Variant Calling in *E. coli*

Smita Sahay [1] Niema Moshiri [2]

Pre-Professional Health Academic Program, California State University, East Bay, Hayward, CA[1]

Department of Bioinformatics, University of California, San Diego, San Diego, CA[1]

Abstract

- Microbial resistance to antibiotics is an issue in the treatment of bacterial infections.
- Knowing exact points in the genome of a bacterial strain where mutations (variants) have caused antibiotic resistance can guide doctors toward better treatment.
- We used Next Generation Sequencing (NGS) and open-source bioinformatics tools such as Trimmomatic, FastQC, BWA MEM, VarScan, and Ensembl to identify variants in bacterial genomes that contribute to antibiotic resistance in patients.

Introduction

Patients can often be sick with bacterial infections. In order to treat these infections, it is useful to understand the bacteria's genome, specifically with regards to variants. With the breakthrough of NGS, we can quickly identify these variants. We can align reads to the reference genome and perform variant calling. We can then predict variant effects by comparing against the genome annotation.

Objectives

- To pre-process paired end whole genome sequencing data.
- To identify variants in bacterial genomes.
- To find the genes that contribute to antibiotic resistance in patients once the variants are found to guide treatment.

Methods

- We generated paired-end whole genome sequencing data using Illumina sequencers.
- We used Trimmomatic to trim out low quality bases in the reads.
- Next, we used FastQC to validate the quality of our resulting reads.
- After this, we used BWA MEM to align the reads to the reference genome.
- Lastly, we used VarScan to detect variants in the bacterial genome.
- We predicted variant effects using the Ensembl Variant Effect Predictor, filtering for coding regions only.

Results

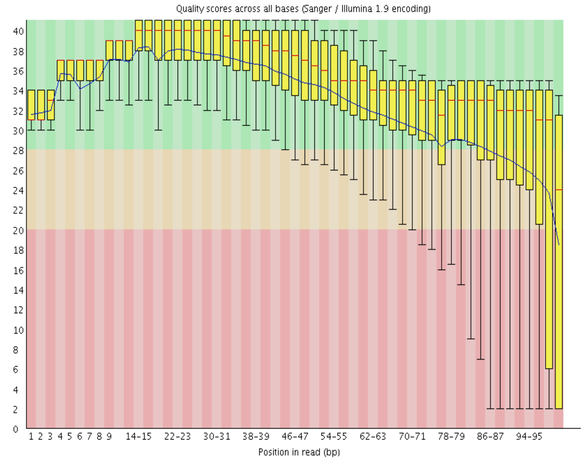

Figure 1: FastQC report for forward reads. Verifies high quality scores and validates reliability of data.

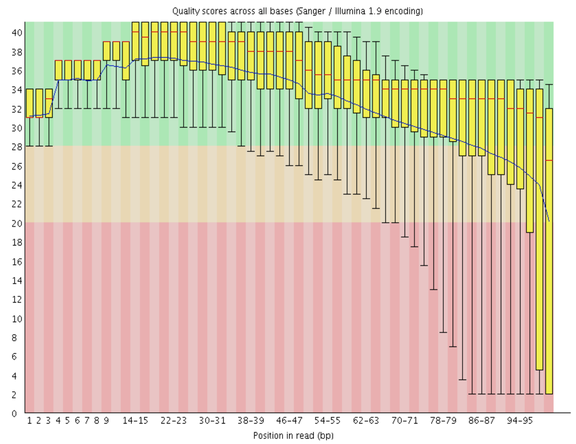

Figure 2: FastQC report for reverse reads. Verifies high quality scores and validates reliability of data.

Position	Reference	Alternate	Gene	Effect
93043	C	G	b0084 ftsI	Missense
482698	T	A	b0462 acrB	Missense
852762	A	G	-	Synonymous
1905761	G	A	b1821 mntP	Missense
3535147	A	C	b3404 envZ	Missense
4390754	G	T	N/A	N/A

Table 1: Shows the position of the SNP, reference allele, the alternative allele, the name of the gene in which the SNP resides, and the effect of the SNP.

Conclusion

- Using NGS techniques and open-source bioinformatics tools, we successfully found target genes associated with antibiotic resistance in the bacterial genome.
- With the power of these tools, we are able to find variants and quickly understand the effects of these variants. This guides the development of better treatments and allows doctors to seek alternative antibiotics for treatment if certain medications are failing to help patients.