Estimation of country-specific tuberculosis antibiograms using a large genomic dataset

Avika Dixit[1,2], Luca Freschi[2], Roger Vargas[2], Matthias Groeschel[2], Michael Chen[2], Alena Skrahina[3], Sabira Tahseen[4], SM Mostofa Kamal[5], Nazir A. Ismail[6], Maha Farhat[2,7]

[1]Boston Children's Hospital, Boston MA; [2]Harvard Medical School, Boston MA; [3]Republican Research and Practical Centre for Pulmonology and Tuberculosis Dolginovski Trakt, Minsk, Belarus; [4]National Tuberculosis Control Programme, Islamabad, Pakistan; [5]National Institute of Diseases of the Chest and Hospital, Dhaka, Bangladesh; [6]National Institute for Communicable Diseases, Sandringham, Gauteng, South Africa, [7]Massachussetts General Hospital, Boston, MA

INTRODUCTION

Background

- High resolution drug resistance estimates in TB are needed to: a) assign patients to MDR-TB treatment regimens and b) improve surveillance of MDR-TB.
- Globally representative whole genome sequencing (WGS) TB data from clinical samples in the public domain can be used to predict DR in TB with good accuracy.

Objective

- To leverage publicly available WGS data and generate country-level estimates of DR to anti-TB agents using a previously developed Random Forest (RF) predictor.

METHODS

Data Curation and Processing

- TB WGS was curated from the published literature, public databases and via sequencing. WGS was processed through a custom bioinformatics pipeline:

PRINSEQ → Kraken → BWA → PICARD → Pilon → RF Predictor

Correction for Sampling of Outbreaks

- A representative isolate was randomly selected from each group of isolates with a genetic distance of <= 10 SNPs for each country.

Correction for Drug Resistance Oversampling

- Proportion of resistant isolates for each antibiotic was calculated by conditioning on WHO estimated rifampin resistance:

The probability (P) of resistance (R) to antibiotic A is given by:

$$P(A=R) = \frac{P(A=R|RR)P(RR)}{P(RR|A=R)} + P(A=R|S)$$

where $P(RR)$ is available from the WHO, and $P(A{=}R \mid RR)$, $P(RR \mid A{=}R)$ and $P(A{=}R|S)$, i.e. mono-resistance to A, were calculated from the curated database.

RESULTS

Dataset Description

- 22,387 isolates had country of origin data available. Of these, 17,710 (79.1%) had phenotypic drug susceptibility testing (DST) metadata.
- Of the 22,387 isolates with geographic metadata, 1,419 isolates were excluded because they were found to have >=10% missing calls at selected SNP sites.
- After applying correction for outbreak sampling, 12,641 isolates were available for further analysis (Figure 1).
- Of those isolates with phenotypic DST data, 16,824 (95%) were tested for resistance against both isoniazid and rifampin, of which 4,313 (25.63%) were MDR. The global distribution of MDR isolates included in the study is shown in figure 2.

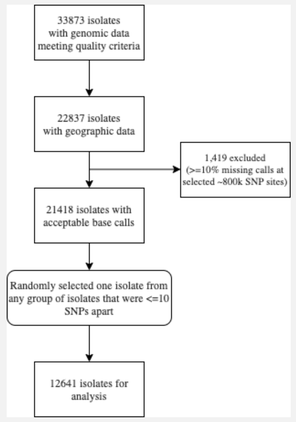

Fig. 1 Isolate filtering and selection

Antibiograms based on in-silico prediction

- Example antibiograms for all isolates (Figure 3) and MDR isolates (Figure 4) are shown.

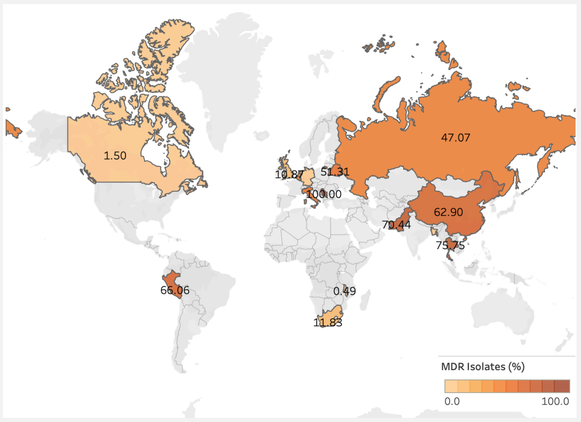

Fig. 2 MDR isolates (%) from each country with at least 100 isolates in the dataset

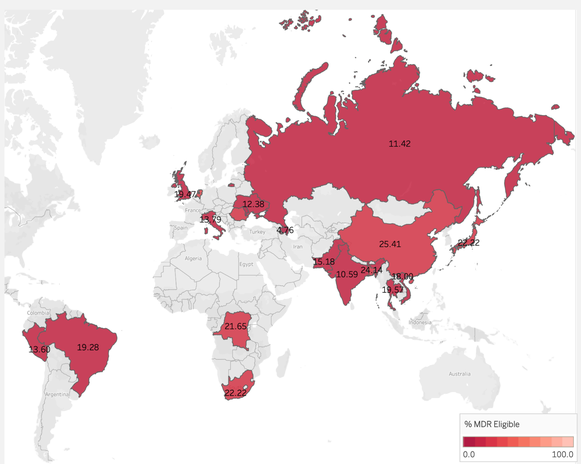

Fig. 5 MDR isolates (%) from each country with at least 50 MDR isolates in the dataset, that were predicted to be susceptible to ethambutol, pyrazinamide, kanamycin and levofloxacin (as a proxy for moxifloxacin).

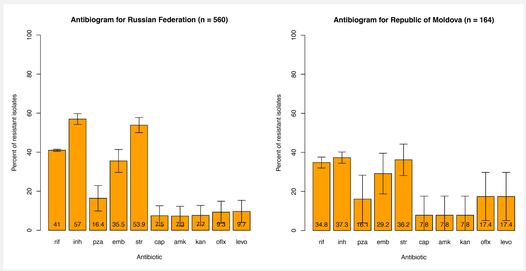

Fig. 3 Genotypic antibiograms for Russian Federation and Republic of Moldova. Rates are conditioned on composite rifampin resistance rates reported by the WHO. rif: rifampin, inh: isoniazid, pza: pyrazinamide, emb: ethambutol, str: streptomycin, cap: capreomycin, amk: amikacin, kan: kanamycin, oflx: ofloxacin, levo: levofloxacin

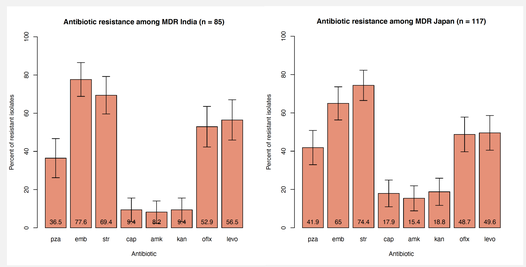

Fig. 4 Genotypic antibiograms for MDR isolates from India and Japan. pza: pyrazinamide, emb: ethambutol, str: streptomycin, cap: capreomycin, amk: amikacin, kan: kanamycin, oflx: ofloxacin, levo: levofloxacin

Eligibility for Bangladesh Regimen

- Among countries with at least 50 isolates available for genotypic prediction, MDR isolates that were sensitive to antibiotics included in the Bangladesh regimen was highest for China and lowest for Azerbaijan (Figure 5).

CONCLUSIONS

Prevalence of DR

- Countries of the former Soviet Union had high rates of estimated DR, which is consistent with data reported to the WHO.

Eligibility for the Bangladesh Regimen

- Empiric use of Bangladesh regimens across most countries is limited by the prevalence of resistance to agents used in this regimen. This supports the phasing out of this regimen in favor of newer bedaquiline containing regimens.

Implications

- Our results highlight the need for faster diagnostics that identify DR with a quick turnaround time as well as newer treatment options.
- Scale up of WGS will allow for improved monitoring and surveillance of DR in TB
- Rapid identification resistance to second line agents in MDR-TB patients will allow for quick and appropriate treatment improving outcomes.

Limitations and Further Work

- Our study is limited by the systematic oversampling of DR that is inherent in the published MTB genomic literature. We have corrected for these by conditioning on RR rates reported by the WHO and are conducting additional analyses including bias correction using Bayesian uncertainty model and validation using systematic DR surveys. These results will provide global estimates that can support policymakers and empiric treatment choices.

ACKNOWLEDGEMENTS

We thank Dr. Anna Dean and the WHO for assistance in curating MTB genomes. AD is supported through the Boston Children's Hospital OFD/BTREC/CTREC Faculty Career Development Fellowship and the Bushrod H. Campbell and Adah F. Hall Charity Fund/Charles A. King Trust Postdoctoral Fellowship.

References available upon request, please email avika.dixit@childrens.harvard.edu