E. coli outbreak investigation

Kupaeva D., Lopatina S.

Abstract

In 2011, Germany experienced the largest outbreak of hemolytic uremic syndrome in recent history, resulting in the death of 53 people. The infection was associated with a pathogenic strain of E coli, which contaminated bean seedlings. In this work, we provide a bioinformatic analysis of the genotype of a pathogenic strain. The de novo genome assembly made it possible to identify the genes of the shiga-like toxin obtained by horizontal gene transfer, which are responsible for the pathogenicity of the strain, as well as antibiotic resistance.

Introduction

Bloody diarrhea and the resulting hemolytic uremic syndrome are serious diseases. Post-diarrheal hemolytic uremic syndrome (D + HUS 1) is characterized by hemolytic anemia, thrombocytopenia, and kidney damage, and has a fatal outcome of 3-5% (1). Often, an outbreak of the syndrome is associated with shiga-toxin-producing Escherichia coli, serotype O157: H7. The usual reservoir for these bacteria is ruminants, especially cattle. Human infection with the shiga toxin-producing E. coli occurs through unintentional ingestion of feces - for example, through contaminated food or water, or through contact with animals or their environment, or, secondarily, through contact with infected people

In May 2011, a major outbreak of hemolytic uremic syndrome associated with rare E. coli occurred in Germany with serotype O104: H4. Research has determined that bean sprouts are the most likely spread of the outbreak (2) . In this study, we analyze the genotype of the E coli strain responsible for the outbreak, how it acquires pathogenicity, and discuss options for antibiotic therapy.

Despite the existence of high-quality annotated E. coli reference sequences, in the case of the analysis of bacterial strains, especially those with new phenotypes, it can be efficient to use not the alignment of the initial data on the reference, but the assembly of de-novo. In bacteria, the genome has high variability as a result of horizontal gene transfer by means of transposons, bacteriophage, and the sexual process of bacteria - the transfer of plasmids. Information about their sequences is often necessary for effective bioinformatics analysis. (3)

Methods

We assemblied a genome of strain using paired-end libraries of short reads (forward reads, reverse reads) and two different pair-end libraries of mate-pair reads (first: forward reads, reverse reads, second: forward reads, reverse reads). We estimated the quality of reads using FastQC (4) tools. Short summary of FastQC result attached in supplementary material 1.

Genome was assembly by SPAdes (5) v.3.14 with standard settings. We perform assembly with only short reads and with short reads and mate reads. Quality of assembly we estimate using QUAST (6). For the next studies, we used assembly included all libraries of reads.

Annotation of genome was performed using prokka (7) v.1.12 with options -centre X --compliant. Ribosomal RNA was identified with help of barrnap (8). For finding the closest relative of E. coli we used blastn web-service, RefSeq Genome Database, Organism: E.coli , Entrez Query: 1900/01/01:2011/01/01[PDAT]. According to our results, the closest relative with a reference genome is Escherichia coli strain 55989. For analyze differences between the reference genome and our assembled genome, we used Mauve (9), progressiveMauve algorithm. Because this annotation has a lot of unannotated proteins, we used RAST (10) annotation. For prediction antibiotic resistance we used ResFinder web-service (11).

All commands which was used in this project are presented in supplementary material 2.

Results

Assembly contained reads from different insert size have increased N50 (48732 vs 54616), but more misassemblies (80 vs 124). Another statistic of assembly presented in table 1.

	only short reads	short reads + mate reads
Genome fraction (%)	90.341	91.225
Duplication ratio	1.003	1.005
genomic features	3817 + 103 part	3856 + 105 part
operons	739 + 63 part	749 + 65 part
Largest alignment	159892	163228
Total aligned length	4195891	4250238
NGA50	48732	54616
LGA50	29	26
misassemblies	80	124
Misassembled contigs length	3709979	4841987
contigs	205	105
Largest contig	300784	698474
Total length	5243761	5350156

Table 1. Basic statistic of genome alignments.

Gene 16S RNA was founded in 7 copies, length 1537. Annotation with help of prokka was prepare with statistics: tRNAs : 80, rRNAs: 0, CRISPRs : 1, CDS: 5064, Unique gene codes : 2923 .

With help of Mauve based on prokka annotation, we investigate gene which coded toxic products. Finding proteins with word “toxin” in description provide find stxA and stxB genes, which coded Shiga toxins. These genes located in a zone which lake in the reference genome. According to annotation. other genes in these parts of genome - different bacteriophage genes, including ninH p-22, Pha and

other. In figure 1 showed alignment with coordinates and gene structure. It points at horizontal transcfer of these genes with bacteriophage.

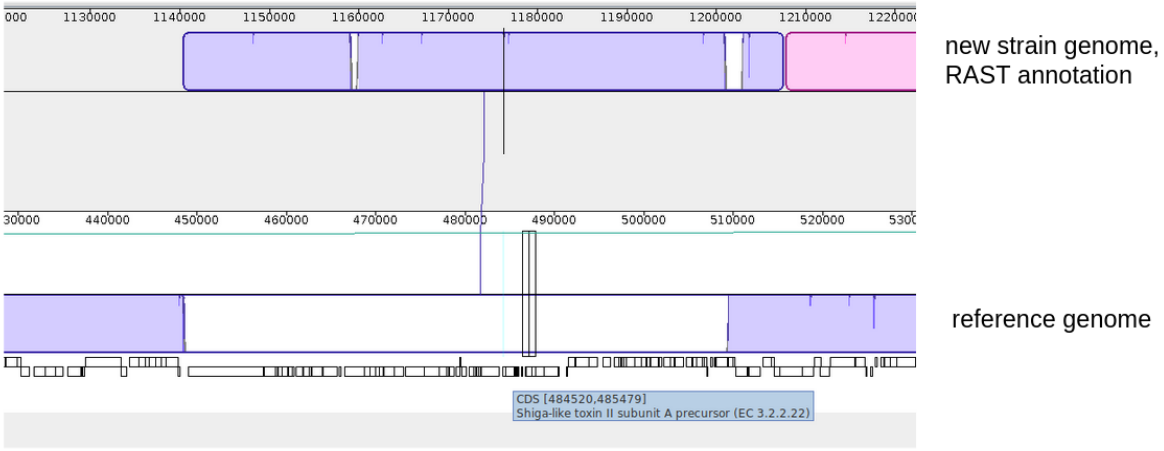

Figure 1.

ResFinder predicts resistant to 4 beta-lactam antibiotic: ceftazidime, cefepime, ampicillin, cefotaxime; two folate pathway antagonist: sulfamethoxazole, trimethoprim; and tetracycline.

We can identify genes bla1 and bla2, which locate in part of genome new strain which has no alignment to reference genome, Beside bla1 and bla2 genes, which response with beta-lactam antibiotics resistance, this zone contains predicted transposon genes, include integrase and resolvase. It points at horizontal transfer these genes with mobile genetic elements.

Another resistance to antibiotics connected with mdf(A) gene, but this gene represented in reference too and probably reference is resistant to tetracycline too.

	Coordinate of start	Coordinate of end	Length
stxA	3483886	3484845	959
stxB	3483605	3483874	269
Pha	348386	349627	1241
Transposon Tn21 resolvae	5113693	5114253	560
TniA putatice transposase	5126568	5127836	1268
Bla1	5119164	5120024	860
Bla2	5376443	5377318	875
MdfA	692826	694058	1232

Table 2. Genes coordinate and length according annotation RAST.

Discussion

Unfortunately, the acquisition of genes for the synthesis of toxins and genes for antibiotic resistance by the E coli strain is not unusual or unexpected. Analytical histories of acquisition of both resistance genes and toxin genes have already been described in the literature. Thus, the Creuzburg article describes the sequence of a bacteriophage containing genes for the synthesis of shiga toxin (12).

Antibiotic resistance genes are distributed in a similar way in the population. In our case, the bla1 and bla2 genes were acquired using a transposon, a mobile genetic element (13). Expression of these genes hydrolyzes the beta-lactam bond in antibiotics of the first type, providing resistance to a number of beta-lactam antibiotics (14). The horizontal transfer of antibiotic resistance genes is a growing public health problem.

Citations

---

1) Tarr, Phillip I., Carrie A. Gordon, and Wayne L. Chandler. "Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome." *The lancet* 365.9464 (2005): 1073-1086.
2) Buchholz, Udo, et al. "German outbreak of Escherichia coli O104: H4 associated with sprouts." *New England Journal of Medicine* 365.19 (2011): 1763-1770.
3) Daubin, Vincent, and Gergely J. Szöllősi. "Horizontal gene transfer and the history of life." *Cold Spring Harbor Perspectives in Biology* 8.4 (2016): a018036.
4) Andrews, S. FastQC A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics (2010). Available at: www.bioinformatics.babraham.ac.uk/projects/fastqc/.
5) Nurk, Sergey, et al. "Assembling genomes and mini-metagenomes from highly chimeric reads." Annual International Conference on Research in Computational Molecular Biology. Springer, Berlin, Heidelberg, 2013.
6) Gurevich, Alexey, et al. "QUAST: quality assessment tool for genome assemblies." *Bioinformatics* 29.8 (2013): 1072-1075.
7) Seemann, Torsten. "Prokka: rapid prokaryotic genome annotation." *Bioinformatics* 30.14 (2014): 2068-2069.
8) Seemann T barrnap 0.9 : rapid ribosomal RNA prediction https://github.com/tseemann/barrnap
9) Darling, Aaron CE, et al. "Mauve: multiple alignment of conserved genomic sequence with rearrangements." *Genome research* 14.7 (2004): 1394-1403.
10) Aziz, Ramy K., et al. "The RAST Server: rapid annotations using subsystems technology." *BMC genomics* 9.1 (2008): 1-15.
11) Bortolaia, Valeria, et al. "ResFinder 4.0 for predictions of phenotypes from genotypes." *Journal of Antimicrobial Chemotherapy* 75.12 (2020): 3491-3500.
12) Creuzburg, Kristina, et al. "The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-like type III effector protein." *Journal of bacteriology* 187.24 (2005): 8494-8498.
13) Naas, Thierry, et al. "Genetic structures at the origin of acquisition of the β-lactamase blaKPC gene." *Antimicrobial agents and chemotherapy* 52.4 (2008): 1257-1263.
14) Walther-Rasmussen, Jan, and Niels Høiby. "Cefotaximases (CTX-M-ases), an expanding family of extended-spectrum β-lactamases." *Canadian journal of microbiology* 50.3 (2004): 137-165.
15) Freedman, Stephen B., et al. "Shiga toxin–producing Escherichia coli infection, antibiotics, and risk of developing hemolytic uremic syndrome: a meta-analysis." *Clinical Infectious Diseases* 62.10 (2016): 1251-1258.

Supplemental resources

Supplementary materials №1: https://docs.google.com/document/d/1A_krnU6QPC8SdZE54urH2CPdm7_iHyVC65ntJi7ApOE/edit?usp=sharing

Supplementary materials №2: https://www.notion.so/project-3-2d6abf3a2b004071b1ab9c0273d8a0af