nature genetics

ARTICLES

https://doi.org/10.1038/s41588-017-0029-0

Corrected: Author Correction

Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*

BCB546X Final project
Group - Animal crossers

Giang Nguyen, Prita Pandya, Sushma Ambekar, Vishesh Bhatia

The question
What are the genetic determinants of resistance to antituberculosis drugs?

-Mycobacterium tuberculosis (Mtb) genome is 4.4Mb with low mutation rate

-Resistance in Mtb is mainly conferred by nucleotide variations (SNPs and indels) in genes encoding drug targets or drug-converting enzymes.

-Current molecular tests for resistance lack high levels of sensitivity, which is addressed by this paper.

Lineages and types of TB

Lineages 1-4

1: Indo-Oceanic (East African–Indian spoligotype families)

2: East Asian (for example, W/Beijing spoligotype families)

3: East African–Indian (for example, Central Asian strain (e.g., CASDELHI) spoligotype families)

4: Euro-American (for example, Latin American–Mediterranean (LAM), Haarlem and the ‘ill-defined’ T spoligotype families)

MDR-TB: Resistance to rifampicin and isoniazid

XDR-TB: Resistance to the fluoroquinolones and any of the injectable drugs (amikacin, kanamycin or capreomycin) used to treat MDR-TB

Study design

Collection of clinical isolates

6,465 samples from more than 30 geographic locations, from 4 Mtb lineages

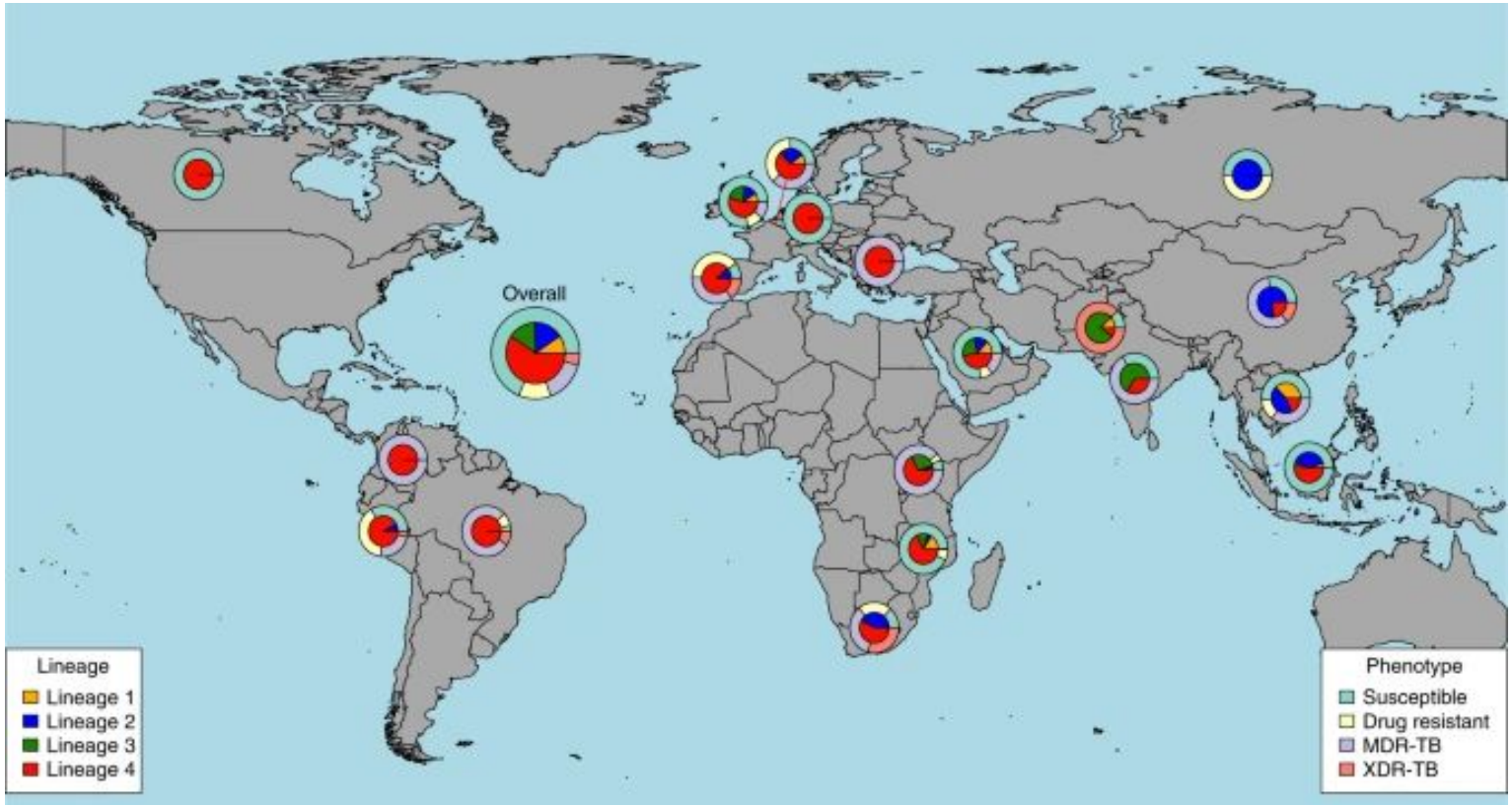
GWAS

Identify the nucleotide variation and loci responsible for resistance

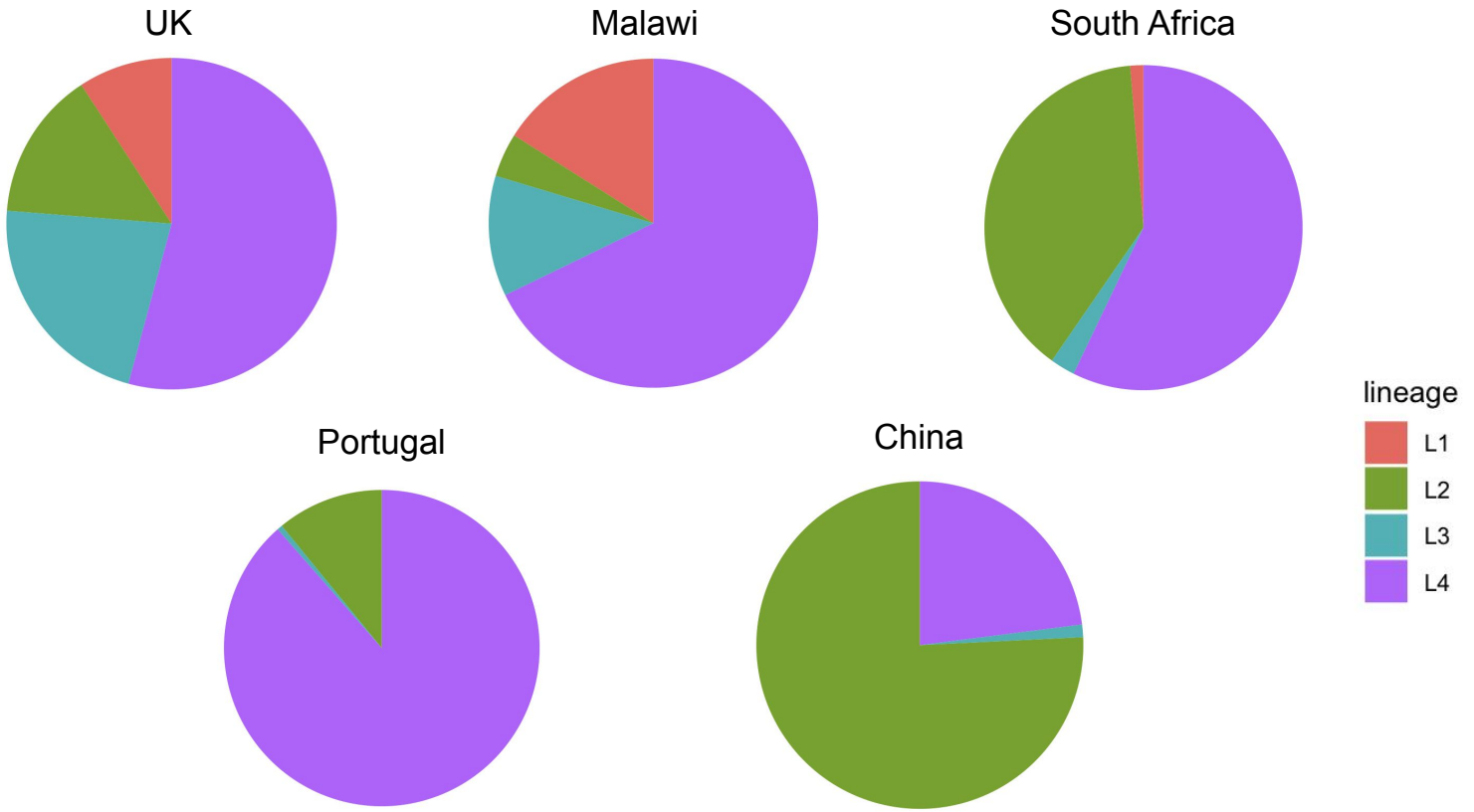
Multiple drug study

Study 14 drugs with available phenotypic data on drug susceptibility testing

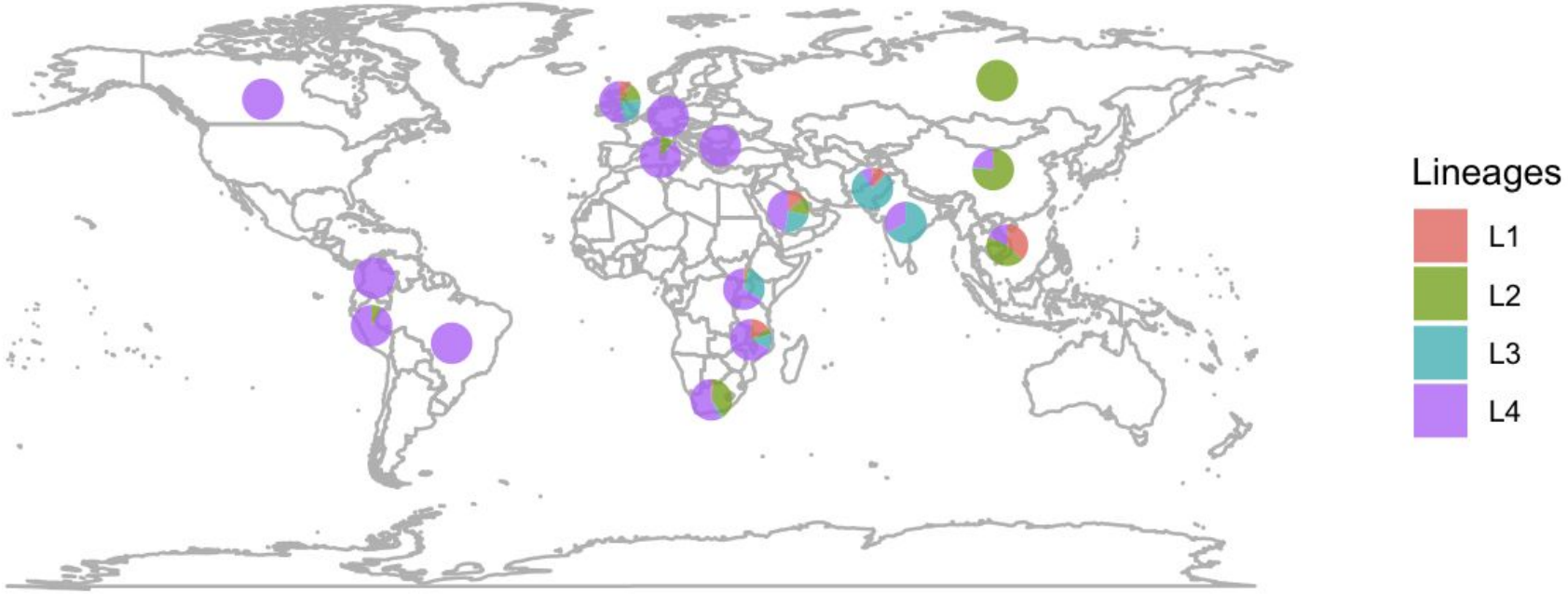
Geographic distribution of the 6,465 *Mycobacterium tuberculosis* isolates analyzed in the study



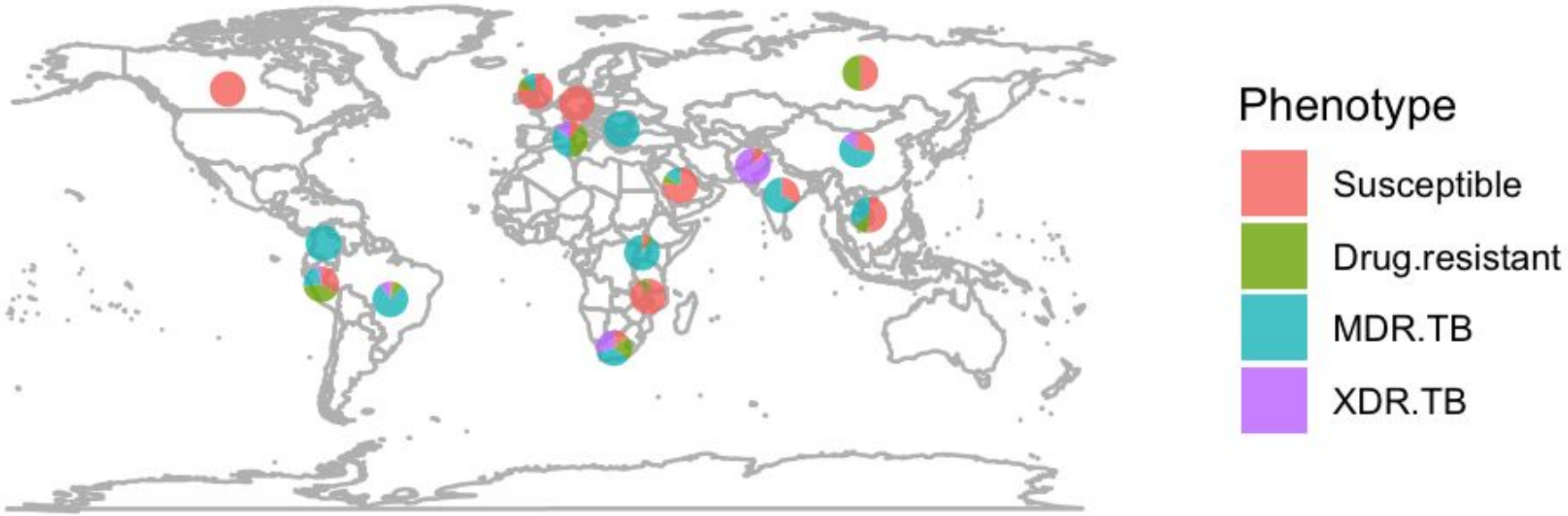
Geographic distribution of lineages in 5 countries with highest TB population



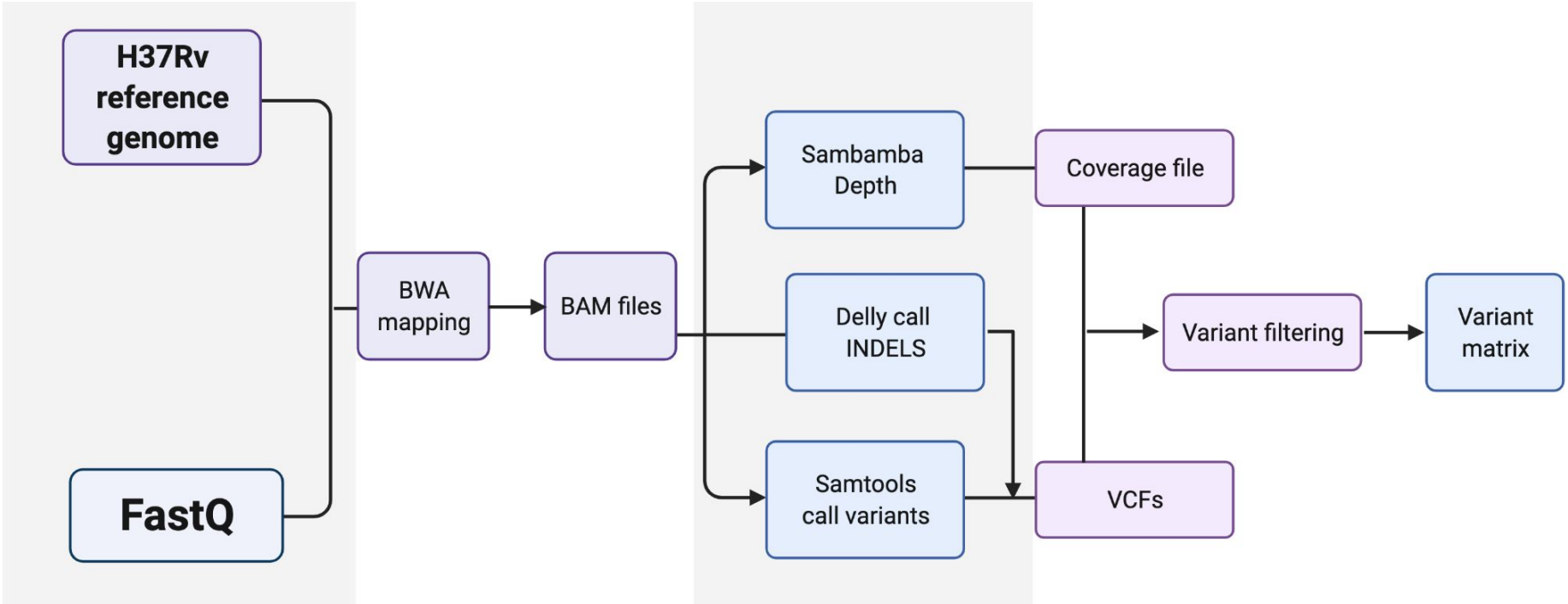
Geographic distribution of lineages generated with scatterpie



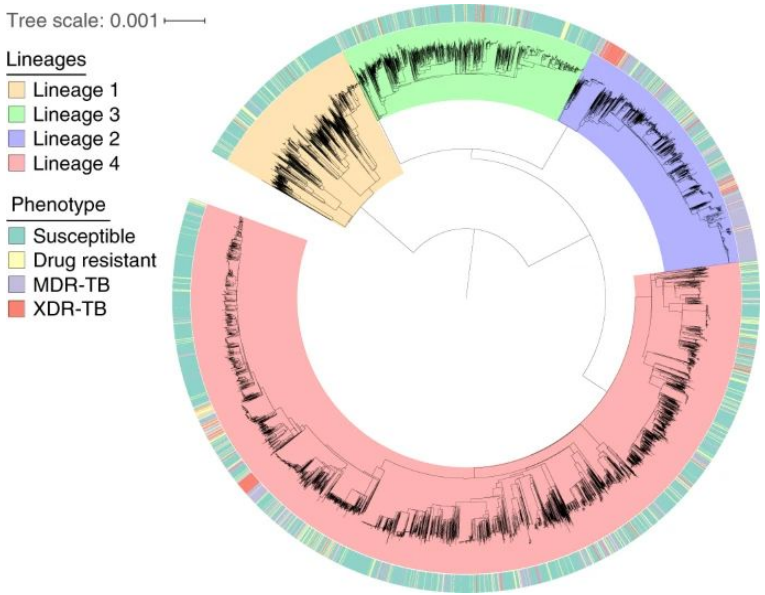
Geographic distribution of TB phenotype generated with scatterpie

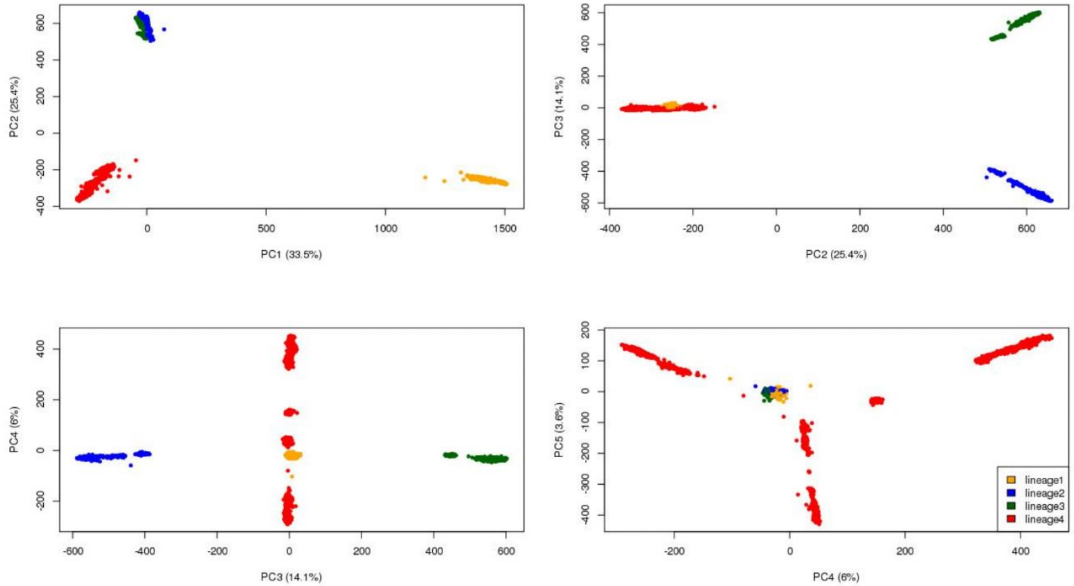


Analytical workflow for the raw sequence data



Genome-wide SNPs showed clustering by lineage





GWAS : Genome wide association studies

GWAS typically involves testing of genetic variants across the genomes of large number of individuals to identify genotype-phenotype associations.

The main goal is to identify the variants - SNPs associated with a trait, usually a disease phenotype

PhyC test

To identify SNPs enriched for convergent evolution, to help find the evolutionary path, the phylogenetics based phyC test was used.

PhyC test is a form of hierarchical clustering method used to construct phylogenetic tree and identify homoplastic variants

Table 1 graph: number of SNPs identified in different groups

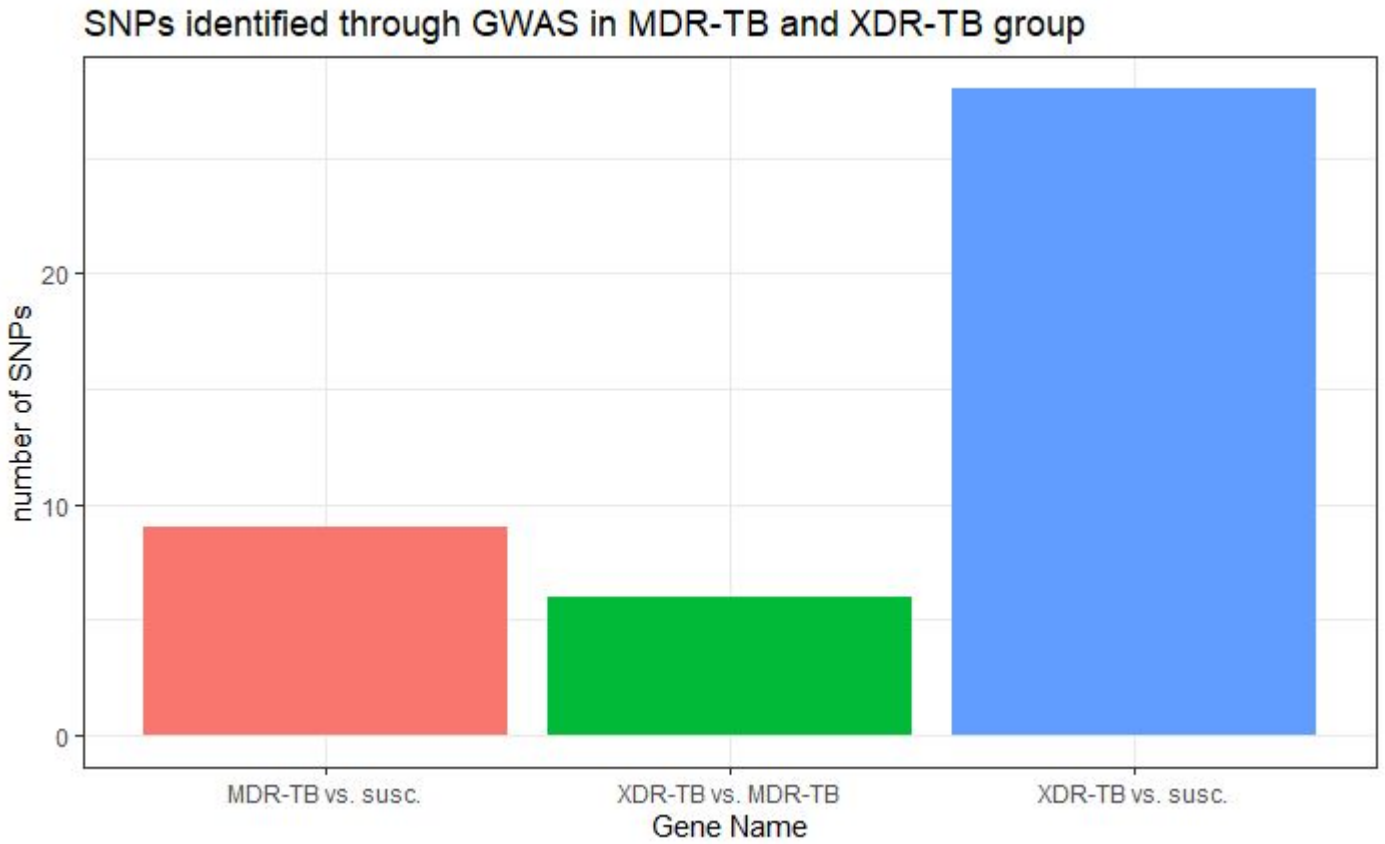


Table 1 graph: number of SNPs identified in different groups

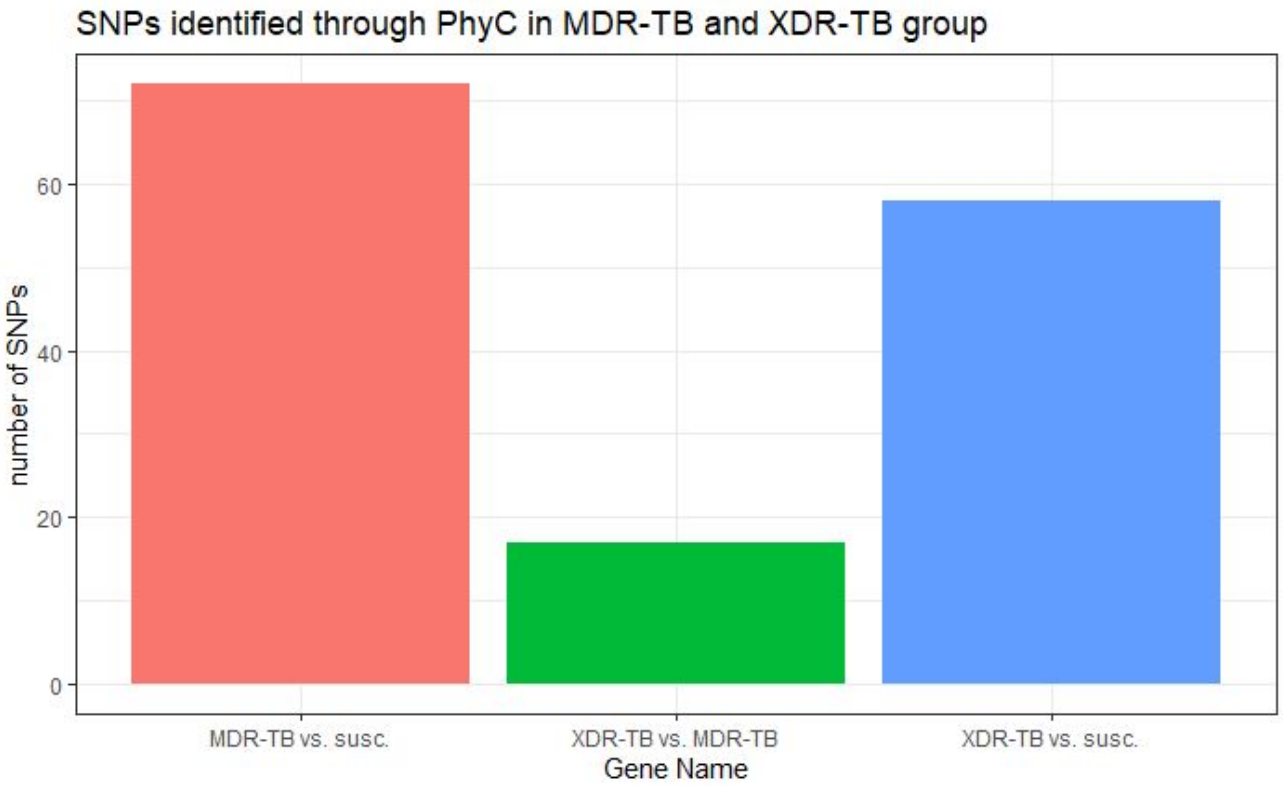


Table 2 plot: SNPs observed in response to different drugs

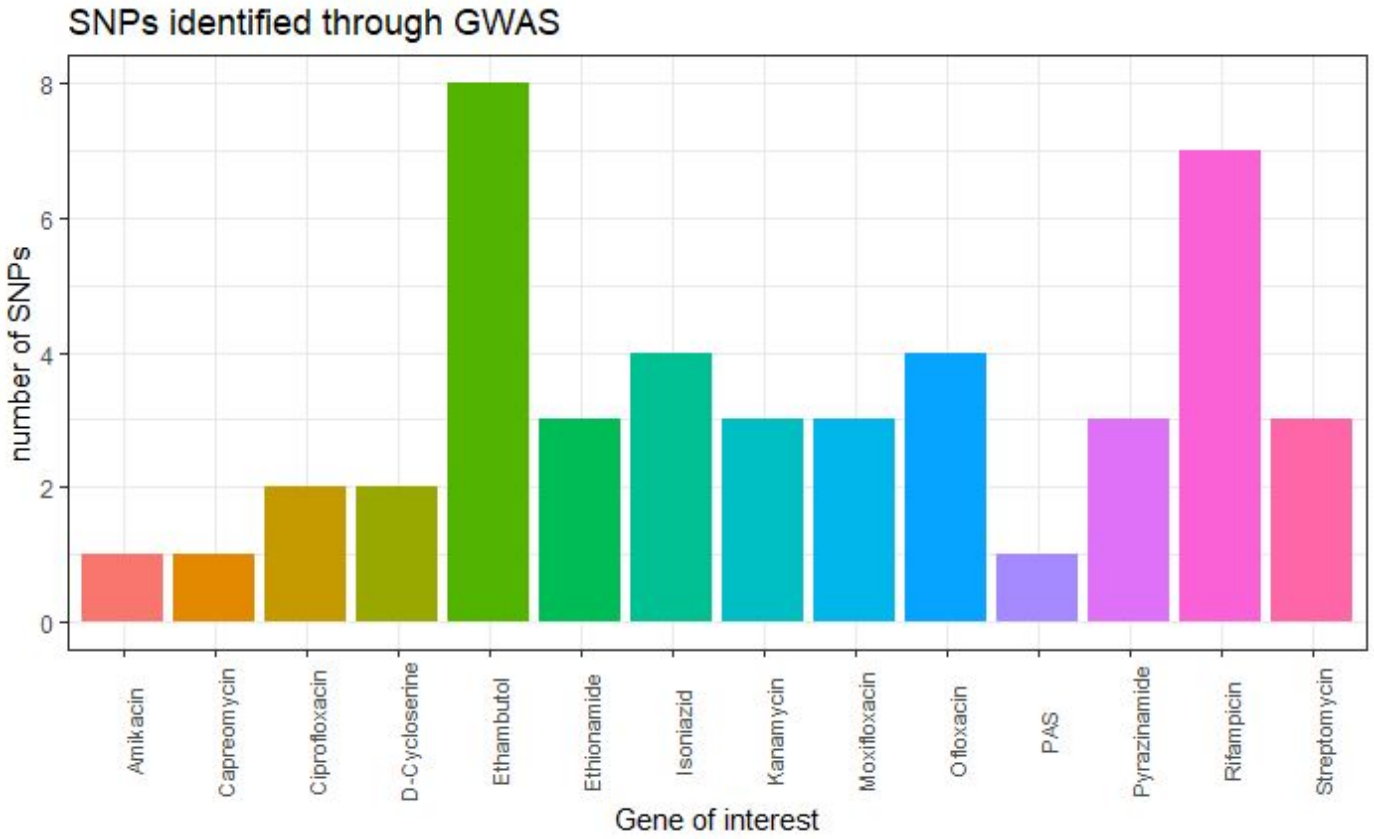
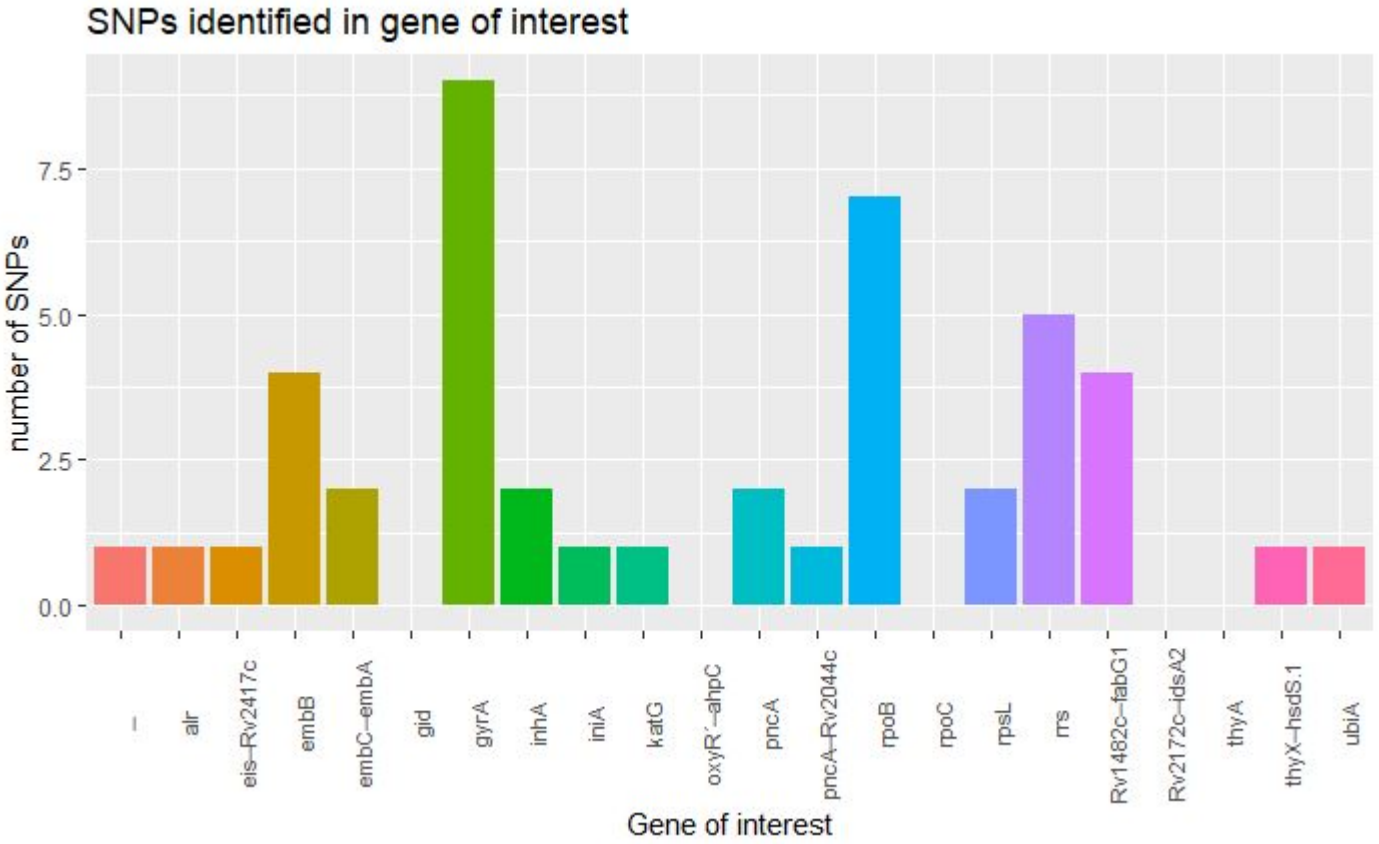
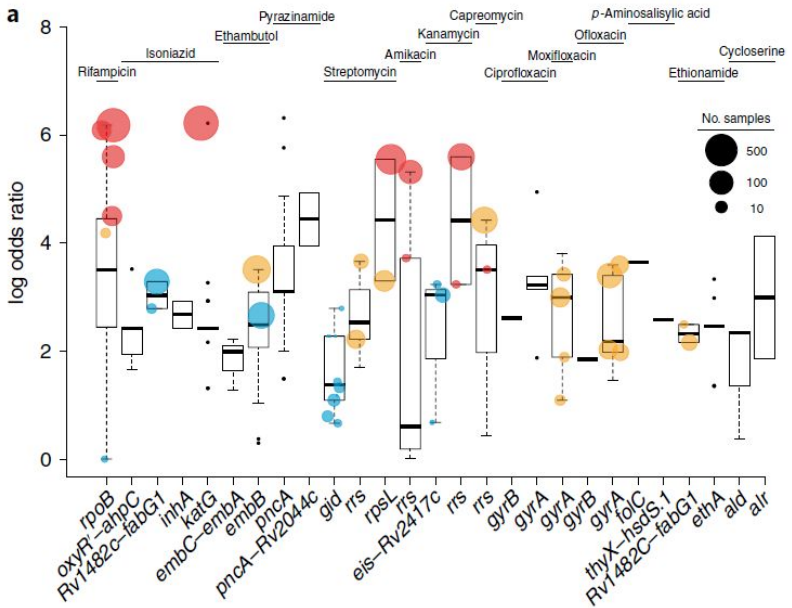
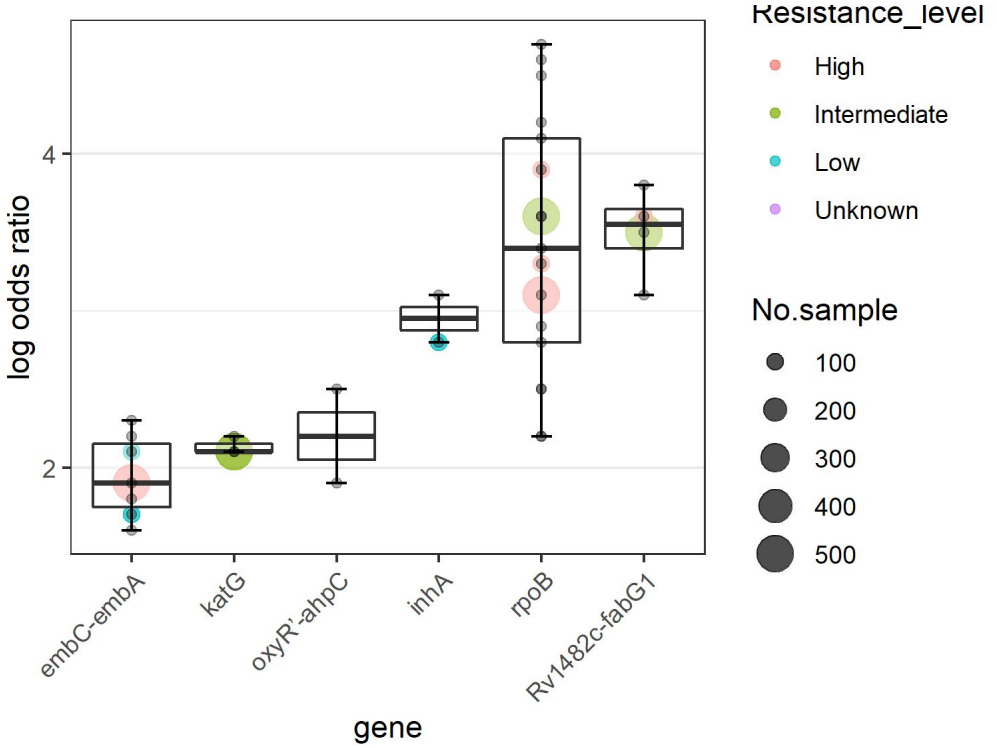


Table 2: number of SNPs observed in different genes



Log odds ratios from SNP–drug resistance associations are a potential surrogate for resistance level



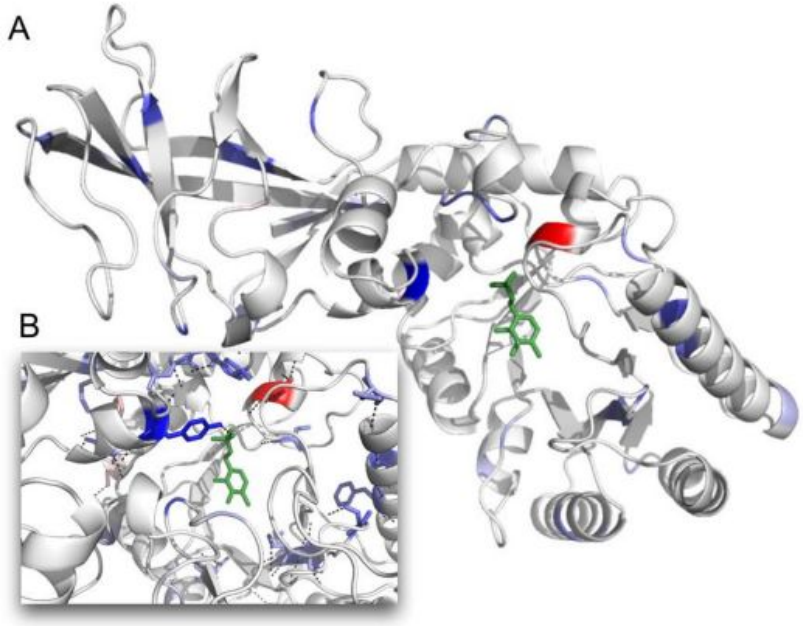


Mutations known to confer low, intermediate or high levels of resistance are represented by points colored blue, yellow and red,respectively, and point size is proportional to mutation frequency

(Note: makeup data for 5 random genes)

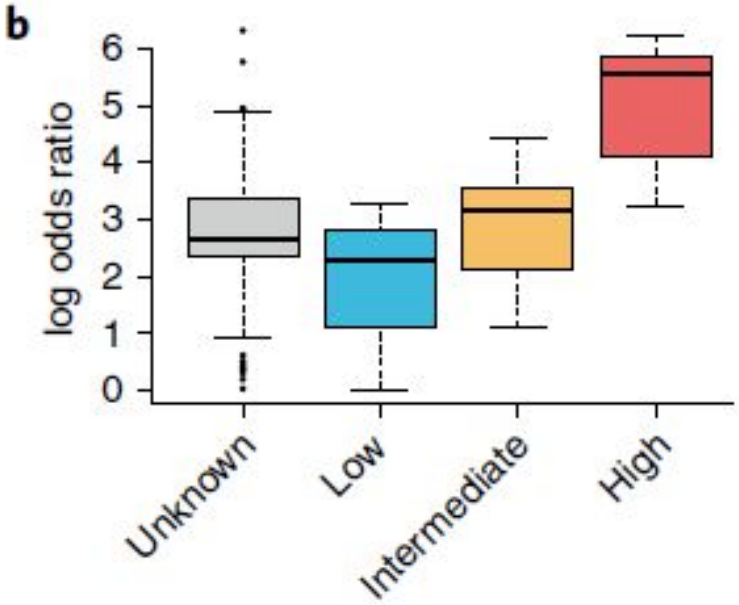
Model the effect of various mutations on protein stability and ligand binding

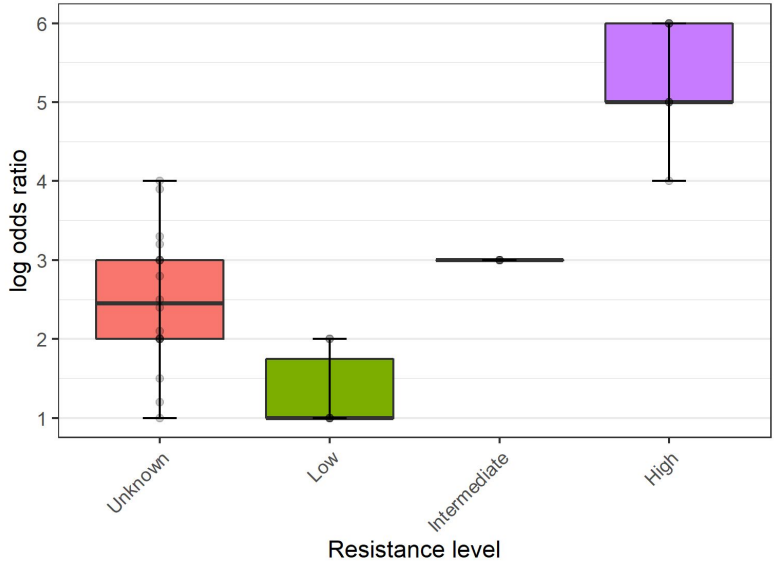
Protein structure for Alr



Unfavourable mutations are depicted in blue and favourable mutations are depicted in red, where color intensity reflects the extent of effect

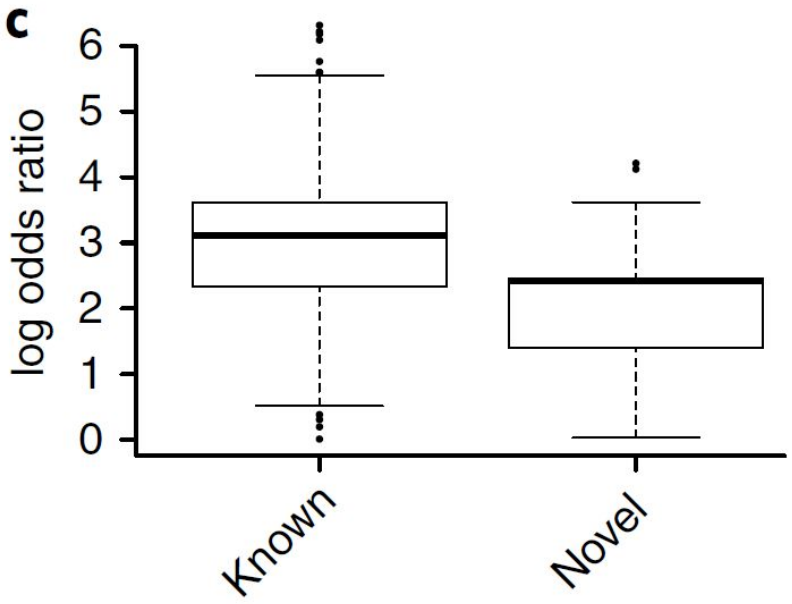
The distribution of log odds ratios for mutations within unknown or known genes conferring low , intermediate, or high levels of resistance.

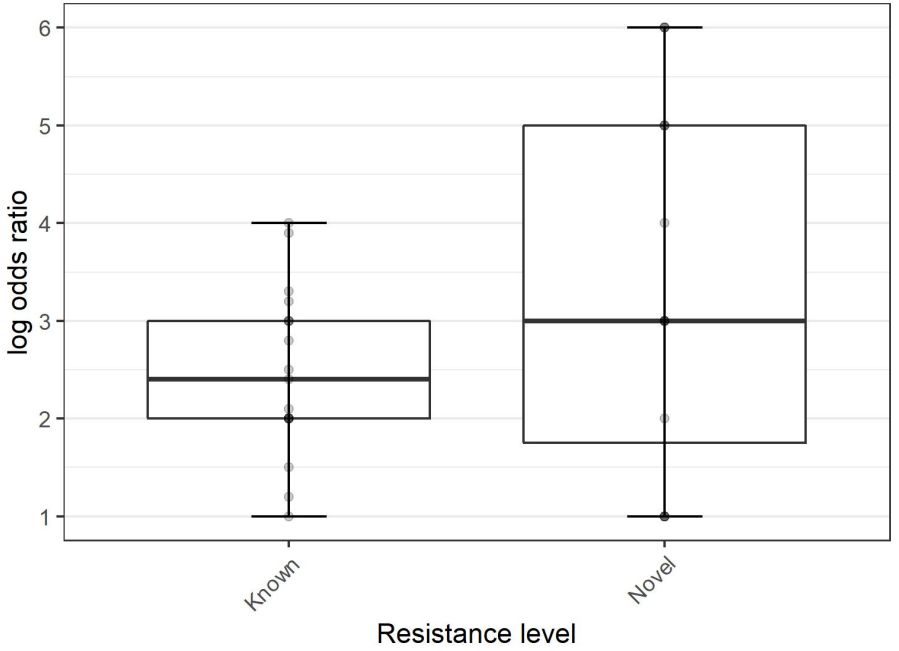




(Note: makeup data)

The distribution of log odds ratios for known and novel drug-resistance-conferring mutations





(Note: makeup data)

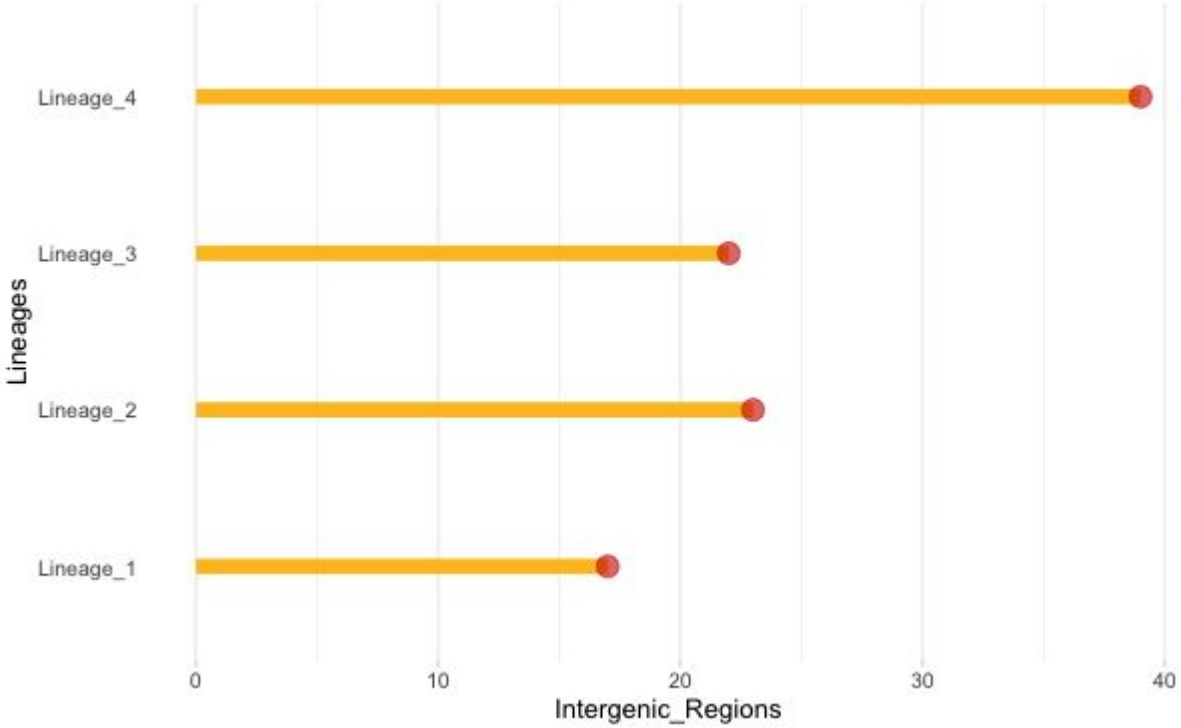
Lineage-Specific Loci Associated with Drug Resistance

Supplementary Table 4

Lineage	Gene / intergenic region	Position	Drug	Min P-value	Susc.	DR	MDR-TB	XDR-TB
4	*gyrA*	7570	X v M or SUS	1.51E-15	0.001	0.005	0.042	0.321
4	*gyrA*	7572	X v SUS	8.92E-21	0.001	0.014	0.013	0.061
2	*gyrA*	7581	X v MDR	7.99E-06	0.001	0.004	0.019	0.120
3,4	*gyrA*	7581	X v SUS	1.17E-21	0.001	0.004	0.019	0.120
2,3	*gyrA*	7582	X v M	3.71E-07	0.003	0.050	0.067	0.349
4	*gyrA*	7582	X v M or SUS	8.52E-07	0.003	0.050	0.067	0.349
2	*rpoB*	760314	M v SUS	4.92E-22	0	0.004	0.006	0
3	*rpoB*	761108	X v SUS	3.44E-14	0	0	0.002	0.018
2-4	*rpoB*	761109	M or X v SUS, RMP	3.34E-28	0	0.011	0.027	0.055
3,4	*rpoB*	761110	X v M, X or M v SUS, RMP	3.35E-85	0	0.007	0.066	0.333
1,2,4	*rpoB*	761139	X or M v SUS	3.46E-16	0.001	0.057	0.087	0.018
1-4	*rpoB*	761139	M v SUS, RMP	1.61E-97	0.001	0.057	0.087	0.018
1,2,4	*rpoB*	761140	M or X v SUS, RMP	2.66E-17	0.001	0.005	0.033	0.014
1-4	*rpoB*	761155	M or X v SUS, RMP	1.17E-219	0.001	0.128	0.358	0.439

Lineage 1: Indo-Oceanic
Lineage 2: East Asian
Lineage 3: East African-indian
Lineage 4: Euro American

Lineage-Specificity for Drug Resistance



Largest number of lineage specific mutations associated with drug resistance found in **Lineage 4.**

Figure 4 (from Supplementary Table 4)

Conclusions

- Four major Mtb Lineages across five continents encompassing a total of 6,465 clinical isolates marks this as the largest geographical study.
- GWAS and PhyC test identified various resistance-conferring loci and compensatory relationships.
- Various further analysis highlighted the importance of indels to drug resistance.
- Relationship between resistance levels and odds ratio from the GWAS could benefit understanding the molecular diagnostic data.
- Enhanced understanding of genetic basis of multi-drug resistance may lead to improved personalized treatment and rates of cure.

Shortcomings of the Research Study

- No repository found online.
- Very limited data was available either through supplementary tables in the paper or URL link to download excel sheets.
- No information available on the R packages used to generate the plots, only the version of R used was provided
- The study links to PathogenSeq- a website for global drug resistance data but this provides no easy access to data as well.

THANK YOU