Researchers release complete de novo E. coli O104 genome, details of their detection kit

June 7 2011

Scientists worldwide have been working on the publicly available genomic sequences of the deadly *E. coli* O104 strain, which is causing the current health crisis in Germany and now spreading throughout Europe. To continue to speed the ongoing international efforts of researchers to assess and halt this growing epidemic, BGI and their collaborators at the University Medical Centre Hamburg-Eppendorf have now released their third version of the assembled genome, which includes new data from this *E. coli* O104.

(ftp://ftp.genomics.org.cn/pub/Ecoli_TY-2482/Escherichia_coli_TY-2482.contig.20110606.fa.gz). In addition, the FTP site contains a file that provides the PCR primer sequences BGI and their collaborators have used to create diagnostic kits for rapid identification of this highly infectious bacterium.

The new assembly includes more than 200x single-end reads from the Illumina HighSeq Platform, which allowed BGI to provide a more complete genome map and to correct any assembly errors from the previous version. More importantly, this version is a completely de novo assembly, whereas the previous versions by BGI and others used a reference-based assembly method to obtain a consensus sequence. The new assembly continues to support the finding that this infectious strain carries disease-causing genes from two types of pathogenic *E. coli*: enteroaggregative *E. coli* (EAEC) and enterohemorrhagic *E. coli* (EHEC).

Taking advantage of this genomic feature, BGI and the Beijing Institute of Microbiology and Epidemiology researchers have developed a straightforward PCR diagnostic protocol for rapid identification of the outbreak strain. The diagnostic method consists of two pairs of amplification primers that target the enteroaggregative- and hemorrhagic-associated genes (more detailed protocol is available on the BGI FTP site). Diagnostic results can be obtained within 2□□ hours after receiving the sample, and thus will be extremely useful for epidemic surveillance and detection of this bacterium.

BGI has assessed the specificity and sensitivity of this kit and protocol through computational analyses of 4,547 strains (from 2,183 species) using publicly available whole-genome sequences, and through experimental analyses of 323 DNA samples (from 93 species, including 55 *E. coli* strains that have different phenotypes and the current infectious strain). The findings demonstrated that the kit and protocol have high specificity: no bacterial strain other than *E. coli* O104 had positive amplification results of both target regions. Sensitivity testing indicated that the kit and protocol could detect this bacterium using a DNA concentration as low as ~1 picogram (10-12 g) in the PCR. Additional validation tests on more patient isolates will be carried out within the week.

Provided by Beijing Genomics Institute at Shenzhen