Antibiotic resistance genotyping in Bacteria

Suchitra Thapa

Background:

Antibiotic resistance has become a global burden. Nevertheless, the emerging resistance towards many antimicrobials has limited the therapeutic options. Since most of the infectious diseases are commonly caused by gram negative bacteria in comparison to gram positive bacteria. Therefore, the antimicrobial resistant (AMR) has also been observed higher in gram negative bacteria. Among the gram negative group, Enterbacteriaceae family has been dominancy over other family, in case of disease causation and AMR gene prevalence. Thus, in this project genotyping of the most common bacteria i.e. *E. coli* has been conducted. This project was conducted with an aim to find the prevalence of beta-lactam, aminoglycoside and tetracycline resistance in different regions of the world.

Methodology:

For the project publically available dataset was downloaded from ENA database. The search was based on WGS of *E.coli* which were paired end data. The dataset was from 5 different countries USA, Hongkong, Australia, Japan and Hungary. The dataset was downloaded from the database and quality of the reads were checked using FastQC tool and then MultiQC tool for combined report. All the analysis was done in google cloudshell terminal. Thereafter, Fastp tool was used for trimming the reads, removing the adapters and so on. The trimmed reads were saved as a different file for assembly. Then those trimmed reads were assembled using spades tool. The output file of the spade tool was downloaded. Finally, the contig fasta file from the output file of spade was used for ResFinder database to find any beta-lactam, aminoglycoside and tetracycline resistance gene within the dataset. The final report was a txt file.

Result:

The quality checking showed some reads needed trimming and adapter removal. So, the fastp tool trimmed the adapter on the reads to improve the quality. The following table shows the trimming detail of each of the reads.

		Total Reads	Total bases	Reads passed filter	Low quality reads	Reads with too many N	Adapter trimmed reads	Bases trimmed due to adapter	Duplication	Insert size peak
SRR9873306	R1	703027	105454050	1401710	4344	0	411222	18617128	0.055%	151
	R2	703027	105454050							
SRR13342195	R1	1154773	150307650	2176524	132524	498	73686	2886047	1.91989%	130
	R2	1154773	164196571							

SRR14160921	R1	1319409	181192916	2522866	114780	1172	25448	604051	0.96%	148
	R2	1319409	181626544							
SRR17859194	R1	514136	140785405	693224	335048		52276	1876666	0.046%	241
	R2	514136	142309116							
SRR13220452	R1	760094	182448563	1500430	19758		6988	140285	0.4396%	335
	R2	760094	182559044							

Further, spades assembler assembled the reads into contigs and then the fasta file was uploaded in the ResFinder software for detection of AMR gene. The ResFinder showed that the assembeled contigs did not have any beta-lactam, aminoglycoside and tetracycline resistance gene.

Conclusion:

The AMR gene of beta-lactam, aminoglycoside and tetracycline antibiotics were not found in any of the dataset of all the five countries in the ResFinder database.