences we identified that bacteria, which has the most similarity with *E coli X* is *E. coli 55989* (NCBI id is NCBI:txid585055). This strain hadn't demonstrated any hemorrhagic symptoms. For finding a difference between the two strains we used mauve. As a result of progressive alignment with using annotation, which had been made by prokka. We found genes, which determine antibiotic-resistance and producing of Shiga toxin. Some significant virulence genes are listed in table 6.

We also checked assembled genome on the ResFinder for identification spectrum of antibiotic resistance. Investigated strain is resistant to the next antibiotics and antiseptics: ampicillin, sulfamethoxazole, trimethoprim, cephalothin piperacillin, tetracycline, amoxicillin, ethidium bromide, cefepime, doxycycline, benzylkonium chloride, ceftriaxone, ceftazidime, ticarcillin, chlorhexidine, cetylpyridinium chloride, cefotaxime, aztreonam and streptomycin.

4 Discussion

4.1 Founded virulence factors genes of *E. coli X*

Shiga toxin (Stx) is a bacterial toxin, founded in *Shigella dysenteriae 1* and in STEC, the most serious manifestation of the disease, the hemolytic uremic syndrome or HUS. The Stxs consist of two major subunits, an A subunit that joins noncovalently to a pentamer of five identical B subunits. The A subunit of the toxin injures the eukaryotic ribosome and halts protein synthesis. The function of the B pentamer is to bind to the cellular receptor, globotriaosylceramide, Gb3, found primarily on endothelial cells [11].

We conclude that presence of Stx led to HUS in affected patients, and this is the main cause of *E. coli X* virulence.

StxA and StxB both locate in small insertions between two synteny blocks. It may give a nudge, that possibly, they came from bacteriophage transfusion.

HokE is an isoform of member Hok/Gef toxin-antitoxin family [12]. We didn't find any evidence that hokE increases virulence and pathogenicity of *E. coli X*.

4.2 Founded antibiotic resistance genes of *E. coli X*

TetA is a membrane protein, which determines tetracycline resistance. TetA functions as a metal-tetracycline/H+ antiporter. The mechanism of tetracycline resistance involves active efflux of the drug; inside-out membrane vesicles prepared from resistant bacteria concentrate tetracycline by a process that requires proton motive force. In each case, the resistance determinant consists of two genes, a resistance gene (tetA) that encodes inner membrane protein (TetA) and a repressor gene (tetR) that encodes a regulatory protein (TetR) [13].

TetA and TetR are both situated between the synteny block and a small non-synteny block. TetR came from transposon. So we can suggest that tetracycline resistance genes in our case came from a mobile genetic element.

TEM-1 β-lactamase is one of the most well-known antibiotic resistance determinants around. β-lactamase catalyzes the hydrolysis of the amide bond in the β-lactam ring to generate ineffective products [14]. So, β-lactamase inactivates β-lactam antibiotics (such as penicillin, ampicillin, ceftriaxone, cefepime e.t.c.) itself.

CTX-Ms is another β-lactamase-coding gene. CTX-Ms and TEMs are likely to co-occur with each other because the acquisition of one does not seem to hinder the acquisition of the other. This suggests that the TEMs are likely to persist in bacterial populations even as CTX-Ms sweep through them [14].

Both β-lactamase-coding genes are situated in a big block, which is bourdered with a synteny block only from one side. Block with our genes of interest also contains genes of transposon, but their location in the middle of the block

Table 6 Reads quality characteristics

Quality charasteristic	paired-end i = 470 bp forward	paired-end i = 470 bp reverse	Mate pair i = 2 kb forward	Mate pair i = 2 kb reverse	Mate pair i = 6 kb forward	Mate pair i = 6 kb reverse
Per base sequence quality	Normal	Normal	Normal	Normal	Normal	Normal
Per sequence quality scores	Normal	Normal	Normal	Normal	Normal	Normal
Per base sequence content	Normal	Normal	Normal	Normal	Normal	Normal
Per sequence GC content	Slightly abnormal	Slightly abnormal	Normal	Normal	Slightly abnormal	Normal
Per base N content	Normal	Normal	Normal	Normal	Normal	Normal
Sequence Length Distribution	Normal	Normal	Normal	Normal	Normal	Normal
Sequence Duplication Levels	Normal	Normal	Normal	Normal	Normal	Normal
Overrepresented sequences	Normal	Normal	Normal	Normal	Normal	Normal
Adapter Content	Normal	Normal	Normal	Normal	Normal	Normal

and big size of the block gave us an idea these genes are located in the plasmid, and maybe appeared in the genome of *E. coli X* in process of conjugation with another bacteria - host of the plasmid.

The dhfrI gene mediated high-level trimethoprim resistance. Mechanism of resistance is related to target site alteration: additional dihydrofolate reductase has lesser sensitivity to trimethoprim.

The dhfrI gene is situated in a small block among two synteny blocks. There is prophage integrase was founded in it, so we conclude, *E. coli X* achieves dhfrI gene by transduction with bacteriophage [15].

4.3 Treatment strategy

Since we have resistance to almost all β-lactams, trimethoprim and tetracycline group, and there are problems associated with phage activation and increasing of Shiga toxin synthesis, we need to use another antibiotic. Theoretically, we may use carbapenems or aminoglycosides, but aminoglycosides may cause renal failure, which in combination with HUS may lead to exitus letalis. So using of meropenem or imipenem (carbapenems group) like most adequate treatment.

Supplementary information.

Acknowledgments. We thank Mikhail Rayko (Bioinformatics Institute) for providing a guide for performance this investigation. We thank Oksana Sidorenko (Bioinformatics Institute) and Daniil Vlasenko (Bioinformatics Institute) for providing instruction for prokka and mauve installation.

Declarations

All authors declare that they have no conflicts of interest.

References

[1] Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M., Finlay, B.B.: Recent advances in understanding enteric pathogenic Escherichia coli. Clin Microbiol Rev **26**(4), 822–880 (2013)

[2] Nataro, J.P., Kaper, J.B.: Diarrheagenic Escherichia coli. Clin Microbiol Rev **11**(1), 142–201 (1998)

[3] Freeman, V.J.: Studies on the virulence of bacteriophage-infected strains of corynebacterium diphtheriae. Journal of bacteriology **61**(6), 675–688 (1951)

[4] Chen, I., Dubnau, D.: DNA uptake during bacterial transformation. Nat Rev Microbiol **2**(3), 241–249 (2004)

[5] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A.: SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**(5), 455–477 (2012)

[6] Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G.: QUAST: quality assessment tool for genome assemblies. Bioinformatics **29**(8), 1072–1075 (2013)

[7] Seemann, T.: Prokka: rapid prokaryotic genome annotation. Bioinformatics **30**(14), 2068–2069 (2014). https://doi.org/10.1093/bioinformatics/btu153

[8] Seemann, T.: Barrnap. github. https://github.com/tseemann/barrnap (2013)

[9] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L.: BLAST+: architecture and applications. BMC Bioinformatics **10**, 421 (2009)

[10] Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T.: Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14**(7), 1394–1403 (2004)

[11] Melton-Celsa, A.R.: Shiga Toxin (Stx) Classification, Structure, and Function. Microbiol Spectr **2**(4), 0024–2013 (2014)

[12] Wilmaerts, D., De Loose, P.J., Vercauteren, S., De Smedt, S., Verstraeten, N., Michiels, J.: Functional analysis of cysteine residues of the Hok/Gef

type I toxins in Escherichia coli. FEMS Microbiol Lett **368**(11) (2021)

[13] Allard, J.D., Bertrand, K.P.: Membrane topology of the pBR322 tetracycline resistance protein. TetA-PhoA gene fusions and implications for the mechanism of TetA membrane insertion. J Biol Chem **267**(25), 17809–17819 (1992)

[14] Palzkill, T.: Beta-lactamase structure and function. Ann N Y Acad Sci **1277**, 91–104 (2013)

[15] Adrian, P.V., DU Plessis, M., Klugman, K.P., Amyes, S.G.: New trimethoprim-resistant dihydrofolate reductase cassette, dfrXV, inserted in a class 1 integron. Antimicrob Agents Chemother **42**(9), 2221–2224 (1998)