

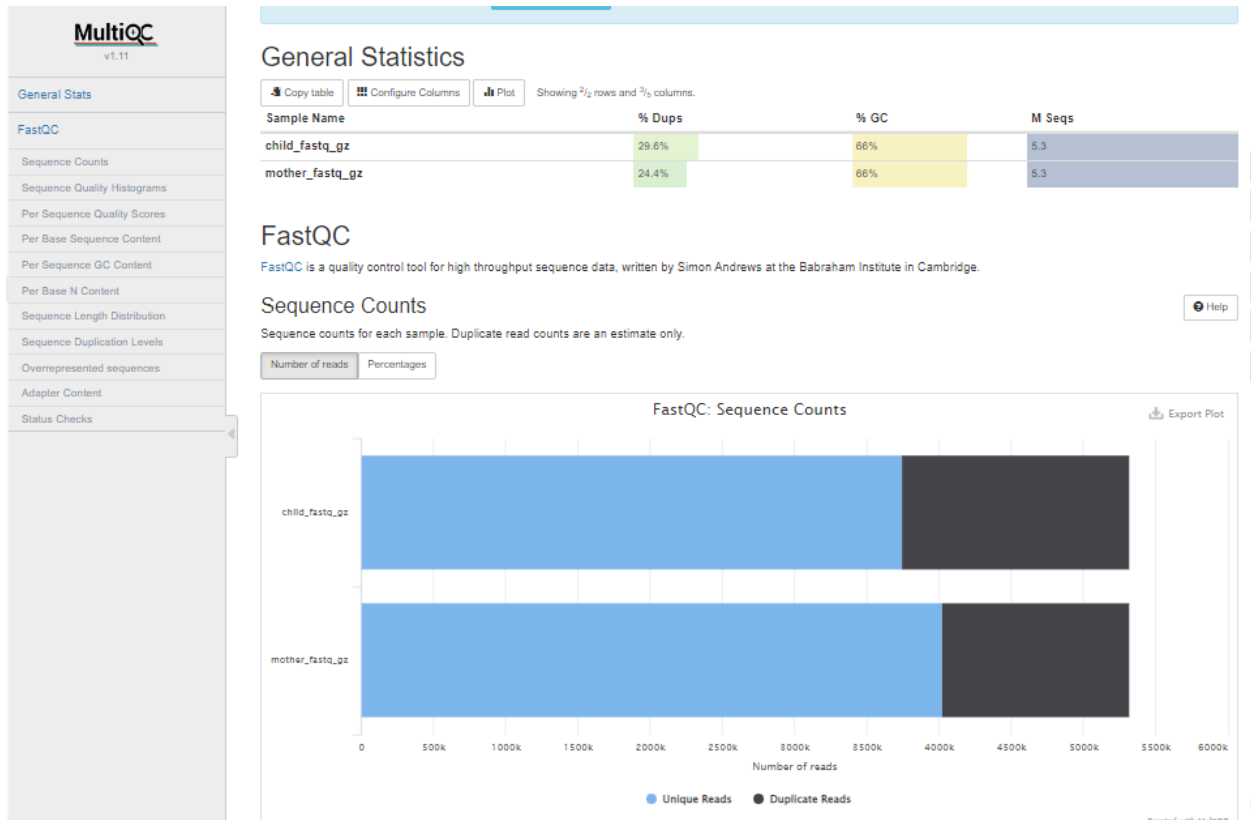
Team Members: Bryan Dang and David Warshawsky

Goal: M. tuberculosis Variant Analysis

Variation in the genome of M. tuberculosis (Mtb) is associated with changes in phenotype, for example drug resistance and virulence. It is also useful for outbreak investigation as the single nucleotide polymorphisms (SNPs) in a sample can be used to build a phylogeny. Therefore the goal of this project is to identify genomic variation in M. tuberculosis through Galaxy and using that to explore drug resistance and other aspects of the bacteria.

| Task Table | | | |
|---|-------------------------------|------------|-------------|
| Tasks | Owner | Due Date | CheckList |
| Running M. tuberculosis Variant Analysis through Galaxy | Bryan Dang & David Warshawsky | 04/27/2023 | X |
| GPTInfomatics_Project_Progress_Report | Bryan Dang & David Warshawsky | 04/27/2023 | X |
| Google Cloud Platform VM Instructions | Bryan Dang & David Warshawsky | 05/06/2023 | In Progress |
| GPTInfomatics_Project_Slides | Bryan Dang & David Warshawsky | 05/06/2023 | |
| GPTInfomatics_Project_Final_Report | Bryan Dang & David Warshawsky | 05/06/2023 | |

Preliminary Results:
MultiQC Report



Tb Profile

TBProfiler report

The following report has been generated by TBProfiler.

Summary

ID: tbprofiler
Date: Tue Apr 25 19:39:56 2023
Strain: lineage4.4.1.1.1
Drug-resistance: MDR-TB
Median Depth: 139

Lineage report

| Lineage | Estimated Fraction | Spoligotype | Rd |
|------------------|--------------------|-------------|------|
| lineage4 | 1.000 | LAM;T;5;X;H | None |
| lineage4.4 | 0.994 | S;T | None |
| lineage4.4.1 | 1.000 | S;T | None |
| lineage4.4.1.1 | 0.998 | S;Orphans | None |
| lineage4.4.1.1.1 | 1.000 | S;Orphans | None |

Resistance report

| Drug | Genotypic Resistance | Mutations |
|--------------------------|----------------------|-------------------------|
| Rifampicin | R | rpoB p.Ser450Leu (0.99) |
| Isoniazid | R | katG p.Ser315Thr (1.00) |
| Ethambutol | R | embB p.Met386Val (1.00) |
| Pyrazinamide | R | pncA p.Tyr34Asp (1.00) |
| Streptomycin | R | rpsL p.Lys43Arg (0.99) |
| Fluoroquinolones | | |
| Moxifloxacin | | |
| Ofloxacin | | |
| Levofloxacin | | |
| Ciprofloxacin | | |
| Aminoglycosides | R | rrs n.1402C>A (0.13) |
| Amikacin | R | rrs n.1402C>A (0.13) |
| Capreomycin | R | rrs n.1402C>A (0.13) |
| Kanamycin | R | rrs n.1402C>A (0.13) |
| Cycloserine | | |
| Ethionamide | | |
| Clofazimine | | |
| Para-aminosalicylic_acid | | |
| Delamanid | | |
| Bedaquiline | | |
| Linezolid | | |

Resistance variants report

| Genome Position | Locus Tag | Gene | Variant Type | Change | Estimated Fraction | Drug | Annotation |
|-----------------|-----------------|------|------------------------------------|-------------|--------------------|---|--|
| 761155 | Rv0667 | rpoB | missense_variant | p.Ser450Leu | 0.991 | rifampicin | type=who_confidence drug=rifampicin who_confidence=Assoc w R |
| 781687 | Rv0682 | rpsL | missense_variant | p.Lys43Arg | 0.987 | streptomycin | type=who_confidence drug=streptomycin who_confidence=Assoc w R |
| 1473247 | EBG080000313325 | rrs | non_coding_transcript_exon_variant | n.1402C>A | 0.128 | kanamycin, capreomycin, aminoglycosides, amikacin | type=who_confidence drug=kanamycin who_confidence=;type=who_confidence drug=capreomycin who_confidence=;type=who_confidence drug=aminoglycosides who_confidence=;type=who_confidence drug=amikacin who_confidence= |
| 2155168 | Rv1908c | katG | missense_variant | p.Ser315Thr | 1.000 | isoniazid | type=who_confidence drug=isoniazid who_confidence=Assoc w R |
| 2289142 | Rv2043c | pncA | missense_variant | p.Tyr34Asp | 1.000 | pyrazinamide | type=who_confidence drug=pyrazinamide who_confidence=Assoc w R |
| 4247429 | Rv3795 | embB | missense_variant | p.Met386Val | 1.000 | ethambutol | type=who_confidence drug=ethambutol who_confidence=Assoc w R |

Variants Report

Texttransformationondata32 Report

| Lineage | Family | Spoligotype | RD | Agreement | | | | | | |
|-------------------|------------------------|-------------|---------|--------------------------------|--------|--------|-------|-------|----------------|---------------------------|
| lineage4 | Euro-American | LAM;T;S;X;H | None | 1.0 | | | | | | |
| lineage4.4 | Euro-American | S;T | None | 0.9937369519832986 | | | | | | |
| lineage4.4.1 | Euro-American (S-type) | S;T | None | 1.0 | | | | | | |
| lineage4.4.1.1 | Euro-American | S;Orphans | None | 0.9977307110438729 | | | | | | |
| lineage4.4.1.1.1 | Euro-American | S;Orphans | None | 1.0 | | | | | | |
| CHR | GENE | LOCUS | PRODUCT | TYPE | EFFECT | POS | REF | ALT | BIOTYPE | NUCLEOTIDE |
| ChromosomeRv0002 | Rv0002 | P9WNJ1 | snp | synonymous_variant | | 2532 | C | T | protein_coding | c.481C>T |
| ChromosomeRv0006 | Rv0006 | P9WGA7 | snp | synonymous_variant | | 9143 | C | T | protein_coding | c.1842C>T |
| ChromosomeRv0010c | Rv0010c | P9WMA3 | snp | synonymous_variant | | 13460 | G | A | protein_coding | c.99C>T |
| ChromosomeRv0012 | Rv0012 | L0T243 | snp | missense_variant | | 14401 | G | A | protein_coding | c.313G>A |
| ChromosomeRv0013 | Rv0013 | P9WN35 | snp | synonymous_variant | | 15036 | G | C | protein_coding | c.123C>G |
| ChromosomeRv0013 | Rv0013 | P9WN35 | snp | missense_variant | | 15117 | G | C | protein_coding | c.204G>C |
| ChromosomeRv0018c | Rv0018c | P9WHW5 | snp | missense_variant | | 21819 | C | A | protein_coding | c.1363G>T |
| ChromosomeRv0018c | Rv0018c | P9WHW5 | snp | missense_variant | | 23174 | A | C | protein_coding | c.8T>G |
| ChromosomeRv0030 | Rv0030 | P9WM95 | snp | synonymous_variant | | 33436 | T | C | protein_coding | c.213C>T |
| ChromosomeRv0032 | Rv0032 | P9WQ85 | snp | synonymous_variant | | 36538 | T | T | protein_coding | c.2244C>T |
| ChromosomeRv0036c | Rv0036c | P9WM91 | snp | synonymous_variant | | 39758 | G | A | protein_coding | c.72C>T |
| ChromosomeRv0039c | Rv0039c | P9WM89 | snp | missense_variant | | 42281 | A | C | protein_coding | c.71T>G |
| ChromosomeRv0041 | Rv0041 | P9WVF1 | snp | synonymous_variant | | 43945 | G | A | protein_coding | c.384G>A |
| ChromosomeRv0046c | Rv0046c | P9WKJ1 | snp | missense_variant | | 50557 | C | T | protein_coding | c.568G>A |
| ChromosomeRv0048c | Rv0048c | P9WM87 | snp | missense_variant | | 51949 | G | A | protein_coding | c.749C>T |
| ChromosomeRv0050 | Rv0050 | P71707 | snp | synonymous_variant | | 54394 | G | A | protein_coding | c.732G>A |
| ChromosomeRv0050 | Rv0050 | P71707 | indel | conservative_inframe_insertion | | 55533 | T | TGCC | protein_coding | c.1888_1890dupCCG |
| ChromosomeRv0051 | Rv0051 | P71708 | snp | missense_variant | | 55813 | G | A | protein_coding | c.118G>A |
| ChromosomeRv0059 | Rv0059 | O53604 | snp | synonymous_variant | | 63214 | C | T | protein_coding | c.15C>T |
| ChromosomeRv0063 | Rv0063 | O53608 | snp | synonymous_variant | | 67792 | C | T | protein_coding | c.870C>T |
| ChromosomeRv0064 | Rv0064 | P9WFL5 | snp | synonymous_variant | | 68688 | T | T | protein_coding | c.69C>T |
| ChromosomeRv0064 | Rv0064 | P9WFL5 | snp | missense_variant | | 70267 | G | T | protein_coding | c.1648T>G |
| ChromosomeRv0075 | Rv0075 | O53620 | snp | synonymous_variant | | 84908 | C | T | protein_coding | c.913C>T |
| ChromosomeRv0080 | Rv0080 | P9WMA5 | snp | missense_variant | | 89200 | T | G | protein_coding | c.179T>G |
| ChromosomeRv0092 | Rv0092 | P9WPU1 | snp | missense_variant | | 102348 | T | C | protein_coding | c.1766T>C |
| ChromosomeRv0093c | Rv0093c | P9WM69 | snp | missense_variant | | 103600 | G | A | protein_coding | c.64C>T |
| ChromosomeRv0095c | Rv0095c | O10891 | indel | missense_variant | | 105007 | CTCAG | GTCAA | protein_coding | c.205_209delCTGAGinsTTGAC |

Drug Resistance Report

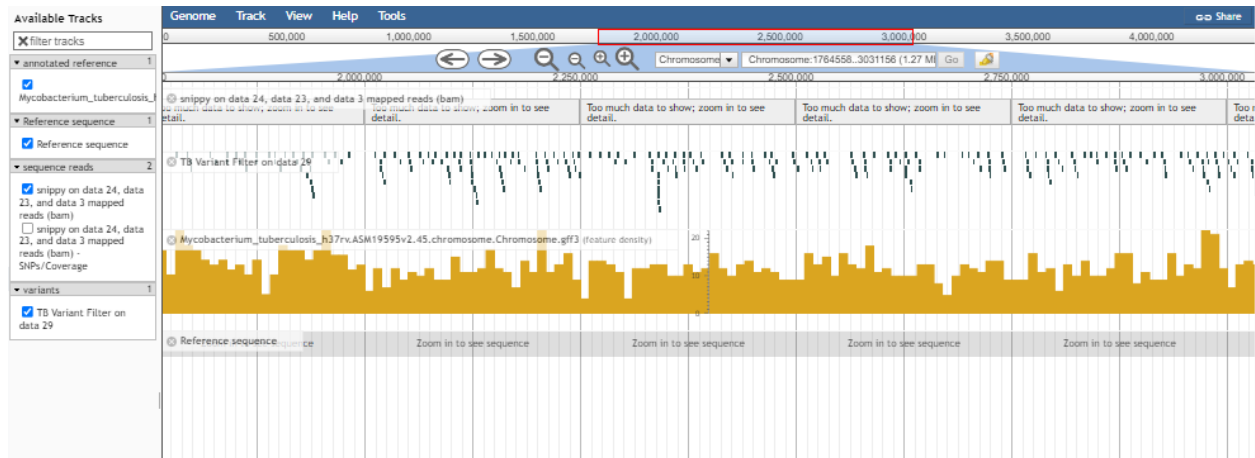
Texttransformationondata32 Report

TBProfiler Drug Resistance Report

| Disclaimer: Drug resistance predictions are for Research Purposes Only and are produced by TBProfiler. | | |
|--|------------|-------------------------|
| Drug | Resistance | Supporting Mutations |
| Isoniazid | R | katG (1.0 p.Ser315Thr) |
| Rifampicin | R | rpoB (0.99 p.Ser450Leu) |
| Ethambutol | R | embB (1.0 p.Met306Val) |
| Pyrazinamide | R | pncA (1.0 p.Tyr34Asp) |
| Streptomycin | R | rpsL (0.99 p.Lys43Arg) |
| Ethionamide | S | |
| Fluoroquinolones | S | |
| Amikacin | R | rrs (0.13 n.1402C>A) |
| Capreomycin | R | rrs (0.13 n.1402C>A) |
| Kanamycin | R | rrs (0.13 n.1402C>A) |
| Para-aminosalicylic acid | S | |
| Linezolid | S | |
| Cycloserine | S | |
| Delamanid | S | |
| Clofazimine | S | |
| Bedaquiline | S | |
| Aminoglycosides | R | rrs (0.13 n.1402C>A) |

\$(function () { var table = \$('#cfreport').DataTable({ lengthChange: false, buttons: [{ extend: 'csv', text: 'Download CSV', filename: 'Texttransformationondata32' }] }); table.buttons().container().appendTo(\$('#cfreport_wrapper').col-md-6.eq(0)); });

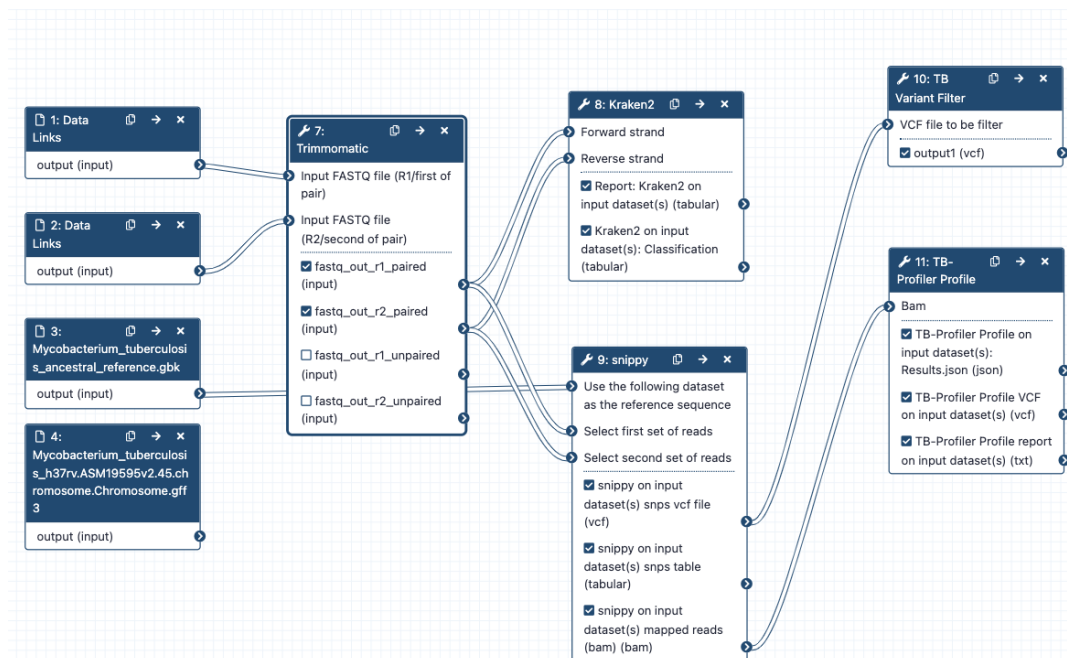
JBrowse Results



Workflow up until the TB-Profiler Profile

Explanation

- Data Retrieval
- EDA(Exploratory Data Analysis) with FastQC→MultiQC
- Quality Trimming through Trimmomatic. Unpaired reads unused.
- Kraken2 contamination analysis binomial nomenclature manual analysis.
- Snippy variant identification with annotated ancestral genome.
- TB Variant Filter to screen variants determined through know high variance genes.
- TB Profiler creates the actual report with drug resistance and lineages.



- TB Variant Report makes the profiler report more readable and summarized.

Command Line

```
(base) davidawarshawsky@dwarshawsky-cs123b:~/sjsu-cs123b/tb-variant-analysis$ pwd
/home/davidawarshawsky/sjsu-cs123b/tb-variant-analysis
(base) davidawarshawsky@dwarshawsky-cs123b:~/sjsu-cs123b/tb-variant-analysis$ ls
004-2_1.fastq.gz  Mycobacterium_tuberculosis_ancestral_reference.gbk
004-2_2.fastq.gz  Mycobacterium_tuberculosis_h37rv.ASM19595v2.45.chromosome.Chromosome.gff3
(base) davidawarshawsky@dwarshawsky-cs123b:~/sjsu-cs123b/tb-variant-analysis$
```