Investigation of the origin and genomic characteristics of the pathogenic Escherichia coli caused Hemolytic-uremic syndrome outbreak

Filippov Mikhail
@pseudocephalus

Herzen University, Saint Petersburg, Russia

Abstract

Hundreds of people were hospitalized with haemolytic uremic syndrome in Germany in April 2011. Unknown strain of E. coli was identified as a cause of infection. This outbreak rapidly grown into an epidemy and caused nearly thousand infections and 53 deaths across Europe. To investigate the origin of an outbreak, characterize the new strain and reveal causes of such pathogenity, bionformatic methods was applied. Goal of this study is to reproduce the workflow of the bionformaticians investigated this case. We used Illumina sequencing data generated in Beijing Genome Institute in attempt to perform de novo assembly of novel strain genome. Then we discovered the closest known strain (as of early 2011) and using it as a reference observed genome features that differ unknown strain from other E.coli strains and made it pathogenic. As a result, we obtained an assembled genome of novel E. coli strain, found out it's origin and revelaed shiga toxin and antibiotic resistance genes present in one's genome.

Introduction

Hemolytic-uremic syndrome

Hemolytic-uremic syndrome is a rare desease characterized by bloody diarrhea, vomiting, which are early symptoms of infection, and appearing later haematuria (blood in urine), oliguria (low amount of urine), thrombocytopenia, kidney failure and lethargy. In some cases there are also hypertension, skin bleeding, janudice and seizures present [1, 2]. There symptoms are severe and can be lethal, which is also a case of researched outbreak.

As there can be different causes of HUS, this disease divided into typical and atypical HUS. Typical HUS is most abundant and caused by ingestion of certain pathogenic bacterias, and atypical is thought to have endogenic causes, such as rare mutation leading to uncontrolled activation of the complement system.

Bacteria causing HUS are usually E. coli strain O157:H7 or one from genus Shigiella, both of which known to produce Shiga or Shiga-like toxins [3]. E. coli strains able to cause HUS called Enterohemorrhagic Escherichia coli (EHEC) or Shiga-toxin producing E. coli (STEC). Number of specific features allow these bacteria to cause infection: beside production of Shiga toxins they are acid resistant and able to generate attaching and effacing lesions in the mucose of rectum, which allow them effectively colonize this part of intestine [4]. In described outbreak resistance to the number of antibiotics also takes place, which is an object of further investigation.

Horizontal gene transfer

Bacteria characterized by ability to obtain genomic elements in noninheritable way, which is known as horizontal gene transfer (HGT). There are two major ways for bacteria to obtain new genes from other organism during life cycle. One is with plasmids - small circular DNA which can be transferred from one cell to another directly from environment (transformation), from another bacterial cell (conjugation) or artificially (transfection). Another is to obtain DNA with bacteriophage infection, when viral genome is embedded in bacterial DNA - also known as viral vector. HGT-associated genetic elements are part of larger group, known as mobile genetic elements, which can be present only in the same cell and not to "infect" others. Such are, for example, transposones. HGT has a great significance for bacteria survival and evolution and one of its important functions is transferring of virulence or resistance associated genes, such as Shiga toxins or antibiotic resistance genes [5].

De novo assembly

There are several cases when de novo assembly instead of reference assembly is preferrable option. One of them is, for shure, is when reference organism is not known or not present. Another one is conservation biology, when scientists intent to preserve genetic information of threatened species. Third one, which is the case of present study, is when we expect high divercity of researched organism from reference, for example, when we suggest that organism of interest had obtained new genetic elements or demonstrate high mutation rate. Assembly itself resolves some sequencing problems like shortness of sequenced fragments, high error rate and contamination, but in turn can bring in another errors.

Methods

Data properties

Three Illumina sequencing libraries were used for genome assembly:

SRR292678 (forward) (reverse) - paired end, insert size 470 base pair

SRR292862 (forward) (reverse) – mate pair, insert size 2000 base pairs

SRR292770 (forward) (reverse) – mate pair, insert size 6000 base pairs

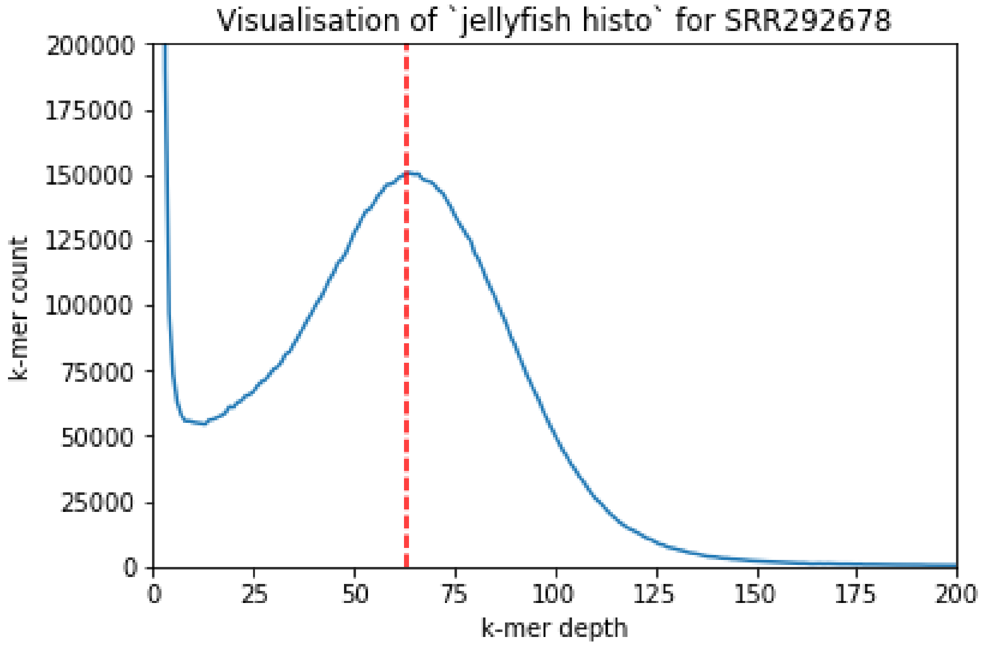

Figure 1: Visualization of **jellyfish histo**. Vertical axis is the k-mer number in genome (limited on 200000), horizontal axis is k-mer's depth (limited on 200). Red dotted line is the k-mer depth of the max k-mer count = 63

FastQC 0.11.9 [6] reported following number of reads foe each library:

SRR292770: 5102041 reads

SRR292862: 5102041 reads

SRR292678: 5499346 reads

Read length is 90 b.p.

K-mer counting

K-mers of the length 31 were counted and plotted with **jellyfish 2.3.0** [7]. For counting we used **jellyfish count** with -s 600000 and for plotting was used **jellyfish histo** (figure 1)

De novo assembly and quality control

SPAdes v3.15.4 [8] was used for de novo genome assembly, which was performed for paired end (SRR292678) library and for all three libraries altogether. N50 and number of contigs were obtained by **QUAST 5.2.0** [9].

Annotation

Assembled genome was then annotated with **Prokka 1.14.6** [10] with parameters *–centre X –compliant*

16S rRNA genes prediction

16S rRNA genes predicted by **Barrnap 0.9** software [11] with standard parameters.

BLAST

Closest relative of researched strain was identificated with **BLAST** [12] on the revealed by Prokka 16S rRNA genes in .fasta format. Nucleotide blast was run with **1900/01/01:2011/01/01[PDAT]**, as we interested in results available for researchers in spring 2011.

Alignment and visualization

Annotated genome in *.gbk* format was aligned on the reference E. coli genome available on NCBI nucleotide database by the means of **Mauve 2.4.0** [13].

Antibiotics resistance prediction

This procedure was performed with ResFinder 4.1 [14], based on *scaffolds.fasta* obtained via SPAdes run on three libraries.

Results

Assembly

There are several metrics for de novo assembly provided by QUAST. SPAdes run on pair end library only resulted in a genome of 5295721 b.p. total length splitted in 201 contigs with N50 = 111860. SPAdes run on all three libraries resulted in genome of 5350156 b.p. total length with 105 contigs with N50 = 335515.

Annotation

Prokka run succeeded with annotated 5064 CDSs, 1 CRISPR, 5144 genes and 80 tRNA

16S rRNA prediction

Barrnap resulted in 6 16 rRNA genes of 1538 b.p. and 1 16S rRNA gene of 406 b.p.

BLAST result

The closest E. coli strain appreaed to be an EHEC Escherichia coli 55989, which was isolated in the Central African Republic from a stool sample obtained from a HIV-positive man with persistent diarrhea *(without signs of HUS)* [15]. It thought to be a parent strain of novel E. coli caused an outbreak, but this one have somehow obtained additional pathogenity.

Genome exploration

First that was noted is the great rate of genomic rearrangements, which can be seen as extremely different order of elements in the reference and investigated genomes. Presence of several genes was noted while observing assembled genome in Mauve and loading assembly in the ResFinder: there are stxA, stxB, blaCTX-M-15, blaTEM-1B, sul1, sul2, aph(3")-Ib, aph(6)-Id and tetA genes located on two unalignet regions of researched genome. stxA and stxB are located in the middle of insertion in one of the bacterial chromosome's fragments, and the rest are located in major genome fragment not linked to reference chromosome.

Discussion

Assembly statistics for different libraries

Notably, when we used three libraries with different insert size, assembly quality is much improved: N50 is nearly tripled, and number of contigs is twice less, both of which are positive changes in overall statistics. This can be explained as result of more effective resolution of genomic repeats: knowing the exact sequences upstream and downstream of the repeat and the length between them (insert size of the library), we can easily differ repeated fragments and place them in right order.

Comparative genomics

Number of significant genetic features differ this unknown strain from reference Escherichia coli 55989 strain. Unlike reference, this bacteria has *stxA* and *stxB* genes, which are known to encode Shiga toxin subunits A and B, which, combined, cause HUS symptoms in infected human [16]. Another differentiating feature is presence of two *bla* genes that responsible for expression of beta-lactamases - enzymes that degrade beta-lactam ring of most beta-lactam antibiotics. Presence of these genes in E. coli is known to cause resistance to such antibiotics [17]. More, there are *aph(3")-Ib* and *aph(6)-Id* genes that can induce streptomycin resistancy in bacteria [18]. *sul1* and *sul2* genes also present in researched genome are the cause of sulfamethoxazole resistance. Finally, there are tetA gene which is also present in the reference genome and causing tetracycline and doxycycline resistance [20]. Overall table representing these results is the Table 1. Thus, novel E. coli is not only enteroagressive, but also demonstrate resistance to several groups of antibiotics, that severely affects effectiveness of treatment.

Gene	Insert origin	Presence in reference	Product	Function
stxA	Viral genome	-	Shiga toxin subunit A	Toxin production
stxB	Viral genome	-	Shiga toxin subunit B	Toxin production
bla1	Plasmid	-	beta-lactamase	Beta-lactam antibiotics resistance
bla2	Plasmid	-	beta-lactamase	Beta-lactam antibiotics resistance
sul1	Plasmid	-	Sulfonamide resistant dihydropteroate synthase	Sulfonamide resistance
sul2	Plasmid	-	Sulfonamide resistant dihydropteroate synthase	Sulfonamide resistance
aph(3")-Ib	Plasmid	-	Aminoglycoside 3'-phosphotransferase	Streptomycin resistance
aph(6)-Id	Plasmid	-	Aminoglycoside O-phosphotransferase	Streptomycin resistance
tet(A)	Plasmid	+	Tetracycline resistance protein	Tetracycline resistance

Table 1: Revealed Shiga toxin and antibiotics resistance genes in novel E.coli strain

Origin of pathogenity

As mentioned above, additional virulence genes are located on unaligned fragments of genome, which have no representation in reference Escherichia coli 55989. Using Prokka annotation, we able to trace the source of such acquisitions. There are two different types of unaligned regions. One of them, including both *stx* genes, is located in the middle of the fragment of bacterial chromosome and contains a lot of viral proteins as well. Thus, best explanation is an acquisition of that region via viral infection. Another one is located aside of bacterial chromosome and contains mobile elements and plasmid genes as well as antibiotic resistance genes. We suggest that acquisition of these genes was the result of plasmid transfer in one of the ways described in Introduction.

Antibiotic therapy

Summarising all of the features of the novel E.coli strain, it can be assumed that number of antibiotics described in "Comparative genomics" section will be unefficient in case of infection with this bacteria. Research shows that norfloxacin, ciprofloxacin, gentamicin and chloramphenicol are trustworthy choice for the treatment of resistant E.coli strains [21].

References

1. "E.coli (Escherichia coli): Symptoms". Centers for Disease Control and Prevention. U.S. Department of Health & Human Services. 2017-11-30.
https://www.cdc.gov/ecoli/ecoli-symptoms.html

2. "Hemolytic uremic syndrome (HUS)". Center for Acute Disease Epidemiology. Iowa Department of Public Health.
https://idph.iowa.gov/cade/disease-information/hus

3. Salvadori M, Bertoni E. Update on hemolytic uremic syndrome: Diagnostic and therapeutic recommendations. World J Nephrol. 2013 Aug 6;2(3):56-76. PMID: 24255888; PMCID: PMC3832913.
doi: 10.5527/wjn.v2.i3.56.

4. Nguyen Y, Sperandio V. Enterohemorrhagic E. coli (EHEC) pathogenesis. Front Cell Infect Microbiol. 2012 Jul 12;2:90. PMID: 22919681; PMCID: PMC3417627.
doi: 10.3389/fcimb.2012.00090.

5. Singh PK, Bourque G, Craig NL, Dubnau JT, Feschotte C, Flasch DA, et al. (2014-11-18). "Mobile genetic elements and genome evolution 2014". Mobile DNA. 5: 26. PMC 4363357. PMID 30117500.
doi:10.1186/1759-8753-5-26.

6. Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data Available online at: `http://www.bioinformatics.babraham.ac.uk/projects/fastqc`

7. Marcais G., Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics (2011) 27(6): 764-770
doi:10.1093/bioinformatics/btr011 `http://www.cbcb.umd.edu/software/jellyfish`

`http://bioinformatics.oxfordjournals.org/content/early/2011/01/07/bioinformatics.btr011`

8. Bankevich A., Nurk S., Antipov D., Gurevich A., Dvorkin M., Kulikov A. S., Lesin V., Nikolenko S., Pham S., Prjibelski A., Pyshkin A., Sirotkin A., Vyahhi N., Tesler G., Alekseyev M. A., Pevzner P. A. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 2012

9. Alla Mikheenko, Andrey Prjibelski, Vladislav Saveliev, Dmitry Antipov, Alexey Gurevich, Versatile genome assembly evaluation with QUAST-LG, Bioinformatics (2018) 34 (13): i142-i150.
doi: 10.1093/bioinformatics/bty266

10. Seemann T. Prokka: rapid prokaryotic genome annotation Bioinformatics 2014 Jul 15;30(14):2068-9. PMID:24642063

11. Seemann T. barrnap 0.9 : rapid ribosomal RNA prediction

`https://github.com/tseemann/barrnap`

12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10. PMID: 2231712.
doi: 10.1016/S0022-2836(05)80360-2.

13. Darling, A. C., Mau, B., Blattner, F. R., Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome research, 14(7), 1394–1403.
https://doi.org/10.1101/gr.2289704

14. Bortolaia V, Kaas RF, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AR, 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, NykÃ¤senoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial Chemotherapy, 75(12),3491-3500

15. Mossoro, C., Glaziou, P., Yassibanda, S., Lan, N. T., Bekondi, C., Minssart, P., Bernier, C., Le Bouguénec, C., & Germani, Y. (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), 3086–3088.
https://doi.org/10.1128/JCM.40.8.3086-3088.2002

16. Melton-Celsa A. R. (2014). Shiga Toxin (Stx) Classification, Structure, and Function. Microbiology spectrum, 2(4), 10.1128/microbiolspec.EHEC-0024-2013.
https://doi.org/10.1128/microbiolspec.EHEC-0024-2013

17. Bajaj, P., Singh, N. S., & Virdi, J. S. (2016). Escherichia coli beta-Lactamases: What Really Matters. Frontiers in microbiology, 7, 417.
https://doi.org/10.3389/fmicb.2016.00417

18. Zeng, L., Jin, S. (2003). aph(3')-IIb, a gene encoding an aminoglycoside-modifying enzyme, is under the positive control of surrogate regulator HpaA. Antimicrobial agents and chemotherapy, 47(12), 3867–3876.
https://doi.org/10.1128/AAC.47.12.3867-3876.2003

19. Antunes P, Machado J, Sousa JC, Peixe L. Dissemination of sulfonamide resistance genes (sul1, sul2, and sul3) in Portuguese Salmonella enterica strains and relation with integrons. Antimicrob Agents Chemother. 2005 Feb;49(2):836-9. PMID: 15673783; PMCID: PMC547296.
doi: 10.1128/AAC.49.2.836-839.2005.

20. Olowe, O. A., Idris, O. J., Taiwo, S. S. (2013). Prevalence of tet genes mediating tetracycline resistance in Escherichia coli clinical isolates in Osun State, Nigeria. European journal of microbiology & immunology, 3(2), 135–140. https://doi.org/10.1556/EuJMI.3.2013.2.7

21. Kibret, M., Abera, B. (2011). Antimicrobial susceptibility patterns of E. coli from clinical sources in northeast Ethiopia. African health sciences, 11 Suppl 1(Suppl 1), S40–S45.
https://doi.org/10.4314/ahs.v11i3.70069