



BIOINFORMATICS
INSTITUTE

FINDING VARIANTS FOR ANTIBIOTICS RESISTANCE IN E. COLI K12

December 9, 2023

Homework number 1

by Sapozhnikov N.A.

1 Abstract

Antibiotic resistance in microbial communities poses a significant challenge in the effective treatment of bacterial infections. Bacteria, driven by the evolutionary force, exhibit diverse mechanisms of resistance to antibiotics, demanding precise understanding by scientists and healthcare professionals. This project focuses on unraveling the genomic basis of antibiotic resistance in a strain of *Escherichia coli* (*E. coli*) resistant to ampicillin, utilizing authentic sequencing data. The objective is to identify the specific mutations responsible for conferring antibiotic resistance properties to *E. coli*.

2 Introduction

Microbial resistance to antibiotics stands as a formidable obstacle in the realm of infectious disease management. The relentless force of evolution empowers bacteria to adapt and survive, often employing various strategies to resist the effects of antibiotics. Understanding the intricacies of these resistance mechanisms is paramount for scientists and healthcare professionals.

This project focuses on addressing the antibiotic resistance puzzle within *Escherichia coli*, a bacterial species notorious for its adaptability. The target antibiotic is ampicillin, and our approach involves the analysis of real sequencing data derived from a strain of ampicillin-resistant *E. coli*. The primary goal is to identify the mutations that drive antibiotic resistance. By deciphering the genomic alterations within key genes, we aim to provide insights into the underlying mechanisms of resistance. Subsequently, this knowledge will serve as a foundation for suggesting alternative antibiotics.

3 Methods

To process our Illumina 1.9 sequencing data, firstly, we perform a quality control using FastQC [1]. Then we trim our paired reads using TrimmomaticPE [2]. We used the following parameters while performing the trimming: ‘LEADING:20’: Removes bases from the start of each read if the quality is below 20; ‘TRAILING:20’: Removes bases from the end of each read if the quality is below 20; ‘SLIDINGWINDOW:10:20’: Performs a sliding window trimming with a window size of 10 and trims when the average quality within the window falls below 20; ‘MINLEN:20’: Drops reads below a length of 20. After achieving satisfactory QC results we performed an alignment on the reference (Genome assembly ASM584v2) using BWA-MEM [3] and samtools [4] programs. To distinguish actual mutations from the sequencing errors we have used samtools mpileup command with default parameters. Then we called variants using the VarScan program [5] with `-min-var-frequency` set to 0.50 (50%).

4 Results

After the trimming we’ve achieved the following results: Input Read Pairs: 455876 Both Surviving: 446259 (97.89%) Forward Only Surviving: 9216 (2.02%) Reverse Only Surviving: 273 (0.06%) Dropped: 128 (0.03%). This data was aligned onto the reference genom with 891649 + 0 mapped (99.87% : N/A) reads, which we are satisfied with. In our work we’ve obtained the .vcf file via VarScan with all variants present in our reads. Sadly, due to the technical problems with snpEff they remain unannotated.

5 Discussion

The data obtained is yet to be annotated and described, which is a massive amount of work that we are yet to finish.

6 Bibliography

References

- [1] S. Andrews, F. Krueger, A. Segonds-Pichon, L. Biggins, C. Krueger, and S. Wingett, “FastQC.” Babraham Institute, Jan. 2012.
- [2] A. M. Bolger, M. Lohse, and B. Usadel, “Trimmomatic: a flexible trimmer for Illumina sequence data,” *Bioinformatics*, vol. 30, pp. 2114–2120, Apr. 2014. __eprint: https://academic.oup.com/bioinformatics/article-pdf/30/15/2114/48924714/bioinformatics_30_15_2114.pdf.
- [3] H. Li and R. Durbin, “Fast and accurate short read alignment with burrows–wheeler transform,” *bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.
- [4] P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li, “Twelve years of SAMtools and BCFtools,” *Giga-Science*, vol. 10, p. giab008, Feb. 2021. __eprint: <https://academic.oup.com/gigascience/article-pdf/10/2/giab008/36332246/giab008.pdf>.
- [5] D. C. Koboldt, K. Chen, T. Wylie, D. E. Larson, M. D. McLellan, E. R. Mardis, G. M. Weinstock, R. K. Wilson, and L. Ding, “VarScan: variant detection in massively parallel sequencing of individual and pooled samples,” *Bioinformatics*, vol. 25, pp. 2283–2285, Sept. 2009.