## Supplementary Notes

### Supplementary Note S1

*Data collection*. WGS and epidemiological data of RR/MDR-TB cases from EU/EEA countries were collected by the EUSeqMyTB consortium (San Raffaele Scientific Institute (OSR), Milan, Italy; National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands; National Reference Center for Mycobacteria, Research Center Borstel (RCB), Germany; and Public Health England (PHE), London United Kingdom).

Raw sequencing data (FASTQ files) generated by individual countries using either the Illumina or Ion Torrent technologies, were shared via secured File transfer protocol (SFTP) by country study coordinators, processed and included in the EUSeqMyTB database. EU/EEA countries without access to WGS services sent DNA or heat-inactivated RR/MDR-MTBC isolates to their appointed EUSeqMyTB consortium laboratory according to the study standard operating procedures. Countries provided, if available, key clinical/epidemiological data (i.e. age, sex, country of birth of the TB case, and site of the disease) and laboratory data associated to each RR/MDR-MTBC isolate (i.e. year of isolation, rifampicin resistance profile, and sequencing method used). Laboratories followed their national guidelines to determine the drug resistance profile of the submitted isolates and the sampling method. Local laboratory codes were pseudo anonymized using a unique European Union Sequencing Typing (EUST) sample identifier to comply with the EU Regulation 2016/679 on General Data Protection Regulation (GDPR). The list of identifiers was communicated to the country study coordinators via SFTP.

*Data validation*. WGS data underwent quality checking and were only included in the study if fulfilling pre-defined quality criteria. Raw sequence data were mapped to the H37Rv genome (GenBank ID: NC_000962.3) and entered into the study database if more than 95% of the reference genome was covered by sequence reads with sufficient sequence data quantity and quality for reliable variant detection. A mean read coverage depth of at least 30x was considered as acceptable. The key pre-analytical, analytical and post-analytical parameters applied in this study were derived from a technical consultation involving experts from different fields, i.e. from microbiology to bioinformatic and public health. These are summarized in Supplementary table 1.

*Sequencing platform comparison.* Out of a total 2,218 isolates submitted to our study, 2,182 (98.4%) and 36 (1.6%) isolates were sequenced using the Illumina and Ion Torrent technology, respectively. Among the Illumina sequencing platforms, a total of 1,896 (86.9%) isolates underwent sequencing by NextSeq, 149 (6.8%) by HiSeq, 115 (5.3%) by MiSeq, and 22 (1.0%) by MiniSeq. Although a detailed comparison of the sequencing platforms performance through the analysis of the error statistics was out of scope of this work, we did compare the obtained coverage depth and coverage breadth. Overall, the isolates sequenced by Ion Torrent had a slightly lower coverage breadth (97.25% vs 98.70%) and a lower mean coverage depth compared to those sequenced by Illumina (Table 1 below). Among the different Illumina sequencing platforms, the isolates sequenced by NextSeq and Miniseq platforms showed a slightly higher mean coverage depth compared to those sequenced using the HiSeq and MiSeq platforms, while the coverage breadth was comparable (Table 1).

**Table 1.** Sequencing platforms comparison.

| | **Ion Torrent** | **Illumina** | | | |
|---|---|---|---|---|---|
| **Mean coverage depth** | **N. isolates (%)** | **NextSeq N. isolates (%)** | **HiSeq N. isolates (%)** | **MiSeq N. isolates (%)** | **MiniSeq N. isolates (%)** |
| >50 | 22 (61.1) | 1869 (98.6) | 135 (90.6) | 101 (87.8) | 22 (100) |
| 30-50 | 11 (30.6) | 22 (1.2) | 13 (8.7) | 13 (11.3) | 0 |
| 20-30 | 3 (8.3) | 5 (0.2) | 1 (0.7) | 0 | 0 |
| <20 | 0 | 0 | 0 | *1 (0.9)** | 0 |
| **Total** | **36** | **1896** | **149** | **115** | **22** |
| **Mean coverage breadth (%)** | **97.25** | **98.76** | **98.45** | **98.23** | **98.27** |

## Supplementary Note S2

*Methods used for WGS-based relatedness analysis.* The cgMLST analysis was performed using the commercially available SeqSphere+ version 6.0.0 software (Ridom GmbH, Münster, Germany) default setting with a scheme of 2,891 core genome genes [1], using a threshold of ≤ 5 alleles to identify clusters.

The SNP-based approach relied on the use of the MTBseq bioinformatic pipeline [2]. Briefly, reads were mapped to the H37Rv genome (GenBank ID: NC_000962.3) with BWA [3]. Alignments were then refined with the GATK [4] and Samtools [5] toolkits for base quality recalibration and alignment corrections for possible PCR and InDel artefact. Variants (SNPs and InDels) were called if the following criteria were met: a minimum coverage of four reads

in both forward and reverse orientation, four reads calling the allele with at least a phred score of 20, and an allele frequency of 75%.

The SNP-based analysis was performed on the pool of MTBC isolates clustering by cgMLST, using a maximum distance threshold ≤ 5 SNPs. Briefly, regions annotated as repetitive elements, InDels, multiple consecutive SNPs in a 12-bp window, and 92 genes implicated in antibiotic resistance are excluded for the phylogenetic reconstruction. In the combined analysis, all genome positions that fulfil the aforementioned criteria for coverage and variant frequency in 95% of all samples in the datasets are considered as valid [6]. From the concatenated sequence alignments, isolates are grouped by agglomerative clustering with a maximum distance threshold ≤ 5 SNPs to the nearest isolate in the same group.

*WGS-based drug resistance prediction.* The screening for drug resistance mutations was performed by switching the MTBseq pipeline [2] into the low frequency detection mode, in which the non-wild type majority base call is used and thresholds set to at least one read in both forward and reverse orientation, at least one read calling the allele with a phred score of at least 20, and 5% allele frequency. Detected variants were annotated with known resistance association of either the mutation itself or its genomic region according to the literature [7-9].

**Supplementary Note S3**

*Data reporting to study participants.* The results of the SNP-based relatedness and drug resistance analysis were made available to the study participants through a dedicated and access controlled external webserver (https://www.euseqmytb.eu/). For each of the submitted isolates, the webserver allowed to assess: i) the related isolates within a distance chosen by the user; ii) the list of resistance related variants; iii) the list of all variants of each isolate; iv) the comparison of the SNPs of a chosen index isolate to the SNPs in any other isolate in the database.

The Minimum Spanning Trees of the related isolates could also be viewed, where isolates from different countries were highlighted in different colours. In addition, Minimum Spanning Networks and UPGMA trees based on groups of isolate chosen by the user could be generated and easily exported.

**Supplementary Note S4**

*RR/MDR-MTBC strains lineage distribution in EU/EEA*. A total of 2,151 RR/MDR-MTBC isolates were included in the study database. The Euro-American lineage represented more than 50% of the RR/MDR-MTBC strains in Croatia (66.7%), France (58.6%), Hungary (72.7%), Ireland (60.9%), Italy (63.8%), Norway (50.0%), Portugal (89.2%), Slovakia (75.0%), Spain (71.2%), and almost the totality of RR/MDR-MTBC strains in Bulgaria (96.9%), Romania (96.0%), and Slovenia (100%). The Beijing lineage was instead more frequently represented in Austria (52.0%), Belgium (57.9%), Czech Republic (76.0%), Germany (52.0%), Finland (66.7%), Lithuania (65.4%), Latvia (50.6%), Poland (58.2%), reaching almost 91% in Estonia. In the Netherlands and Sweden, the Euro-American and Beijing lineages were equally represented. Denmark was the only country where the Delhi-CAS lineage constituted more than 60% of the MDR-TB strains (overrepresentation due to a national MDR-TB cluster), while in the United Kingdom, the Euro-American, Beijing and Delhi-CAS lineages were equally represented.

**Supplementary Note S5**

*Drug resistance profile of MTBC isolates collected in the study*. The WGS-based analysis revealed mutations predicting resistance to rifampicin (R) in 2,151 (97.0%; 95%CI 96.2%; 97.7%) isolates including 1,962 isolates (91.2%; 95%CI 89.9%; 92.3%) with additional resistance to isoniazid (H) (i.e. MDR-TB cases). Pyrazinamide (Z) resistance was predicted in 1,286 isolates, including 1,280 (59.5%; 95%CI 57.4%; 61.6%) rifampicin resistant and six (9.1%; 95%CI 4.2%; 18.5%) rifampicin susceptible-isoniazid resistant isolates. Mutations predicting resistance to fluoroquinolones (FQs) were detected in 578 (26.8%; 95%CI 25.0%; 28.7%) of the rifampicin resistant isolates and three rifampicin susceptible-isoniazid resistant isolates. A total of 696 (31.4%; 95%CI 30.4; 34.4%) isolates were predicted resistant to any of the second line injectable drugs among the rifampicin resistant isolates including 315 (14.7%; 95%CI 13.2%; 16.2%) RR/MDR-MTBC isolates carried mutations predicting resistance to amikacin, 677 (31.8%; 95%CI 29.6%; 33.2%) to kanamycin, and 338 (15.7%; 95%CI 14.2%; 17.3%) to capreomycin. A total of 26 (1.2%; 95%CI 0.8%;1.8%) RR/MDR-MTBC isolates carried mutations (point mutations, frameshifts and large deletions) predicting resistance to bedaquiline [9] or were shown to be phenotypically resistant to this drug.

**Supplementary Note References**

1. Kohl TA. Harmsen D. Rothgänger J. Walker T. Diel R. Niemann S. Harmonized Genome Wide Typing of Tubercle Bacilli Using a Web-Based Gene-By-Gene Nomenclature System. *EBioMedicine* 2018: 34: 131-138.
2. Kohl TA. Utpatel C. Schleusener V. De Filippo MR. Beckert P. Cirillo DM. Niemann S. MTBseq: a comprehensive pipeline for whole genome sequence analysis of. *PeerJ* 2018: 6: e5895.
3. Li H. Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009: 25(14): 1754-1760.
4. McKenna A. Hanna M. Banks E. Sivachenko A. Cibulskis K. Kernytsky A. Garimella K. Altshuler D. Gabriel S. Daly M. DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010: 20(9): 1297-1303.
5. Li H. Handsaker B. Wysoker A. Fennell T. Ruan J. Homer N. Marth G. Abecasis G. Durbin R. Subgroup GPDP. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009: 25(16): 2078-2079.
6. Merker M. Kohl TA. Barilar I. Andres S. Fowler PW. Chryssanthou E. Ängeby K. Jureen P. Moradigaravand D. Parkhill J. Peacock SJ. Schön T. Maurer FP. Walker T. Köser C. Niemann S. Phylogenetically informative mutations in genes implicated in antibiotic resistance in Mycobacterium tuberculosis complex. *Genome Med* 2020: 12(1): 27.
7. Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, Hanna D, Kim PS, Liwski R, Zignol M, Gilpin C, Niemann S, Denkinger CM, Fleming J, Warren RM, Crook D, Posey J, Gagneux S, Hoffner S, Rodrigues C, Comas I, Engelthaler DM, Murray M, Alland D, Rigouts L, Lange C, Dheda K, Hasan R, Ranganathan UDK, McNerney R, Ezewudo M, Cirillo DM, Schito M, Köser CU, Rodwell TC. A standardised method for interpreting the association between mutations and phenotypic drug resistance in. *Eur Respir J* 2017: 50(6).
8. CRyPTIC consortium, on behalf of 100,000 Genomes Project. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *N Engl J Med* 2018: 379(15): 1403-1415.
9. Kadura S, King N, Nakhoul M, Zhu H, Theron G, Köser CU, Farhat M. Systematic review of mutations associated with resistance to the new and repurposed Mycobacterium tuberculosis drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. *J Antimicrob Chemother*. 2020 May 3. pii: dkaa136. doi: 10.1093/jac/dkaa136.

## Supplementary Tables and Figures

**Supplementary Table S1.** Summary of the pre-analytical, analytical and post-analytical parameters and methods agreed upon during the technical expert consultation.

| | |
|---|---|
| **Pre-analytical** | |
| **Inclusion criteria** | *M. tuberculosis* complex (MTBC) isolates with confirmed resistance to rifampicin based on either genotypic- or phenotypic-based drug susceptibility testing. Subsequent isolates from the same TB patient can be included in the study as long as the isolation date is more than six months after the isolation date of the previous isolate. |
| **Material collected** | Well-grown MTBC isolates on either solid or liquid media. |
| **MTBC inactivation method** | Heat-inactivation at 95°C for 30 minutes in thermal block or water bath. |
| **Genomic DNA extraction method** | Any method suitable for extraction of high-quality genomic DNA, including any commercially available (para)magnetic or column-based system, or chemical method (e.g. N-cetyl-N,N,N-trimethyl ammonium bromide[CTAB] /NaCl protocol). |
| **Genomic DNA inclusion criteria** | Genomic DNA must be quantifiable by fluorometric-based method (e.g. using Qubit® Fluorometer) and must have a purity, measured by UV absorbance method, falling within the range of absorbance: $OD_{260/280}$: 1.8 - 2.0 and $OD_{260/230}$: 2.0-2.2. |
| **Timeline for sample collection** | Samples should be sent to the EUSeqMyTB consortium in batches every two to three months for the entire duration of the project. |
| **Analytical** | |
| **Sequencing platforms** | Any sequencing platform. This study is technology-agnostic. |
| **FASTQ files inclusion criteria** | Any set of paired FASTQ files with an unzipped file size > 350 MB. More than 95% of the reference genome (i.e. H37Rv, NC_000962.3) should be covered by sequence read. |
| **Analysis pipeline** | The relatedness analysis is performed using two sequential approaches: i) core genome multilocus sequence typing (cgMLST); ii) single nucleotide polymorphism (SNP)-based calculation of distances. All WGS data are analysed using the same analytical pipeline to ensure data comparability. |
| **Genomic regions excluded from relatedness analysis** | Regions of the genome that are known to assemble poorly (i.e. repetitive elements) and resistance-associated genes are excluded from the relatedness analysis. |
| **Post-analytical** | |
| **Cluster definition** | A threshold of maximum 5 SNPs/alleles is used to indicate high-confidence transmission, while a SNP/allele difference between 6 and 12 SNPs indicates a more putative (less recent) transmission. Beyond 12 SNPs/alleles, transmission is considered unlikely. |
| **Drug resistance prediction** | The list of drug resistance associated SNPs provided by the Relational Sequencing TB Data Platform (ReSeq-TB) is used as reference. |

MB: megabyte; MTBC: M. *tuberculosis* complex; SNP: single nucleotide polymorphism; UV: Ultraviolet; WGS: whole genome sequencing.

**Supplementary Table S2.** *Mycobacterium tuberculosis* isolates submitted in 2018 stratified by drug resistance profile as determined by WGS analysis (n=948)

| Country | Number of isolates submitted in 2018 | | | MDR-TB cases ECDC data 2018 | Coverage by country |
|---|---|---|---|---|---|
| | RS-TB | RR-TB | MDR-TB | | |
| Austria | 0 | 0 | 19 | 18 | 105.6 |
| Belgium | 0 | 1 | 7 | 7 | 100 |
| Bulgaria | 0 | 3 | 21 | 24 | 87.5 |
| Croatia | 0 | 0 | 2 | 2 | 100 |
| Cyprus | 0 | 0 | 0 | 0 | n.a. |
| Czech Republic | 0 | 0 | 9 | 12 | 75.0 |
| Denmark | 0 | 0 | 4 | 4 | 100 |
| Estonia | 0 | 2 | 23 | 30 | 76.7 |
| Finland | 0 | 0 | 4 | 4 | 100 |
| France | 1 | 3 | 62 | 82 | 75.6 |
| Germany | 1 | 5 | 118 | 116 | 101.7 |
| Hungary | 1 | 0 | 12 | 12 | 100 |
| Ireland | 0 | 3 | 4 | 5 | 80.0 |
| Italy | 1 | 8 | 35 | *53* [a] | 66.0 |
| Latvia | 0 | 0 | 33 | *46* [b] | 71.7 |
| Lithuania | 3 | 3 | 54 | 170 | 31.8 |
| Luxemburg | 0 | 0 | 0 | 1 | 0 |
| Malta | 0 | 0 | 0 | 0 | n.a. |
| Netherlands | 0 | 0 | 6 | 6 | 100 |
| Norway | 0 | 2 | 4 | 4 | 100 |
| Poland | 0 | 6 | 51 | 48 | 106.3 |
| Portugal | 0 | 1 | 18 | 10 | 180 |
| Romania | 20 | 39 | 276 | 354 | 78.0 |
| Slovakia | 0 | 0 | 4 | 2 | 200 |
| Slovenia | 0 | 0 | 0 | 0 | n.a. |
| Spain | 0 | 4 | 34 | 33 | 103 |
| Sweden | 0 | 0 | 11 | 13 | 84.6 |
| United Kingdom | 0 | 3 | 27 | 37 | 73.0 |
| **Total** | **27** | **83** | **838** | **1093** | **76.7** |

RS-TB: rifampicin susceptible TB; RR-TB: rifampicin resistant TB; MDR-TB: multidrug resistant TB; n.a.: not applicable.

[a] Italian NIH notification 2020 – 2018 data

[b] ECDC notification 2019 – 2017 data

**Supplementary Table S3.** List of genomic regions considered for drug resistance analysis.

| Drug | Genomic Region | Gene |
|---|---|---|
| Rifampicin | Rv0667 | *rpoB* |
| Isoniazid | Rv1483 | *inhA* |
| | Rv1908c | *katG* |
| Ethambutol | Rv3795 | *embB* |
| Pyrazinamide | Rv2043c | *pncA* |
| Fluoroquinolones | Rv0006 | *gyrA* |
| | Rv0005 | *gyrB* |
| Amikacin | Rvnr01 | *Rrs* |
| Kanamycin | Rvnr01 | *Rrs* |
| | Rv2416c | *eis* |
| Capreomycin | Rvnr01 | *Rrs* |
| | Rv1694 | *tlyA* |
| Bedaquiline | Rv0678 | *mmpR* |
| | Rv1305 | *atpE* |

**Supplementary Table S4.** Cross-border clusters identified using a cgMLST-based approach.

| **cgMLST cross-border clusters involving recent transmission events** | | |
|---|---|---|
| **Cross-border cluster name** | **Isolates (N.)** | **Countries involved (N.)** |
| cgCL 1 | 36 | 4 |
| cgCL 2 | 34 | 4 |
| cgCL 3 | 32 | 3 |
| cgCL 4 | 29 | 8 |
| cgCL 5 | 20 | 2 |
| cgCL 7 | 17 | 3 |
| cgCL 8 | 16 | 3 |
| cgCL 9 | 14 | 2 |
| cgCL 10 | 13 | 2 |
| cgCL 12 | 12 | 4 |
| cgCL 13 | 12 | 2 |
| cgCL 14 | 11 | 3 |
| cgCL 15 | 11 | 3 |
| cgCL 16 | 11 | 5 |
| cgCL 17 | 10 | 4 |
| cgCL 18 | 10 | 4 |
| cgCL 19 | 10 | 2 |
| cgCL 20 | 9 | 3 |
| cgCL 21 | 9 | 4 |
| cgCL 27 | 7 | 2 |
| cgCL 28 | 6 | 2 |
| cgCL 31 | 6 | 3 |
| cgCL 33 | 6 | 2 |
| cgCL 40 | 5 | 2 |
| cgCL 48 | 5 | 2 |
| cgCL 55 | 4 | 3 |
| cgCL 57 | 5 | 2 |
| cgCL 64 | 5 | 2 |
| cgCL 68 | 4 | 2 |
| cgCL 72 -135 [a] | 3 | 2 |
| cgCL 136 -307 [b] | 2 | 2 |

cgCL: cgMLST-based cluster; N: numbers.

[a] Twelve (12) clusters comprising three isolates from two countries;

[b] Twenty-two (22) clusters comprising two isolates from two countries.

To each cluster identified by cgMLST, either national or cross-border, a sequential number from 1 to 307 was assigned based on the size of the cluster starting from the largest one. Only cross-border clusters are reported in the table

**Supplementary Table S5.** Major cross-border clusters identified in the study.

| Cluster name | Number of isolates in cluster | Lineage classification | WGS-based drug resistance profile (gene; mutation; number of isolates with mutation) | Country of isolation (number of isolates by country) | Country of birth of RR/MDR-TB cases (number of TB cases) | Site of disease (number of RR/MDR-TB cases) |
|---|---|---|---|---|---|---|
| **snpCL1** | 30 | 4.8 | R-R (*rpoB;* S450L; n=30);<br>H-R (*katG;* S315T; n=30; *inhA* prom; c-15t; n=30);<br>E-R (*embB;* M306I; n=30);<br>Z-R (*pncA;* A146V; n=30);<br>FQ-R (*gyrA*; D94Y; n=11; *gyrB*; A504V; n=3);<br>BDQ-R (Rv0678; large deletion; 5/30) | Romania (n=16)<br>Italy (n=12)<br>United Kingdom (n=2) | Romania (n=22)<br>Italy (n=5)<br>Albania (n=1)<br>Brazil (n=1)<br>Ukraine (n=1) | pulmonary (n=28)<br>unknown (n=2) |
| **snpCL3** | 16 | 4.6.2 | R-R (*rpoB;* S450L; n=16);<br>H-R (*katG;* S315T; n=16);<br>E-R (*embB;* M306I; n=16);<br>Z-R (*pncA;* W68C; n=16);<br>CAP-R (*tlyA*; N236K; n=16); | Germany (n=9)<br>Austria (n=4)<br>Italy (n=2)<br>France (n=1) | Somalia (n=3)<br>Sudan (n=1)<br>unknown (n=12) | pulmonary (n=3)<br>unknown (n=13) |
| **snpCL8** | 12 | 4.2.2 | R-R (*rpoB;* S450L; n=12);<br>H-R (*katG;* S315T; n=12);<br>E-R (*embB;* G406A; n=12);<br>Z-R (*pncA;* T76P; n=12); | Germany (n=5)<br>Italy (n=5)<br>Netherlands (n=1)<br>Sweden (n=1) | Somalia (n=7)<br>unknown (n=5) | pulmonary (n=3)<br>unknown (n=9) |

R: rifampicin; H: isoniazid; E: ethambutol; Z: pyrazinamide; FQ: fluoroquinolones; BDQ: bedaquiline; CAP: capreomycin; R: resistance; n: numbers.
All isolates were isolated between 2017 and 2019.

**Supplementary Figure S1**. Lineage classification of the 316 clustered RR/MDR *Mycobacterium tuberculosis* isolates

**a**

| Lineage | Number of strains | Percentage |
|---|---|---|
| 2.2.1 | 72 | 22.8 |
| 3 | 2 | 0.6 |
| 4.1.1.1 | 3 | 0.9 |
| 4.1.2 | 2 | 0.6 |
| 4.1.2.1 | 29 | 9.2 |
| 4.2.1 | 37 | 11.7 |
| 4.2.2 | 27 | 8.5 |
| 4.2.2.1 | 5 | 1.6 |
| 4.3 | 30 | 9.5 |
| 4.6.2 | 16 | 5.1 |
| 4.7; 4.8 | 91 | 28.8 |
| 4.9 | 2 | 0.6 |

**b**

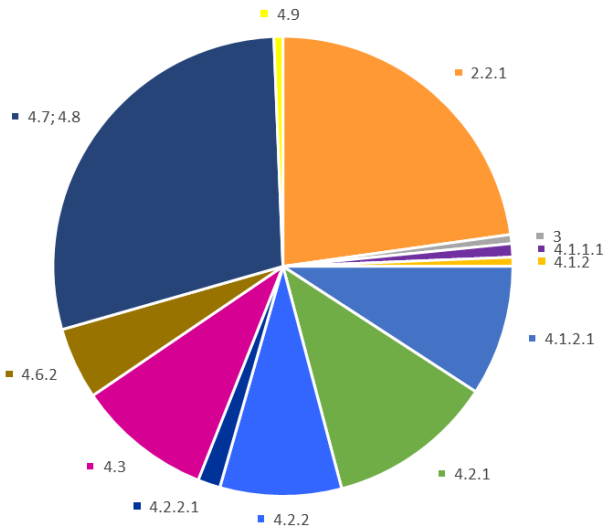
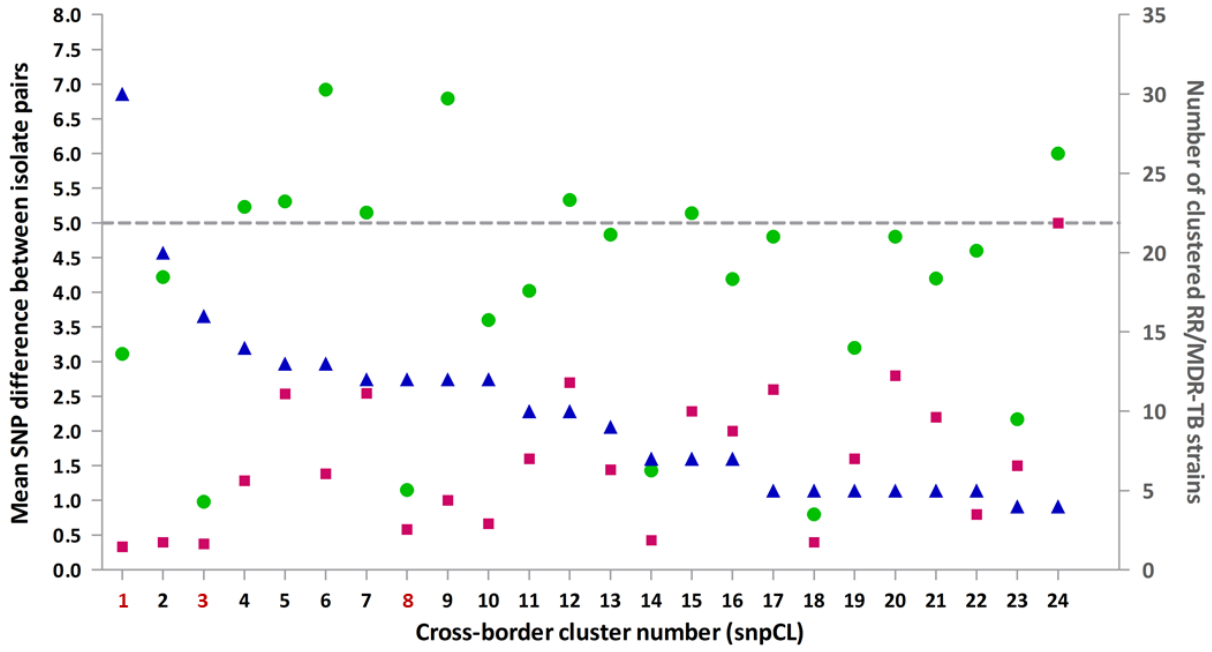



Table (a) and chart pie (b) showing the lineage distribution of the 316 RR/MDR-MTBC isolates in cross-border clusters. Lineages: 2.2.1: Beijing; 3 Delhi-CAS; 4.1.1.1: X-type; 4.1.2: Euro-American; 4.1.2.1: Haarlem; 4.2.1: Ural; 4.2.2: Euro-American; 4.2.2.1: TUR; 4.3: LAM; 4.6.2: Euro-American; 4.7: manly T; 4.8: mainly T; 4.9: H37Rv-like.

**Supplementary Figure S2**




Blue triangles (▲) represent the number of RR/MDR-MTBC isolates included in each SNP-based cross-border cluster, green dots (●) represent the mean SNP difference between each isolate pair of cluster (i.e. "cluster distribution"), red squares (■) represent the mean SNP difference between each isolate and its next closest isolate; Y left axis reports the mean SNP difference between isolate pairs; Y right axis reports the number of clustered isolates; X axis reports the identified SNP-based cross-border clusters (only those including at least four MTBC isolates are shown). The dashed line at 5 SNPs indicates the threshold selected to identify cross-border clusters with increased likelihood of recent transmission. Cross-border clusters number in red (snpCL1, snpCL3 and snpCL8) are described in the result section and Table 6.