Identification of Antimicrobial Resistance Genes - A technical report submitted to HackBio in partial fulfillment of the Genomics workshop completion requirement.

By Aanuoluwa Adekoya - Team Crick/Project1

Introduction

Antimicrobial resistance occurs when microbes - bacteria, fungi, viruses, and parasites are able to evade the effectiveness of antimicrobials used to combat the infections caused by these microbes. It takes full effect when a mutation occurs in the genome of the microbe and it is able to synthesize proteins that can perform this function. According to WHO, antimicrobial resistance remains a threat to the management of diseases as it threatens the effective prevention and treatment of an ever-increasing range of infections caused by microbes. Despite active research in this area of biology, microbes have developed resistance to almost all of the commonly used antimicrobials and the mechanism used is specific to each microbe.

Aim: This project aims to identify the beta-lactam, tetracycline, and aminoglycoside resistance pattern present in *Staphylococcus aureus* in five countries in the world. Our results will provide clues on the resistance pattern in each of these countries and the types of antibiotics that should be prescribed to patients infected by *S. aureus* in these regions.

Importance: Despite the relief that the discovery of antimicrobials and the advent of antimicrobial therapy brought, antimicrobial resistance has remained a threat to the core of modern medicine and global public health. Knowledge of the prevalence of antimicrobial resistance genes will help in optimizing antimicrobials usage in medicine.

Methodology

Data collection: Whole Genome Sequencing (WGS) of *S. aureus* from five countries - Argentina, Canada, China, Nigeria, and Russia were retrieved from the sequence reads archive (SRA) of NCBI (https://www.ncbi.nlm.nih.gov/guide/sitemap/#_V) a public biological database. These countries represent the most populated or second most populated countries in five continents - South America, North America, Asia, Africa, and Europe respectively and ninety-eight reads were retrieved in total.

Computation: (FastP, SPAdes, ResFinder, Matplotlib-Python)
Retrieved data were screened for quality control and fastp was used to trim adapter sequences. The trimmed sequences were assembled using the genome assembler - SPAdes. Post assembly, the contigs of each sequence were uploaded to ResFinder (cge.cbs.dtu.dk/services/ResFinder/), an open online resource for identification of antimicrobial resistance genes in high-throughput sequencing data and prediction of phenotypes from genotypes. A 90% ID threshold was used alongside a 60% minimum length. The resistance genes for beta-lactam, tetracycline, and aminoglycosides were compiled into a CSV file.

Results and Discussion:

This analysis revealed a prevalence of *mecA*, a beta-lactam resistant gene against cefoxitin, and *blaZ*, a beta-lactam resistant gene against penicillin, and variants of *tet* genes against tetracycline in the five countries. The prevalence of penicillin-resistant genes confirms the report of previous studies that have addressed the reduced potency of penicillin as a drug of choice. Reads from China contained a high amount (up to 50%) of aminoglycoside resistance genes which were specific to amikacin, tobramycin, and gentamicin when compared to the other countries (Fig. 1). The project showed that while the aminoglycosides can be used as a drug of choice in Russia, Canada, Argentina, and Nigeria, the use of beta-lactams should be reduced to avoid the spread of antimicrobial resistance genes. Table 2 shows the occurrence of each resistance gene in all of the sample datasets. It is worthy to note that the resistance genes in each country's dataset were of the same or similar origin and these had between 99.89% and 100% identity.
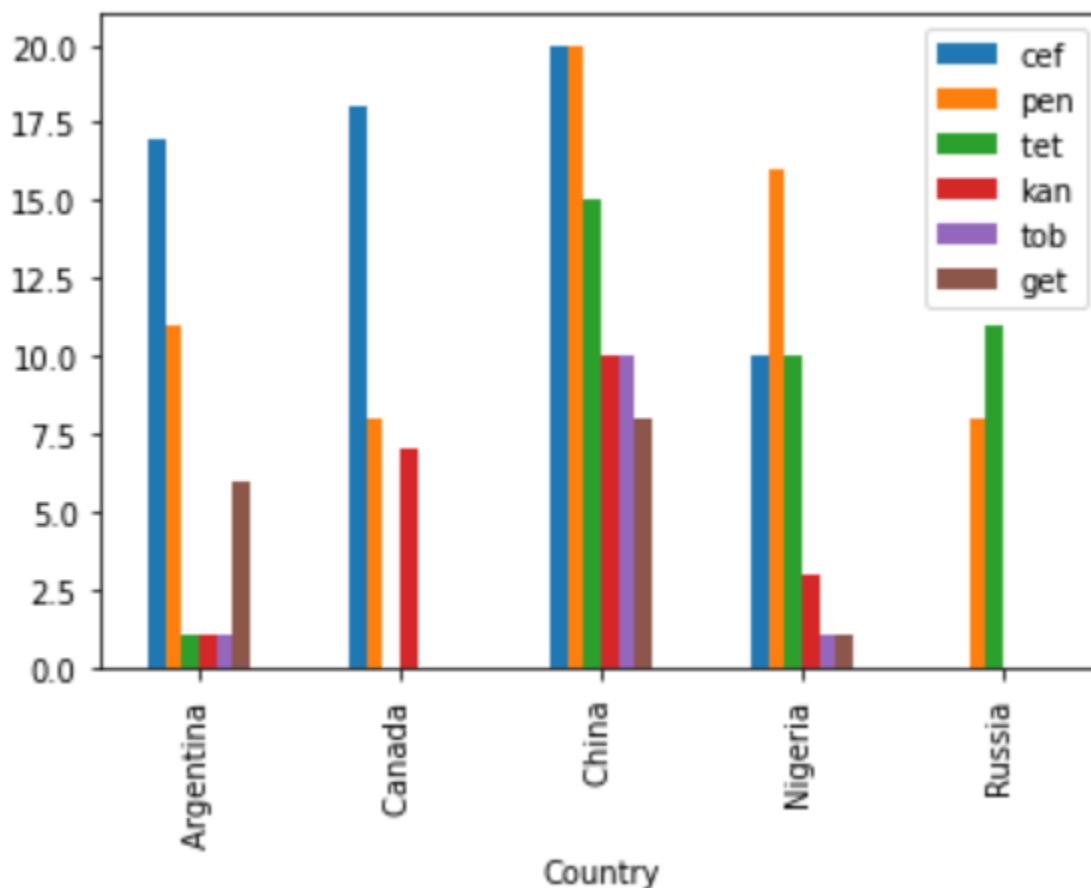

Fig. 1: Prevalence of antimicrobial resistance genes in sample reads. Cef=Cefoxitin, Pen=Penicillin, Tet=Tetracycline, Ami=Amikacin, Tob=Tobramycin, Get=Gentamicin

Link to GitHub Repo

Link to SRR IDs of the dataset and the distribution of the ARGs where 1 means positive and 0 means negative

Link to the table showing country by the country distribution of genes