
Supplementary information

Prisons as ecological drivers of fitness-compensated multidrug-resistant *Mycobacterium tuberculosis*

In the format provided by the authors and unedited

1 **1. Supplementary results**

2 **Whole genome sequencing summary statistics**

3 The final dataset included 1,613 strains with a median coverage of 104 X (25th percentile =
4 80, 75th percentile = 126, interquartile range (IQR) = 46). We observed a high percentage of reads that
5 mapped to the reference genome (median = 99.00 %, 25th percentile = 98.73, 75th percentile = 99.09,
6 IQR = 0.36), indicating the absence of major contaminants. The median percentage of the genome
7 with less than 7 X coverage was 0.89 % (25th percentile = 0.82, 75th percentile = 0.98, IQR = 0.16).
8 The median number of fixed mutations per genome was 898 (25th percentile = 891, 75th percentile =
9 903, IQR = 13) and the median number of unfixed mutations was 31 (25th percentile = 27, 75th
10 percentile = 36, IQR = 9).

11

12 **2. Supplementary tables**

13 **Supplementary Table S1 Sampling coverages per year**

14

15

Sampling year	WHO culture-confirmed MDR cases	Included in study	Proportion (%)
2011	475	166	34.9
2012	346	287	82.9
2013	400	211	52.8
2014	384	397	103.4
2015	368	324	88.0
2016	319	228	71.5
Total	2292	1613	70.4

16

17

18

19

20

21

22 **Supplementary Table S3** List of drug resistance mutations compiled from various studies.

Genomic position	Ref. base	Alt. base	Locus	Substitution	Drug‡	Ref.
1472358	C	T	rrs	C513T	SM	1
1472359	A	C	rrs	A514C	SM	1
1472361	C	T	rrs	C516T	SM	2
1472362	C	T	rrs	C517T	SM	1
1472749	T	G	rrs	C904G	SM	3
1472749	T	A	rrs	C904A	SM	3
1472751	A	G	rrs	A906G	SM	4
1472752	A	T	rrs	A907T	SM	4
1473246	A	G	rrs	A1401G	AK;CAP; KAN	1
1473247	C	T	rrs	C1402T	CAP	5
1473329	G	T	rrs	G1484T	AK;CAP	1
1472359	A	T	rrs	A514T	SM	1
1472750	C	G	rrs	C905G	SM	6
6620	G	A	Rv0005	D461N	FQ	7
6734	A	G	Rv0005	N499D	FQ	7
7563	G	T	Rv0006	G88C	FQ	S*
7564	G	C	Rv0006	G88A	FQ	8
7570	C	T	Rv0006	A90V	FQ	1
7572	T	C	Rv0006	S91P	FQ	1
7581	G	A	Rv0006	D94N	FQ	1
7581	G	C	Rv0006	D94H	FQ	S*
7581	G	T	Rv0006	D94Y	FQ	9
7582	A	G	Rv0006	D94G	FQ	1
7582	A	C	Rv0006	D94A	FQ	1

760314	G	T	Rv0667	V170F	RIF	1
760882	T	C	Rv0667	V359A	RIF	1
761100	C	G	Rv0667	Q432E	RIF	S*
761100	C	A	Rv0667	Q432K	RIF	1
761101	A	C	Rv0667	Q432P	RIF	10
761101	A	T	Rv0667	Q432L	RIF	S*
761109	G	T	Rv0667	D435Y	RIF	S*
761110	A	G	Rv0667	D435G	RIF	10
761110	A	T	Rv0667	D435V	RIF	1
761128	C	T	Rv0667	S441L	RIF	S*
761139	C	G	Rv0667	H445D	RIF	1
761139	C	T	Rv0667	H445Y	RIF	1
761139	C	A	Rv0667	H445N	RIF	1
761140	A	T	Rv0667	H445L	RIF	10
761140	A	C	Rv0667	H445P	RIF	11
761140	A	G	Rv0667	H445R	RIF	1
761149	G	A	Rv0667	R448Q	RIF	S*
761155	C	T	Rv0667	S450L	RIF	1
761155	C	G	Rv0667	S450W	RIF	1
761161	T	C	Rv0667	L452P	RIF	1
761277	A	T	Rv0667	I491F	RIF	1
761155/761156	C/G	T/C	Rv0667	S450F	RIF	1
761139/761140	C/A	T/T	Rv0667	H445S	RIF	10
761139/761140	C/A	T/G	Rv0667	H445C	RIF	10
761109/761110	G/A	T/T	Rv0667	D435F	RIF	1
761095	T	C	Rv0667	L430P	RIF	10
761095	T	G	Rv0667	L430R	RIF	10
761277	A	G	Rv0667	I491V	RIF	1

781687	A	G	Rv0682	K43R	SM	1
781822	A	G	Rv0682	K88R	SM	1
1673423	G	T	Rv1483	G-17T	INH;ETH	1
1673425	C	T	Rv1483	C-15T	INH;ETH	1
1673432	T	C	Rv1483	T-8C	INH;ETH	1
1674263	T	C	Rv1484	I21T	INH;ETH	1
1674481	T	G	Rv1484	S94A	INH;ETH	1
1674782	T	C	Rv1484	I194T	INH	1
1917857	C	T	Rv1694	C-83T	CAP	1
1918647	T	A	Rv1694	N236K	CAP	10
1918647	T	G	Rv1694	N236K	CAP	10
2154214	A	G	Rv1908c	V633A	INH	1
2154218	G	A	Rv1908c	R632C	INH	12
2154488	C	G	Rv1908c	D542H	INH	12
2154853	A	G	Rv1908c	M420T	INH	12
2154857	C	G	Rv1908c	D419H	INH	12
2155059	C	T	Rv1908c	W351STOP	INH	S*
2155060	C	T	Rv1908c	W351STOP	INH	S*
2155126	T	G	Rv1908c	D329A	INH	6
2155129	C	A	Rv1908c	W328L	INH	1
2155167	G	T	Rv1908c	S315R	INH	12
2155167	G	C	Rv1908c	S315R	INH	12
2155168	C	A	Rv1908c	S315I	INH	S*
2155168	C	G	Rv1908c	S315T	INH	1
2155168	C	T	Rv1908c	S315N	INH	1
2155169	T	G	Rv1908c	S315R	INH	12
2155212	C	A	Rv1908c	W300C	INH	1
2155212	C	G	Rv1908c	W300C	INH	1

2155289	T	G	Rv1908c	T275P	INH	6
2155417	G	C	Rv1908c	P232R	INH	6
2155541	A	T	Rv1908c	W191R	INH	1
2155541	A	G	Rv1908c	W191R	INH	1
2155573	G	T	Rv1908c	T180K	INH	1
2155636	A	G	Rv1908c	L159P	INH	1
2155648	T	G	Rv1908c	Y155S	INH	6
2155814	G	T	Rv1908c	P100T	INH	6
2155844	A	T	Rv1908c	W90R	INH	1
2155844	A	G	Rv1908c	W90R	INH	1
2155169	T	C	Rv1908c	S315G	INH	13
2289252	T	C	Rv2043c	A-11G	PZA	1
2289253	A	G	Rv2043c	T-12C	PZA	1
2289252	T	G	Rv2043c	A-11C	PZA	14
2715338	C	A	Rv2416c	G-6T	KAN	15
2715342	C	T	Rv2416c	G-10A	KAN	1
2715342	C	G	Rv2416c	G-10C	KAN	16
2715344	G	A	Rv2416c	C-12T	KAN	16
2715345	T	C	Rv2416c	A-13G	KAN	17
2715346	G	A	Rv2416c	C-14T	KAN;CAP	10
2715347	G	A	Rv2416c	C-15T	KAN	15
2715369	C	A	Rv2416c	G-37T	KAN	17
2726136	C	T	Rv2428	C-57T	INH	1
2747141	T	C	Rv2447c	E153G	PAS	18
2747144	A	G	Rv2447c	F152S	PAS	18
2747145	A	G	Rv2447c	F152L	PAS	18
2747149	G	C	Rv2447c	S150R	PAS	18
2747151	T	C	Rv2447c	S150G	PAS	18

2747381	T	C	Rv2447c	N73S	PAS	18
2747480	T	C	Rv2447c	E40G	PAS	18
2747052	C	G	Rv2447c	A183P	PAS	18
2747454	G	A	Rv2447c	R49W	PAS	18
2747541	T	G	Rv2447c	T20P	PAS	18
2747481	C	G	Rv2447c	E40Q	PAS	18
2747471	A	G	Rv2447c	I43T	PAS	18
2747471	A	C	Rv2447c	I43S	PAS	18
2747453	C	G	Rv2447c	R49P	PAS	18
2747433	G	C	Rv2447c	L56V	PAS	18
2747328	G	A	Rv2447c	R91W	PAS	18
2747151	T	A	Rv2447c	S150C	PAS	18
2747141	T	G	Rv2447c	E153A	PAS	18
2986860	G	C	Rv2671	G8R	PAS	18
3073808	G	C	Rv2764c	R222G	PAS	18
3074171	A	G	Rv2764c	W101R	PAS	18
3074249	G	T	Rv2764c	H75N	PAS	18
3074408	T	C	Rv2764c	T22A	PAS	18
3074365	T	C	Rv2764c	Y36C	PAS	18
3074246	C	A	Rv2764c	G76*	PAS	18
3074243	C	A	Rv2764c	V77F	PAS	18
3074223	C	A	Rv2764c	W83C	PAS	18
3074224	C	T	Rv2764c	W83*	PAS	18
3074201	C	T	Rv2764c	G91R	PAS	18
3074178	C	T	Rv2764c	W98*	PAS	18
3074159	A	G	Rv2764c	S105P	PAS	18
3074095	C	T	Rv2764c	R126Q	PAS	18
3074018	A	C	Rv2764c	F152V	PAS	18

3073990	C	T	Rv2764c	C161Y	PAS	18
3073852	T	C	Rv2764c	H207R	PAS	18
3073801	G	A	Rv2764c	P224L	PAS	18
3073768	C	G	Rv2764c	R235P	PAS	18
3074211	TG	T	Rv2764c	NA	PAS	18
3074360	AA	A	Rv2764c	NA	PAS	18
3074254	T	TGTG CTCGT G	Rv2764c	NA	PAS	18
3074099	A	A	Rv2764c	NA	PAS	18
3073999	CG	C	Rv2764c	NA	PAS	18
4243217	C	G	Rv3794	C-16G	EMB	1
4243221	C	T	Rv3794	C-12T	EMB	1
4243217	C	T	Rv3794	C-16T	EMB	19
4243217	C	A	Rv3794	C-16A	EMB	19
4247429	A	T	Rv3795	M306L	EMB	20
4247429	A	C	Rv3795	M306L	EMB	20
4247429	A	G	Rv3795	M306V	EMB	1
4247431	G	C	Rv3795	M306I	EMB	1
4247431	G	A	Rv3795	M306I	EMB	1
4247431	G	T	Rv3795	M306I	EMB	1
4247469	A	C	Rv3795	Y319S	EMB	6
4247495	G	T	Rv3795	D328Y	EMB	6
4247496	A	G	Rv3795	D328G	EMB	6
4247574	A	C	Rv3795	D354A	EMB	1
4247728	G	T	Rv3795	E405D	EMB	6
4247728	G	C	Rv3795	E405D	EMB	6
4247729	G	T	Rv3795	G406C	EMB	20
4247730	G	A	Rv3795	G406D	EMB	6

4247730	G	C	Rv3795	G406A	EMB	6
4248002	C	A	Rv3795	Q497K	EMB	1
4248003	A	T	Rv3795	Q497L	EMB	6
4248003	A	G	Rv3795	Q497R	EMB	1
4248027	C	T	Rv3795	A505V	EMB	6
4249512	T	G	Rv3795	M1000R	EMB	6
4249583	G	A	Rv3795	D1024N	EMB	6
4249757	A	G	Rv3795	T1082A	EMB	6
4248003	A	C	Rv3795	Q497P	EMB	6
4247469	A	G	Rv3795	Y319C	EMB	6
4407554	G	C	Rv3919c	R217G	SM	6
4407598	A	T	Rv3919c	A202E	SM	1
4407604	G	T	Rv3919c	A200E	SM	1
4407730	C	A	Rv3919c	R158L	SM	6
4407756	G	T	Rv3919c	S149R	SM	6
4407756	G	C	Rv3919c	S149R	SM	6
4407758	T	G	Rv3919c	S149R	SM	6
4407790	G	A	Rv3919c	A138V	SM	1
4407791	C	T	Rv3919c	A138T	SM	1
4407794	G	A	Rv3919c	R137W	SM	1
4407802	G	T	Rv3919c	A134E	SM	1
4407880	A	C	Rv3919c	L108R	SM	6
4407922	A	G	Rv3919c	L94P	SM	1
4407931	A	G	Rv3919c	L91P	SM	1
4407940	A	G	Rv3919c	V88A	SM	1
4407952	G	A	Rv3919c	P84L	SM	6
4407965	C	G	Rv3919c	A80P	SM	1
4407973	A	C	Rv3919c	V77G	SM	6

4407979	G	A	Rv3919c	P75L	SM	1
4407985	C	T	Rv3919c	G73E	SM	6
4407985	C	G	Rv3919c	G73A	SM	6
4407994	C	T	Rv3919c	S70N	SM	1
4407997	C	T	Rv3919c	G69D	SM	1
4408003	T	C	Rv3919c	D67G	SM	6
4408007	C	A	Rv3919c	V66L	SM	6
4408007	C	G	Rv3919c	V66L	SM	6
4408009	A	C	Rv3919c	V65G	SM	1
4408061	G	T	Rv3919c	H48N	SM	1
4407952	G	C	Rv3919c	P84R	SM	6

23

24 ‡. Abbreviations of drug names are as follows: SM: Streptomycin, AK: Amikacin, CAP:
25 Capreomycin, KAN: Kanamycin, FQ: Fluoroquinolones, RIF: Rifampicin, INH: Isoniazid,
26 ETH: Ethionamide, PZA: Pyrazinamide, PAS: Para-aminosalicylic acid, EMB: Ethambutol
27 *S: Spontaneous mutant generated in our laboratory by plating drug sensitive strains on
28 medium containing the corresponding drug.

29

30 **Supplementary Table S5** List of putative compensatory mutations in the genes *rpoA*, *rpoB* and *rpoC*,
 31 identified in this study, after applying filtering for phylogenetic markers.

Gene	AA Substitution	Genomic position(s)	Ref. Allele(s)	Alt. Allele(s)
rpoA_Rv3457c	p.Ala180Val	3877969	G	A
rpoA_Rv3457c	p.Arg182Trp	3877964	G	A
rpoA_Rv3457c	p.Glu184Asp	3877956	C	A
rpoA_Rv3457c	p.Gly31Ala	3878416	C	G
rpoA_Rv3457c	p.Gly31Ser	3878417	C	T
rpoA_Rv3457c	p.Thr187Ala	3877949	T	C
rpoA_Rv3457c	p.Val183Ala	3877960	A	G
rpoA_Rv3457c	p.Val183Gly	3877960	A	C
rpoB_Rv0667	p.Arg552Cys	761460	C	T
rpoB_Rv0667	p.Arg552His	761461	G	A
rpoB_Rv0667	p.Arg827Cys	762285	C	T
rpoB_Rv0667	p.Arg827His	762286	G	A
rpoB_Rv0667	p.Arg827Leu	762286	G	T
rpoB_Rv0667	p.Asn437Asp	761115	A	G
rpoB_Rv0667	p.Gln409Arg	761032	A	G
rpoB_Rv0667	p.Glu761Asp	762089	G	C
rpoB_Rv0667	p.His445Cys	761139/761140	C/A	T/G
rpoB_Rv0667	p.His445Ser	761139/761140	C/A	A/G
rpoB_Rv0667	p.His445Ser	761139/761140	C/A	T/C
rpoB_Rv0667	p.His835Arg	762310	A	G
rpoB_Rv0667	p.Leu731Pro	761998	T	C
rpoB_Rv0667	p.Lys446Gln	761142	A	C
rpoB_Rv0667	p.Met434Ile	761108	G	C
rpoB_Rv0667	p.Met434Val	761106	A	G
rpoB_Rv0667	p.Pro45Arg	759940	C	G
rpoB_Rv0667	p.Pro45Leu	759940	C	T
rpoB_Rv0667	p.Pro768Leu	762109	C	T
rpoB_Rv0667	p.Ser450Ala	761154	T	G
rpoB_Rv0667	p.Ser450Gly	761154/761155	T/C	G/G
rpoB_Rv0667	p.Ser450Phe	761155/761156	C/G	T/C
rpoB_Rv0667	p.Thr427Ser	761086	C	G
rpoB_Rv0667	p.Val496Ala	761293	T	C
rpoC_Rv0668	p.Ala492Val	764844	C	T
rpoC_Rv0668	p.Ala521Asp	764931	C	A
rpoC_Rv0668	p.Arg459Trp	764744	C	T
rpoC_Rv0668	p.Arg770His	765678	G	A
rpoC_Rv0668	p.Asn698His	765461	A	C
rpoC_Rv0668	p.Asn698Lys	765463	C	A
rpoC_Rv0668	p.Asn698Ser	765462	A	G
rpoC_Rv0668	p.Asp485Asn	764822	G	A
rpoC_Rv0668	p.Asp485Tyr	764822	G	T
rpoC_Rv0668	p.Asp747Ala	765609	A	C
rpoC_Rv0668	p.Asp747Asn	765608	G	A
rpoC_Rv0668	p.Gln523Lys	764936	C	A
rpoC_Rv0668	p.Gly332Arg	764363	G	C
rpoC_Rv0668	p.Gly332Cys	764363	G	T
rpoC_Rv0668	p.Gly332Ser	764363	G	A
rpoC_Rv0668	p.Gly433Ser	764666	G	A

rpoC_Rv0668	p.Gly945Val	766203	G	T
rpoC_Rv0668	p.His525Asn	764942	C	A
rpoC_Rv0668	p.His525Gln	764944	C	A
rpoC_Rv0668	p.His525Gln	764944	C	G
rpoC_Rv0668	p.Ile491Thr	764841	T	C
rpoC_Rv0668	p.Ile491Val	764840	A	G
rpoC_Rv0668	p.Leu449Val	764714	C	G
rpoC_Rv0668	p.Leu516Pro	764916	T	C
rpoC_Rv0668	p.Leu527Val	764948	T	G
rpoC_Rv0668	p.Lys445Arg	764703	A	G
rpoC_Rv0668	p.Phe452Cys	764724	T	G
rpoC_Rv0668	p.Phe831Leu	765862	C	G
rpoC_Rv0668	p.Phe831Ser	765861	T	C
rpoC_Rv0668	p.Pro1040Arg	766488	C	G
rpoC_Rv0668	p.Pro434Leu	764670	C	T
rpoC_Rv0668	p.Trp484Gly	764819	T	G
rpoC_Rv0668	p.Val1252Leu	767123	G	T
rpoC_Rv0668	p.Val1252Met	767123	G	A
rpoC_Rv0668	p.Val431Met	764660	G	A
rpoC_Rv0668	p.Val483Ala	764817	T	C
rpoC_Rv0668	p.Val483Gly	764817	T	G
rpoC_Rv0668	p.Val517Ala	764919	T	C
rpoC_Rv0668	p.Val517Leu	764918	G	C

32

33

34

Supplementary Table S6 Overview of compensatory mutations in the three genes encoding the RNA polymerase subunits by lineage. *Proportion of the total number of Lineage 2 (N =1,386) and Lineage 4 (N = 277) *Mtb* strains harbouring compensatory mutations. Lineage 2 strains are more likely to harbor compensatory mutations, compared to Lineage 4 strains ($p < 0.001$, two-tailed χ^2 test).

Lineage	Compensated	Gene harbouring compensatory mut.		
		<i>rpoA</i>	<i>rpoB</i>	<i>rpoC</i>
L2	1041 (75%)*	93	381	567
L4	147 (35 %)*	5	17	58

Supplementary Table S8 Priors for mutation rate, the gamma-distributed generation and sampling time used in the phybreak analysis. Mean generation time (μ_g), Mean sampling time (μ_s), Shape parameter generation time (k_g), Shape parameter sampling time (k_s), Rate parameter generation time (β_g), Rate parameter sampling time (β_s). Mutation rate given as mutations per genome per year. Text in bold indicates the priors used for the main results of the study.

Prior set	Source	μ_g (days)	μ_s (days)	k_g	k_s	β_g	β_s	Mutation rate
A	This study/ Behr <i>et al.</i>¹⁷	257	189	4.66	3.05	0.0181	0.0161	1/4.4*10⁶/year
B	This study/ Behr <i>et al.</i> ¹⁷	257	189	4.66	3.05	0.0181	0.0161	0.5/4.4*10 ⁶ /year
C	This study/ Behr <i>et al.</i> ¹⁷	257	189	4.66	3.05	0.0181	0.0161	2/4.4*10 ⁶ /year
D	This study/ Behr <i>et al.</i> ¹⁷	151	173	3.05	4.66	0.0201	0.0268	1/4.4*10 ⁶ /year
E	Ayabina <i>et al.</i> ³⁹	1578	935	1.3	0.8	8.27*10 ⁻⁴	8.56*10 ⁻⁴	1/4.4*10 ⁶ /year
F	Walter <i>et al.</i> ⁴⁰	365	365	10	10	0.0274	0.0274	1/4.4*10 ⁶ /year

Supplementary Table S14 Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set B (Supplementary Table S8).

		Dependent variable: Rate of secondary cases				
		All isolates with metadata (PP > 0.5) N = 1263	Excluding isolates from incarcerated individuals (PP > 0.5) n = 1092		All isolates with metadata N = 1263	
Explanatory variables	Levels	Total	Univariable IRR (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)
Putative compensatory mutation in <i>rpoABC</i>	0	389 (30.80)	-	-	-	-
	1	874 (69.20)	1.65 (1.32 - 2.10, p < 0.001)	1.28 (1.01 - 1.65, p = 0.045)	1.27 (0.98 - 1.66, p = 0.071)	1.39 (1.13-1.73, p = 0.002)
Incarcerated individual	0	1092 (86.46)	-	-	-	-
	1	171 (13.54)	2.15 (1.72 - 2.66, p < 0.001)	1.43 (1.12 - 1.83, p = 0.004)	-	1.50 (1.21 - 1.85, p < 0.001)
Lineage 2 strain	0	180 (14.25)	-	-	-	-
	1	1083 (85.75)	3.07 (2.05 - 4.86, p < 0.001)	2.45 (1.59 - 3.97, p < 0.001)	2.80 (1.70 - 4.94, p < 0.001)	2.69 (1.84 - 4.08, p < 0.001)
Age	Mean (SD)	38.54 (13.75)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)
Sex Female	0	981 (77.67)	-	-	-	-
	1	282 (22.33)	0.63 (0.48 - 0.82, p = 0.001)	0.69 (0.52 - 0.90, p = 0.008)	0.70 (0.52 - 0.91, p = 0.011)	0.93 (0.74 - 1.15, p = 0.503)
Number of additional drug resistance mut.	Mean (SD)	2.02 (0.68)	1.22 (1.06 - 1.40, p = 0.006)	1.08 (0.90 - 1.29, p = 0.422)	1.07 (0.88 - 1.30, p = 0.464)	1.02 (0.87 - 1.19, p = 0.825)
Drug resistance profile	MDR	512 (40.54)	-	-	-	-
	Pre-XDR	534 (42.28)	1.17 (0.95 - 1.45, p = 0.149)	0.92 (0.74 - 1.15, p = 0.459)	0.95 (0.73 - 1.23, p = 0.674)	0.89 (0.74 - 1.08, p = 0.238)
	XDR	217 (17.18)	1.29 (0.99 - 1.68, p = 0.056)	1.19 (0.90 - 1.57, p = 0.214)	1.16 (0.84 - 1.57, p = 0.360)	0.97 (0.76 - 1.24, p = 0.809)
TB diagnosis in the past	0	716 (56.69)	-	-	-	-
	1	547 (43.31)	0.99 (0.82 - 1.20, p = 0.938)	0.94 (0.77 - 1.14, p = 0.524)	0.89 (0.70 - 1.13, p = 0.343)	0.94 (0.79 - 1.11, p = 0.465)

[‡]Incidence rate ratios were estimated by multivariable Poisson regression adjusting for the presence of putative compensatory mutations, incarceration status, *Mtb* lineage, patient age, patient sex, number of additional drug resistance mutations, drug resistance profile, TB diagnosis in the past. We corrected for unequal observation time of the isolates by including the isolation time as a category. For the 6 year study period we included 12 half-year categories. SD = standard deviation. PP = posterior probability

Supplementary Table S15 Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set C (Supplementary Table S8).

		Dependent variable: Rate of secondary cases					
Explanatory variables	Levels	All isolates with metadata (PP > 0.5) N = 1263		Excluding isolates from incarcerated individuals (PP > 0.5) n = 1092		All isolates with metadata N = 1263	
		Total	Univariable IRR (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	
Putative compensatory mutation in <i>rpoABC</i>	0	389 (30.80)	-	-	-	-	
	1	874 (69.20)	1.63 (1.30 - 2.07, p < 0.001)	1.27 (1.00 - 1.63, p = 0.055)	1.21 (0.94 - 1.58, p = 0.150)	1.37 (1.11 - 1.70, p = 0.003)	
Incarcerated individual	0	1092 (86.46)	-	-	-	-	
	1	171 (13.54)	2.20 (1.76 - 2.73, p < 0.001)	1.45 (1.13 - 1.85, p = 0.003)	-	1.51 (1.22 - 1.87, p < 0.001)	
Lineage 2 strain	0	180 (14.25)	-	-	-	-	
	1	1083 (85.75)	3.20 (2.12 - 5.12, p < 0.001)	2.60 (1.68 - 4.26, p < 0.001)	2.80 (1.70 - 4.94, p < 0.001)	2.64 (1.82 - 3.98, p < 0.001)	
Age	Mean (SD)	38.54 (13.75)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	
Sex Female	0	981 (77.67)	-	-	-	-	
	1	282 (22.33)	0.64 (0.49 - 0.83, p = 0.001)	0.69 (0.52 - 0.91, p = 0.010)	0.71 (0.53 - 0.93, p = 0.014)	0.96 (0.77 - 1.19, p = 0.735)	
Number of additional drug resistance mut.	Mean (SD)	2.02 (0.68)	1.21 (1.05 - 1.39, p = 0.009)	1.07 (0.89 - 1.28, p = 0.485)	1.08 (0.89 - 1.31, p = 0.432)	1.00 (0.85 - 1.16, p = 0.983)	
Drug resistance profile	MDR	512 (40.54)	-	-	-	-	
	pre-XDR	534 (42.28)	1.13 (0.91 - 1.40, p = 0.258)	0.89 (0.71 - 1.11, p = 0.286)	0.91 (0.70 - 1.18, p = 0.464)	0.91 (0.75 - 1.10, p = 0.314)	
	XDR	217 (17.18)	1.29 (0.98 - 1.67, p = 0.062)	1.18 (0.89 - 1.56, p = 0.233)	1.19 (0.87 - 1.61, p = 0.284)	0.98 (0.76 - 1.25, p = 0.859)	
TB diagnosis in the past	0	716 (56.69)	-	-	-	-	
	1	547 (43.31)	1.01 (0.83 - 1.22, p = 0.927)	0.95 (0.78 - 1.17, p = 0.647)	0.91 (0.71 - 1.15, p = 0.411)	0.95 (0.80 - 1.13, p = 0.566)	

[‡]Incidence rate ratios were estimated by multivariable Poisson regression adjusting for the presence of putative compensatory mutations, incarceration status, *Mtb* lineage, patient age, patient sex, number of additional drug resistance mutations, drug resistance profile, TB diagnosis in the past. We corrected for unequal observation time of the isolates by including the isolation time as a category. For the 6 year study period we included 12 half-year categories. SD = standard deviation. PP = posterior probability

Supplementary Table S16 Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set D (Supplementary Table S8).

		Dependent variable: Rate of secondary cases				
Explanatory variables	Levels	Total	All isolates with metadata (PP > 0.5) N = 1263		Excluding isolates from incarcerated individuals (PP > 0.5) n = 1092	All isolates with metadata N = 1263
			Univariable IRR (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)
Putative compensatory mutation in <i>rpoABC</i>	0	389 (30.80)	-	-	-	-
	1	874 (69.20)	1.58 (1.26 - 2.02, p < 0.001)	1.27 (0.99 - 1.63, p = 0.065)	1.22 (0.94 - 1.59, p = 0.146)	1.42 (1.16 - 1.77, p = 0.001)
Incarcerated individual	0	1092 (86.46)	-	-	-	-
	1	171 (13.54)	1.96 (1.54 - 2.45, p < 0.001)	1.37 (1.05 - 1.77, p = 0.019)	-	1.56 (1.26 - 1.92, p < 0.001)
Lineage 2 strain	0	180 (14.25)	-	-	-	-
	1	1083 (85.75)	2.61 (1.77 - 4.05, p < 0.001)	2.09 (1.37 - 3.33, p = 0.001)	2.27 (1.42 - 3.84, p = 0.001)	2.60 (1.79 - 3.92, p < 0.001)
Age	Mean (SD)	38.54 (13.75)	0.98 (0.97 - 0.98, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)
Sex Female	0	981 (77.67)	-	-	-	-
	1	282 (22.33)	0.76 (0.59 - 0.98, p = 0.036)	0.79 (0.60 - 1.03, p = 0.095)	0.81 (0.61 - 1.06, p = 0.133)	0.96 (0.77 - 1.19, p = 0.736)
Number of additional drug resistance mut.	Mean (SD)	2.02 (0.68)	1.20 (1.04 - 1.39, p = 0.015)	1.06 (0.88 - 1.27, p = 0.556)	1.05 (0.86 - 1.27, p = 0.625)	1.00 (0.85 - 1.16, p = 0.971)
Drug resistance profile	MDR	512 (40.54)	-	-	-	-
	Pre-XDR	534 (42.28)	1.17 (0.94 - 1.47, p = 0.156)	0.95 (0.76 - 1.21, p = 0.692)	0.94 (0.72 - 1.23, p = 0.652)	0.93 (0.77 - 1.13, p = 0.464)
	XDR	217 (17.18)	1.37 (1.04 - 1.79, p = 0.023)	1.29 (0.96 - 1.71, p = 0.085)	1.27 (0.93 - 1.74, p = 0.134)	0.99 (0.77 - 1.26, p = 0.910)
TB diagnosis in the past	0	716 (56.69)	-	-	-	-
	1	547 (43.31)	0.88 (0.72 - 1.07, p = 0.202)	0.84 (0.68 - 1.03, p = 0.097)	0.84 (0.66 - 1.06, p = 0.150)	1.04 (0.87 - 1.23, p = 0.677)

[‡]Incidence rate ratios were estimated by multivariable Poisson regression adjusting for the presence of putative compensatory mutations, incarceration status, *Mtb* lineage, patient age, patient sex, number of additional drug resistance mutations, drug resistance profile, TB diagnosis in the past. We corrected for unequal observation time of the isolates by including the isolation time as a category. For the 6 year study period we included 12 half-year categories. SD = standard deviation. PP = posterior probability

Supplementary Table S17 Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set E (Supplementary Table S8).

		Dependent variable: Rate of secondary cases						
Explanatory variables	Levels	Total	All isolates with metadata (PP > 0.5) N = 1263		Excluding isolates from incarcerated individuals (PP > 0.5) n = 1092		All isolates with metadata n = 1263	
			Univariable IRR (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)		
Putative compensatory mutation in <i>rpoABC</i>	0	389 (30.80)	-	-	-	-	-	-
	1	874 (69.20)	1.26 (0.99 - 1.64, p = 0.067)	1.09 (0.84 - 1.43, p = 0.536)	1.17 (0.88 - 1.58, p = 0.277)	1.19 (0.96 - 1.48, p = 0.115)		
Incarcerated individual	0	1092 (86.46)	-	-	-	-	-	-
	1	171 (13.54)	1.69 (1.28 - 2.21, p < 0.001)	1.15 (0.85 - 1.56, p = 0.354)	-	1.46 (1.17 - 1.82, p = 0.001)		
Lineage 2 strain	0	180 (14.25)	-	-	-	-	-	-
	1	1083 (85.75)	1.52 (1.07 - 2.25, p = 0.026)	1.38 (0.93 - 2.12, p = 0.127)	1.80 (1.13 - 2.98, p = 0.017)	1.85 (1.32 - 2.67, p = 0.001)		
Age	Mean (SD)	38.54 (13.75)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.98 - 0.99, p = 0.001)	0.99 (0.98 - 0.99, p = 0.002)	0.99 (0.98 - 0.99, p < 0.001)		
Sex Female	0	981 (77.67)	-	-	-	-	-	-
	1	282 (22.33)	0.98 (0.74 - 1.27, p = 0.857)	0.98 (0.74 - 1.30, p = 0.913)	0.99 (0.74 - 1.32, p = 0.966)	1.03 (0.82 - 1.29, p = 0.793)		
Number of additional drug resistance mut.	Mean (SD)	2.02 (0.68)	1.04 (0.89 - 1.23, p = 0.620)	0.89 (0.73 - 1.09, p = 0.253)	0.85 (0.68 - 1.05, p = 0.132)	0.92 (0.79 - 1.08, p = 0.334)		
Drug resistance profile	MDR	512 (40.54)	-	-	-	-	-	-
	Pre-XDR	534 (42.28)	1.37 (1.07 - 1.77, p = 0.013)	1.28 (0.98 - 1.68, p = 0.067)	1.15 (0.85 - 1.55, p = 0.363)	1.19 (0.98 - 1.45, p = 0.085)		
	XDR	217 (17.18)	1.41 (1.02 - 1.92, p = 0.033)	1.60 (1.14 - 2.23, p = 0.006)	1.43 (0.99 - 2.06, p = 0.054)	1.14 (0.87 - 1.49, p = 0.323)		
TB diagnosis in the past	0	716 (56.69)	-	-	-	-	-	-
	1	547 (43.31)	0.83 (0.66 - 1.04, p = 0.101)	0.85 (0.67 - 1.07, p = 0.168)	0.72 (0.55 - 0.95, p = 0.021)	1.06 (0.89 - 1.26, p = 0.534)		

[‡]Incidence rate ratios were estimated by multivariable Poisson regression adjusting for the presence of putative compensatory mutations, incarceration status, *Mtb* lineage, patient age, patient sex, number of additional drug resistance mutations, drug resistance profile, TB diagnosis in the past. We corrected for unequal observation time of the isolates by including the isolation time as a category. For the 6 year study period we included 12 half-year categories. SD = standard deviation. PP = posterior probability

Supplementary Table S18 Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set F (Supplementary Table S8).

		Dependent variable: Rate of secondary cases				
Explanatory variables	Levels	Total	All isolates with metadata (PP > 0.5) N = 1263		Excluding isolates from incarcerated individuals (PP > 0.5) n = 1092	All isolates with metadata N = 1263
			Univariable IRR (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)	Multivariable [‡]IRR_{adj} (CI₉₅, p-value)
Putative compensatory mutation in <i>rpoABC</i>	0	389 (30.80)	-	-	-	-
	1	874 (69.20)	1.79 (1.42 - 2.29, p < 0.001)	1.44 (1.12 - 1.86, p = 0.005)	1.44 (1.11 - 1.90, p = 0.008)	1.40 (1.13 - 1.74, p = 0.002)
Incarcerated individual	0	1092 (86.46)	-	-	-	-
	1	171 (13.54)	2.09 (1.67 - 2.61, p < 0.001)	1.45 (1.12 - 1.86, p = 0.004)	-	1.67 (1.35 - 2.07, p < 0.001)
Lineage 2 strain	0	180 (14.25)	-	-	-	-
	1	1083 (85.75)	3.48 (2.26 - 5.71, p < 0.001)	2.94 (1.86 - 4.94, p < 0.001)	3.40 (2.01 - 6.25, p < 0.001)	2.25 (1.57 - 3.34, p < 0.001)
Age	Mean (SD)	38.54 (13.75)	0.98 (0.97 - 0.98, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.97 - 0.99, p < 0.001)	0.98 (0.98 - 0.99, p < 0.001)
Sex Female	0	981 (77.67)	-	-	-	-
	1	282 (22.33)	0.84 (0.65 - 1.06, p = 0.155)	0.91 (0.70 - 1.18, p = 0.495)	0.94 (0.72 - 1.22, p = 0.656)	0.91 (0.73 - 1.14, p = 0.437)
Number of additional drug resistance mut.	Mean (SD)	2.02 (0.68)	1.18 (1.02 - 1.36, p = 0.023)	1.01 (0.84 - 1.21, p = 0.945)	0.99 (0.81 - 1.21, p = 0.937)	1.00 (0.85 - 1.17, p = 0.994)
Drug resistance profile	MDR	512 (40.54)	-	-	-	-
	Pre-XDR	534 (42.28)	1.08 (0.87 - 1.34, p = 0.480)	0.83 (0.66 - 1.04, p = 0.103)	0.89 (0.69 - 1.16, p = 0.390)	0.93 (0.77 - 1.12, p = 0.434)
	XDR	217 (17.18)	1.21 (0.92 - 1.58, p = 0.164)	1.10 (0.82 - 1.45, p = 0.525)	1.11 (0.80 - 1.51, p = 0.529)	1.01 (0.78 - 1.29, p = 0.952)
TB diagnosis in the past	0	716 (56.69)	-	-	-	-
	1	547 (43.31)	0.87 (0.72 - 1.06, p = 0.179)	0.85 (0.69 - 1.04, p = 0.113)	0.80 (0.63 - 1.01, p = 0.067)	0.88 (0.74 - 1.05, p = 0.158)

[‡]Incidence rate ratios were estimated by multivariable Poisson regression adjusting for the presence of putative compensatory mutations, incarceration status, *Mtb* lineage, patient age, patient sex, number of additional drug resistance mutations, drug resistance profile, TB diagnosis in the past. We corrected for unequal observation time of the isolates by including the isolation time as a category. For the 6-year study period we included 12 half-year categories. SD = standard deviation. PP = posterior probability

Supplementary Table S19 Secondary case rates and secondary case rate ratios stratified by index case being an incarcerated individual or a non-incarcerated individual. I = Incarcerated individual, NI = Non-incarcerated individual

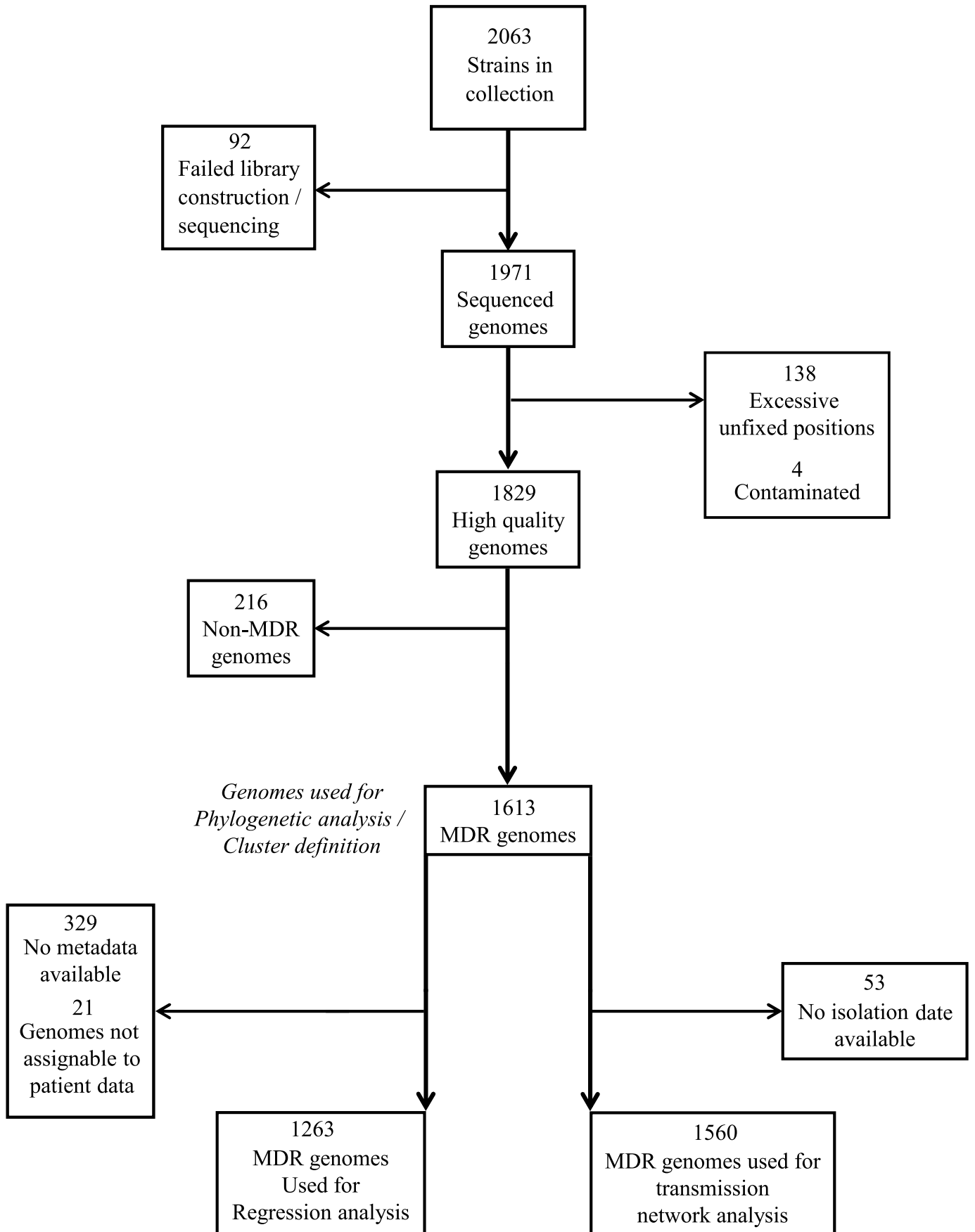
Class	n	(%)	Rate	Rate ratio	CI ₉₅	P-value
I-derived NI case	181	11.2	3.6	10.6	8.7-13.1	< 0.01
NI-derived NI case	183	11.3	0.3			
Compensated I-derived NI case	177	11.0	5.2	13.6	10.8-17.1	< 0.01
Compensated NI-derived NI case	126	7.8	0.4			
Not compensated I-derived NI case	4	0.24	0.2	1	0.3-2.4	0.94
Not compensated NI-derived NI case	54	3.3	0.2			
Total	1613	100				

43 3. Supplementary references

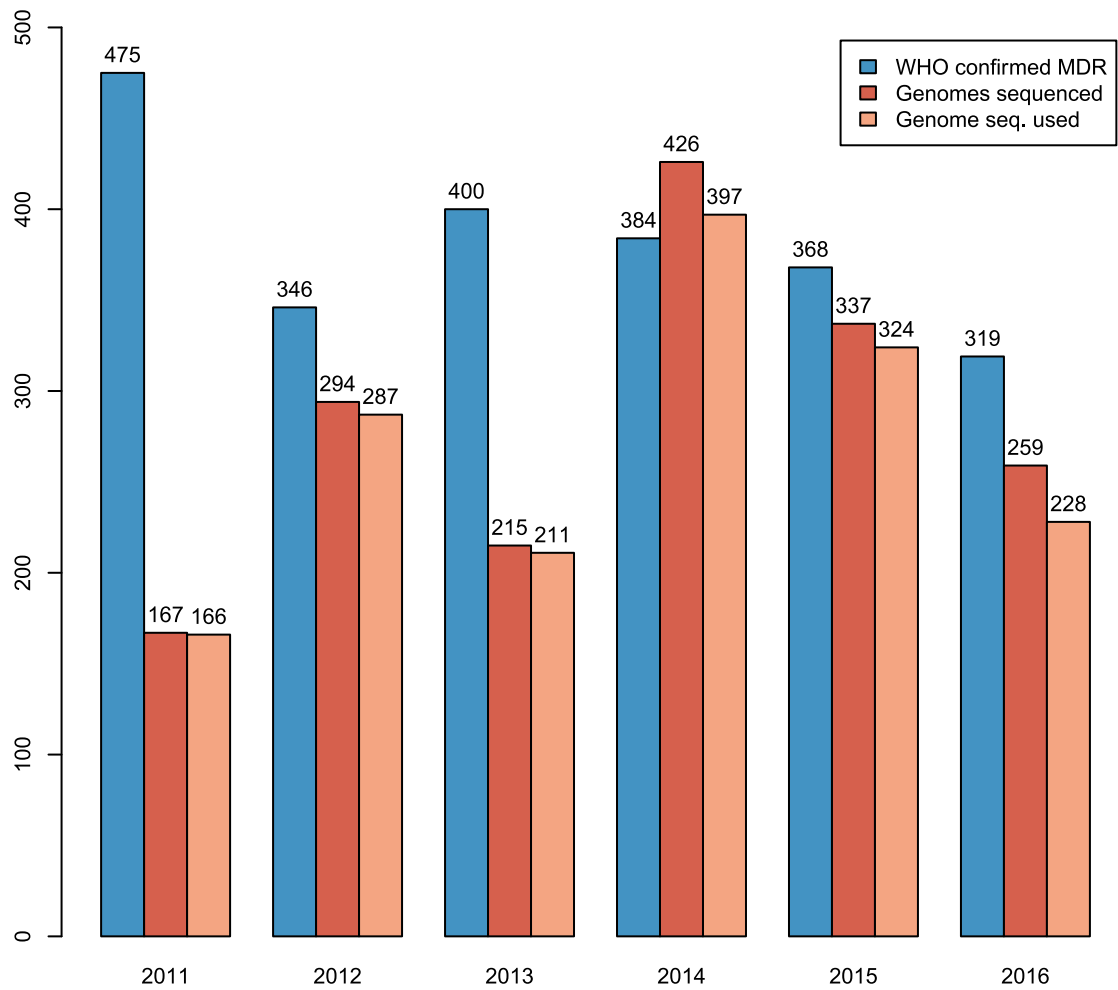
- 44 1. Walker, T. M. *et al.* Whole-genome sequencing for prediction of *Mycobacterium tuberculosis*
45 drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect. Dis.* **15**, 1193–1202
46 (2015).
- 47 2. Tudo, G. *et al.* Characterization of mutations in streptomycin-resistant *Mycobacterium*
48 *tuberculosis* clinical isolates in the area of Barcelona. *J. Antimicrob. Chemother.* **65**, 2341–2346
49 (2010).
- 50 3. Honore, N. & Cole, S. T. Streptomycin resistance in mycobacteria. *Antimicrob. Agents*
51 *Chemother.* **38**, 238–242 (1994).
- 52 4. Jagielski, T. *et al.* Screening for Streptomycin Resistance-Confering Mutations in
53 *Mycobacterium tuberculosis* Clinical Isolates from Poland. *PLoS ONE* **9**, e100078 (2014).
- 54 5. Maus, C. E., Plikaytis, B. B. & Shinnick, T. M. Molecular Analysis of Cross-Resistance to
55 Capreomycin, Kanamycin, Amikacin, and Viomycin in *Mycobacterium tuberculosis*. *Antimicrob.*
56 *Agents Chemother.* **49**, 3192–3197 (2005).
- 57 6. Gygli, S. M. *et al.* Whole-Genome Sequencing for Drug Resistance Profile Prediction in
58 *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **63**, e02175-18 (2019).

- 59 7. Malik, S., Willby, M., Sikes, D., Tsodikov, O. V. & Posey, J. E. New Insights into
60 Fluoroquinolone Resistance in *Mycobacterium tuberculosis*: Functional Genetic Analysis of gyrA
61 and gyrB Mutations. *PLoS ONE* **7**, e39754 (2012).
- 62 8. Matrat, S. *et al.* Functional analysis of DNA gyrase mutant enzymes carrying mutations at
63 position 88 in the A subunit found in clinical strains of *Mycobacterium tuberculosis* resistant to
64 fluoroquinolones. *Antimicrob. Agents Chemother.* **50**, 4170–4173 (2006).
- 65 9. Bloemberg, G. V. *et al.* Acquired Resistance to Bedaquiline and Delamanid in Therapy for
66 Tuberculosis. *N. Engl. J. Med.* **373**, 1986–1988 (2015).
- 67 10. Miotto, P. *et al.* A standardised method for interpreting the association between mutations and
68 phenotypic drug resistance in *Mycobacterium tuberculosis*. *Eur. Respir. J.* **50**, 1701354 (2017).
- 69 11. Gagneux, S. *et al.* The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*.
70 *Science* **312**, 1944–6 (2006).
- 71 12. Ando, H. *et al.* Identification of katG Mutations Associated with High-Level Isoniazid Resistance
72 in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **54**, 1793–1799 (2010).
- 73 13. Suarez, J., Rangelova, K., Schelvis, J. P. M. & Magliozzo, R. S. Antibiotic resistance in
74 *Mycobacterium tuberculosis*. Peroxidase intermediate bypass causes poor isoniazid activation by
75 the S315G mutant of *M. tuberculosis* catalase-peroxidase (KatG). *J. Biol. Chem.* **284**, 16146–
76 16155 (2009).
- 77 14. Ramirez-Busby, S. M. & Valafar, F. Systematic review of mutations in pyrazinamidase
78 associated with pyrazinamide resistance in *mycobacterium tuberculosis* clinical isolates.
79 *Antimicrob. Agents Chemother.* **59**, 5267–5277 (2015).
- 80 15. Engström, A., Perskvist, N., Werngren, J., Hoffner, S. E. & Juréen, P. Comparison of clinical
81 isolates and in vitro selected mutants reveals that tlyA is not a sensitive genetic marker for
82 capreomycin resistance in *Mycobacterium tuberculosis*. *J. Antimicrob. Chemother.* **66**, 1247–
83 1254 (2011).
- 84 16. Kambli, P. *et al.* Correlating rrs and eis promoter mutations in clinical isolates of *Mycobacterium*
85 *tuberculosis* with phenotypic susceptibility levels to the second-line injectables. *Int. J.*
86 *Mycobacteriology* **5**, 1–6 (2016).

- 87 17. Zaunbrecher, M. A., Sikes, R. D., Metchock, B., Shinnick, T. M. & Posey, J. E. Overexpression
88 of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin
89 resistance in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* **106**, 20004–20009 (2009).
- 90 18. Coll, F. *et al.* Rapid determination of anti-tuberculosis drug resistance from whole-genome
91 sequences. *Genome Med.* **7**, 51 (2015).
- 92 19. Cui, Z. *et al.* Mutations in the embC-embA intergenic region contribute to mycobacterium
93 tuberculosis resistance to ethambutol. *Antimicrob. Agents Chemother.* **58**, 6837–6843 (2014).
- 94 20. Safi, H. *et al.* Evolution of high-level ethambutol-resistant tuberculosis through interacting
95 mutations in decaprenylphosphoryl- β -D-arabinose biosynthetic and utilization pathway genes.
96 *Nat. Genet.* **45**, 1190–1197 (2013).
- 97

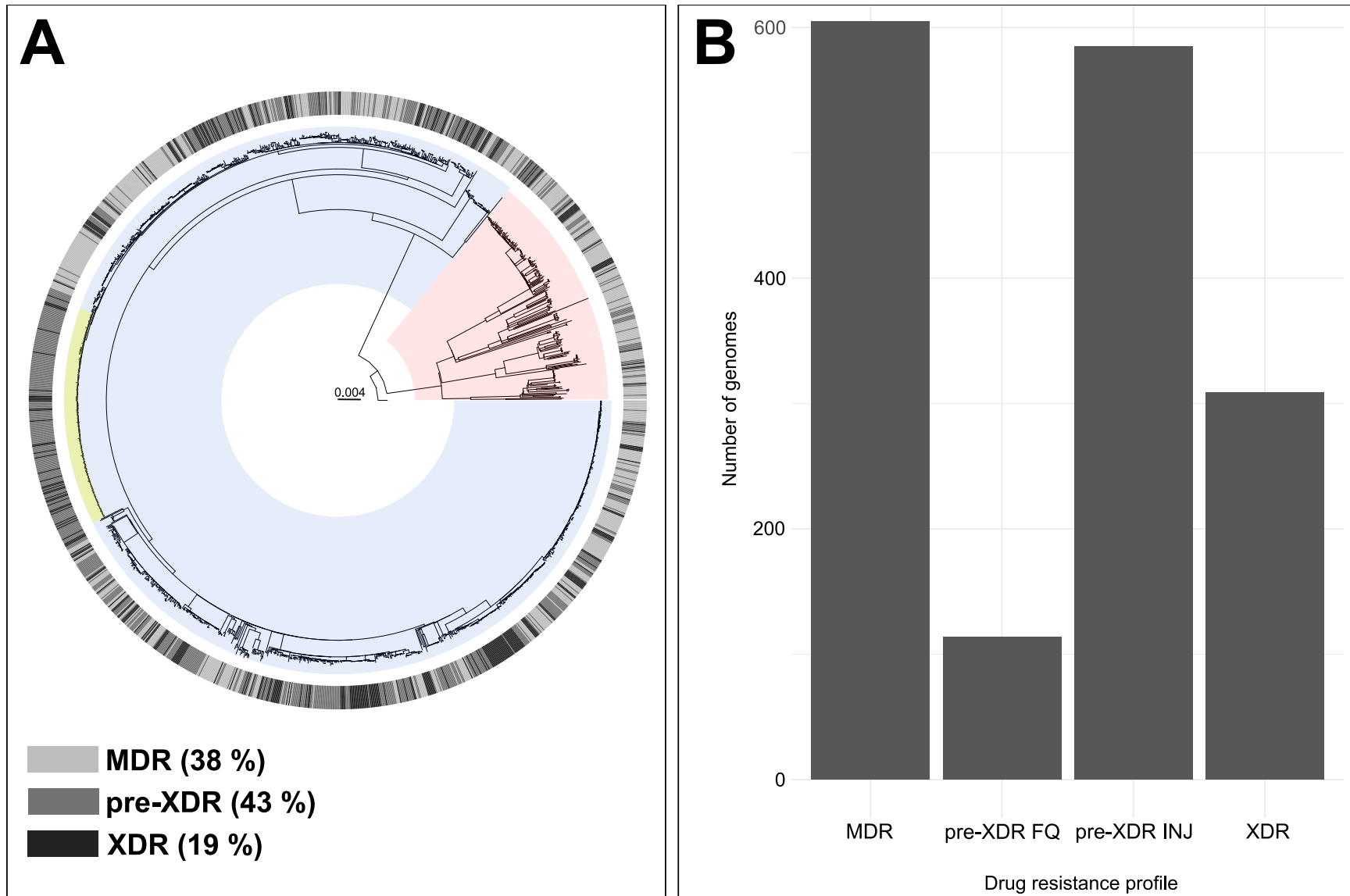


Supplementary Figure S1. Flow diagram of samples included in the study

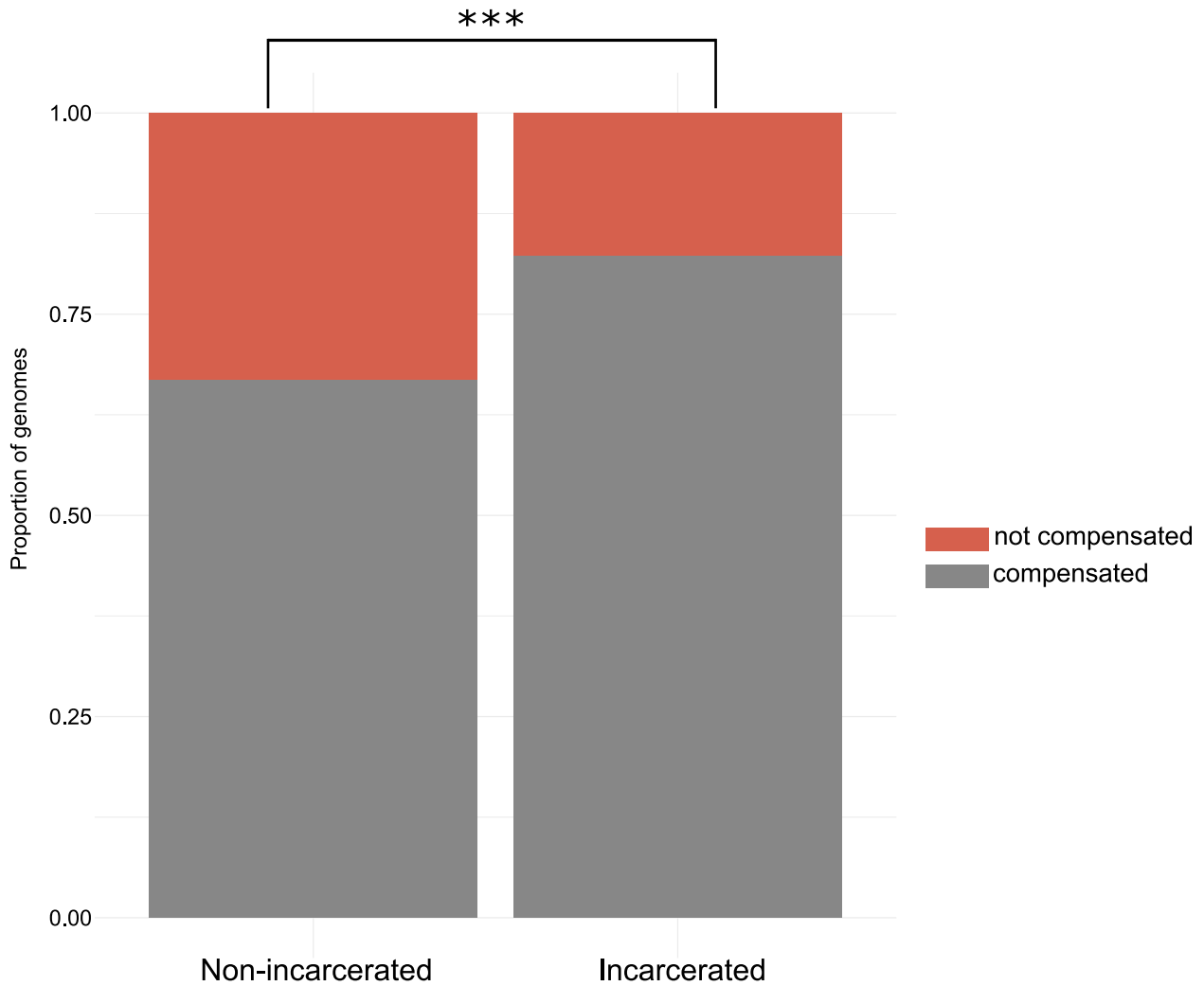


Supplementary Figure S2. Analysed MDR TB strains vs. notified MDR TB strains

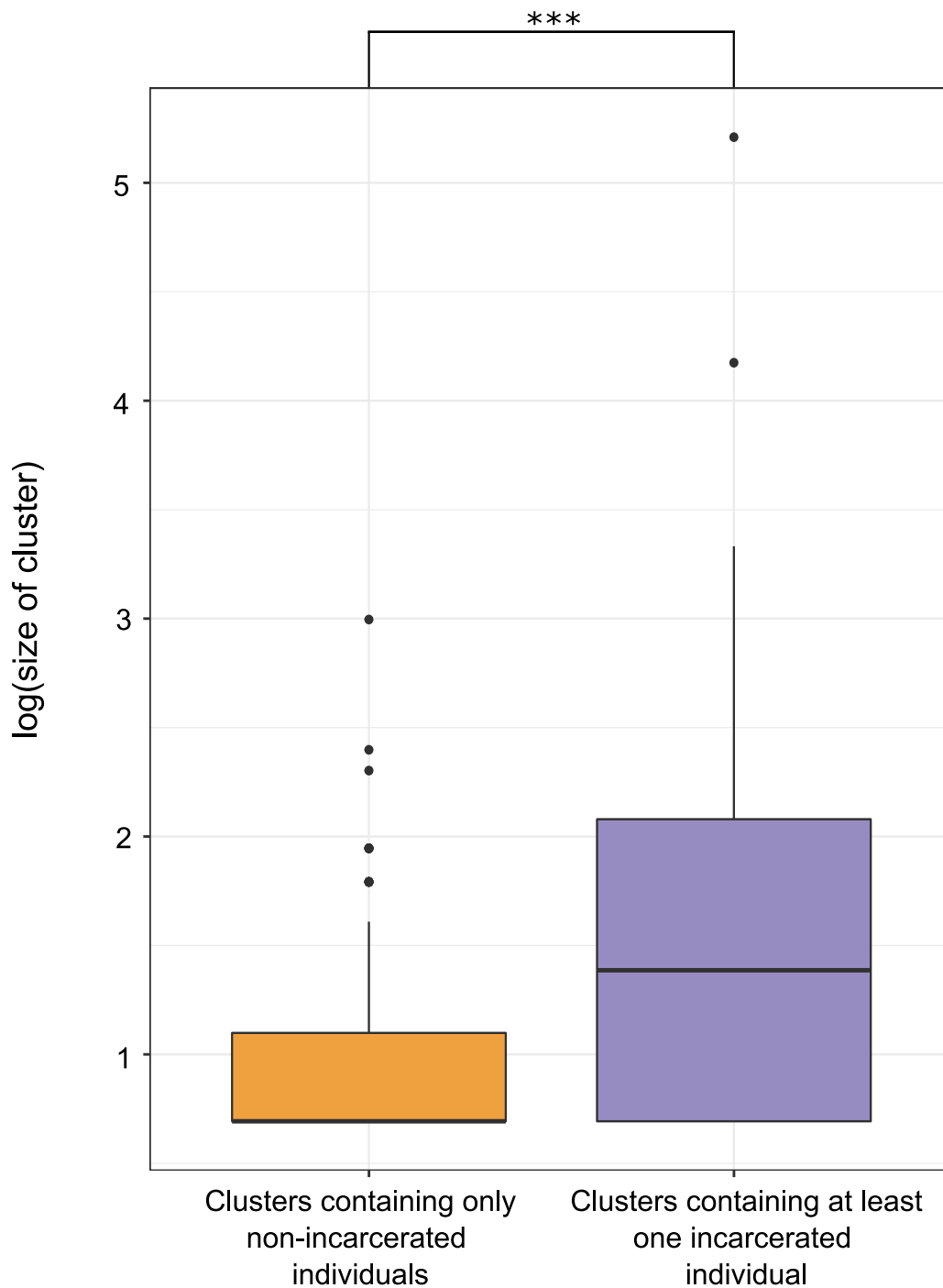
Total number of sequenced strains or used sequences in relation to the WHO reported culture-confirmed MDR TB cases in Georgia during 2011-2016.



Supplementary Figure S3 Drug resistance profiles of the strains included in the study A. Drug resistance profiles mapped to the phylogenetic tree of all samples included in the study. Lineage 2 strains are shaded in blue, Lineage 4 strains in red. The largest cluster based on an average SNV distance of 5 is shaded in yellow. **B.** Number of strains displaying and MDR (isoniazid + rifampicin resistance), pre-XDR FQ (MDR + fluoroquinolone resistance), pre-XDR INJ (MDR + amikacin, kanamycin or capreomycin resistance) or XDR resistance profile).



Supplementary Figure S4. Proportion of compensated strains among incarcerated and non-incarcerated individuals Proportions of strains carrying compensatory mutations stratified according to incarceration status (yes/no). Isolates from incarcerated individuals were more likely to carry a compensatory mutation compared to isolates from non-incarcerated individuals ($p = 5.844 \times 10^{-5}$, two-tailed χ^2 -test).

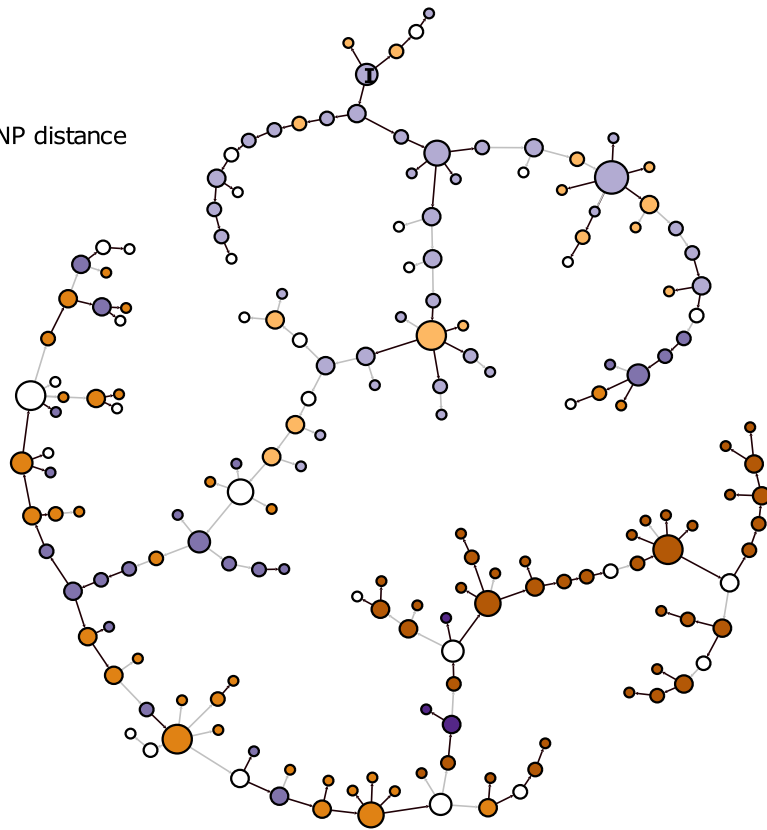


Supplementary Figure S5. Sizes of clusters containing only non-incarcerated individuals (n = 155) vs. clusters containing at least one incarcerated individual (n = 43) The median size of clusters containing at least one incarcerated individual (median = 4, 25th percentile = 2, 75th percentile = 8, interquartile range = 6) was larger compared to clusters containing only non-incarcerated individuals (median = 2, 25th percentile = 2, 75th percentile = 3, interquartile range = 1) ($p = 0.0001$, two-tailed Wilcoxon rank-sum test).

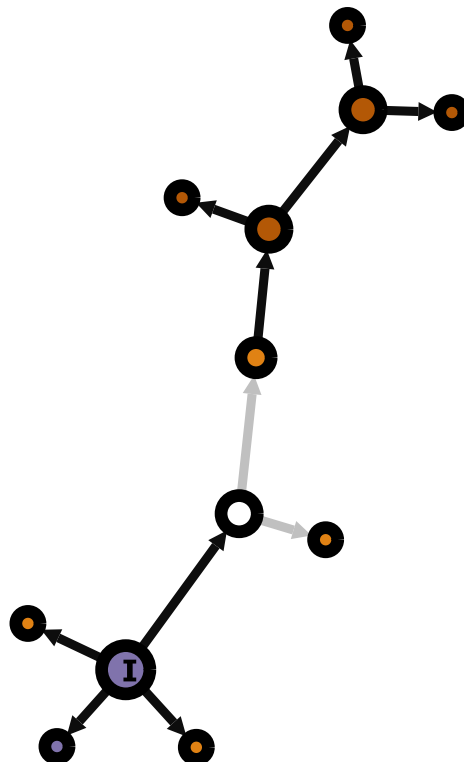
Supplementary Figure S6. Transmission networks

Transmission networks for all clusters identified in the study with more than 9 members

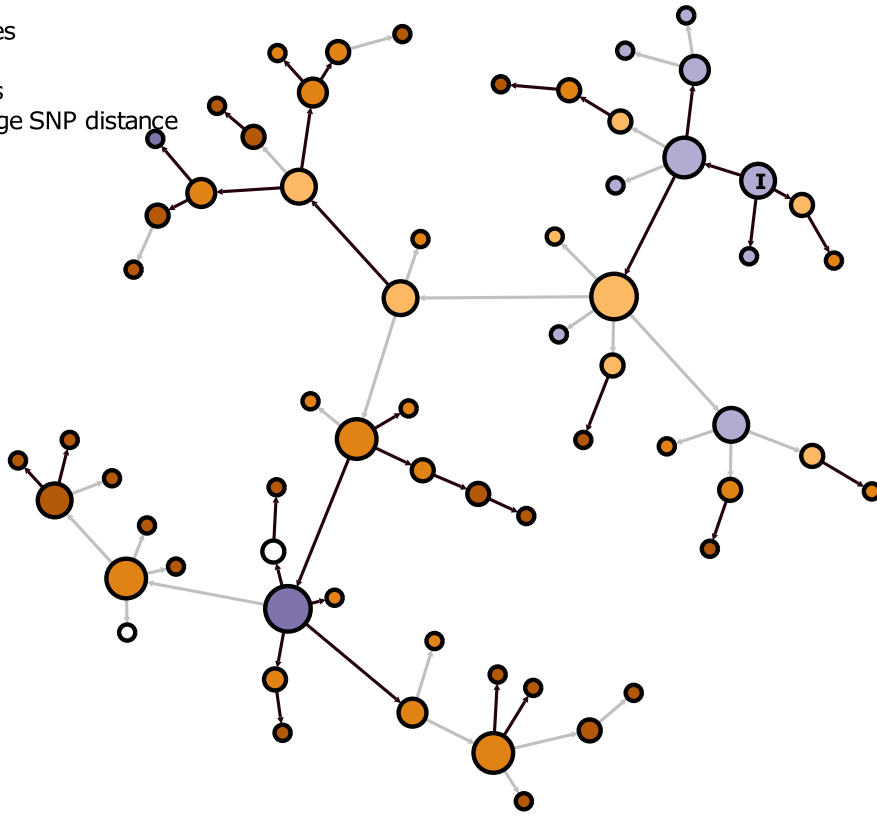
Cluster 4
183 genomes
177 nodes
42293 SNPs
2.54 average SNP distance
Lineage 2



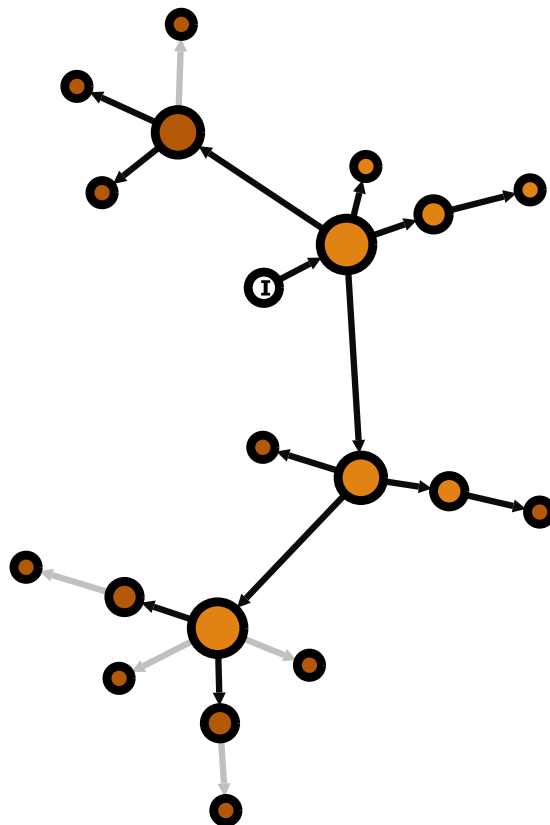
Cluster 9
13 genomes
12 nodes
73 SNPs
0.94 average SNP distance
Lineage 2



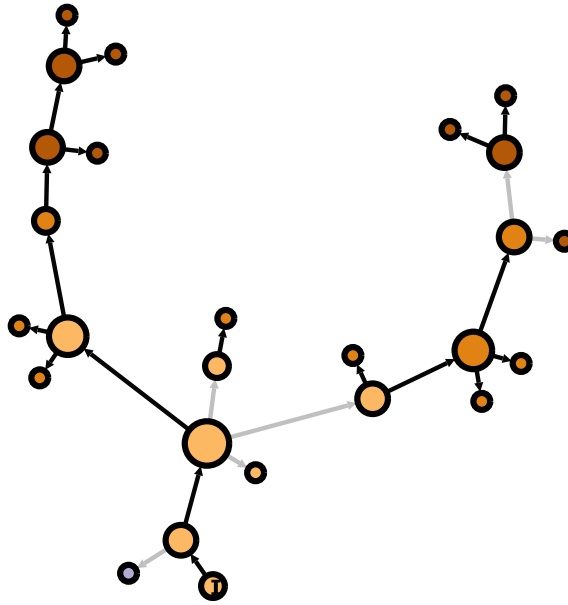
Cluster 10
65 genomes
64 nodes
7211 SNPs
3.47 average SNP distance
Lineage 2



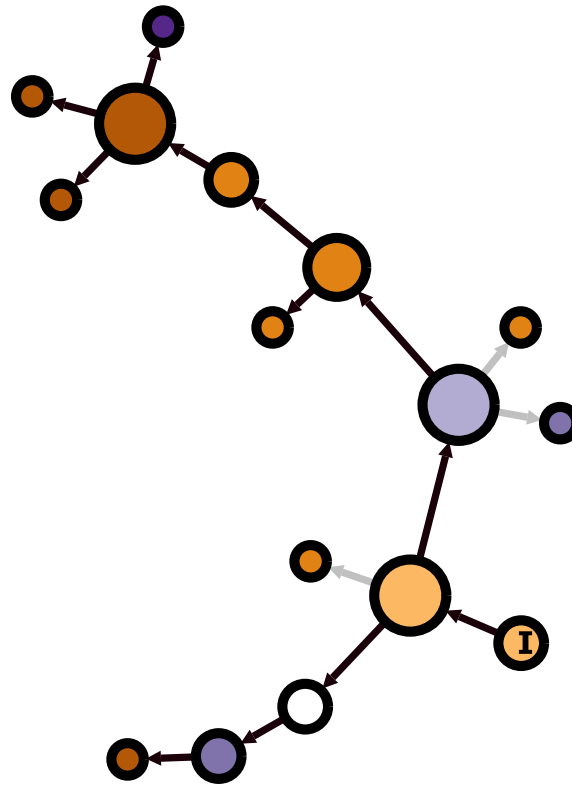
Cluster 15
20 genomes
20 nodes
623 SNPs
3.28 average SNP distance
Lineage 2



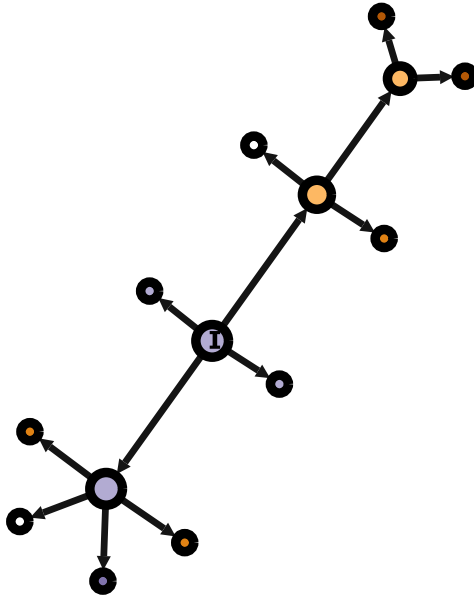
Cluster 16
26 genomes
26 nodes
867 SNPs
2.67 average SNP distance
Lineage 2



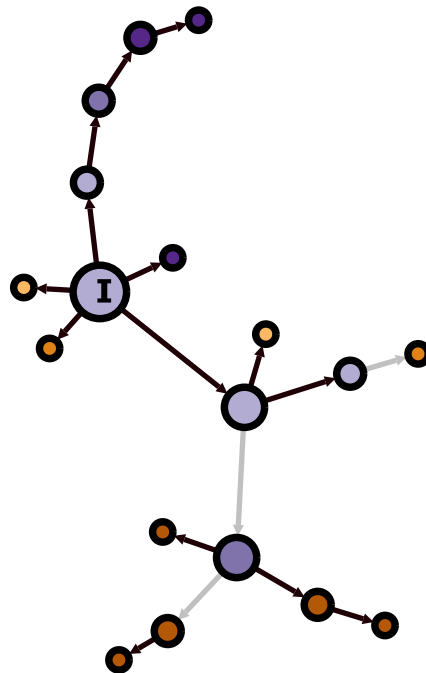
Cluster 18
16 genomes
16 nodes
410 SNPs
3.42 average SNP distance
Lineage 2



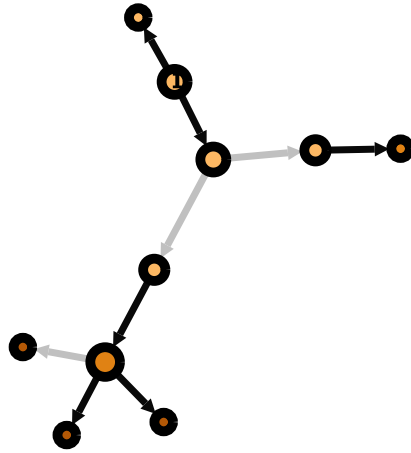
Cluster 19
15 genomes
14 nodes
292 SNPs
2.79 average SNP distance
Lineage 2



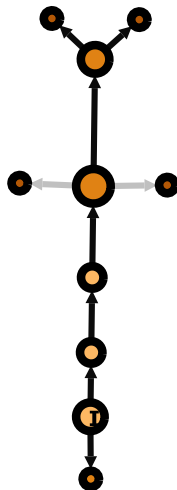
Cluster 24
18 genomes
18 nodes
418 SNPs
2.73 average SNP distance
Lineage 2



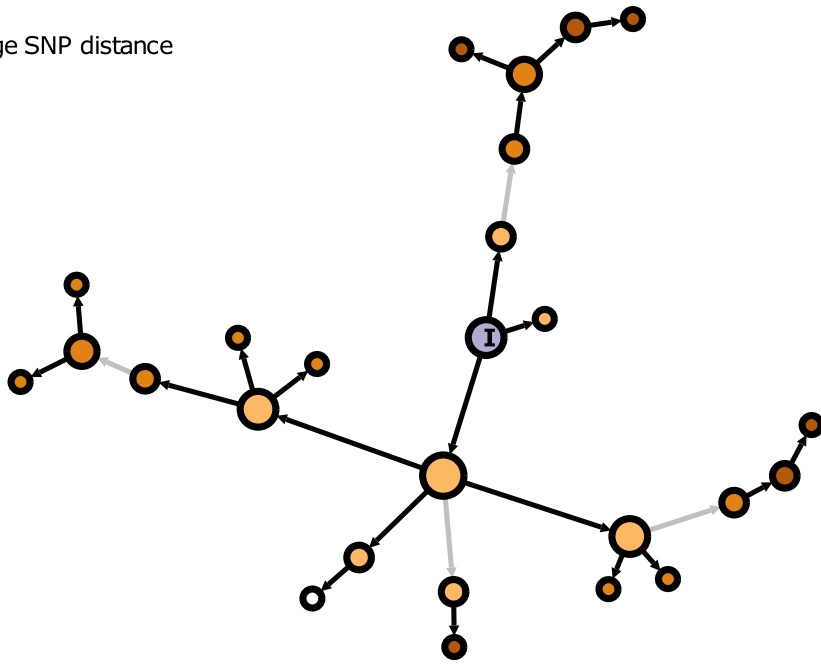
Cluster 84
11 genomes
10 nodes
166 SNPs
3.02 average SNP distance
Lineage 2

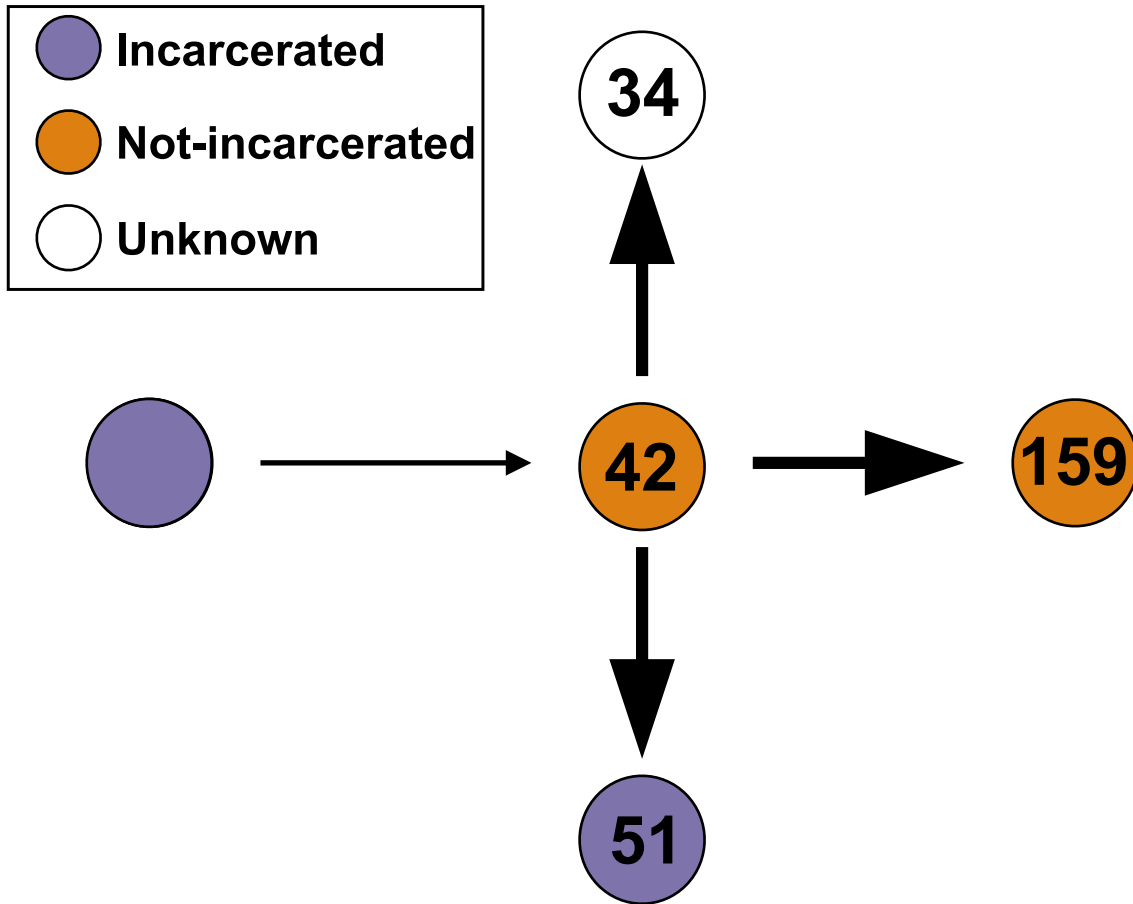


Cluster 87
10 genomes
10 nodes
122 SNPs
2.71 average SNP distance
Lineage 2

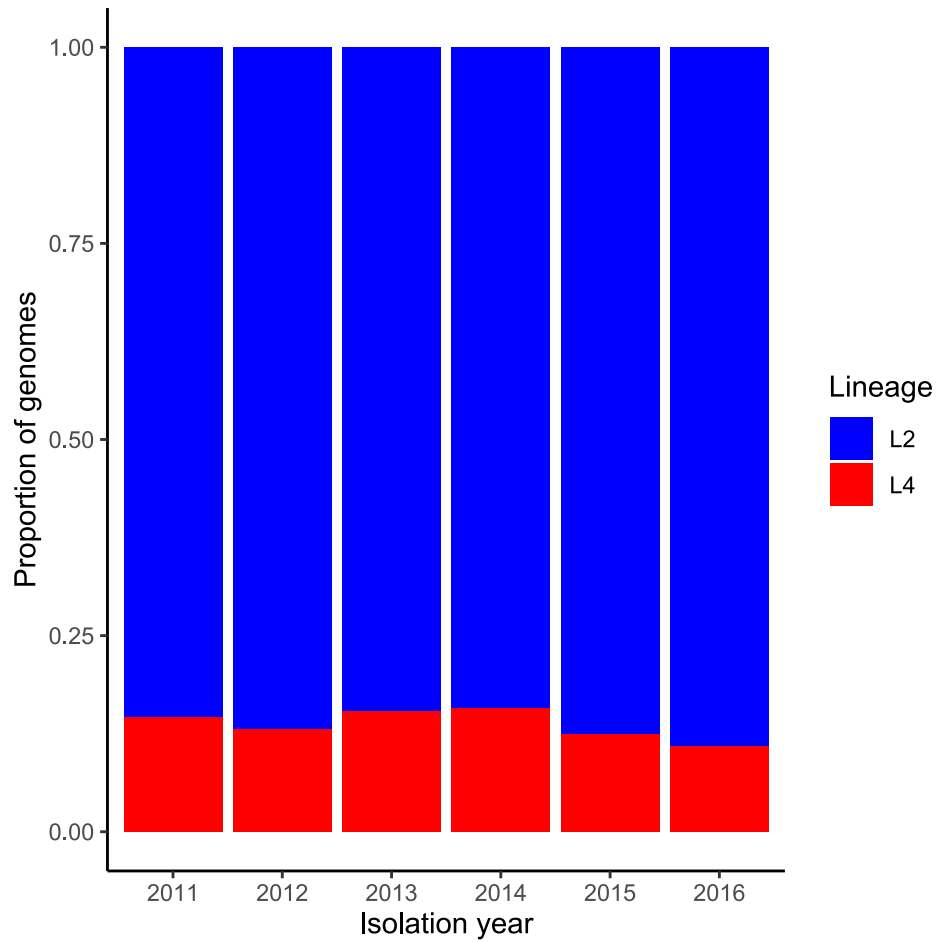


Cluster 124
28 genomes
26 nodes
869 SNPs
2.30 average SNP distance
Lineage 2





Supplementary Figure S7. Transmission network analysis Number within the nodes represent the number of transmission events (posterior probability > 0.5 for transmission events from incarcerated to non-incarcerated individuals). The open node represents transmission events for which no information on the incarceration status was available.



Supplementary Figure S8. Lineage proportions throughout the sampling timeframe. Lineage 4 proportions: 2011: 15 % ± 5, 2012: 13 % ± 4, 2013: 15 ± 5, 2014: 16 ± 4, 2015: 12 % ± 4, 2016: 11 ± 4; There was no statistically significant difference in lineage proportions per year ($p = 0.58$, two-sided Z-test).

Supplementary information inventory and table & figure captions

1. Supplementary results

a. Whole genome sequencing summary statistics

2. Supplementary Tables

- a. **Supplementary Table S1** Sampling coverages per year
- b. **Supplementary Table S2** Collated metadata including patient and *M. tuberculosis* isolate information.
- c. **Supplementary Table S3** List of drug resistance mutations compiled from various studies.
- d. **Supplementary Table S4** List of drug resistance mutations per strain analysed (SM = streptomycin, INH = isoniazid, EMB = ethambutol, KAN = kanamycin, CAP = capreomycin, AK = amikacin, RIF = rifampicin, FQ = fluoroquinolones, PZA = pyrazinamide, ETH = ethionamide).
- e. **Supplementary Table S5** List of putative compensatory mutations in the genes *rpoA*, *rpoB* and *rpoC*, identified in this study after applying filtering for phylogenetic markers.
- f. **Supplementary Table S6** Overview of compensatory mutations in the three genes encoding the RNA polymerase subunits by lineage. *Proportion of the total number of Lineage 2 (n =1386) and Lineage 4 (n = 277) *Mtb* strains harbouring compensatory mutations. Lineage 2 strains are more likely to harbor compensatory mutations, compared to Lineage 4 strains (p < 0.001, two-tailed χ^2 -test).
- g. **Supplementary Table S7** Infector-infectee relationships, support values, estimated infection times, effective sample size (ESS) inferred by phybreak using prior set A (Supplementary Table S8).
- h. **Supplementary Table S8** Priors for mutation rate, the gamma-distributed generation and sampling time used in the phybreak analysis. Mean generation time (μ g), Mean sampling time (μ s), Shape parameter generation time (κ g), Shape parameter sampling time (κ s), Rate parameter generation time (β g), Rate parameter sampling time (β s). Mutation rate given as mutations per genome per year. Text in bold indicates the priors used for the main results of the study.
- i. **Supplementary Table S9** Infector-infectee relationships, support values, estimated infection times, effective sample size (ESS) inferred by phybreak using prior set B (Supplementary Table S8).
- j. **Supplementary Table S10** Infector-infectee relationships, support values, estimated infection times, effective sample size (ESS) inferred by phybreak using prior set C (Supplementary Table S8).
- k. **Supplementary Table S11** Infector-infectee relationships, support values, estimated infection times, effective sample size (ESS) inferred by phybreak using prior set D (Supplementary Table S8)
- l. **Supplementary Table S12** Infector-infectee relationships, support values, estimated infection times, effective sample size (ESS) inferred by phybreak using prior set E (Supplementary Table S8).

- m. **Supplementary Table S13** Infector-infectee relationships, support values, estimated infection times, effective sample size (ESS) inferred by phybreak using prior set F (Supplementary Table S8).
- n. **Supplementary Table S14** Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set B (Supplementary Table S8).
- o. **Supplementary Table S15** Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set C (Supplementary Table S8).
- p. **Supplementary Table S16** Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set D (Supplementary Table S8).
- q. **Supplementary Table S17** Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set E (Supplementary Table S8).
- r. **Supplementary Table S18** Estimated association between bacterial and patient factors and the rate of secondary cases generated based on the phybreak output with prior set F (Supplementary Table S8).
- s. **Supplementary Table S19** Secondary case rates and secondary case rate ratios stratified by index case being an incarcerated or not-incarcerated individual.

3. Supplementary Figures

- a. **Supplementary Figure S1. Flow diagram of samples included in the study**
- b. **Supplementary Figure S2. Analyzed MDR TB strains vs. notified MDR TB strains** Total number of sequenced strains or used sequenced in relation to the WHO reported culture-confirmed MDR TB cases in Georgia during 2011-2016.
- c. **Supplementary Figure S3. Drug resistance profiles of the strains included in the study** **A.** Drug resistance profiles mapped to the phylogenetic tree of all samples included in the study. Lineage 2 strains are shaded in blue, Lineage 4 strains in red. The largest cluster (based on an average SNV distance of 5) is shaded in yellow. **B.** Number of strains displaying an MDR (isoniazid + rifampicin resistance), pre-XDR FQ (MDR + fluoroquinolone resistance), pre-XDR INJ (MDR + amikacin, kanamycin or capreomycin resistance) or an XDR (MDR + FQ + INJ) resistance profile.
- d. **Supplementary Figure S4. Proportion of compensated strains among incarcerated and non-incarcerated individuals** Proportions of strains carrying compensatory mutations stratified according to incarceration status (yes/no). Isolates from incarcerated individuals were more likely to carry a compensatory mutation compared to isolates from non-incarcerated individuals ($p = 5.844 \times 10^{-5}$, two-tailed χ^2 -test).
- e. **Supplementary Figure S5. Sizes of clusters containing only non-incarcerated individuals (n = 155) vs. clusters containing at least one**

incarcerated individual (n = 43) The median size of clusters containing at least one incarcerated individual (median = 4, 25th percentile = 2, 75th percentile = 8, interquartile range = 6) was larger compared to clusters containing only non-incarcerated individuals (median = 2, 25th percentile = 2, 75th percentile = 3, interquartile range = 1) ($p = 0.0001$, two-tailed Wilcoxon rank-sum test).

- f. **Supplementary Figure S6. Transmission networks** Transmission networks of all clusters identified in the study with more than 9 members.
- g. **Supplementary Figure S7. Transmission network analysis** Number within the nodes represent the number of transmission events (posterior probability > 0.5 for transmission events from incarcerated to non-incarcerated individuals). The open node represents transmission events for which no information on the incarceration status was available.
- h. **Supplementary Figure S8. Lineage proportions throughout the sampling timeframe.** Lineage 4 proportions: 2011: 15 % \pm 5, 2012: 13 % \pm 4, 2013: 15 \pm 5, 2014: 16 \pm 4, 2015: 12 % \pm 4, 2016: 11 \pm 4; There was no statistically significant difference in lineage proportions per year ($p = 0.58$, two-sided Z-test).