Deciphering within-host microevolution of *Mycobacterium tuberculosis* through whole genome sequencing: the phenotypic impact and way forward

Ley SD[a], de Vos M[a], Van Rie A[b*], Warren RM[a*#]

[a] DST-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town.

[b]Department of Epidemiology and Social Medicine, Faculty of Medicine and Health Sciences, University of Antwerp, Belgium

corresponding author: rw1@sun.ac.za

* shared senior last authors

SUPPLEMENTAL MATERIAL

Dataset S1: Patient and isolate data extracted from the included publications

see separate file

Dataset S2: Comparison of genes found to micro-evolve within the host

see separate file

Table S1: Search strategy and electronic databases screened

07.08.2018	**PubMed**	**No. of hits**
search order/strategy		
(DP = 2010 - present) AND	tuberculosis[MeSH terms] OR tuberculosis OR TB OR mycobacterium tuberculosis[MeSH terms] OR mycobacterium tuberculosis OR Mtb OR M. tuberculosis	69044
1 AND	((full genome OR complete genome OR entire genome OR next generation OR whole genome OR deep OR genome) AND (sequenc*)) OR NGS OR WGS	2203
1 AND 2 AND	within-host OR within host OR within patient OR within-patient OR intra patient OR intra-patient OR in patient OR single patient OR during treatment	1043
NOT PT= review AND 1 AND 2 AND 3 AND	serial* OR consecutive OR multiple OR several OR follow up OR emerg* OR acquired OR acquisition	314

07.08.2018	**Web of Science Core Collection [v 5.29]**	**No. of hits**
search order/strategy	For all searches only hits published between 2010 - 2018 included	
1	tuberculosis[MeSH terms] OR tuberculosis OR TB OR mycobacterium tuberculosis[MeSH terms] OR mycobacterium tuberculosis OR Mtb OR M. tuberculosis	84340
within 1	((full genome OR complete genome OR entire genome OR next generation OR whole genome OR deep OR genome) AND (sequenc*)) OR NGS OR WGS	2056
within 1 AND 2	within-host OR within host OR within patient OR within-patient OR intra patient OR intra-patient OR in patient OR single patient OR during treatment	437
within 1 AND 2 AND 3	serial* OR consecutive OR multiple OR several OR follow up OR emerg* OR acquired OR acquisition	210

07.08.2018	Scopus	No. of hits
search order/strategy	For all searches only hits published between 2010 - 2018 included	
1	tuberculosis OR TB OR mycobacterium tuberculosis OR Mtb OR M. tuberculosis	198833
within 1	((full genome OR complete genome OR entire genome OR next generation OR whole genome OR deep OR genome) AND (sequenc*)) OR NGS OR WGS	10267
within 1 AND 2	within-host OR within host OR within patient OR within-patient OR intra patient OR intra-patient OR in patient OR single patient OR during treatment	128
within 1 AND 2 AND 3	serial* OR consecutive OR multiple OR several OR follow up OR emerg* OR acquired OR acquisition	89
within 1 AND 2 AND 3 AND 4	EXCLUDE publication type = review	63

07.08.2018	MEDLINE	No. of hits
search order/strategy	For all searches only hits published between 2010-2018 (DP) included and publication type = review excluded	
1	tuberculosis[MeSH terms] OR tuberculosis OR TB OR mycobacterium tuberculosis[MeSH terms] OR mycobacterium tuberculosis OR Mtb OR M. tuberculosis	78711
1 AND	((full genome OR complete genome OR entire genome OR next generation OR whole genome OR deep OR genome) AND (sequenc*)) OR NGS OR WGS	1873
1 AND 2 AND	within-host OR within host OR within patient OR within-patient OR intra patient OR intra-patient OR in patient OR single patient OR during treatment	231
1 AND 2 AND 3 AND	serial* OR consecutive OR multiple OR several OR follow up OR emerg* OR acquired OR acquisition	117

07.08.2018	CINAHL	No. of hits
search order/strategy	For all searches only hits published between 2010 - 2018 (DT) included and publication type = review excluded	
1	tuberculosis[MeSH terms] OR tuberculosis OR TB OR mycobacterium tuberculosis[MeSH terms] OR mycobacterium tuberculosis OR Mtb OR M. tuberculosis	5271
1 AND	((full genome OR complete genome OR entire genome OR next generation OR whole genome OR deep OR genome) AND (sequenc*)) OR NGS OR WGS	24
1 AND 2 AND	within-host OR within host OR within patient OR within-patient OR intra patient OR intra-patient OR in patient OR single patient OR during treatment	9
1 AND 2 AND 3 AND	serial* OR consecutive OR multiple OR several OR follow up OR emerg* OR acquired OR acquisition	6

Table S2: Inclusion and exclusion criteria for the comparison of micro-evolved loci between studies

	Author & Reference	**Included in gene analysis** (yes/no)	**Data set used**	**Reason for exclusion**
1	Saunders *et al* (1)	yes	SNVs within article	
2	Comas *et al* (2)	yes	Suppl. Table 2	SNVs from *in vitro* single colonies excluded (see inclusion criteria)
3	Walker *et al* (3)	no		SNVs between serial isolates not described
4	Sun *et al* (4)	yes	Table S2	
5	Farhat *et al* (5)	yes	Suppl. Table 17; IDs 6 - 13	
6	Bryant *et al* (6)	yes	Suppl. Table 1	
7	Clark *et al* (7)	no		SNVs could not be assigned to specific isolates, so unclear which ones are from serials and which ones from non-serial isolates
8	Merker *et al* (8)	yes	Yellow highlighted genes in Table S1 (Patient A) and Table S3 (Patient B)	Table S5 (Patient C) excluded as exogenous re-infection, not within-patient microevolution
9	Pérez-Lago *et al* (9)	no		Only nucleotide change indicated, but no gene name or loci (Table 2). Within main text only 2 genes mentioned.
10	Eldholm *et al* (10)	yes	Suppl. Table 13059_2014_490_MOESM1_ESM.xlsx	
11	Guerra-Assunçao *et al* (11)	no		Suppl. Tables indicate relapse associated SNVs, but these are not necessarily all SNVs detected between pairs. Data not assignable to specific isolate (i.e. gain/loss, present in both isolates of a patient?)
12	Witney *et al* (12)	yes	Table A1; only data of Case 2	Complete set of SNVs between paired isolates only given for Case 2; no heterozygous SNVs though…
13	Guerra-Assunçao *et al* (13)	no		Same data than Guerra-Assunçao *et al* (11)
14	Black *et al* (14)	yes	Table 5	SNV between *in vitro* grown single colonies excluded (see inclusion criteria)
15	O'Neill *et al* (15)	no		Meta-analysis of data sets of Merker *et al,* Sun *et a*l and Eldholm *et al* (4, 8, 10)
16	Stinear *et al* (16)	yes	Suppl. Table 1	
17	Pérez-Lago *et al* (17)	no		No high confidence SNVs detected between serial isolates
18	Bloemberg *et al* (18)	no		Drug resistance-conferring mutations only
19	Liu *et al* (19)	yes	Suppl. Table S1 B	Suppl. Table S1 A excluded, as these SNVs did not appear *de novo* or disappear completely over time (see inclusion criteria)
20	Silva Feliciano *et al* (20)	no		Only a sub-set of drug resistance-conferring mutations analyzed and reported; unclear if these are all SNVs detected or not.

	Author & Reference	**Included in gene analysis** (yes/no)	**Data set used**	**Reason for exclusion**
21	Korhonen *et al* (21)	yes	Suppl. Table	Not ALL SNVs compared to reference reported, but all *de novo* emerged or disappeared SNVs reported
22	Faksri *et al* (22)	yes	SNVs described in main text	
23	Ssengooba *et al* (23)	yes	additional file 2	Parallel isolates (sputum and blood)
24	Zhang *et al* (24)	no		Only SNVs in the genes *gyrA* and *gyrB* reported
25	Casali *et al* (25)	no		Only drug resistance-conferring and compensatory mutations indicated, but no other SNVs between serial isolates
26	Lieberman *et al* (26)	yes		Parallel isolates across several anatomical sites
27	Wollenberg *et al* (27)	no		Drug resistance-conferring mutations only
28	Datta *et al* (28)	yes	Suppl.Tables 2, 3 and 4	
29	Manson *et al* (29)	no		Drug resistance-conferring mutations only
30	Dheda *et al* (30)	no		No SNVs reported
31	Witney *et al* (31)	yes	Table 4	
32	Trauner *et al* (32)	yes	Additional File 3_v-SNPs	SNVs present at all timepoints of a patient were excluded (see inclusion criteria)
33	Navarro *et al* (33)	yes	Table 1	SNVs from *in vitro* single colonies excluded (see inclusion criteria)
34	Nsofor *et al* (34)	no		Drug resistance-conferring mutations only
35	Senghore *et al* (35)	no		Drug resistance-conferring mutations and compensatory mutations only
36	Leung *et al* (36)	yes	Table 5	Only Sanger sequencing confirmed SNVs included
37	Herranz *et al* (37)	yes	Supplementary table 1	
38	Dheda *et al* (38)	yes	Suppl. Table e6: SNPs gained at t= 1 All new non-synonymous SNPs	SNPs not assignable to specific patient, but as all patients have parallel/serial isolates, and reported SNPs are all SNPs gained during treatment, still included in analysis

Table S3: Sequencing methods and analysis parameters of each study

Study	Sequencing Platform	Average Depth (range)	Read length	Sequencing Type	Reference strain (Accession Number)	Aligner	Variant caller	Min. coverage (at specific site)	Hetero-frequency cut-off (%)	Genomic regions excluded from analysis	Visual/manual inspection of sequences (yes/no)	Validation by Sanger /targeted PCR (yes/no)	Additional (non-WGS based) genotyping methods used
(1)	Illumina Solexa	60-100x	50bp	paired-end	*De novo* assembly of pooled data from 3 clinical strains investigated	Velvet	Mapping & Assembly with Quality (MAQ)	*not indicated*	*not indicated*	transposable & repetitive elements	no	no	RFLP, 15-locus MIRU-VNTR
(2)	Illumina	302x (77 - 512x)	76bp	paired-end	Common ancestor	Mapping & Assembly with Quality (MAQ)	Mapping & Assembly with Quality (MAQ)	5	100	mobile genetic elements, PE/PPE, PE-PGRS, heterozygous calls	no	no	*none*
(3)	Illumina HiSeq	88.5x	75bp	paired-end	H37Rv (NC000962.2)	Stampy	SAMtools	5	75	repetitive regions, SNVs within 12bp of another SNV	no	no	24-loci MIRU-VNTR
(4)	Illumina HiSeq 1000 & Illumina GenomeAnalyzer	55-269x	115bp + 95bp	paired-end	H37Rv (AL123456) CCDC5079 (CP001641)	BWA	SAMtools	*not indicated*	5	mobile genetic elements, indels, PE/PPE, PE-PGRS	no	no	*none*
(5)	Illumina Genome Analyser IIx	*not attributable to individual isolate*	**P1:** *not indicated* **P2,4,5,10:** 50bp **P2,6,8:** 35bp + 75bp	single-end	H37Rv	MAQ v0.6.6	MAQ	20	*not indicated*	SNVs within 5bp indels or adjacent consensus quality of <20. Reads aligning with >3 mismatches in the first 24bp and those aligning to multiple loci.	no	no	24-loci MIRU-VNTR, spoligotyping
(6)	Illumina HiSeq	120x	100bp	paired-end	corrected H37Rv	SMALT	SAMtools & BCFtools	4	5	SNVs within 200bp of other heterozygous sites	yes	no	24-loci MIRU-VNTR
(7)	Illumina HiSeq 2000	314x	76bp	paired-end	corrected H37Rv	BWA	SAMtools & BCFtools	10	*not indicated*	PE/PPE/PGRS	no	no	*none*
(8)	Illumina	*not available for whole genome*	72 and 46bp	paired-end	H37Rv (NC_000962.2)	SARUMAN	In-house perl script	10	75	high GC content regions, PE/PPE/PGRS gene families, ESAT-6 like, *lppA*, *lppB*, *pks12*	no	yes	Sanger sequencing
(9)	Illumina HiSeq	473.34x (105.3 - 3885.9x)	51-101bp	paired-end	Common ancestor *(Comas et al)*	BWA and MAQ	SAMtools	10	*not indicated*	positions with only heterozygous calls in all patients	no	yes	24-loci MIRU-VNTR, RFLP
(10)	Illumina HiSeq & MiSeq	210x	50-150bp	paired-end	H37Rv	SeqMan Ngen	SeqMan Pro	50	4	repetitive sequences and SNVs with distance <10	no	yes "verified by mutation -specific PCR"	24-loci MIRU-VNTR, RFLP

Study	Sequencing Platform	Average Depth (range)	Read length	Sequencing Type	Reference strain (Accession Number)	Aligner	Variant caller	Min. coverage (at specific site)	Hetero-frequency cut-off (%)	Genomic regions excluded from analysis	Visual/manual inspection of sequences (yes/no)	Validation by Sanger /targeted PCR (yes/no)	Additional (non-WGS based) genotyping methods used
(11)	Illumina HiSeq 2000	88x (median) 127x (mean)	100bp	paired-end	H37Rv (AL123456.3)	BWA-mem	SAMtools	20	75	highly repetitive and variable regions (e.g. PE/PPE)	no	no	Spoligotyping, RFLP
(12)	Ion Torrent personal genome machine (Ion PGM)	maximum 200x	200bp	*not indicated*	H37Rv (NC_000962.3)	TMAP	SAMtools	4	75 or 100 (*unclear formulation)*	heterozygous loci	no	no	*none*
(13)	Ilumina HiSeq 2000	88x (median) 127x (mean)	100bp	paired-end	H37Rv (AL123456.3)	BWA-mem	SAMtools	20	75	highly repetitive and variable regions (e.g. PE/PPE)	no	no	Spoligotyping, RFLP
(14)	Illumina HiSeq 2000	137	100bp	paired-end	H37Rv (AL123456)	BWA, Novoalign, SMALT	GATK	50	30	mobile genetic elements, repetitive regions, PE/PPE, PE-PGRS	yes	yes	Sanger sequencing, spoligotyping
(15)	Illumina HiSeq	96 - 192x	min 35bp	paired-end	H37Rv (NC_000962.3)	BWA MEM	GATK	50	1	poorly mapped sequences, SNVs with strand-bias or tail-distance-bias	yes	no	*none*
(16)	Illumina MiSeq	12-215x	250bp	paired-end	H37Rv (NC_000962.3)	Snippy v2.6	Nesoni v0.130	30	10	repetitive sequences	no	no	*none*
(17)	Illumina HiSeq 2000	715 to 1,252x	51bp	paired-end	Ancestral	BWA	SAMtools, Varscan	10	95 *(screening heterozygous positions in a 2nd step)*	*not indicated*	yes	no	Spoligotyping, 24-loci MIRU-VNTR, RFLP
(18)	Illumina	154x (median)	*not indicated*	*not indicated*	H37Rv (NC_000962)	MAQ	bre-seq pipeline	*not indicated*	10	repetitive regions	no	yes	*none*
(19)	Illumina HiSeq 2000	1000x	*not indicated*	*not indicated*	H37Rv (NC_000962.3) CCDC5079 (NC_021251.1)	BWA	SAMtools	*not indicated*	5	Mobile genetic elements, PE/PPE, PE-PGRS	no	yes	*none*
(20)	Illumina MiSeq	244x and 469x	*not indicated*	*not indicated*	H37Rv	BWA v0.7.5	SAMtools V0.1.19	*not indicated*	*not indicated*	*not indicated*	no	no	Line probe assay
(21)	Illumina MiSeq	*not indicated*	*not indicated*	paired-end	H37Rv	PhyResSE	PhyResSE, Mycrobe	2	70	repeat regions and regions difficult to map	no	no	*none*
(22)	Illumina MiSeq	64x	250bp	paired-end	H37Rv (NC_000962.3)	BWA-mem	GATK, SAMtools	10	75	repetitive elements (PE/PPE, PE-PGRS, phages, transposases, integrases)	no	yes	24-loci MIRU-VNTR, Spoligotyping

Study	Sequencing Platform	Average Depth (range)	Read length	Sequencing Type	Reference strain (Accession Number)	Aligner	Variant caller	Min. coverage (at specific site)	Hetero-frequency cut-off (%)	Genomic regions excluded from analysis	Visual/manual inspection of sequences (yes/no)	Validation by Sanger /targeted PCR (yes/no)	Additional (non-WGS based) genotyping methods used
(23)	Illumina HiSeq	*not indicated*	min 45bp	*not indicated*	H37Rv (NC_018143.1)	nesoni bowtie tool	nensoni consensus	10	*not indicated*	*not indicated*	no	no	Spoligotyping, 24-loci MIRU-VNTR
(24)	Illumina HiSeq	495 - 728x	101bp	paired-end	H37Rv (NC_000962.3)	BWA v0.5.9	Pilon	*not indicated*	0	highly repetitive regions	yes	yes	Sanger sequencing
(25)	Illumina HiSeq	37-126x	100bp	paired-end	re-sequenced (corrected) H37Rv Cole *et al*)	SMALT	SAMtools	2	70	repetitive regions including PE-PPE genes	yes	no	Spoligotyping, 24-loci MIRU-VNTR, RFLP
(26)	Illumina HiSeq 2000	36-303x	100bp	paired-end	Ancestral Genome (Comas *et al*) & H37Rv (NC_000962.3)	*not indicated*	SAMtools	4	10	removed genomic positions suspected to have false polymorphisms based on the distribution of allele frequencies across isolates of the same patient	no	no	*none*
(27)	Illumina HiSeq 2000	140x	101bp	paired-end	H37Rv (CP003248.2)	BWA	Pilon v1.5	*not indicated*	0	*not indicated*	no	no	*none*
(28)	SOLiD 3	155-234x	*not indicated*	*not indicated*	H37Rv	Lifescope Genome analysis software	SAMtools v0.1.18 & bcftools v 0.1.17-dev	20	*not indicated*	repeat regions (PE/PPE and PE-PGRS gene family)	no	yes *(katG)*	Spoligotyping, RFLP
(29)	Illumina HiSeq	130 +/- 79x	*not indicated*	paired-end	H37Rv (CP003248.2)	BWA v0.5.9	Pilon v1.5	*not indicated*	*not indicated*	*not indicated*	no	no	Spoligotyping
(30)	Illumina HiSeq 2000	*not indicated*	*not indicated*	paired-end	H37Rv (AL123456.3)	bwa-mem v0.7	SAMtools (v1.2) GATK (v3.3-0)	10	*not indicated*	highly repetitive and variable regions (e.g. PE/PPE)	no	no	Spoligotyping; *IS6110* RFLP
(31)	Illumina HiSeq 2500 & MiSeq	>20x	100bp & 250bp	paired-end	H37Rv (NC_000962.3)	BWA-mem v0.7.3a-r367	SAMtools v0.1.19	4	75 or 100 (*unclear formulation)*	*not indicated*	no	no	24-loci MIRU-VNTR
(32)	Illumina HiSeq 2000	850x	*not indicated*	paired-end	CCDC5079 (GenBank CP001641) H37Rv (GenBank AL123456)	BWA	SAMtools	5	1.5	repetitive regions including PPE/PE-PGRS genes, insertion elements...	no	no	*none*
(33)	Illumina HiSeq 2000 & MiSeq	*not indicated*	101-51bp	paired-end	Ancestral Genome *(Comas et al)*	BWA	SAMtools; Varscan	10	*not indicated*	*not indicated*	no	yes	24-loci MIRU-VNTR, allele specific PCR, ligation-mediated PCR

Study	**Sequencing Platform**	**Average Depth (range)**	**Read length**	**Sequencing Type**	**Reference strain** (Accession Number)	**Aligner**	**Variant caller**	**Min. coverage** (at specific site)	**Hetero-frequency cut-off** (%)	**Genomic regions excluded from analysis**	**Visual/manual inspection of sequences** (yes/no)	**Validation by Sanger /targeted PCR** (yes/no)	**Additional (non-WGS based) genotyping methods used**
(34)	Illumina HiSeq 2000	*not indicated*	300bp	paired-end	*not indicated*	*not indicated*	In-house Perl script	*not indicated*	*not indicated*	*not indicated*	no	no	12-loci MIRU-VNTR, RT PCR melting curve assay
(35)	Illumina MiSeq	49.6x (24.3 - 78.4x)	250bp	paired-end	H37Rv (NC_000962.3)	unclear (TMAP? But that is for Ion Torrent)	SAMtools	4	75	highly repetitive regions and regions of deletions	no	no	*none*
(36)	SMRT PacBio	**Isol. 1:** 307x **Isol. 2:** 241x	10-20kbp	*not indicated*	H37Rv (AL123456.3) Beijing reference (NZ_CP011510.1)	Mauve aligner (v2.3.1)	Quiver	40	*not indicated*	none	no	yes	24-loci MIRU-VNTR, Sanger sequencing
(37)	Illumina MiSeq	87x (62-113x)	*not indicated*	paired-end	hypothetical ancestral genome *(Comas et al)*	BWA and MAQ	SAMtools	20	*not indicated*	repeats, phages, PE/PPE, variants near indels, variants in regions with anomalous accumulation of SNVs	yes	no	24-locus MIRU-VNTR
(38)	Illumina HiSeq 2500	24 - 153x	100bp	paired-end	H37Rv (NC000962)	CLC Genomic workbench	CLC Genomic workbench	*not indicated*	*not indicated*	*not indicated*	no	no	*none*

Table S4: Genes micro-evolving in more than two studies

Gene	Alter. name	No. of studies reporting variants in that gene	Study	No. of patients with a variant in that gene /Total no. of isolates available of that patient		No. of different variants found	Isolates in which the variant was present	Drug Resistance Level	Lineage	Variant (Amino acid change at codon)	PROVEAN prediction (cut-off -2.5)	hetero-frequency	gain/loss of variant over time	Product & Protein Function (39)	Gene function class
Rv0118c	*oxcA*	3	(4)	1	3	1	**Patient 1:** Isol. a1	Scpt.	L2	A185P	deleterious -2.52	0.111	loss	Probable oxalyl-CoA decarboxylase OxcA; Involved in catabolism of oxalic acid	intermediary metabolism & respiration
Rv0118c	*oxcA*		(26)	1	93	1	**Patient 44**: Spleen: B2	Scpt.	L2	A284A	neutral	1.00	N/A		
Rv0118c	*oxcA*		(32)	1	2	1	**Patient 4:** Isol. 1 (timepoint 0)	Scpt.	L2	A573A	neutral	0.1262	loss		
Rv0278c	*PE-PGRS3*	3	(26)	1	117	1	**Patient 34:** Eta: A2,A3,C1,C4,C5 Lung1:A1-A5,B1-B5,C1-C5 Lung2:A1,A3-A4,B1-B5,C2,C4-C6 Lung3:A1,A3-A5,B1-B5,C1-C5 Lung4:B3 Lung5:A1-A3,A5,B1-B5,C3-C5 Lung6:A1-A5,B-B5,C1-C5 Liver:A1-A5,B1,C2-C5 Spleen:A1,A2,B1-B5,C2-C5	Scpt.	L2	P681T	neutral -0.38	1 0.58 - 1 0.88 - 1 0.25 - 1 0.07 0.13-0.9 0.64-1 0.85-1 0.25-1	N/A	Function unknown; PE-PGRS family protein PE_PGRS3	PE/PPE family
Rv0278c	*PE-PGRS3*		(5)	1	8	1	Isol. 3: P3-ITL Isol. 7: P8-ITL	MDR	L4	R807G	neutral 2.15	*not indicated*	transient		
Rv0278c	*PE-PGRS3*		(38)	*not assignable*	*not assignable*	1	*not indicated*	MDR - XDR	*not indicated*	A557? *(nucleotide change not indicated)*	N/A	*not indicated*	gain		
Rv0457c	---	3	(4)	1	3	1	**Patient_a:** Isol. 1	Scpt.	L2	E17Q	deleterious -2.92	0.12	loss	Function unknown; probable peptidase	Intermediary metabolism & respiration
Rv0457c	---		(32)	1	4	1	**Patient 2:** Isol. 2 (timepoint 2)	Scpt.	L2	A292T	neutral	0.0983	transient		

Gene	Alter. name	No. of studies reporting variants in that gene	Study	No. of patients with a variant in that gene /Total no. of isolates available of that patient		No. of different variants found	Isolates in which the variant was present	Drug Resistance Level	Lineage	Variant (Amino acid change at codon)	PROVEAN prediction (cut-off -2.5)	hetero-frequency	gain/loss of variant over time	Product & Protein Function (39)	Gene function class
Rv0457c	---		(37)	1	2	1	**Case 5:** Isol. 2	Scpt.	*not indicated*	L16R	deleterious -5.84	*not indicated*	gain		
Rv0565c	---	3	(4)	1	2	1	**Patient_c**: Isol. 2			Y245C	neutral -2.12	0.074	gain		
Rv0565c	---		(8)	1	3	1	**Patient A:** Isol. 1 (mixed infection) Isol. 2 Isol. 3	poly/MDR preXDR XDR	L2	R59H	deleterious -5.0	0.08 1 0.35	transient		
Rv0565c	---		(38)	*not assign able*	*N/A*	1	*not indicated*	MDR-XDR	*not indicated*	G347? *(nucleotide change not indicated)*	N/A	*not indicated*	gain	Probable monooxigenase	Intermediary metabolism & respiration
Rv0726c	---	3	(6)	1	2	*not indicated*	**Patient 4**	Scpt.	L4	*not indicated*	N/A	*not indicated*	N/A		
Rv0726c	---		(26)	1	123	1	**Patient 30**: Lung1: A5 Lung4: A5 Lung6: B3	Scpt.	L4	V58L	neutral 0.89	0.01 0.02 0.26	N/A	Putative S-adenosyl-L-methionine-dependent methyltransferase	lipid metabolism
Rv0726c	---		(32)	1	4	1	**Patient 10:** Isol. 3 (timepoint week 6)	MDR	L2	P128S	deleterious -7.14	0.0724	transient		
Rv0746	*PE-PGRS9*	3	(12)	1	2	1	**Patient 2:** Isol. 2b	XDR	L2	T320A	neutral 1.61	0.75-1	gain		
Rv0746	*PE-PGRS9*		(5)	1	8	4	Isol. P6-ITL, Isol. P8-ITL Isol. P1-P3-ITL, P5 - P6-ITL, P8-ITL, P10-ITL Isol. P3-ITL Isol. P1-3-ITL, P5-6-ITL, P10-ITL	MDR Scpt-MDR MDR Scpt-MDR	L4	E191G T320A T252A N280D	neutral 3.65 neutral 1.61 neutral 0.30 neutral -0.26	*not indicated*	transient		
Rv0746	*PE-PGRS9*		(38)	*not assign able*	*N/A*	1	*in 4 different isolates?*	MDR-XDR	*not indicated*	V739? *(nucleotide change not indicated)*	N/A	*not indicated*	gain	Function unknown; PE-PGRS family protein PE_PGRS9	PE/PPE family

Gene	Alter. name	No. of studies reporting variants in that gene	Study	No. of patients with a variant in that gene /Total no. of isolates available of that patient		No. of different variants found	Isolates in which the variant was present	Drug Resistance Level	Lineage	Variant (Amino acid change at codon)	PROVEAN prediction (cut-off -2.5)	hetero-frequency	gain/loss of variant over time	Product & Protein Function (39)	Gene function class
Rv1129c	*prpR*	3	(6)	1	2	*not indicated*	**Patient 3**	Scpt.	L2	*not indicated*	N/A	*not indicated*	N/A	HTH-type transcriptional regulator PrpR; Involved in transcriptional mechanism	regulatory proteins
Rv1129c	*prpR*		(26)	1	98	1	**Patient 15:** Lung3: C1 Lung4: A5, B2, B4 Lung6: B3	Scpt.	L4	S357R	deleterious -5.0	0.01 - 0.13	N/A		
Rv1129c	*prpR*		(32)	1	5	1	**Patient 11:** Isol. 1, 3, 4	preXDR	L4	D387G	deleterious -7.0	0.88 0.87 0.67	loss		
Rv1181	*pks4, msl3*	4	(31)	1	2	1	**Patient 031** Isol. 2	Scpt.	L2	P1330P	neutral	*not indicated*	gain	Mycolipanoate synthase; Polyketide synthase involved in the biosynthesis of methyl-branched fatty acids	lipid metabolism
Rv1181	*pks4, msl3*		(23)	1	2	1	**Patient 6,** blood isolate	Scpt.	L4	S1512S	neutral	*not indicated*	N/A		
Rv1181	*pks4, msl3*		(26)	2	9 49	2	**Patient 9:** Lymph Liver **Patient 29:** Liver: B4, B5 Spleen: B4	MDR Scpt.	L4 L4	M824T L301P	neutral -2.34 deleterious -5.64	**P9:** 0.01 0.12 **P29:** 0.06, 1.0 0.01	N/A		
Rv1181	*pks4, msl3*		(32)	2	4 5	2	**Patient 2:** Isol. 1 (timepoint 0) **Patient 8:** Isol. 3 (timepoint week 4)	Scpt. INH-mono	L2 L2	S846S R630L	neutral deleterious -4.49	**P2:** 0.0744 **P8:** 0.0387	loss transient		
Rv1527c	*pks5*	3	(33)	1	2	1	**Patient 1:** Isol. 2	*not indicated*	*not indicated*	*no position indicated*	N/A	*not indicated*	gain	Mycocerosic acid synthase-like polyketide synthase; Polyketide synthase likely involved in the biosynthesis of a polymethyl-branched fatty acid	lipid metabolism

Gene	Alter. name	No. of studies reporting variants in that gene	Study	No. of patients with a variant in that gene /Total no. of isolates available of that patient		No. of different variants found	Isolates in which the variant was present	Drug Resistance Level	Lineage	Variant (Amino acid change at codon)	PROVEAN prediction (cut-off -2.5)	hetero-frequency	gain/loss of variant over time	Product & Protein Function (39)	Gene function class
Rv1527c	*pks5*		(26)	3	105 116 97	3	**Patient 25:** Eta: C1,C3 Lung1: A2,A3,A4,B1,B2,C4,C5 Lung2: A1,A2,B1,C3,C4 Lung3: A3,B4,B5 **Patient 28:** Lung4: B2 Lung5: A2,A4,B1 Liver: A4 **Patient 38:** Lung3: B1,B5 Liver: C1	Scpt. Scpt. Scpt.	L2 L2/L4 *(mixed)* L4	M672I N1510K V1504I	deleterious -3.59 deleterious -2.82 neutral -0.96	**P25:** Eta: 0.24,0.01 Lung1: 0.01-1.00 Lung2: 0.03-0.18 Lung3: 0.02-0.41 **P28:** Lung4: 0.06 Lung5: 0.01-0.25 Liver-A4: 0.04 **P38:** Lung3: 0.02,0.09 Liver-C1: 1.00	N/A		
Rv1527c	*pks5*		(32)	1	4	1	**Patient 10:** Isol. 1	MDR	L2	P261L	deleterious -3.68	0.0162	loss		
Rv2024c	---	3	(4)	1	2	1	**Patient_b:** Isol. 1	INH, SM	L2	I145T	deleterious -4.53	0.089	loss		
Rv2024c	---		(26)	1	8	1	**Patient 5:** Eta Lung4	Scpt.	L4	Y42Y	neutral	0.03 0.18	N/A		
Rv2024c	---		(5)	1	8	2	Isol. P1-P2-ITL,P4-6ITL,P8-ITL,P10-ITL Isol. P1-P2-ITL,P4-6ITL,P8-ITL,P10-ITL	MDR	L4	W47R D154G	neutral 4.48 neutral -0.37	*not indicated*	transient	Function unknown	conserved hypothetical
Rv2187	*fadD15*	3	(4)	1	3	1	**Patient_a:** isolate 2	INH-mono	L2	F230S	deleterious -7.91	0.085	transient	Long-chain-fatty-acid-CoA ligase FadD15; Catalyzes the activation of long-chain fatty acids as acyl-coenzyme A (acyl-CoA)	lipid metabolism

Gene	Alter. name	No. of studies reporting variants in that gene	Study	No. of patients with a variant in that gene /Total no. of isolates available of that patient		No. of different variants found	Isolates in which the variant was present	Drug Resistance Level	Lineage	Variant (Amino acid change at codon)	PROVEAN prediction (cut-off -2.5)	hetero-frequency	gain/loss of variant over time	Product & Protein Function (39)	Gene function class
Rv2187	*fadD15*		(26)	1	105	1	**Patient 25:** detected in 54 isolates: Eta: A3,B5,B6,C1,C3,C4 Lung1:A2,A3,B1,B3,B4,C1,C2, C4 Lung2:A1,A2,A4,B1,B5,C2-C5 Lung3:A3, A4, B3,B4,B5,C1-C5 Lung5:A4,+H29B3,C2,D3-D5 Lung6:A1,A2,A3,B3,B5 Liver:A5,C1,C4,B2 Spleen:A4,A5,B2,B3,B5,C2	Scpt.	L2	L503L	neutral	0.07 - 1.00 0.01 - 0.50 0.05 - 0.62 0.04 - 0.42 0.02 - 1.00 0.02 - 0.91 0.01 - 0.09 0.03 - 0.25	N/A		
Rv2187	*fadD15*		(32)	1	5	1	**Patient 6:** Isol. 3 (timepoint week 4)	INH-mono	L2	D570D	neutral	0.1909	transient		
Rv2477c	*ettA*	3	(28)	1	2	1	**Patient 3** (C21/C27) C --> T mutation happens in same isolate than the increase to INH-resistance	INH-mono	L4	R287C	deleterious -7.97	*not indicated*	gain	Energy-dependent translational throttle protein EttA;translation factor; essential gene for *in vitro* growth	cell wall & cell processes
Rv2477c	*ettA*		(22)	1	2	1	**Patient 2:** isolate 2	preXDR	L2	W135G	deleterious -12.76	≥ 0.75	gain		
Rv2477c	*ettA*		(21)	1	2	1	**Patient 1:** isolate 2	INH, STR	L4	P547P	neutral	≥ 0.70	gain		
Rv2931	*ppsA*	3	(26)	1	98	1	**Patient 35**: Eta: A3, A5, B2 Lung1: C1 Lung2: A3, C1 Lung3: B4 Lung4: B4, C5 Lung5: A2	Scpt.	L4	V61I	neutral -0.46	Eta: 0.01;0.03;0.02 Lung1: 0.67 Lung 2: 0.04; 0.01 Lung 3: 0.01 Lung 4: 0.03; 0.02 Lung 5: 0.02	N/A	Phenolpthiocerol synthesis type-I polyketide synthase PpsA. Involved in phenolpthiocerol & phthiocerol dimycocerosate (dim) biosynthesis.	lipid metabolism
Rv2931	*ppsA*		(32)	1	5	1	**Patient 11:** isol. 3 (timepoint 4)	MDR	L4	H13P	neutral -1.58	0.015	transient		

Gene	Alter. name	No. of studies reporting variants in that gene	Study	No. of patients with a variant in that gene /Total no. of isolates available of that patient		No. of different variants found	Isolates in which the variant was present	Drug Resistance Level	Lineage	Variant (Amino acid change at codon)	PROVEAN prediction (cut-off -2.5)	hetero-frequency	gain/loss of variant over time	Product & Protein Function (39)	Gene function class
Rv2931	*ppsA*		(5)	1	8	2	Isol. P5-ITL, P6-ITL, P8-ITL Isol. P6-ITL, P8-ITL	MDR MDR	L4	A803T R877H	neutral -0.58 neutral -1.93	*not indicated*	transient transient		
Rv2935	*ppsE*	4	(14)	1	2	1	**Patient 2:** isolate 2	MDR	L2	C582R	deleterious -11.13	0.69	gain	Phthiocerol synthesis polyketide synthase type I PpsE; Involved in the elongation of fatty acids	lipid metabolism
Rv2935	*ppsE*		(4)	1	2	1	**Patient C**: isolate 2	MDR	L2	A718A	neutral	0.1	gain		
Rv2935	*ppsE*		(2)	1	2	1	**Patient 8:** isolate 2	MDR	L4	A478E	neutral 1.26	1.00	gain		
Rv2935	*ppsE*		(32)	1	5	1	**Patient 12:** isolate 1 (timepoint 0)	MDR	L2	A1220G	neutral -0.58	0.0178	loss		
Rv2973	*recG*	3	(2)	1	2	1	**Patient 10**: isolate 2	MDR	L4	M657R	neutral -1.09	1.00	gain	Probable ATP-dependent DNA helicase RecG; Critical role in recombination and DNA repair	information pathways
Rv2973	*recG*		(10)	1	9	1	Present in isolate SF8	XDR	L4	V521A	deleterious -3.98	0.53	transient		
Rv2973	*recG*		(32)	1	4	1	**Patient 02:** isolate 2 (timepoint 2)	Scpt.	L2	S475A	deleterious -2.68	0.2665	trans.		
Rv3303c	*lpdA*	3	(6)	1	2	1	**Patient 13:** isolate 2	Scpt.	L4	*not indicated*	N/A	*not indicated*	gain	NAD(P)H quinone reductase LpdA; Involved in energy metabolism. Protects against oxidative stress	intermediary metabolism & respiration
Rv3303c	*lpdA*		(32)	1	5	1	**Patient 12:** isolates 2-5 (timepoints 2,4,6,8)	MDR	L2	L470L	neutral	0.0156 - 0.2795	gain		
Rv3303c	*lpdA*		(36)	1	2	1	**Patient 1:** isolate 2	preXDR	L2	V44I	neutral -0.91	*not indicated*	gain		

Gene	Alter. name	No. of studies reporting variants in that gene	Study	No. of patients with a variant in that gene /Total no. of isolates available of that patient		No. of different variants found	Isolates in which the variant was present	Drug Resistance Level	Lineage	Variant (Amino acid change at codon)	PROVEAN prediction (cut-off -2.5)	hetero-frequency	gain/loss of variant over time	Product & Protein Function (39)	Gene function class
Rv3696	*glpK*	6	(14)	1	2	2	**Patient 1:** isolate 1 **Patient 1:** isolate 2	RIF-mono MDR	L4	**Isolate 1:** ins AC *(position not indicated)* **Isolate 2:** T91I	deleterious -5.98	0.73 0.8	loss gain	Probable glycerol kinase GlpK. Acts in rate-limiting step in glycerol utilization. Key enzyme in the regulation of glycerol uptake & metabolism	Intermediary metabolism & respiration
Rv3696	*glpK*		(10)	1	9	2	isolate S9 isolate S9	XDR XDR	L4	W481G W114C	deleterious -12.75 deleterious -12.96	0.3146 0.375	gain gain		
Rv3696	*glpK*		(22)	1	2	1	**Patient 1:** isolate 2	XDR	L2	S26R	deleterious -4.95	≥ 0.75	gain		
Rv3696	*glpK*		(21)	1	2	1	**Patient 3:** isolate 1	Scpt.	L4	N312N	neutral	≥ 0.70	loss		
Rv3696	*glpK*		(2)	1	2	1	**Patient 6:** isolate 2	RIF-mono	L4	C435R	deleterious -3.53	1.00	gain		
Rv3696	*glpK*		(32)	1	5	5	**Patient 12:** isolate 1 (timepoint 0)	MDR	L2	W491S V460L T389P M266V G89A	neutral -2.21 neutral -0.98 deleterious -3.08 neutral -0.22 deleterious -5.98	0.0166 0.0208 0.0258 0.0321 0.020	loss loss loss loss loss		
Rv3879c	*espK*	3	(28)	1	2	1	**Patient 2:** Isol. 2 (C1)	INH-mono	L2	D86N	neutral -0.86	*not indicated*	gain	ESX-1 secretion-associated protein EspK	cell wall & cell processes
Rv3879c	*espK*		(4)	1	2	1	**Patient_c:** isol. 1	MDR	L2	Y543STOP	deleterious -318.82	0.247	loss		

Gene	Alter. name	No. of studies reporting variants in that gene	Study	No. of patients with a variant in that gene /Total no. of isolates available of that patient		No. of different variants found	Isolates in which the variant was present	Drug Resistance Level	Lineage	Variant (Amino acid change at codon)	PROVEAN prediction (cut-off -2.5)	hetero-frequency	gain/loss of variant over time	Product & Protein Function (39)	Gene function class
Rv3879c	*espK*		(26)	1	98	1	**Patient 15:** Lung4: B2 Lung5: A2 Lung6: B3 Liver: C5 Eta: C1	Scpt.	L4	S156S	neutral	0.01 0.03 0.99 0.01 0.03	N/A		
Rv3910	*mviN*	4	(4)	1	2	1	**Patient C:** isolate 2	MDR	L2	G708G	neutral	0.05	gain	Probable peptidoglycan biosynthesis protein MviN Essential for cell growth and peptidoglycan synthesis.	cell wall & cell processes
Rv3910	*mviN*		(21)	1	2	1	**Patient 15**: isolate 1	Scpt.	L4	P1040P	neutral	≥ 0.70	loss		
Rv3910	*mviN*		(26)	2	70 82	2	**Patient 17:** Lung6: A1,A4 **Patient 22:** Lung1: A1 Lung3: B3	Scpt. Scpt.	L4 L4	L344L T228N	neutral deleterious -2.83	0.14, 0.08 1.00 0.01	N/A		
Rv3910	*mviN*		(32)	1	4	1	**Patient 10:** isolate 1 (timepoint 0)	MDR	L2	V549A	neutral 0.64	0.0172	loss		

Patient and isolate numbering is according to the original publication, except for Casali *et al* (43)
Scpt. = susceptible; MDR = multi-drug resistant; XDR = extensively drug resistant; INH = isoniazid; RIF = rifampicin; STR = streptomycin; Alter. = alternative

REFERENCES

1. Saunders NJ, Trivedi UH, Thomson ML, Doig C, Laurenson IF, Blaxter ML. 2011. Deep resequencing of serial sputum isolates of Mycobacterium tuberculosis during therapeutic failure due to poor compliance reveals stepwise mutation of key resistance genes on an otherwise stable genetic background. J Infect 62:212–217.

2. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S. 2012. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 44:106–110.

3. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TEA. 2013. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 13:137–146.

4. Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, Zheng H, Tian W, Wang S, Barry CE, Mei J, Gao Q. 2012. Dynamic population changes in Mycobacterium tuberculosis during acquisition and fixation of drug resistance in patients. J Infect Dis 206:1724–1733.

5. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PKC, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat Genet 45:1183–1189.

6. Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, van Helden P, Pym A, Mahayiddin AA, Chuchottaworn C, Sanne IM, Louw C, Boeree MJ, Hoelscher M, McHugh TD, Bateson ALC, Hunt RD, Mwaigwisya S, Wright L, Gillespie SH, Bentley SD. 2013. Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: a retrospective observational study. Lancet Respir Med 1:786–792.

7. Clark TG, Mallard K, Coll F, Preston M, Assefa S, Harris D, Ogwang S, Mumbowa F, Kirenga B, O'Sullivan DM, Okwera A, Eisenach KD, Joloba M, Bentley SD, Ellner JJ, Parkhill J, Jones-López EC, McNerney R. 2013. Elucidating emergence and transmission of multidrug-resistant tuberculosis in treatment experienced patients by whole genome sequencing. PloS One 8:e83012.

8. Merker M, Kohl TA, Roetzer A, Truebe L, Richter E, Rüsch-Gerdes S, Fattorini L, Oggioni MR, Cox H, Varaine F, Niemann S. 2013. Whole genome sequencing reveals complex evolution patterns of multidrug-resistant Mycobacterium tuberculosis Beijing strains in patients. PloS One 8:e82551.

9. Pérez-Lago L, Comas I, Navarro Y, González-Candelas F, Herranz M, Bouza E, García-de-Viedma D. 2014. Whole genome sequencing analysis of intrapatient microevolution in Mycobacterium tuberculosis: potential impact on the inference of tuberculosis transmission. J Infect Dis 209:98–108.

10. Eldholm V, Norheim G, von der Lippe B, Kinander W, Dahle UR, Caugant DA, Mannsåker T, Mengshoel AT, Dyrhol-Riise AM, Balloux F. 2014. Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient. Genome Biol 15:490.

11. Guerra-Assunção JA, Houben RMGJ, Crampin AC, Mzembe T, Mallard K, Coll F, Khan P, Banda L, Chiwaya A, Pereira RPA, McNerney R, Harris D, Parkhill J, Clark TG, Glynn JR. 2015. Recurrence due to Relapse or Reinfection With *Mycobacterium tuberculosis* : A Whole-Genome Sequencing Approach in a Large, Population-Based Cohort With a High HIV Infection Prevalence and Active Follow-up. J Infect Dis 211:1154–1163.

12. Witney AA, Gould KA, Arnold A, Coleman D, Delgado R, Dhillon J, Pond MJ, Pope CF, Planche TD, Stoker NG, Cosgrove CA, Butcher PD, Harrison TS, Hinds J. 2015. Clinical application of whole-genome sequencing to inform treatment for multidrug-resistant tuberculosis cases. J Clin Microbiol 53:1473–1483.

13. Guerra-Assunção JA, Crampin AC, Houben RMGJ, Mzembe T, Mallard K, Coll F, Khan P, Banda L, Chiwaya A, Pereira RPA, McNerney R, Fine PEM, Parkhill J, Clark TG, Glynn JR. 2015. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. eLife 4.

14. Black PA, de Vos M, Louw GE, van der Merwe RG, Dippenaar A, Streicher EM, Abdallah AM, Sampson SL, Victor TC, Dolby T, Simpson JA, van Helden PD, Warren RM, Pain A. 2015. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in Mycobacterium tuberculosis isolates. BMC Genomics 16:857.

15. O'Neill MB, Mortimer TD, Pepperell CS. 2015. Diversity of Mycobacterium tuberculosis across Evolutionary Scales. PLoS Pathog 11:e1005257.

16. Stinear TP, Denholm J, Globan M, Leslie D, Seemann T, Meumann EM, Porter JL, Fyfe JAM. 2015. Genome sequence comparisons of serial multi-drug-resistant Mycobacterium tuberculosis isolates over 21 years of infection in a single patient. Microb Genomics 1.

17. Pérez-Lago L, Navarro Y, Montilla P, Comas I, Herranz M, Rodríguez-Gallego C, Ruiz Serrano MJ, Bouza E, García de Viedma D. 2015. Persistent Infection by a Mycobacterium tuberculosis Strain That Was Theorized To Have Advantageous Properties, as It Was Responsible for a Massive Outbreak. J Clin Microbiol 53:3423–3429.

18. Bloemberg GV, Keller PM, Stucki D, Stuckia D, Trauner A, Borrell S, Latshang T, Coscolla M, Rothe T, Hömke R, Ritter C, Feldmann J, Schulthess B, Gagneux S, Böttger EC. 2015. Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis. N Engl J Med 373:1986–1988.

19. Liu Q, Via LE, Luo T, Liang L, Liu X, Wu S, Shen Q, Wei W, Ruan X, Yuan X, Zhang G, Barry CE, Gao Q. 2015. Within patient microevolution of Mycobacterium tuberculosis correlates with heterogeneous responses to treatment. Sci Rep 5:17507.

20. Silva Feliciano C, Rodrigues Plaça J, Peronni K, Araújo Silva W, Roberto Bollela V. 2016. Evaluation of resistance acquisition during tuberculosis treatment using whole genome sequencing. Braz J Infect Dis Off Publ Braz Soc Infect Dis 20:290–293.

21. Korhonen V, Smit PW, Haanperä M, Casali N, Ruutu P, Vasankari T, Soini H. 2016. Whole genome analysis of Mycobacterium tuberculosis isolates from recurrent episodes of tuberculosis, Finland, 1995-2013. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis 22:549–554.

22. Faksri K, Tan JH, Disratthakit A, Xia E, Prammananan T, Suriyaphol P, Khor CC, Teo Y-Y, Ong RT-H, Chaiprasert A. 2016. Whole-Genome Sequencing Analysis of Serially Isolated Multi-Drug and Extensively Drug Resistant Mycobacterium tuberculosis from Thai Patients. PloS One 11:e0160992.

23. Ssengooba W, de Jong BC, Joloba ML, Cobelens FG, Meehan CJ. 2016. Whole genome sequencing reveals mycobacterial microevolution among concurrent isolates from sputum and blood in HIV infected TB patients. BMC Infect Dis 16.

24. Zhang D, Gomez JE, Chien J-Y, Haseley N, Desjardins CA, Earl AM, Hsueh P-R, Hung DT. 2016. Genomic analysis of the evolution of fluoroquinolone resistance in Mycobacterium tuberculosis prior to tuberculosis diagnosis. Antimicrob Agents Chemother.

25. Casali N, Broda A, Harris SR, Parkhill J, Brown T, Drobniewski F. 2016. Whole Genome Sequence Analysis of a Large Isoniazid-Resistant Tuberculosis Outbreak in London: A Retrospective Observational Study. PLOS Med 13:e1002137.

26. Lieberman TD, Wilson D, Misra R, Xiong LL, Moodley P, Cohen T, Kishony R. 2016. Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated Mycobacterium tuberculosis. Nat Med.

27. Wollenberg KR, Desjardins CA, Zalutskaya A, Slodovnikova V, Oler AJ, Quiñones M, Abeel T, Chapman SB, Tartakovsky M, Gabrielian A, Hoffner S, Skrahin A, Birren BW, Rosenthal A, Skrahina A, Earl AM. 2016. Whole genome sequencing of Mycobacterium tuberculosis provides insight into the evolution and genetic composition of drug-resistant tuberculosis in Belarus. J Clin Microbiol JCM.02116-16.

28. Datta G, Nieto LM, Davidson RM, Mehaffy C, Pederson C, Dobos KM, Strong M. 2016. Longitudinal whole genome analysis of pre and post drug treatment Mycobacterium tuberculosis isolates reveals progressive steps to drug resistance. Tuberc Edinb Scotl 98:50–55.

29. Manson AL, Abeel T, Galagan JE, Sundaramurthi JC, Salazar A, Gehrmann T, Shanmugam SK, Palaniyandi K, Narayanan S, Swaminathan S, Earl AM. 2017. Mycobacterium tuberculosis Whole Genome Sequences From Southern India Suggest Novel Resistance Mechanisms and the Need for Region-Specific Diagnostics. Clin Infect Dis 64:1494–1501.

30. Dheda K, Limberis JD, Pietersen E, Phelan J, Esmail A, Lesosky M, Fennelly KP, Te Riele J, Mastrapa B, Streicher EM, Dolby T, Abdallah AM, Ben-Rached F, Simpson J, Smith L, Gumbo T, van Helden P, Sirgel FA, McNerney R, Theron G, Pain A, Clark TG, Warren RM. 2017. Outcomes, infectiousness, and transmission dynamics of patients with extensively drug-resistant tuberculosis and home-discharged patients with programmatically incurable tuberculosis: a prospective cohort study. Lancet Respir Med.

31. Witney AA, Bateson ALE, Jindani A, Phillips PPJ, Coleman D, Stoker NG, Butcher PD, McHugh TD, RIFAQUIN Study Team. 2017. Use of whole-genome sequencing to distinguish relapse from reinfection in a completed tuberculosis clinical trial. BMC Med 15:71.

32. Trauner A, Liu Q, Via LE, Liu X, Ruan X, Liang L, Shi H, Chen Y, Wang Z, Liang R, Zhang W, Wei W, Gao J, Sun G, Brites D, England K, Zhang G, Gagneux S, Barry CE, Gao Q. 2017. The within-host population dynamics of Mycobacterium tuberculosis vary with treatment efficacy. Genome Biol 18:71.

33. Navarro Y, Pérez-Lago L, Herranz M, Sierra O, Comas I, Sicilia J, Bouza E, García de Viedma D. 2017. In-Depth Characterization and Functional Analysis of Clonal Variants in a Mycobacterium tuberculosis Strain Prone to Microevolution. Front Microbiol 8:694.

34. Nsofor CA, Jiang Q, Wu J, Gan M, Liu Q, Zuo T, Zhu G, Gao Q. 2017. Transmission is a Noticeable Cause of Resistance Among Treated Tuberculosis Patients in Shanghai, China. Sci Rep 7:7691.

35. Senghore M, Otu J, Witney A, Gehre F, Doughty EL, Kay GL, Butcher P, Salako K, Kehinde A, Onyejepu N, Idigbe E, Corrah T, de Jong B, Pallen MJ, Antonio M. 2017. Whole-genome sequencing illuminates the evolution and spread of multidrug-resistant tuberculosis in Southwest Nigeria. PloS One 12:e0184510.

36. Leung KS-S, Siu GK-H, Tam KK-G, To SW-C, Rajwani R, Ho P-L, Wong SS-Y, Zhao WW, Ma OC-K, Yam W-C. 2017. Comparative Genomic Analysis of Two Clonally Related Multidrug Resistant Mycobacterium tuberculosis by Single Molecule Real Time Sequencing. Front Cell Infect Microbiol 7:478.

37. Herranz M, Pole I, Ozere I, Chiner-Oms Á, Martínez-Lirola M, Pérez-García F, Gijón P, Serrano MJR, Romero LC, Cuevas O, Comas I, Bouza E, Pérez-Lago L, García-de-Viedma D. 2017. Mycobacterium tuberculosis Acquires Limited Genetic Diversity in Prolonged Infections, Reactivations and Transmissions Involving Multiple Hosts. Front Microbiol 8:2661.

38. Dheda K, Lenders L, Magombedze G, Srivastava S, Raj P, Arning E, Ashcraft P, Bottiglieri T, Wainwright H, Pennel T, Linegar A, Moodley L, Pooran A, Pasipanodya JG, Sirgel FA, van Helden PD, Wakeland E, Warren RM, Gumbo T. 2018. Drug Penetration Gradients Associated with Acquired Drug Resistance in Tuberculosis Patients. Am J Respir Crit Care Med.

39. Kapopoulou A, Lew JM, Cole ST. 2011. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. Tuberc Edinb Scotl 91:8–13.

40. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. 2014. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun 5:4812.

41. 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. 2017. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. Eur Respir J 50.