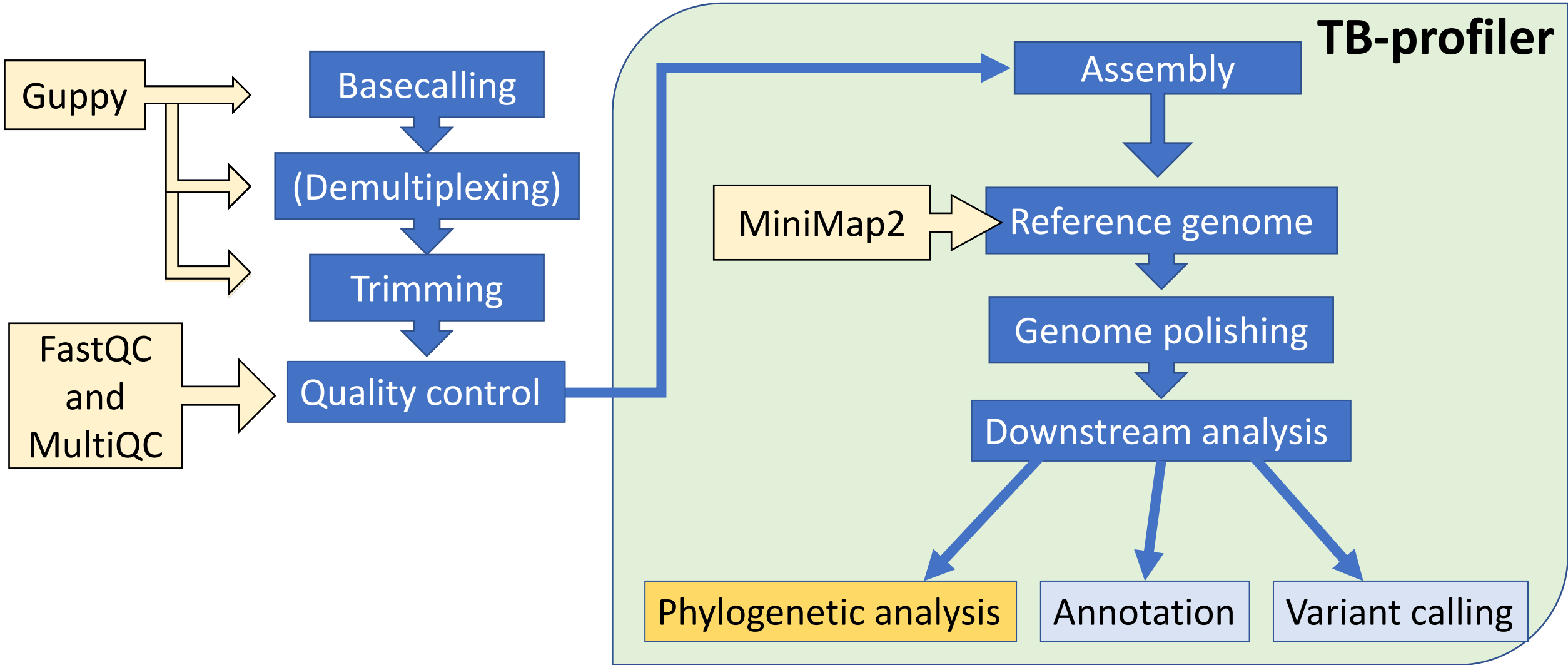
Common file types and using TBProfiler

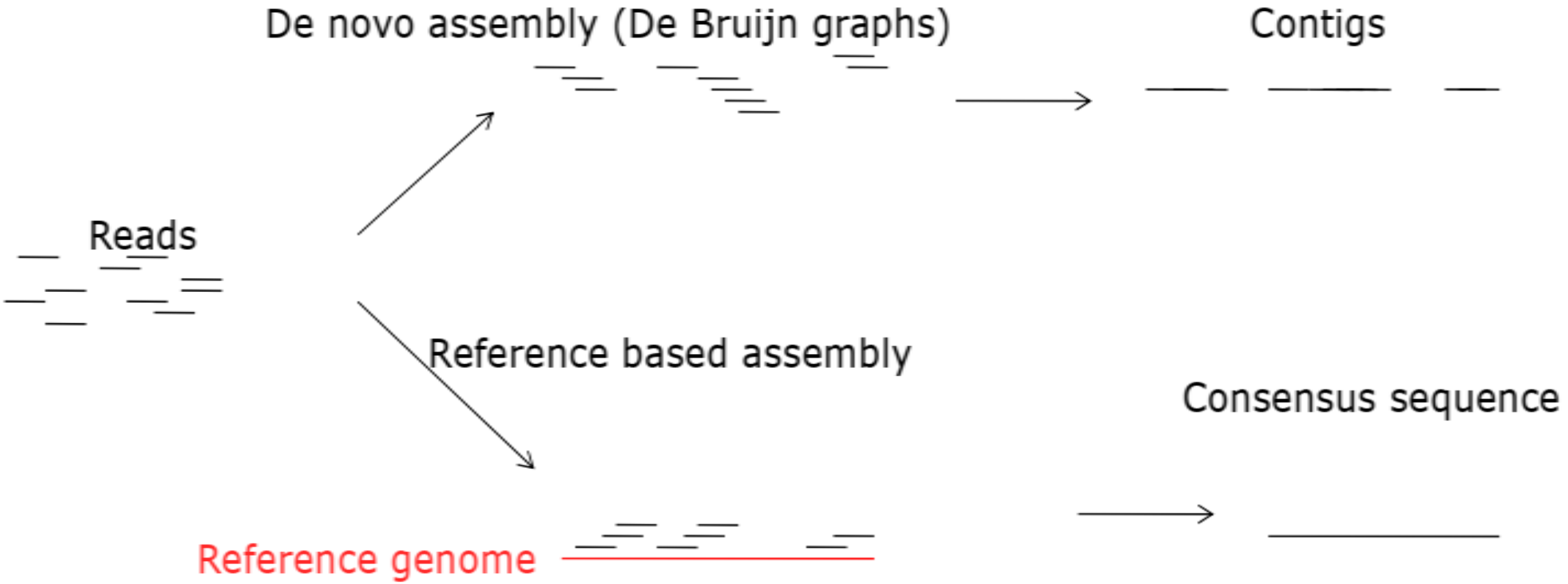
Dr Linzy Elton

We will record these sessions and put them online so you can refer back to them later on

We will also put the slides up online so you can access the notes (links and image credits)

ONT MTB Bioinformatics Pipeline





Reference Genome:
MtbH37Rv_AL123456.embl

Some common file types you might encounter:

- Fast5 (.fast5) – discussed in previous session
- Fastq (.fastq) – discussed in previous session
- **SAM (.sam)**
- **BAM (.bam)**
- **Fasta (.fna or .fasta)**
- **VCF (.vcf)**

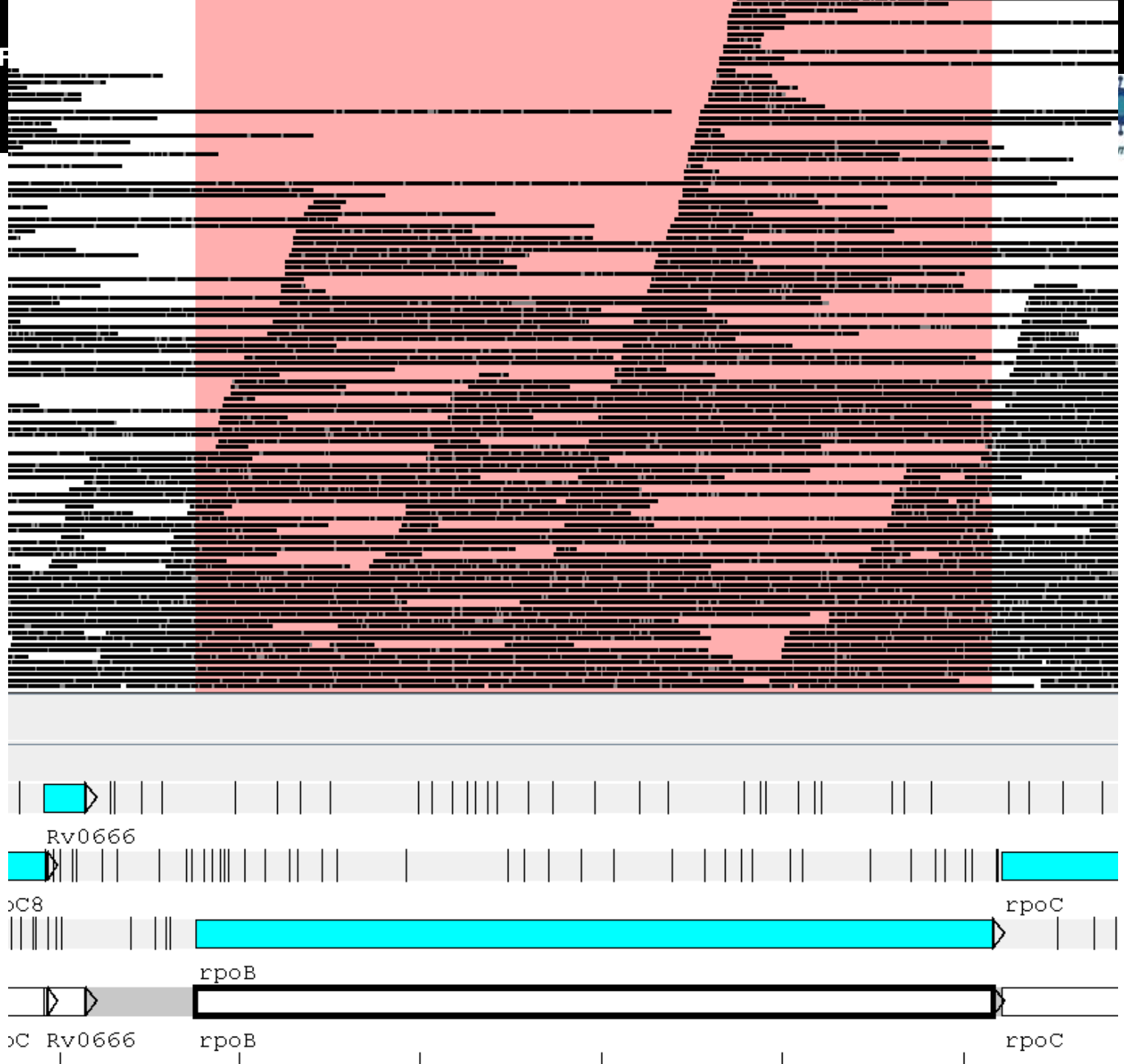
SAM/BAM files

SAM

- The Sequence Alignment/Map file (.sam) is a file format to save alignment information of reads mapped against reference genome sequence
- It usually starts with a header section followed by alignment information as tab separated lines for each read

BAM

- The Binary Alignment/Map file (. bam) is the compressed binary version of a SAM file that is used to represent aligned sequences
- BAM file has an index file ".bam.bai" (this file acts like an external table of contents, and allows programs to jump directly to specific parts of the bam file without reading through all the sequences)
- Not Human readable but you can open it on e.g. Artemis (a Genome Viewer by Sanger)


Rv0666
rpoC
rpoB
rpoB
rpoC

FASTA files

- File extension .fna or .fasta
- This file format is used for reference genomes or reference database sequences
- A FASTA file is the consensus sequence formed from a SAM/BAM file (the calculated order of most frequent residues at each position)
- In this file format, each sequence read is defined in 2 lines: header + sequence code

Sequence header/ID

```
>NC_015758.1 Mycobacterium tuberculosis variant africanum GM041182, complete genome
TTGACCGATGACCCCGGTTCAGGCTTCACCACAGTGTGGAACGCGGTCGTCTCCGAACTTAACGGCGACC
CTAAGGTTGACGACGGACCCAGCAGTGATGCTAATCTCAGCGCTCCGCTGACCCCTCAGCAAAGGGCTTG
```

VCF

- Variant calling format (.vcf) (you will need to unzip it if it is zipped)
- Used to identify SNPs (Single Nucleotide Polymorphism) and Insertions and/or deletions (INDELs)
- VCF files can be viewed in genome viewers or in Notepad
- Developed for large scale data storage
- You can set filter the output so you only see e.g. SNPs or insertions, as well as changing the quality threshold

```
##bcftools_viewVersion=1.10.2+htslib-1.10.2
##bcftools_viewCommand=view -c1 -a -Oz -o pipeline/callers/tbprofiler_test_run/TB-profiler/
##bcftools_viewCommand=view -v snps pipeline/callers/tbprofiler_test_run/TB-profiler/4e84cd
##bcftools/csqVersion=1.10.2+htslib-1.10.2
##bcftools/csqCommand=csq -p m -f /lustre/projects/CCM_shared_NGS_data/sylvia/.snakemake/co
##INFO=<ID=BCSQ,Number=.,Type=String,Description="Haplotype-aware consequence annotation fr
##FORMAT=<ID=BCSQ,Number=.,Type=Integer,Description="Bitmask of indexes to INFO/BCSQ, with
##bcftools_concatCommand=concat 2c672f8c-6cde-440b-9860-bfa1927f541b.vcf e8ee7b0d-7222-4b88-
#CHROM  POS ID  REF ALT QUAL    FILTER  INFO    FORMAT  ERR1664619
Chromosome  6140    .   G   T   225 PASS    DP=40;VDB=0.970148;SGB=-0.693145;MQSB=1;MQ0F=0;A
Chromosome  7362    .   G   C   225 PASS    DP=46;VDB=0.682479;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  7582    .   A   C   225 PASS    DP=58;VDB=0.999803;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  7585    .   G   C   225 PASS    DP=58;VDB=0.999706;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  9304    .   G   A   225 PASS    DP=48;VDB=0.261952;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  761155  .   C   T   225 PASS    DP=35;VDB=0.662388;SGB=-0.693136;MQSB=1;MQ0F=0;A
Chromosome  761998  .   T   C   228 PASS    DP=56;VDB=0.159622;SGB=-0.693147;RPB=1;MQB=1;MQS
Chromosome  764995  .   C   G   225 PASS    DP=54;VDB=0.998895;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  781395  .   T   C   225 PASS    DP=43;VDB=0.7789;SGB=-0.693146;MQSB=1;MQ0F=0;AC=
Chromosome  1473246 .   A   G   225 PASS    DP=47;VDB=0.846133;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  1673425 .   C   T   225 PASS    DP=31;VDB=0.442172;SGB=-0.69311;MQSB=1;MQ0F=0;AC
Chromosome  1674782 .   T   C   225 PASS    DP=39;VDB=0.923491;SGB=-0.693144;MQSB=1;MQ0F=0;A
Chromosome  1917972 .   A   G   225 PASS    DP=37;VDB=0.860765;SGB=-0.693141;MQSB=1;MQ0F=0;A
Chromosome  2288868 .   A   C   225 PASS    DP=35;VDB=0.788686;SGB=-0.693136;MQSB=1;MQ0F=0;A
Chromosome  3067464 .   A   G   225 PASS    DP=49;VDB=0.35383;SGB=-0.693147;MQSB=1;MQ0F=0;AC
```

```
##bcftools_viewVersion=1.10.2+htslib-1.10.2
##bcftools_viewCommand=view -c1 -a -Oz -o pipeline/callers/tbprofiler_test_run/TB-profiler/
##bcftools_viewCommand=view -v snps pipeline/callers/tbprofiler_test_run/TB-profiler/4e84cd
##bcftools/csqVersion=1.10.2+htslib-1.10.2
##bcftools/csqCommand=csq -p m -f /lustre/projects/CCM_shared_NGS_data/sylvia/.snakemake/co
##INFO=<ID=BCSQ,Number=.,Type=String,Description="Haplotype-aware consequence annotation fr
##FORMAT=<ID=BCSQ,Number=.,Type=Integer,Description="Bitmask of indexes to INFO/BCSQ, with
##bcftools_concatCommand=concat 2c672f8c-6cde-440b-9860-bfa1927f541b.vcf e8ee7b0d-7222-4b88-
#CHROM  POS     ID      REF ALT QUAL    FILTER  INFO    FORMAT  ERR1664619
Chromosome      6140    .       G       T       225     PASS    DP=40;VDB=0.970148;SGB=-0.693145;MQSB=1;MQ0F=0;A
Chromosome      7362    .       G       C       225     PASS    DP=46;VDB=0.682479;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome      7582    .       A       C       225     PASS    DP=58;VDB=0.999803;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome      7585    .       G       C       225     PASS    DP=58;VDB=0.999706;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome      9304    .       G       A       225     PASS    DP=48;VDB=0.261952;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome      761155  .       C       T       225     PASS    DP=35;VDB=0.662388;SGB=-0.693136;MQSB=1;MQ0F=0;A
Chromosome      761998  .       T       C       228     PASS    DP=56;VDB=0.159622;SGB=-0.693147;RPB=1;MQB=1;MQS
Chromosome      764995  .       C       G       225     PASS    DP=54;VDB=0.998895;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome      781395  .       T       C       225     PASS    DP=43;VDB=0.7789;SGB=-0.693146;MQSB=1;MQ0F=0;AC=
Chromosome      1473246 .       A       G       225     PASS    DP=47;VDB=0.846133;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome      1673425 .       C       T       225     PASS    DP=31;VDB=0.442172;SGB=-0.69311;MQSB=1;MQ0F=0;AC
Chromosome      1674782 .       T       C       225     PASS    DP=39;VDB=0.923491;SGB=-0.693144;MQSB=1;MQ0F=0;A
Chromosome      1917972 .       A       G       225     PASS    DP=37;VDB=0.860765;SGB=-0.693141;MQSB=1;MQ0F=0;A
Chromosome      2288868 .       A       C       225     PASS    DP=35;VDB=0.788686;SGB=-0.693136;MQSB=1;MQ0F=0;A
Chromosome      3067464 .       A       G       225     PASS    DP=49;VDB=0.35383;SGB=-0.693147;MQSB=1;MQ0F=0;AC
```

```
##bcftools_viewVersion=1.10.2+htslib-1.10.2
##bcftools_viewCommand=view -c1 -a -Oz -o pipeline/callers/tbprofiler_test_run/TB-profiler/
##bcftools_viewCommand=view -v snps pipeline/callers/tbprofiler_test_run/TB-profiler/4e84cd
##bcftools/csqVersion=1.10.2+htslib-1.10.2
##bcftools/csqCommand=csq -p m -f /lustre/projects/CCM_shared_NGS_data/sylvia/.snakemake/co
##INFO=<ID=BCSQ,Number=.,Type=String,Description="Haplotype-aware consequence annotation fr
##FORMAT=<ID=BCSQ,Number=.,Type=Integer,Description="Bitmask of indexes to INFO/BCSQ, with
##bcftools_concatCommand=concat 2c672f8c-6cde-440b-9860-bfa1927f541b.vcf e8ee7b0d-7222-4b88-
#CHROM  POS ID  REF ALT QUAL    FILTER  INFO    FORMAT  ERR1664619
Chromosome  6140    .   G   T   225 PASS    DP=40;VDB=0.970148;SGB=-0.693145;MQSB=1;MQ0F=0;A
Chromosome  7362    .   G   C   225 PASS    DP=46;VDB=0.682479;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  7582    .   A   C   225 PASS    DP=58;VDB=0.999803;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  7585    .   G   C   225 PASS    DP=58;VDB=0.999706;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  9304    .   G   A   225 PASS    DP=48;VDB=0.261952;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  761155  .   C   T   225 PASS    DP=35;VDB=0.662388;SGB=-0.693136;MQSB=1;MQ0F=0;A
Chromosome  761998  .   T   C   228 PASS    DP=56;VDB=0.159622;SGB=-0.693147;RPB=1;MQB=1;MQS
Chromosome  764995  .   C   G   225 PASS    DP=54;VDB=0.998895;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  781395  .   T   C   225 PASS    DP=43;VDB=0.7789;SGB=-0.693146;MQSB=1;MQ0F=0;AC=
Chromosome  1473246 .   A   G   225 PASS    DP=47;VDB=0.846133;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  1673425 .   C   T   225 PASS    DP=31;VDB=0.442172;SGB=-0.69311;MQSB=1;MQ0F=0;AC
Chromosome  1674782 .   T   C   225 PASS    DP=39;VDB=0.923491;SGB=-0.693144;MQSB=1;MQ0F=0;A
Chromosome  1917972 .   A   G   225 PASS    DP=37;VDB=0.860765;SGB=-0.693141;MQSB=1;MQ0F=0;A
Chromosome  2288868 .   A   C   225 PASS    DP=35;VDB=0.788686;SGB=-0.693136;MQSB=1;MQ0F=0;A
Chromosome  3067464 .   A   G   225 PASS    DP=49;VDB=0.35383;SGB=-0.693147;MQSB=1;MQ0F=0;AC
```

```
##bcftools_viewVersion=1.10.2+htslib-1.10.2
##bcftools_viewCommand=view -c1 -a -Oz -o pipeline/callers/tbprofiler_test_run/TB-profiler/
##bcftools_viewCommand=view -v snps pipeline/callers/tbprofiler_test_run/TB-profiler/4e84cd
##bcftools/csqVersion=1.10.2+htslib-1.10.2
##bcftools/csqCommand=csq -p m -f /lustre/projects/CCM_shared_NGS_data/sylvia/.snakemake/co
##INFO=<ID=BCSQ,Number=.,Type=String,Description="Haplotype-aware consequence annotation fr
##FORMAT=<ID=BCSQ,Number=.,Type=Integer,Description="Bitmask of indexes to INFO/BCSQ, with
##bcftools_concatCommand=concat 2c672f8c-6cde-440b-9860-bfa1927f541b.vcf e8ee7b0d-7222-4b88-
#CHROM  POS ID  REF ALT QUAL    FILTER  INFO    FORMAT  ERR1664619
Chromosome  6140    .   G   T   225 PASS    DP=40;VDB=0.970148;SGB=-0.693145;MQSB=1;MQ0F=0;A
Chromosome  7362    .   G   C   225 PASS    DP=46;VDB=0.682479;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  7582    .   A   C   225 PASS    DP=58;VDB=0.999803;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  7585    .   G   C   225 PASS    DP=58;VDB=0.999706;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  9304    .   G   A   225 PASS    DP=48;VDB=0.261952;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  761155  .   C   T   225 PASS    DP=35;VDB=0.662388;SGB=-0.693136;MQSB=1;MQ0F=0;A
Chromosome  761998  .   T   C   228 PASS    DP=56;VDB=0.159622;SGB=-0.693147;RPB=1;MQB=1;MQS
Chromosome  764995  .   C   G   225 PASS    DP=54;VDB=0.998895;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  781395  .   T   C   225 PASS    DP=43;VDB=0.7789;SGB=-0.693146;MQSB=1;MQ0F=0;AC=
Chromosome  1473246 .   A   G   225 PASS    DP=47;VDB=0.846133;SGB=-0.693147;MQSB=1;MQ0F=0;A
Chromosome  1673425 .   C   T   225 PASS    DP=31;VDB=0.442172;SGB=-0.69311;MQSB=1;MQ0F=0;AC
Chromosome  1674782 .   T   C   225 PASS    DP=39;VDB=0.923491;SGB=-0.693144;MQSB=1;MQ0F=0;A
Chromosome  1917972 .   A   G   225 PASS    DP=37;VDB=0.860765;SGB=-0.693141;MQSB=1;MQ0F=0;A
Chromosome  2288868 .   A   C   225 PASS    DP=35;VDB=0.788686;SGB=-0.693136;MQSB=1;MQ0F=0;A
Chromosome  3067464 .   A   G   225 PASS    DP=49;VDB=0.35383;SGB=-0.693147;MQSB=1;MQ0F=0;AC
```

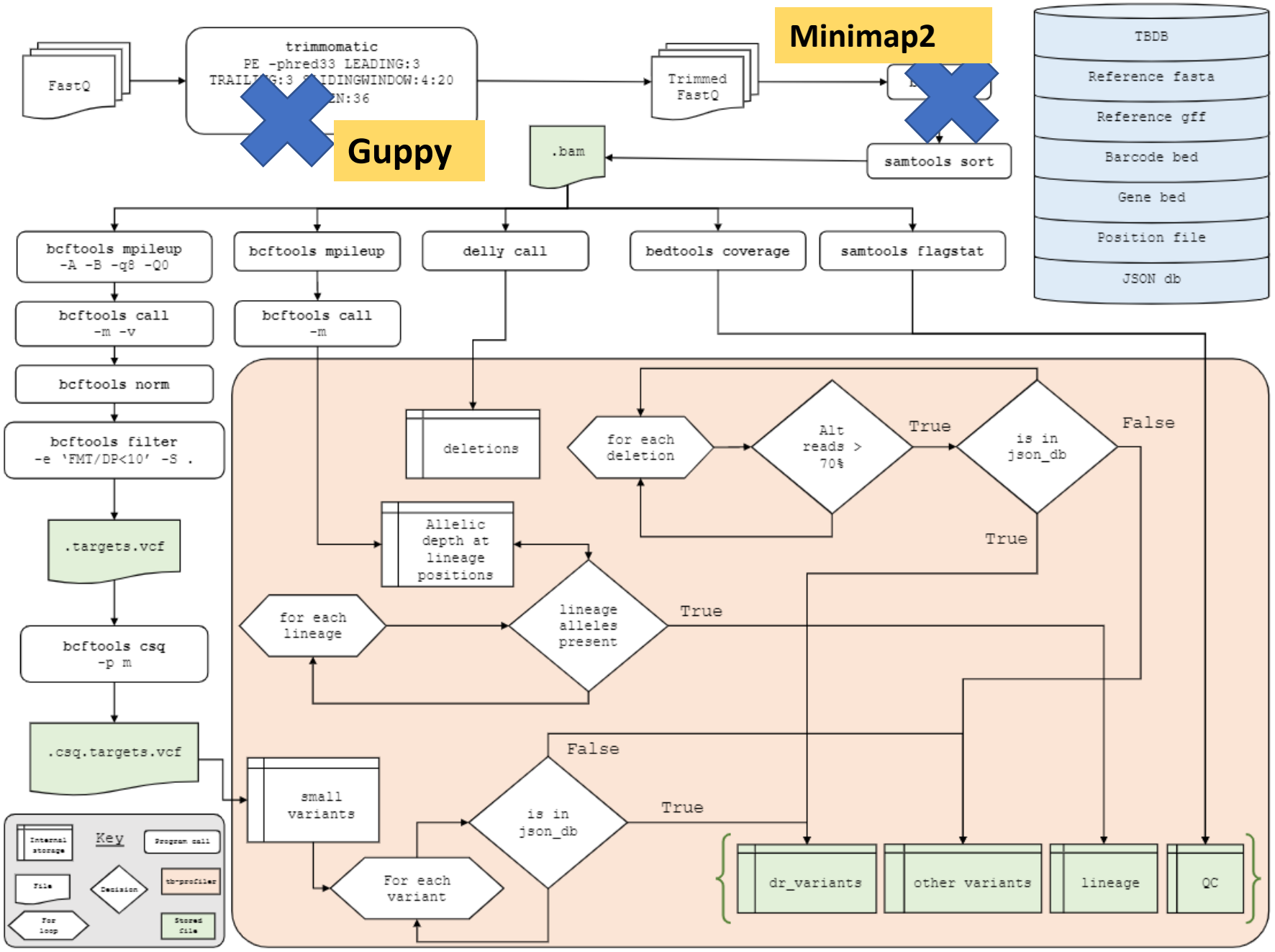
```
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##bcftoolsVersion=1.10.2+htslib-1.10.2
##bcftoolsCommand=mpileup -f /lustre/projects/CCM_shared_NGS_data/sylvia/.snakemake/conda/2c3a442eb1474c51b4a96ffb80e4b315/share/tbprofiler/tbdb.fasta -
##reference=file:///lustre/projects/CCM_shared_NGS_data/sylvia/.snakemake/conda/2c3a442eb1474c51b4a96ffb80e4b315/share/tbprofiler/tbdb.fasta
##contig=<ID=Chromosome,length=4411532>
##ALT=<ID=*,Description="Represents allele(s) other than observed.">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=IDV,Number=1,Type=Integer,Description="Maximum number of raw reads supporting an indel">
##INFO=<ID=IMF,Number=1,Type=Float,Description="Maximum fraction of raw reads supporting an indel">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=<ID=VDB,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts in RNA-seq data (bigger is better)",Version="3
##INFO=<ID=RPB,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
##INFO=<ID=MQB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
##INFO=<ID=BQB,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
##INFO=<ID=MQSB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
##INFO=<ID=SGB,Number=1,Type=Float,Description="Segregation based metric.">
##INFO=<ID=MQ0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Number of high-quality bases">
##FORMAT=<ID=AD,Number=R,Type=Integer,Description="Allelic depths (high-quality bases)">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=ICB,Number=1,Type=Float,Description="Inbreeding Coefficient Binomial test (bigger is better)">
##INFO=<ID=HOB,Number=1,Type=Float,Description="Bias in the number of HOMs number (smaller is better)">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="Number of high-quality ref-forward , ref-reverse, alt-forward and alt-reverse bases">
##INFO=<ID=MQ,Number=1,Type=Integer,Description="Average mapping quality">
##bcftools_callVersion=1.10.2+htslib-1.10.2
##bcftools_callCommand=call -mv; Date=Wed Sep 29 14:19:35 2021
##bcftools_normVersion=1.10.2+htslib-1.10.2
##bcftools_normCommand=norm -f /lustre/projects/CCM_shared_NGS_data/sylvia/.snakemake/conda/2c3a442eb1474c51b4a96ffb80e4b315/share/tbprofiler/tbdb.fasta
##bcftools_filterVersion=1.10.2+htslib-1.10.2
##bcftools_filterCommand=filter -t Chromosome:4998-7267 -e FMT/DP<10 -S . -Oz -o pipeline/callers/tbprofiler_test_run/TB-profiler/4e84cd95-3486-4efc-af2
##bcftools_concatVersion=1.10.2+htslib-1.10.2
##bcftools_concatCommand=concat -aD pipeline/callers/tbprofiler_test_run/TB-profiler/4e84cd95-3486-4efc-af21-63affd161753_Chromosome_4998_7267.vcf.gz pi
```

Normal text file | length : 15,069 lines : 72 | Ln : 1 Col : 1 Pos : 1 | Unix (LF) | UTF-8 | INS

TBProfiler

- Can detect drug resistance-conferring mutations and predict drug-resistant phenotypes
- Two versions: a browser-based one and a command line one (you will need Linux or MacOS for this)
- The pipeline is rapidly developing so make sure you use the latest version (currently v.4.4.1)
- Works with Nanopore data and paired reads from Illumina
- Analysis pipelines is pretty much standardised; the database is manually curated database hosted on separate repository constantly updated, based on emergence of new evidence for the inclusion or exclusion of mutations affect the Accuracy of the AMR prediction

TB-profiler Pipeline workflow sketch (Illumina)

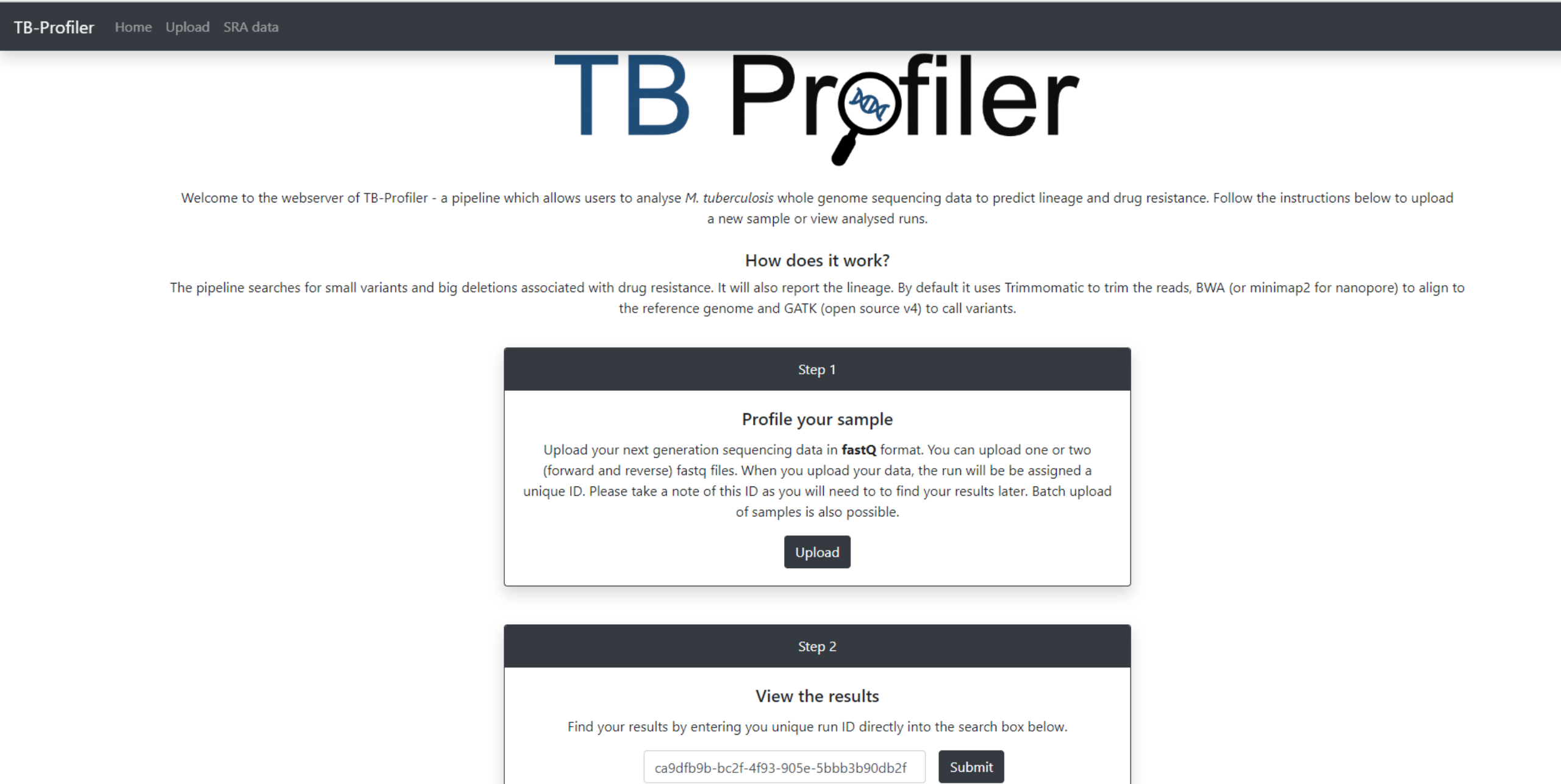


Important! The mutations and resulting library files are in reference to the H37Rv (NC_000962.3/AL123456.3) reference genome

TBProfiler versions

Web-based		Command line	
Pros	**Cons**	**Pros**	**Cons**
Simple to use	Limit on the size of the files is 4GB	Once downloaded, doesn't require internet	Linux or MacOS only
User friendly, interactive interface	Requires stable internet connection	Fast (although depends on computer power)	Must understand line command to use
Creates a browser-based, easy to interpret output file		Unlimited file size	Output .json file is less easy to interpret

https://tbdr.lshtm.ac.uk/



TB-Profiler Home Upload SRA data

Upload your next generation sequencing data in **fastQ** format. Files will be processed using the tb-profiler pipeline with default parameters. You may select the technology used to generate the data (Illumina or Oxford Nanopore). Samples will be process with a first in first out policy so please be patient as there may be runs waiting to be processed before yours.

Please note: at the moment we can only accomodate uploads of files under 4GB. If you have files which are larger or you require many isolates to be processed and have access to a linux or osX operating system then it might be worthwhile to run the commandline version of tb-profiler. For more information on this please visit the github repository.

Batch upload

You can upload multiple samples together to batch process them. By default, paired files must use **_1.fastq.gz** and **_2.fastq.gz** as the file suffix in order for them to be paired correctly. The suffix can be changed in the advanced options.

Drop files here to upload

Drug class definitions

Samples can be classed into different types using the following definitions

Type	Drugs resistance
Sensitive	No drug resistance
Pre-MDR	Rifampicin **or** isonisazid
MDR	Rifampicin **and** isoniazid
Pre-XDR	MDR **and** any fluoriquinolone
XDR	MDR **and** (any fluoriquinolone **and** any group A drug)
Other	Resistance to any drug but none of the above categories

Mutation format

Mutations must follow the HGVS nomenclature. Information on this format can be found here. The following types of mutations are currently allowed:

- Amino acid substitutions. *Example: S450L in rpoB would be p.Ser450Leu*
- Deletions in genes. *Example: Deletion of nucleotide 758 in tlyA would be c.758del*
- Insertion in genes. *Example: Insertion of GT between nucleotide 1850 and 1851 in katG would be c.1850_1851insGT*
- SNPs in non-coding RNAs. *Example: A to G at position 1401 in rrs would be r.1401a>g*
- SNPs in gene promoters. *Example: A to G 7 bases 5' of the start codon in pncA c.-7A>G*

Important! The mutations and resulting library files are in reference to the H37Rv (NC_000962.3/AL123456.3) reference genome

Let’s have a look at TB-profiler output reports

Browser-based report (768345cd-b4b1-4d22-a41b-2881dcdf3af3)

TB-Profiler Home Upload SRA data

TB-Profiler result

Run: 768345cd-b4b1-4d22-a41b-2881dcdf3af3

Summary

Run ID: 768345cd-b4b1-4d22-a41b-2881dcdf3af3

Sample name:

Date: 26-01-2023 15:39:34

Number of reads: 30203

Percentage reads mapped: 95.41

Strain: lineage2.2.2

Spoligotype: 000000000000000

Drug-resistance: Pre-XDR-TB

Drug	Resistance	Supporting mutations
Rifampicin	R	rpoB p.Ser450Leu (0.91)
Isoniazid		
Ethambutol	R	embB p.Met306Ile (1.00)
Pyrazinamide	R	pncA p.Cys14Arg (1.00)
Streptomycin	R	gid p.Leu79Ser (1.00)
Fluoroquinolones	R	gyrA p.Asp94Gly (1.00)
Moxifloxacin	R	gyrA p.Asp94Gly (1.00)
Ofloxacin	R	gyrA p.Asp94Gly (1.00)
Levofloxacin	R	gyrA p.Asp94Gly (1.00)

Lineage Table: The lineage is inferred by analysing lineage specific SNPs

Lineage	Family	Main Spoligotype	RDs	Frequency
lineage2	East-Asian	Beijing	RD105	1.0
lineage2.2	East-Asian (Beijing)	Beijing-RD207	RD105;RD207	1.0
lineage2.2.2	East-Asian (Beijing)	Beijing-RD105/RD207	RD105;RD207	1.0

Drug resistance-Associated Mutations: This table reports mutations found in candidate resistance genes which have been associated with drug resistance

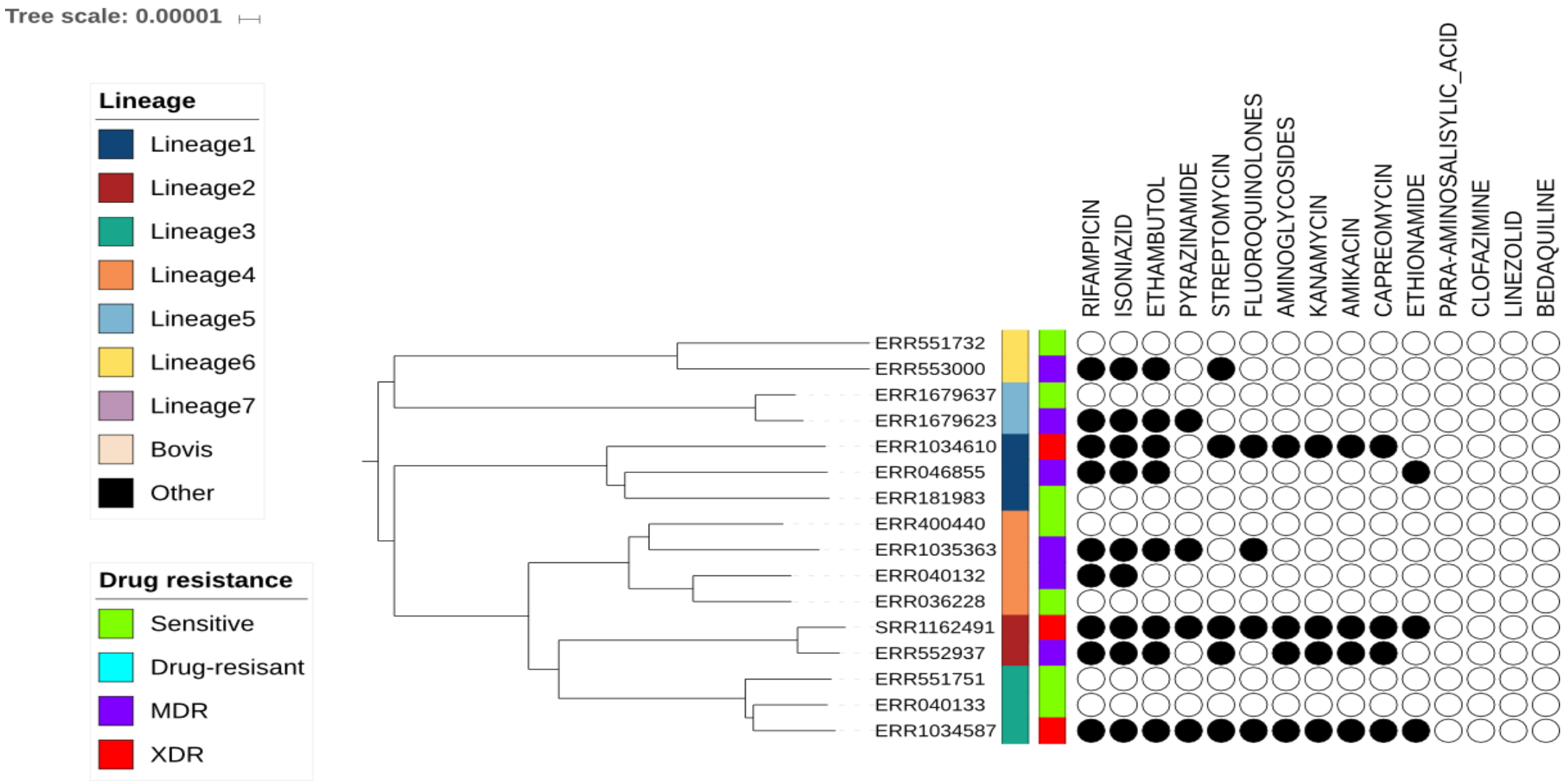
Gene	Chromosome position	Mutation	Type	Estimated fraction	Drugs

JSON report


```
lineage:
  0: {…}
  1: {…}
  2: {…}
  3:
    lin: "lineage4.1.2.1"
    family: "Euro-American (Haarlem)"
    spoligotype: "T1;H1"
    rd: "RD182"
    frac: 0.9984939759036144
  main_lin: "lineage4"
  sublin: "lineage4.1.2.1"
dr_variants:
  0:
    chrom: "Chromosome"
    genome_pos: 761155
    ref: "C"
    alt: "T"
    freq: 1
    feature_id: "CCP43410.1"
    type: "missense_variant"
    nucleotide_change: "c.1349C>T"
    protein_change: "p.Ser450Leu"
    annotation:
      0:
        type: "resistance_association_confidence"
        drug: "rifampicin"
  sublin: "lineage2.2.2"
dr_variants:
  0: {…}
  1:
    chrom: "Chromosome"
    genome_pos: 761155
    ref: "C"
    alt: "T"
    depth: 11
    freq: 0.9090909090909091
    feature_id: "CCP43410"
    type: "missense_variant"
    nucleotide_change: "c.1349C>T"
    protein_change: "p.Ser450Leu"
    annotation: […]
    alternate_consequences: []
    change: "p.Ser450Leu"
    locus_tag: "Rv0667"
    gene: "rpoB"
    drugs: […]
    gene_associated_drugs:
      0: "rifampicin"
  2: {…}
  3: {…}
  4: {…}
other_variants: […]
drtype: "Pre-XDR-TB"
```

TXT report

```
TBProfiler report
=================

The following report has been generated by TBProfiler.

Summary
-------
ID: pan346
Date: Fri Feb  4 14:16:27 2022
Strain: lineage4.1.2.1
Drug-resistance: MDR

Lineage report
--------------
Lineage Estimated Fraction      Family  Spoligotype     Rd
lineage4          1.000   Euro-American   LAM;T;S;X;H     None
lineage4.1        1.000   Euro-American   T;X;H   None
lineage4.1.2      0.998   Euro-American   T;H     None
lineage4.1.2.1    0.998   Euro-American (Haarlem) T1;H1   RD182


Resistance report

Drug    Genotypic Resistance    Mutations
Rifampicin      R       rpoB p.Ser450Leu (1.00)
Isoniazid       R       katG p.Ser315Thr (1.00)
Ethambutol      R       embB p.Met306Ile (1.00)
Pyrazinamide    R       pncA p.Leu85Pro (1.00)
Streptomycin    R       rpsL p.Lys43Arg (1.00)
Fluoroquinolones
Amikacin
Capreomycin
Kanamycin
```


Tree scale: 0.00001
Lineage
Lineage1
Lineage2
Lineage3
Lineage4
Lineage5
Lineage6
Lineage7
Bovis
Other
Drug resistance
Sensitive
Drug-resisant
MDR
XDR
RIFAMPICIN
ISONIAZID
ETHAMBUTOL
PYRAZINAMIDE
STREPTOMYCIN
FLUOROQUINOLONES
AMINOGLYCOSIDES
KANAMYCIN
AMIKACIN
CAPREOMYCIN
ETHIONAMIDE
PARA-AMINOSALISYLIC_ACID
CLOFAZIMINE
LINEZOLID
BEDAQUILINE
ERR551732
ERR553000
ERR1679637
ERR1679623
ERR1034610
ERR046855
ERR181983
ERR400440
ERR1035363
ERR040132
ERR036228
SRR1162491
ERR552937
ERR551751
ERR040133
ERR1034587


Thank You