

Causes of the E.Coli outbreak in Germany in April 2011.

Darya Grechishkina

Abstract.

Part of *E. coli* O104:H4 Genome Analysis Crowdsourcing(1) research was repeated for educational purposes: E.Coli X genome(2, 3, 4, 5, 6, 7) was assembled and annotated, closest relative (as for April 2011) of the strain was found. Possible way of becoming pathogenic and receiving antibiotic resistance proposed. Several pieces of advice about threatening methods were given.

Introduction.

E.Coli strain that produces Shiga toxin causes HUS. Infection begins with consuming contaminated food or water, E.Coli survives in low Ph of the stomach due to acid-resistance. Bacteria causes diarrhea and intestinal inflammation by attaching to intestinal epithelium and rearranging cytoskeletal actin.(8)

Horizontal gene transfer - is a movement of genetic material from one organism to another in other ways than reproduction. There are three mechanisms of horizontal gene transfer: *transformation* (uptake of the naked DNA from environment), *transduction* (transfer of genetic material by bacteriophage from donor to recipient) and *conjugation* (involves physical contact between donor and recipient).(9)

When to use *de novo* assembly, not alignment to reference? When it's crucial to recreate original sequence (e.g. looking for large insertions).(10)

Data.

Libraries from TY2482 sample:

1. SRR292678: forward(2) and reversed(3) -- paired end library
2. SRR292862: forward(4) and reversed(5) -- mate pair library (2kb)
3. SRR292770: forward(6) and reversed(7) -- mate pair library (6kb)

Methods.

Reads were analyzed using FastQC(11). Paired end library was assembled *de novo* alone and with mate pair libraries using SPAdes(12). K-mers were profiled using Jellyfish(13) on raw SRR292678 and on SRR292678 corrected by SPAdes with kmer size 31.

QUAST(14) was runned on both single-library assembly and three-library assembly. Three-library assembly was annotated using Prokka(15).

To find the closest relative of the new *E. Coli* strain 16S ribosomal RNA was located in the assembled genome using Barrnap(16). 16S rRNA sequences were extracted from assembly using Samtools(17), and searched using BLASTN(18) to find the closest relative. The closest relative's genome was used as a reference sequence for alignment using Maeve(18). To research possible acquisition of antibiotic resistance ResFinder(19) was used on both reference and analyzed genomes.

For a detailed pipeline and scripts please refer to lab journal(20).

Results.

Estimated genome size for corrected reads is significantly smaller than for raw reads (Table 1).

Genome assembly using mate-pair libraries alongside a pair-end library has a larger N50 score than assembly using just a pair-end library. (Table 2)

Escherichia coli 55989 (NC_011748.1) is the closest relative to the new strain. It's genome sequence was used as a reference sequence for alignment with the *E.Coli* X genome. Was found that the new strain contains an insertion of two Shiga toxin-related genes: *stxB*[484239-484508] and *stxA*[484520-485479]. Nearest genes([481843-483753] and [481525-481704]) encodes phage proteins.

According to ResFinder *E.Coli* X resistant to cefepime, ampicillin, cefotaxime, sulfamethoxazole, trimethoprim, ceftazidime, to which reference strain is vulnerable to.

New strain has insertions of beta-lactamase encoding genes in [5119264-5120024] and [5376443-5377318], both insertion regions contain phage-related genes.

Discussion.

I suppose, that insertions of both Shiga toxin-related genes and beta-lactamase related genes were obtained by *E.Coli* X via horizontal gene transfer (more precisely, transduction) as the phage genome was found in the insertions.

The antibiotic resistance mechanism most likely is *inactivation of the antibiotic molecule* by producing beta-lactamase. So the threatening advice would be to use non-beta-lactam antibiotics (e.g. vancomycin).

Supplementary materials.

SRR292678	Estimated genome size
raw	5 453 897
corrected	5 155 220

Table 1. Estimated genome size received using Samtools(17).
Refer to lab journal(20) for script.

	Single-library assembly	Three-library assembly
Number of contigs	205	105
N50	105 346	335 515

Table 2. Comparison of single-library and three-library assemblies using QUAST(14).

Citations.

1. <https://github.com/ehec-outbreak-crowdsourced/BGI-data-analysis/wiki>
2. https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292678sub_S1_L001_R1_001.fastq.gz
3. https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292678sub_S1_L001_R2_001.fastq.gz
4. https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292862_S2_L001_R1_001.fastq.gz
5. https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292862_S2_L001_R2_001.fastq.gz
6. https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292770_S1_L001_R1_001.fastq.gz
7. https://d28rh4a8wq0iu5.cloudfront.net/bioinfo/SRR292770_S1_L001_R2_001.fastq.gz
8. Palermo MS, Exeni RA, Fernández GC. Hemolytic uremic syndrome: pathogenesis and update of interventions. Expert Rev Anti Infect Ther. 2009 Aug;7(6):697-707. doi: 10.1586/eri.09.49. PMID: 19681698.

9. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 2000 May 18;405(6784):299-304. doi: 10.1038/35012500. PMID: 10830951.
10. <http://beat.wolf.home.hefr.ch/documents/prague.pdf>
11. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012 May;19(5):455-77. doi: 10.1089/cmb.2012.0021. Epub 2012 Apr 16. PMID: 22506599; PMCID: PMC3342519.
13. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011 Mar 15;27(6):764-70. doi: 10.1093/bioinformatics/btr011. Epub 2011 Jan 7. PMID: 21217122; PMCID: PMC3051319.
14. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013 Apr 15;29(8):1072-5. doi: 10.1093/bioinformatics/btt086. Epub 2013 Feb 19. PMID: 23422339; PMCID: PMC3624806.
15. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014 Jul 15;30(14):2068-9. doi: 10.1093/bioinformatics/btu153. Epub 2014 Mar 18. PMID: 24642063.
16. Seemann T. barrnap 0.9 : rapid ribosomal RNA prediction
<https://github.com/tseemann/barrnap>
17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. Epub 2009 Jun 8. PMID: 19505943; PMCID: PMC2723002.
18. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 2010 Jun 25;5(6):e11147. doi: 10.1371/journal.pone.0011147. PMID: 20593022; PMCID: PMC2892488.
19. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O,

Aarestrup FM. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020 Dec 1;75(12):3491-3500. doi: 10.1093/jac/dkaa345. PMID: 32780112; PMCID: PMC7662176.

20. https://github.com/DespairedController/bioinf_lab_journal/tree/main/E_COLI_OUTBREAK