Coll, F; Phelan, J; Hill-Cawthorne, GA; Nair, MB; Mallard, K; Ali, S; Abdallah, AM; Alghamdi, S; Alsomali, M; Ahmed, AO; Portelli, S; Oppong, Y; Alves, A; Bessa, TB; Campino, S; Caws, M; Chatterjee, A; Crampin, AC; Dheda, K; Furnham, N; Glynn, JR; Grandjean, L; Minh Ha, D; Hasan, R; Hasan, Z; Hibberd, ML; Joloba, M; Jones-Lpez, EC; Matsumoto, T; Miranda, A; Moore, DJ; Mocillo, N; Panaiotov, S; Parkhill, J; Penha, C; Perdigo, J; Portugal, I; Rchiad, Z; Robledo, J; Sheen, P; Shesha, NT; Sirgel, FA; Sola, C; Oliveira Sousa, E; Streicher, EM; Helden, PV; Viveiros, M; Warren, RM; McNerney, R; Pain, A; Clark, TG (2018) Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nature genetics. ISSN 1061-4036 DOI: https://doi.org/10.1038/s41588-017-0029-0

Downloaded from: http://researchonline.lshtm.ac.uk/4646256/

DOI: 10.1038/s41588-017-0029-0

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/

Supplementary Table 1

Population	N	L1	L2	L3	L4	Susc.	DR	MDR-TB	XDR-TB	ENA Accession
Canada	11	0	0	0	11	11	0	0	0	SRA020129
Brazil	108	0	0	0	108	4	9	85	10	**PRJEB10385**
Colombia	15	0	0	0	15	0	0	15	0	**PRJEB10385**
Peru	78	0	6	0	72	26	31	18	3	**PRJEB10385**
Bulgaria	2	0	0	0	2	0	0	2	0	**PRJEB10385**
Germany	20	0	0	0	20	20	0	0	0	ERP006619
Portugal	183	0	20	1	162	19	71	67	26	**ERP002611**
Russia	2	0	2	0	0	1	1	0	0	ERP000192
China	161	0	122	2	37	44	0	94	23	SRP018402
Vietnam	43	16	19	0	8	22	6	15	0	**PRJEB10385**
India	3	0	0	2	1	1	0	2	0	**PRJEB10385**
Pakistan	42	5	0	33	4	5	0	0	37	**ERP008770**
Saudi Arabia	74	10	11	18	35	57	6	11	0	***
Malawi	1646	264	71	195	1116	1526	112	8	0	**ERP000436**
South Africa	594	8	231	15	340	81	131	208	174	****
Uganda	45	1	1	13	30	3	2	40	0	**ERP000520**
WHO*	138	14	34	4	86	35	51	52	0	*****
Mixed**	96	4	38	4	50	96	0	0	0	ERP001037
UK	3204	295	466	706	1737	2500	343	356	5	******
Total	**6465**	**617**	**1021**	**993**	**3834**	**4451**	**763**	**973**	**278**	
%	**100**	**9.5**	**15.8**	**15.4**	**59.3**	**68.8**	**11.8**	**15.1**	**4.3**	

Populations contributing to the analysis

Abbreviations: L1, lineage 1; L2, lineage 2; L3, lineage 3; L4, lineage 4; Susc., susceptible; DR, resistant to at least one drug but not MDR-TB/XDR-TB. Notes: *Bangladesh (8), China (1), Nepal (4), Pakistan (1), Philippines (4), South Korea (39), Thailand (1), Cameroon (1), Central African Republic (1), Equatorial Guinea (1), Guinea (1), Morocco (4), Niger (1), Nigeria (1), Democratic Republic of Congo (4), Rwanda (15), Gemany (12), Kazakstan (1), Portugal (1), Spain (2), Brazil (7), Colombia (1), Domican Republic (1), Peru (31); ** Malaysia, South Africa, and Thailand (96); *** PRJEB10950, PRJEB10385; **** ERP013054, PRJEB10950; ***** PRJNA183624, PRJNA235615, PRJEB10385; ****** PRJEB2221, PRJEB5162, PRJEB6273, PRJEB6276, PRJEB7281, PRJEB7727, PRJEB9680, PRJNA282721; **bolded ENA accession numbers** include sequencing performed as part of the TB Global Drug Resistance Collaboration (http://pathogenseq.lshtm.ac.uk/#tuberculosis).

Supplementary Table 2

Drug	Lineage 1	Lineage 2	Lineage 3	Lineage 4	Total Resistant	(%)
Rifampicin (RIF)	31/616	580/1002	81/992	649/3748	1341	(21.1)
Isoniazid (INH)	93/615	602/1012	161/992	892/3765	1748	(27.4)
Ethambutol (EMB)	17/405	366/914	37/844	317/2862	737	(14.7)
Pyrazinamide (PZA)	20/394	270/666	41/798	222/2360	553	(13.1)
Streptomycin (STR)	28/234	449/758	47/300	491/2170	1015	(29.3)
Capreomycin (CAP)	1/19	131/378	10/67	128/687	270	(23.5)
Amikacin (AMK)	5/19	129/257	28/72	66/553	228	(25.3)
Kanamycin (KAN)	6/27	139/397	28/72	156/731	329	(26.8)
Moxifloxacin (MOX)	0/18	67/235	2/40	22/357	91	(14.0)
Ofloxacin (OFL)	3/23	167/465	5/65	228/849	403	(28.7)
Ethionamide (ETH)	2/7	108/301	3/36	171/386	284	(38.9)
Ciprofloxacin (CIP)	5/48	2/24	32/108	23/176	62	(17.4)
PAS	0/0	7/119	0/0	13/136	20	(7.8)
D-Cycloserine (CYS)	0/0	39/117	0/0	17/131	56	(22.6)

Phenotype	Lineage 1 N (%)	Lineage 2 N (%)	Lineage 3 N (%)	Lineage 4 N (%)	Total	(%)
Susceptible	510 (82.7)	368 (36.0)	815 (82.1)	2758 (71.9)	4451	(68.8)
DR	78 (12.6)	103 (10.1)	103 (10.4)	479 (12.5)	763	(11.8)
MDR-TB	23 (3.7)	430 (42.1)	47 (4.7)	473 (12.3)	973	(15.1)
XDR-TB	6 (1.0)	120 (11.8)	28 (2.8)	124 (3.2)	278	(4.3)
Total	617	1021	993	3834	6465	
(%)	(9.5)	(15.8)	(15.4)	(59.3)	(100)	

Drugs susceptibility test data (resistant/tested) and the phenotypes considered

Abbreviations: DR, resistant to at least 1 drug but not MDR-TB/XDR-TB; MDR-TB, multidrug-resistant tuberculosis; XDR-TB, extensive drug-resistant tuberculosis; PAS Para-aminosalicylic acid. Notes: MOX and OFL are fluoroquinolones (FLQ); CAP, KAN and AMK are second-line injectables drugs.

Supplementary Table 3

Gene / Intergenic region	Mutation	Susceptible %	DR %	MDR-TB allele frequency %	XDR-TB allele frequency %
katG	S315T	0.4	36.9	71.7	73.4
rpoB	S450L	0.1	12.1	64.2	56.1
rpsL	K43R	0.6	11.5	42.2	23.5
embB	M306V	0.0	6.2	31.2	39.4
embB	M306I	0.2	6.2	17.4	35.2
Rv1482c-fabG1	C-15T	0.3	21.9	16.9	36.3
rrs	A1401G	0.0	4.4	10.1	72.3
rrs	A514C	0.2	2.4	5.3	27.4
gyrA	A90V	0.0	0.5	4.1	32.1
gyrA	D94G	0.3	4.0	4.5	26.5
gid	L79S	0.0	0.7	2.1	27.5
rpoB	L452P	0.0	1.6	3.2	18.7
ethA-ethR	T-65C	0.0	0.4	4.4	17.8
Rv1482c-fabG1	T-8A	0.0	0.8	3.5	17.3
rpoB	D435V	0.0	0.4	6.2	14.9
rpoB	D435G	0.0	0.1	0.2	18.5
ubiA	V188A	0.0	0.0	0.2	17.3
rpoB	I1106T	0.0	0.0	0.1	17.3
inhA	S94A	0.0	5.7	4.3	8.6
Rv1482c-fabG1	G-17T	0.0	1.1	0.9	11.2
inhA	I194T	0.0	2.8	3.1	9.0
ubiA	A249T	0.0	0.1	0.9	10.8
PPE52-nuoA	G-314T	0.0	0.1	1.0	11.2
eis-Rv2417c	C-10T	0.0	2.2	3.4	6.9
rpsL	K88R	0.1	3.7	5.3	1.4
iniA	H42R	0.0	0.1	0.6	10.8
gyrA	D94A	0.0	0.9	2.2	8.5
alr	L113R	0.0	0.3	0.9	9.9
pncA	Q10*	0.0	0.0	0.3	0.0
embB	Q497R	0.0	1.4	5.3	1.8
rpoB	D435Y	0.0	1.1	2.7	5.5
rpoB	H445Y	0.0	1.8	4.3	1.4
gyrA	S91P	0.1	1.4	1.3	6.1
embC-embA	C-12T	0.0	0.1	3.0	2.5
embB	G406A	0.0	0.5	3.1	2.5
pncA	Q10P	0.0	0.0	4.8	1.1
eis-Rv2417c	G-12A	0.0	0.0	4.7	0.7
embC-embA	C-16T	0.0	1.6	2.0	2.9
embC-embA	C-16G	0.0	0.4	1.5	4.0
rpoB	H445D	0.0	1.3	3.9	0.4
gyrA	D94Y	0.0	0.3	1.0	4.4
thyX-hsdS.1	G-16A	0.0	0.9	2.3	2.2
rpoB	L731P	0.0	2.1	1.3	2.5
embB	G406D	0.0	1.2	2.4	2.2
pncA	V125G	0.0	2.1	1.4	2.6
embC-embA	C-11A	0.0	1.2	0.9	4.0
katG	S315R	0.0	1.1	1.7	2.2

pncA-Rv2044c	T-11C	0.0	0.1	2.7	2.3
katG	S315N	0.0	1.3	1.8	2.2
gyrA	D94N	0.1	0.1	0.5	5.5
embB	M423T	0.0	2.1	1.0	2.5
gid	A80P	0.0	2.1	1.0	3.1
embC-embA	G-43C	0.0	0.3	1.7	1.4
embB	D354A	0.0	0.4	2.3	0.7
Rv2172c-idsA2	A-65G	0.0	1.2	0.8	3.3
embC-embA	C-12A	0.0	1.2	0.8	3.2
embB	P397T	0.0	1.2	0.8	3.2
rrs	C517T	0.0	0.7	1.9	0.7
eis-Rv2417c	G-14A	0.0	0.1	1.1	2.2
embB	G406S	0.0	0.7	1.9	0.4
rpoB	H445R	0.1	0.1	1.9	0.4
embB	D1024N	0.0	0.4	1.0	2.9
oxyR'-ahpC	G-48A	0.0	0.5	1.0	1.8
alr	M343T	0.0	0.9	0.5	2.5
rpoB	S450W	0.0	0.7	1.8	0.0
oxyR'-ahpC	C-52T	0.1	0.3	1.7	0.4
rpoB	H445L	0.0	0.4	1.2	1.1
pncA	V139M	0.0	0.3	0.4	1.9
rpoB	H445N	0.1	2.5	0.5	0.0
rpoB	L430P	0.1	1.1	1.2	0.0
embC-embA	C-8T	0.0	0.1	0.4	1.8
rpoB	I491F	0.2	2.5	0.8	0.0
pncA	W68*	0.0	0.0	0.1	0.0
Rv1482c-fabG1	T-8C	0.4	0.4	0.9	0.4
pncA	Q141P	0.0	0.4	1.3	0.4
gyrA	D94H	0.0	0.0	0.4	2.2
rrs	A514T	0.0	1.9	0.2	0.4
rpoB	M434I	0.0	0.0	0.1	0.0

Allele frequency of resistance mutations

Abbreviations: DR, Resistant to at least 1 drug but not MDR-TB/XDR-TB; MDR-TB, multidrug-resistant tuberculosis; XDR-TB, extensively drug-resistant tuberculosis; IGR, intergenic region;

* stop codon

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
2,4	*rpoB*	761161	M or X v SUS	9.67E-18	0	0.016	0.032	0.187
4	*rpoB*	763123	X v M or SUS	1.13E-17	0	0	0.001	0.173
3	*rpoC*	764666	X or M v SUS, RMP	3.74E-29	0	0	0.006	0.011
2	*rpoC*	764819	M v SUS	3.33E-18	0	0.001	0.011	0
4	*rpoC*	766823	X v SUS	1.64E-06	0	0.012	0.008	0.032
1	*rpoC*	767123	MDR or XDR v SUS, RMP	3.91E-24	0	0	0.013	0.011
2-4	*rpsL*	781687	XDR v SUS, STM	1.65E-45	0.006	0.115	0.422	0.235
2-4	*rpsL*	781822	STM	4.16E-10	0.002	0.038	0.060	0.014
1	*rrs*	1472358	STM	5.12E-06	0	0.008	0.005	0
4	*rrs*	1472359	STM	2.66E-13	0.002	0.043	0.056	0.278
1,3	*rrs*	1472359	M or X v SUS	5.71E-18	0.002	0.043	0.056	0.278
1	*rrs*	1472362	M or X v SUS, STM	3.52E-71	0	0.007	0.019	0.007
3	*rrs*	1472751	X v SUS	2.28E-10	0	0.003	0.004	0.011
2,4	*rrs*	1473246	AMK, CAP, KAN	6.68E-42	0	0.044	0.101	0.277
3	*rrs*	1473246	STM	3.05E-09	0	0.044	0.101	0.277
2-4	*rrs*	1473246	X v SUS or M	7.73E-246	0	0.044	0.101	0.277
1	*pncB1*	1499617	PZA	3.20E-06	0	0	0.005	0.004
1	*echA12*	1660232	X v M or SUS	3.91E-24	0	0	0.005	0.004
2	*Rv1482c-fabG1*	1673425	ETH	1.91E-04	0.003	0.219	0.169	0.363
1,3,4	*Rv1482c-fabG1*	1673425	M or X v SUS, INH	4.07E-56	0.003	0.219	0.169	0.363
4	*Rv1482c-fabG1*	1673432	X v SUS	2.63E-15	0.004	0.012	0.046	0.183
4	*inhA*	1674481	X v SUS	8.54E-46	0	0.057	0.043	0.086
1,4	*inhA*	1674782	M or X v SUS	3.91E-24	0	0.028	0.031	0.09
1-4	*katG*	2155168	M or X v SUS, INH	3.26E-286	0.004	0.382	0.283	0.266
4	*pncA*	2288868	X v SUS, PZA	1.60E-14	0	0.022	0.016	0.026
1	*pncA*	2288952	M or X v SUS, PZA	3.91E-24	0	0.003	0.01	0.004

2	*pncA-Rv2044c*	2289252	PZA	1.12E-08	0	0.001	0.029	0.027
4	*eis-Rv2417c*	2715342	KAN	1.93E-08	0	0.022	0.037	0.069
2	*oxyR'-ahpC*	2726141	X v SUS	4.53E-08	0.001	0.004	0.024	0.004
2	*alr*	3841083	Cycloserine	1.67E-08	0	0.003	0.009	0.099
2,4	*embC-embA*	4243217	X or M v SUS, EMB	2.09E-14	0	0.021	0.043	0.068
3,4	*embC-embA*	4243221	X v SUS, EMB	1.70E-32	0	0.013	0.038	0.058
4	*embC-embA*	4243222	X v SUS, EMB	2.62E-10	0	0.012	0.011	0.040
1-4	*embB*	4247429	M or X v SUS, EMB	1.28E-47	0.001	0.064	0.317	0.401
1-4	*embB*	4247431	M or X v SUS, EMB	1.58E-51	0.002	0.062	0.174	0.363
1	*embB*	4247574	X or M v SUS, EMB	3.63E-07	0	0.004	0.023	0.007
4	*embB*	4247702	X v SUS, EMB	2.62E-08	0	0.012	0.008	0.032
4	*embB*	4247729	X or M v SUS, EMB	8.81E-13	0	0.009	0.02	0.004
2,4	*embB*	4247730	X v SUS, EMB	3.31E-12	0	0.017	0.055	0.047
4	*embB*	4247781	X v SUS	9.20E-10	0	0.021	0.010	0.025
1,3,4	*embB*	4248003	M or X v SUS, EMB	1.33E-26	0	0.017	0.057	0.025
3	*embB*	4249583	X v SUS, EMB	5.84E-23	0	0.004	0.010	0.029
4	*ubiA*	4269271	X v M or SUS	1.01E-16	0	0	0.002	0.173
3	*ethA*	4326435	X v SUS	3.44E-14	0	0	0	0.018
4	*ethA-ethR*	4327484	X v SUS	9.24E-44	0	0.004	0.044	0.178
4	*ethR*	4328127	X v SUS	9.30E-10	0	0.021	0.010	0.025
4	*gid*	4407965	X v SUS	6.27E-10	0	0.021	0.010	0.031

Lineage specific SNP associations

Abbreviations: X, XDR-TB; M, MDR-TB; Susc., Pan susceptible; DR, resistant to at least one drug but not MDR-TB/XDR-TB; RIF, rifampicin; INH, isoniazid; ETH, ethionamide; EMB, ethambutol; KAN, kanamycin.

Supplementary Table 5

Drug	Resistance gene	Co-occurring gene	Fisher exact test p-value
Rifampicin	*rpoB*	*rpoC**	< 2.2e-16
Rifampicin	*rpoB*	*rpoA**	6.0e-09
Isoniazid	*katG*	*ahpC**	< 2.2e-16
Pyrazinamide	*pncA*	*pncB2*	1.4e-13
Ethambutol	*embB*	*ubiA*	< 2.2e-16
PAS	*thyA*	*thyX-hsdS.1*	< 2.2e-16

Detected co-occurrence of mutations at drug resistance associated loci (Fisher exact test $P<10^{-8}$)

Abbreviations: PAS, para-aminosalicylic acid. Note: underlying overall and lineage data are presented in Supplementary Table 6; * known compensatory mechanisms

Supplementary Table 6

		rpoB (81-bp rifampicin resistance-determining region)									
		Overall		Lineage 1		Lineage 2		Lineage 3		Lineage 4	
		WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.
rpoC	WT	4964	873	557	30	417	322	888	49	3102	472
	Mut.	138	477	15	15	25	251	28	28	70	183
rpoA	WT	5060	1308	564	45	439	553	915	76	3142	634
	Mut.	43	42	8	0	3	20	1	1	31	21

		katG									
		Overall		Lineage 1		Lineage 2		Lineage 3		Lineage 4	
		WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.
ahpC promoter	WT	4959	1390	554	58	472	525	826	156	3107	651
	Mut.	35	62	4	0	5	16	5	5	21	41

		pncA									
		Overall		Lineage 1		Lineage 2		Lineage 3		Lineage 4	
		WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.
pncB2	WT	5608	599	560	23	704	280	953	18	3391	278
	Mut.	116	59	24	0	9	0	13	0	70	59
pncB1	WT	5576	647	528	15	701	280	927	17	3420	335
	Mut.	147	11	58	8	12	0	37	1	40	2

		ethA									
		Overall		Lineage 1		Lineage 2		Lineage 3		Lineage 4	
		WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.
pyrG	WT	5922	285	541	33	914	59	933	43	3534	150
	Mut.	143	15	38	1	7	2	14	0	84	12
Rv0565c	WT	5969	292	562	34	914	56	915	42	3578	160
	Mut.	90	8	17	0	6	5	30	1	37	2

		embB									
		Overall		Lineage 1		Lineage 2		Lineage 3		Lineage 4	
		WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.
ubiA	WT	5033	1281	502	91	489	475	886	97	3156	618
	Mut.	45	104	21	3	3	54	9	1	12	46

		thyA									
		Overall		Lineage 1		Lineage 2		Lineage 3		Lineage 4	
		WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.	WT	Mut.
thyX-hsdS1	WT	6332	36	600	4	982	10	973	14	3777	8
	Mut.	67	21	13	0	14	14	6	0	34	7

Co-occurrence of mutations at drug resistance associated loci with a breakdown by lineage

Each table contains the number of isolates with and without mutations ('mutant' (Mut) & 'wild type' (WT) respectively) at each pair of drug resistance associated loci effects identified or known compensatory effects. 'Mutant' refers to isolates with SNP and indel non-synonymous amino acid changes. Synonymous amino acid changes and deep phylogenetic mutations were discarded. Cells with grey background show statistically significant correlations (Fisher exact test $P<0.02$), i.e. pairs of genes frequently mutated in the same isolates, whereas white background indicates lack of statistical significance. This analysis points to putative epistatic and compensatory relationships.

Supplementary Table 7

Genomic position	Mutation	Overall Mutation Frequency	Resist. Freq.	mCSM *	DUET *	mCSM-Lig **	Distance from CYS **	mCSM -PPI ***
3840259	Y388D	0.0009	0	-3.369	-3.384	-3.737	2.682	-2.819
3840258	Y388C	0.0002	0	-1.889	-1.704	-1.938	2.682	-2.489
3840393	M343T_B	0.0031	0.0358	-2.118	-2.085	0.368	3.636	-0.195
3840708	S238L	0.0002	0	0.611	1.192	0.69	4.246	-0.551
3840952	K157E	0.0003	0	-1.483	-1.455	-1.841	4.474	-0.075
3840636	P262Q	0.0012	0	-2.015	-2.069	0.279	4.987	-0.863
3840717	S235W	0.0002	0	-0.807	-1.460	0.706	5.212	-0.588
3840402	R340L_B	0.0003	0	-0.57	0.616	0.160	5.389	-0.629
3840643	L260V	0.0002	0	-1.244	-1.554	-2.467	6.992	-0.419
3840639	S261N	0.0002	0	-1.443	-1.606	-0.482	7.116	-0.248
3841083	**L113R**	**0.0057**	**0.4461**	**-0.961**	**-0.956**	**-1.721**	**8.477**	**-0.423**

Protein structural modelling of *alr* reveals low frequency mutations conferring higher instability

We applied four measures to quantify the enthalpic effects (the change in Gibbs free energy - ΔΔG) of point mutations on overall protein structure stability (mCSM and DUET), protein-protein interactions (mCSM-PPI) and interaction with substrate/drug (mCSM-Lig). Negative values indicate a destabilising effect, with the most destabilising highlighted in grey, and positive values indicating an increase in stability. The geometrical distance from the mutation to the drug binding position is also provided. The mutation that was statistically significant with the largest resistance frequency (L113R) has a relatively large destabilising effect both on the overall protein structure and in drug binding, yet it is the furthest from the site of drug interaction. Abbreviations: CYS, D-cycloserine. Notes: * protein stability; ** drug binding, *** protein-protein interactions; bolded the mutation that was statistically significant; grey – less stability.

Supplementary Table 8

Drug	Gene	indels/Kb	Total No. positions	Length Median (bp)	Length Range (bp)	Assoc. P-value
MDR-TB vs. Susc.	*embCAB promoter*	72.29	6	1	1-2	8.34E-09
MDR-TB vs. Susc.	*pncA*	44.72	25	1	1-15	2.09E-08
MDR-TB vs. Susc.	*rpoB*	2.27	7	6	3-9	3.91E-03
XDR-TB vs. Susc.	*ethA*	25.89	38	1	1-10	9.76E-79
XDR-TB vs. Susc.	*pncA*	44.72	25	1	1-15	1.20E-23
XDR-TB vs. Susc.	*rpoB*	2.27	7	6	3-9	4.63E-14
XDR-TB vs. Susc.	*embCAB promoter*	72.29	6	1	1-2	6.93E-07
XDR- vs. MDR-TB	*pncA*	44.72	25	1	1-15	3.80E-06
XDR- vs. MDR-TB	*ald*	10.77	12	1	1-5	4.85E-04
XDR- vs. MDR-TB	*rrs*	2.61	4	1	1-1	2.71E-03
Isoniazid	*katG*	5.40	12	1.5	1-12	2.82E-05
Rifampicin	*rpoB*	2.27	7	6	3-9	1.25E-10
Ethionamide	*ethA*	25.89	38	1	1-10	7.22E-09
Capreomycin	*tlyA*	3.73	3	2	2-10	1.21E-12
Capreomycin	*rrs*	2.61	4	1	1-1	2.37E-10
Streptomycin	*gid*	35.66	24	1	1-14	1.45E-09
Pyrazinamide	*pncA*	44.72	25	1	1-15	5.27E-38
Cycloserine	*ald*	10.77	12	1	1-5	5.35E-03
Kanamycin	*rrs*	2.61	4	1	1-1	9.29E-05

Gene-based small insertion and deletion (indel) associations

Abbreviations: Susc., susceptible; MDR-TB, multidrug-resistant tuberculosis; XDR-TB, extensively drug-resistant tuberculosis.

Supplementary Table 9

Gene	No. samples	Drug	No. DR	No. XDR-TB	Mean size (bp)	Size range (bp)
dfrA/thyA	5	PAS	1	3	6,396	2,825-7,912
pncA	12	PZA	1	3	1,402	446-4,670
ethA/ethR	7	ETH	3	3	3,667	1,513-5,271
katG	3	INH	3	0	5,729	4,789-7,608

Large deletions in candidate drug resistance regions

Abbreviations: DR, resistant to at least one drug but not MDR/XDR-TB; XDR-TB, extensively drug-resistance tuberculosis; PAS, para-aminosalicylic acid; ETH, ethionamide; PZA, pyrazinamide; INH, isoniazid.