E. coli outbreak investigation

Abstract

The humankind has observed some *Escherichia coli* outbreaks causing infectious bacterial strains. However, in 2011 European population faced large-scale outbreaks of hemolytic uremic syndrome (HUS) when, within a few months, the bacterium had infected thousands and killed 53 people. We discovered that the new infectious strain originated from *E. coli* 55989 due to bacteriophages transfer of Shiga toxin. According to our research, Escherichia phage P13803 transferred the Shiga toxin gene from the virulent strain *E. coli* O157:H7 to the harmless *E. coli* 55989 strain and its phenomenon led to outbreak.

Introduction

Most *E. coli* strains are harmless, but some serotypes (for instance, O157:H7) can cause serious food poisoning in their hosts, and are occasionally responsible for food contamination incidents [1]. One such disease is hemolytic uremic syndrome (HUS) [2]. The mechanism is associated with Shiga toxins. These compounds are secrets by particular strain of *E. coli* and *Shigell*, then the toxin attaches to specific receptors on the cells' surface and enters the cell [3]. After entering a cell, this protein cleaves a specific adenine base of the 28S RNA of the 60S subunit of the ribosome, therefore protein synthesis is stopped [4]. Usually it happens in the lining of the blood vessels hemorrhage occurs.

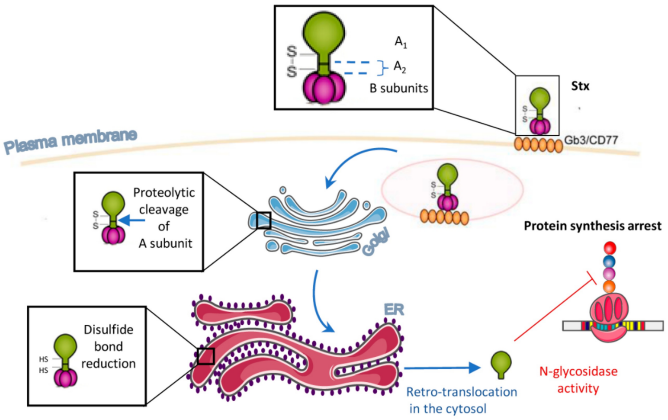

Figure 1 Overview of the intracellular move of Shiga toxins (according to [4])

However, due to horizontal genetic transfer (HGT) genes encoded Shiga toxin can be transferred to other harmless organisms. HGT is not limited to transfer of DNA between organisms of different types. HGT can be released via plasmids, phages and mobile elements exchange [5]. In this case alien genes insert and become a part of the host genome.

The investigation of similar changes is connected with assembling the genome *de novo*. In this case, even small changes can be detected in order to determine the origin of new features of species and strains.

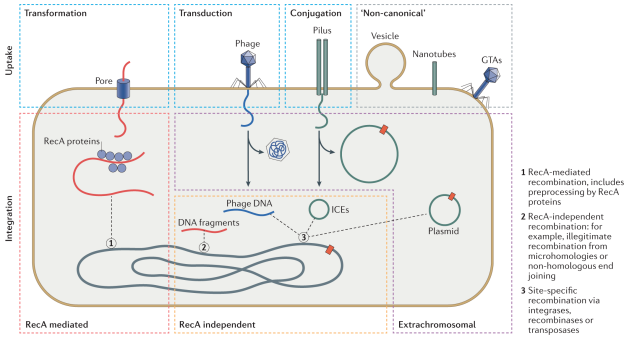

Figure 2 Options of horizontal gene transfer (according to [5])

In our research we expected to find new genes inserted in *E. coli* genome and examined reasons for their origin.

Materials and methods

Materials Raw whole-genome sequencing read for samples of *TY2482* were obtained from [6] other samples may be found in the supplementary. It's paired end and mate pair forward reverse reads.

Alignment and coverage To check the quality of reads, FASTQC was used. To count coverage and genome size, we used Jellyfish and Python [7] according to method on [8].

Annotation and assembling quality check We accomplished assembling of *E. Coli* genome with SPAdes (–careful option) and checked quality with QUAST. Annotation of the resulting data was conducted with Prokka.

Closest relative After that, we located 16S RNA in acquired data. For this we have to use Barrnap and NCBI Blast.

Visualization For visualization of received data and comparing scaffolds with the reference, we used Mauve and Geneious Prime. Visualization of an assembly graph is conducted via Bandage.

Antibiotic annotation Analysis of antibiotic resistance was conducted through annotation of scaffolds in ResFinder 4.1 program.

Results

First of all, we analyzed the number of reads for each of the 6 data sets of Illumina sequencing. We estimated the quality of reads through FASTQC. Table 1 shows the number of reads at every library.

We calculated coverage and approximate genome size of the investigated strain via Jellyfish. In addition, the distribution of k-mer is presented in Figure 3.

For assembling the genome, we used different libraries with different insert sizes in the SPAdes algorithm and estimated the quality of assembling using QUAST. Data presented in Table 2.

After Prokka annotation, we commenced to analyze received data.

$$Coverage = \frac{ML}{L-K+1} = \frac{17*90}{90-31+1} = 25.5$$
$$Genome\ size = \frac{Totalbases}{Coverage} = \frac{97555725}{25.5} = 3825714$$

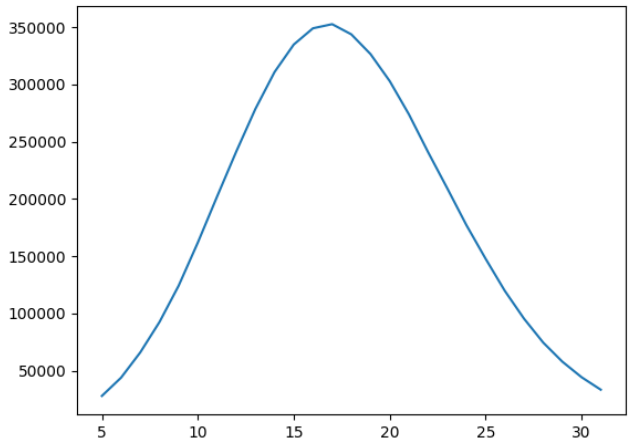

Figure 3 Smoothed peak in a k-mer distribution histogram (k-mer Profile)

Table 1 Three libraries from TY2482

Dataset	Forward reads	Reverse reads
Paired end 470bp	5499346	5499346
Mate pair 2kb	5102041	5102041
Mate pair 6kb	5102041	5102041

Discussion

After genome annotation, we investigated the origin of the new pathogenic *E. coli* strain. The BLAST analysis of conservative 16S RNA showed that this strain occurred from *E. coli* 559 that had early been suspected as a cause of gastroenteritis. [9]

Table 2 QUAST report data

Data	N50	Number of contigs
Pair end	95748	884
All three	769785	1007

Table 3 rRNA data

rRNA type	number	size
5S rRNA	8	109
16S rRNA	7	1537
23S rRNA	7	2838

These data are very crucial for research for inconsistencies between initial and daughter strains. The difference between the two genomes was associated with a region containing a lot of phage genes, Shiga-like toxin 2 subunit A and subunit B.

Protein BLAST demonstrated Shiga-like toxin 2 has a significant level of homology with similar protein of *E. coli* O157:H7 rRNA N-glycosylase and toxins transmitted by phages.

For whole reconstruction happened changes we compared genes surrounded Shiga toxins gene. We analyzed DUF1737 domain-containing protein, Rha protein CDS, phage tail fibre and searched intersection among alignments. Using such approach we discovered that a transferred agent was Escherichia phage P13803. Hence, originally phage P13803 infected *E. coli* O157:H7 and during replication some phage particles "catch" the gene of rRNA N-glycosylase. After the passage of time this phage infected *E. coli* 559 and this event made a new dangerous strain.

Additional danger was revealed when we discovered that the virulent strain stored genes of multiple-resistance especially genes encoded enzymes destroying antibiotics (including streptomycin, amoxicillin, ampicillin, tetracycline).

Surprisingly, the gene family *bla* encoded beta-lactamase is located in the cluster contained genes of mobile elements (such as transposase) and plasmid conjugation (transfer protein TraC), therefore, we assumed that resistant of the new strain occurred via movement of mobile genetic elements from caught early plasmid.

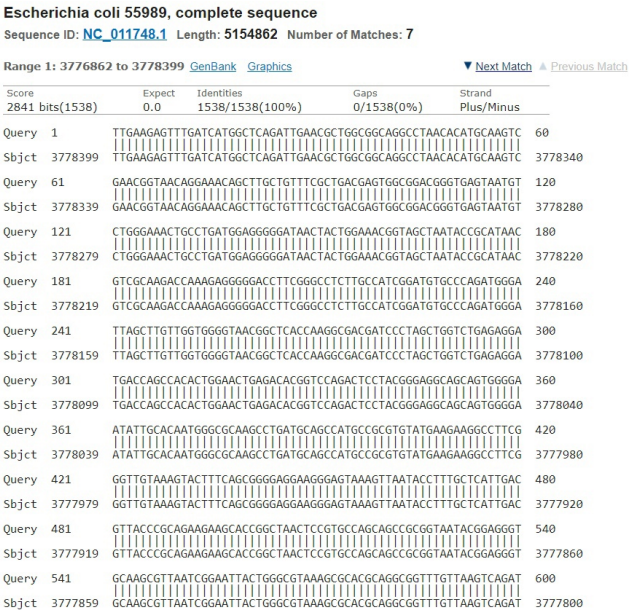

Figure 4 Alignment of 16S rRNA

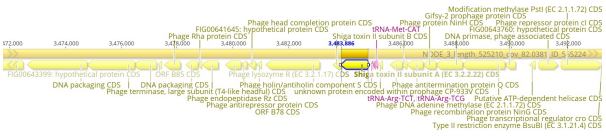

Figure 5 Region of genome exposed the rearrangement

Figure 6 Structure of Shiga toxin

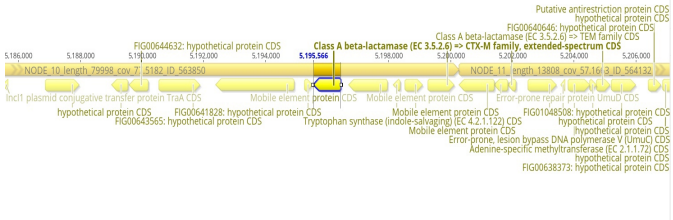

Figure 7 Cluster of genome contained beta-lactamase

Conclusion

In this work, we found out what reasons caused the outbreak in 2011. We discovered phage transfer of Shiga-like toxin from *E. coli* O157:H7 to *E. coli* 559 caused the appearance of a new pathogen. Moreover, the new strain is multi-antibiotic resistant due to transfer genes coding enzymes cleaving antibiotics. However, the new strain is susceptible to ampicillin + clavulanic acid mix, so medicine could struggle with this bacteria using such a drug combination to suppress beta-lactamase and, after that, suppress transpeptidase.

Acknowledgments

We thank Mike Raiko for providing with reads from sequencing of an *E. coli* strain.

References

[1] Dippold L. Vogt R. “Escherichia coli O157:H7 outbreak associated with consumption of ground beef”. In: *Public Health Rep* 120(2) (2005), pp. 174–8.

[2] Yassibanda S. Mossoro C. Glaziou P. “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”. In: *J Clin Microbiol* 40(8) (2002), pp. 3086–8.

[3] Keusch G. Donohue-Rolfe A. Acheson D. “Shiga toxin: purification, structure, and function”. In: *Rev Infect Dis* 4 (1991), pp. 293–7.

[4] Wiels J. Robert A. “Shiga Toxins as Antitumor Tools”. In: *Toxins* 13 (2021), p. 690.

[5] Hanage W. Arnold B. Huang I. “Horizontal gene transfer and adaptive evolution in bacteria”. In: *Nat Rev Microbiol* 20 (2022), pp. 206–218.

[6] ENA-EMBL. URL: https://www.google.com/url?q=https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292678sub_S1_L001_R1_001.fastq.gz.

[7] Report authors. URL: https://github.com/stockman2/BIOINF_E_Coli_3.

[8] Kanazawa. URL: https://koke.asrc.kanazawa-u.ac.jp/HOWTO/kmer-genomesize.html.

[9] Tenaillon O. Touchon M. Hoede C. "Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths". In: *PLoS Genet* 5(1) (2009).

Supplementary

The Illumina sequencing reads from sequencing of *E. coli* new strain sample available via the links:
https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292678sub_S1_L001_R1_001.fastq.gz
https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292678sub_S1_L001_R2_001.fastq.gz
https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292862_S2_L001_R1_001.fastq.gz
https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292862_S2_L001_R2_001.fastq.gz
https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292770_S1_L001_R1_001.fastq.gz
https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292770_S1_L001_R2_001.fastq.gz.

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
amikacin	aminoglycoside	No resistance	
tigecycline	tetracycline	No resistance	
tobramycin	aminoglycoside	No resistance	
cefepime	beta-lactam	Resistant	blaCTX-M-15 (blaCTX-M-15_AY044436)
chloramphenicol	amphenicol	No resistance	
piperacillin+tazobactam	beta-lactam	No resistance	
cefoxitin	beta-lactam	No resistance	
ampicillin	beta-lactam	Resistant	blaTEM-1B (blaTEM-1B_AY458016), blaCTX-M-15 (blaCTX-M-15_AY044436)
ampicillin+clavulanic acid	beta-lactam	No resistance	
cefotaxime	beta-lactam	Resistant	blaCTX-M-15 (blaCTX-M-15_AY044436)
ciprofloxacin	quinolone	No resistance	
colistin	polymyxin	No resistance	
sulfamethoxazole	folate pathway antagonist	Resistant	sul1 (sul1_AY522923), sul1 (sul1_DQ914960), sul1 (sul1_U12338), sul1 (sul1_AY115475), sul2 (sul2_HQ840942)
imipenem	beta-lactam	No resistance	
trimethoprim	folate pathway antagonist	Resistant	dfrA7 (dfrA7_AB161450)
nalidixic acid	quinolone	No resistance	
ertapenem	beta-lactam	No resistance	
tetracycline	tetracycline	Resistant	tet(A) (tet(A)_AJ517790)
fosfomycin	fosfomycin	No resistance	
ceftazidime	beta-lactam	Resistant	blaCTX-M-15 (blaCTX-M-15_AY044436)
temocillin	beta-lactam	No resistance	
gentamicin	aminoglycoside	No resistance	
meropenem	beta-lactam	No resistance	
azithromycin	macrolide	No resistance	

Figure 10 The table of new strains genes associated with antibiotic resistance

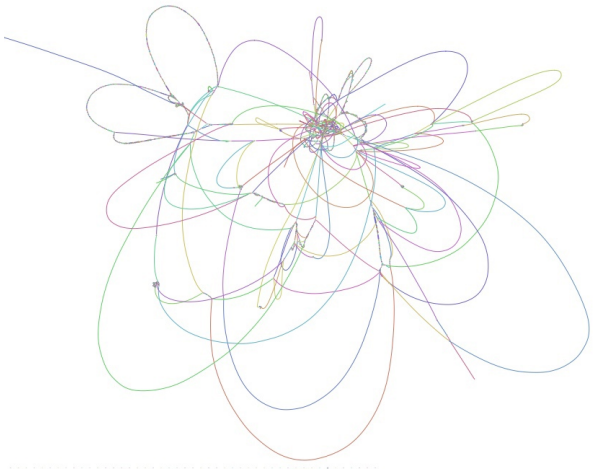

Figure 8 Visualization assembly graphs via Bandage

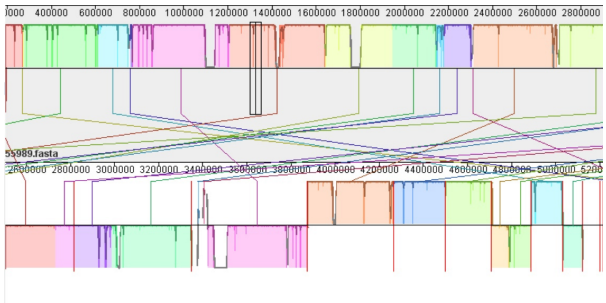

Figure 9 Visualization genome rearrangements of scaffolds via Mauve

Modification methylase PstI (EC 2.1.1.72) CDS	CDS	3,491,758	3,493,227	1,470	forward
Phage repressor protein cI CDS	CDS	3,490,950	3,491,663	714	forward
Phage transcriptional regulator cro CDS	CDS	3,490,651	3,490,854	204	reverse
FIG00643760: hypothetical protein CDS	CDS	3,489,742	3,490,119	378	reverse
Putative ATP-dependent helicase CDS	CDS	3,488,228	3,489,748	1,521	reverse
DNA primase, phage associated CDS	CDS	3,487,267	3,488,238	972	reverse
Gifsy-2 prophage protein CDS	CDS	3,486,818	3,487,267	450	reverse
Phage recombination protein NinG CDS	CDS	3,486,247	3,486,810	564	reverse
Phage protein NinH CDS	CDS	3,486,056	3,486,250	195	reverse
Phage antitermination protein Q CDS	CDS	3,485,629	3,486,063	435	reverse
Phage DNA adenine methylase (EC 2.1.1.72) CDS	CDS	3,485,228	3,485,380	153	reverse
tRNA-Met-CAT	tRNA	3,485,112	3,485,187	76	reverse
tRNA-Arg-TCG	tRNA	3,485,026	3,485,102	77	reverse
tRNA-Arg-TCT	tRNA	3,484,936	3,485,012	77	reverse
Shiga toxin II subunit A (EC 3.2.2.22) CDS	CDS	3,483,886	3,484,845	960	reverse
Shiga toxin II subunit B CDS	CDS	3,483,605	3,483,874	270	reverse
unknown protein encoded within prophage CP-933V CDS	CDS	3,481,182	3,483,119	1,938	reverse
ORF B78 CDS	CDS	3,480,866	3,481,045	180	reverse
Phage head completion protein CDS	CDS	3,480,580	3,480,825	246	reverse
Phage holin/antiholin component S CDS	CDS	3,480,287	3,480,502	216	reverse
Phage lysozyme R (EC 3.2.1.17) CDS	CDS	3,479,749	3,480,282	534	reverse
Phage antirepressor protein CDS	CDS	3,478,906	3,479,475	570	reverse
FIG00641645: hypothetical protein CDS	CDS	3,478,757	3,478,891	135	reverse
Phage endopeptidase Rz CDS	CDS	3,478,317	3,478,745	429	reverse
Phage Rha protein CDS	CDS	3,477,563	3,478,114	552	reverse
ORF B8S CDS	CDS	3,477,326	3,477,439	114	reverse
DNA packaging CDS	CDS	3,476,464	3,477,270	807	reverse
Phage terminase, large subunit (T4-like headful) CDS	CDS	3,474,777	3,476,483	1,707	reverse

Figure 11 Genes surrounded Shiga-like toxin 2

hypothetical protein CDS	CDS	5,204,724	5,205,158	435	forward
hypothetical protein CDS	CDS	5,204,489	5,204,710	222	forward
Adenine-specific methyltransferase (EC 2.1.1.72) CDS	CDS	5,203,805	5,204,488	684	forward
hypothetical protein CDS	CDS	5,203,572	5,203,697	126	reverse
FIG01048508: hypothetical protein CDS	CDS	5,202,494	5,203,420	927	forward
FIG00640646: hypothetical protein CDS	CDS	5,201,851	5,202,099	249	reverse
Error-prone repair protein UmuD CDS	CDS	5,201,417	5,201,854	438	reverse
Error-prone, lesion bypass DNA polymerase V (UmuC) CDS	CDS	5,200,272	5,201,417	1,146	reverse
NODE_11_length_13808_cov_57.1663_ID_564132	fasta_record	5,200,271	5,214,078	13,808	forward
Class A beta-lactamase (EC 3.5.2.6) => TEM family CDS	CDS	5,199,263	5,200,123	861	forward
Mobile element protein CDS	CDS	5,198,523	5,199,080	558	forward
Mobile element protein CDS	CDS	5,198,141	5,198,359	219	reverse
Mobile element protein CDS	CDS	5,196,697	5,197,959	1,263	reverse
Class A beta-lactamase (EC 3.5.2.6) => CTX-M family, extended-spectrum CDS	CDS	5,195,566	5,196,441	876	reverse
Tryptophan synthase (indole-salvaging) (EC 4.2.1.122) CDS	CDS	5,195,247	5,195,519	273	forward
Mobile element protein CDS	CDS	5,192,378	5,194,927	2,550	reverse
FIG00641828: hypothetical protein CDS	CDS	5,190,544	5,191,890	1,347	forward
FIG00644632: hypothetical protein CDS	CDS	5,189,589	5,190,191	603	reverse
FIG00643565: hypothetical protein CDS	CDS	5,189,024	5,189,533	510	reverse
hypothetical protein CDS	CDS	5,186,878	5,187,954	1,077	forward
IncI1 plasmid conjugative transfer protein TraA CDS	CDS	5,185,482	5,185,667	186	reverse
IncI1 plasmid conjugative transfer NusG-type transcription antiterminator TraB CDS	CDS	5,184,507	5,185,040	534	reverse
IncI1 plasmid conjugative transfer protein TraC CDS	CDS	5,183,570	5,184,253	684	reverse
IncI1 plasmid conjugative transfer protein PilI CDS	CDS	5,183,144	5,183,398	255	reverse
IncI1 plasmid conjugative transfer protein PilK CDS	CDS	5,182,426	5,183,019	594	reverse
hypothetical protein CDS	CDS	5,182,297	5,182,416	120	forward
IncI1 plasmid conjugative transfer protein PilL CDS	CDS	5,181,030	5,182,097	1,068	reverse

Figure 12 Genes surrounded beta-lactamase