The relationship between transmission time and clustering methods in *Mycobacterium tuberculosis* epidemiology

Conor J Meehan[1,*], Pieter Moris[1,2,3], Thomas A. Kohl[4,5], Jūlija Pečerska[6], Suriya Akter[1], Matthias Merker [4,5], Christian Utpatel[4,5], Patrick Beckert[4,5], Florian Gehre[1,7,8], Pauline Lempens[1], Tanja Stadler[6], Michel K. Kaswa[1,10], Denise Kühnert[9], Stefan Niemann[4,5,&], Bouke C de Jong[1,&]

Affiliations

[1]Unit of Mycobacteriology, Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium

[2]Advanced Database Research and Modelling (ADReM), Department of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium

[3]Biomedical Informatics Research Network Antwerp (biomina), University of Antwerp, Antwerp, Belgium

[4]German Center for Infection Research, Partner Site Hamburg-Lübeck-Borstel-Riems, Borstel, Germany

[5]Molecular and Experimental Mycobacteriology, Priority Area Infections, Research Center Borstel, Borstel, Germany

[6]Department of Biosystems Science and Engineering, ETH Zürich, Switzerland

[7]Vaccines and Immunity Theme, Medical Research Council Unit The Gambia

[8]Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany

[9]University Hospital Zurich, Zurich, Switzerland

[10]National Tuberculosis Program, Kinshasa, Democratic Republic of Congo

[&]equal contribution; *corresponding author: cmeehan@itg.be

Abstract

Tracking recent transmission is a vital part of controlling widespread pathogens such as *Mycobacterium tuberculosis*. Multiple approaches exist for detecting recent transmission chains, usually by clustering strains based on the similarity of their genotyping results. However, each method gives varying estimates of transmission cluster sizes and inferring when transmission events within these clusters occurred is almost impossible. This study combines whole genome sequence (WGS) data derived from a high endemic setting with phylodynamics to unveil the timing of transmission events posited by a variety of standard genotyping methods. Our results suggest that clusters based on spoligotyping could encompass transmission events that occurred hundreds of years prior to sampling while 24-loci-MIRU-VNTR often represented decades of transmission. Instead, WGS based genotyping applying a low SNP thresholds allows for estimation of recent transmission events. These findings can guide the selection of appropriate clustering methods for uncovering relevant transmission chains within a given time-period.

Introduction

Despite the large global efforts at curbing the spread of *Mycobacterium tuberculosis* complex (Mtbc) strains, 10.4 million new patients develop tuberculosis (TB) every year (WHO, 2018). In addition, the prevalence of multidrug resistant Mtbc strains (MDR-TB) is increasing (WHO, 2018), predominantly through ongoing transmission within large populations (Kendall et al., 2015; Merker et al., 2015). The tracking and timing of recent transmission chains allows TB control programs to effectively pinpoint transmission hotspots and employ targeted intervention measures. This is especially important for the transmission of drug resistant strains as it appears that drug resistance may be transmitted more frequently than acquired

(Kendall et al., 2015; Trauer, Denholm, & McBryde, 2014). Thus, interrupting transmission is key for the control of MDR-TB (Klopper et al., 2013; Merker et al., 2015; Shah et al., 2017). For the development of the most effective control strategies, there is a strong need for (i) appropriate identification of relevant transmission chains, risk factors and hotspots and (ii) robust timing of when outbreaks first arose.

Epidemiological TB studies often apply genotyping methods to Mtbc strains to determine whether two or more patients are linked within a transmission chain (molecular epidemiology) (Merker, Kohl, Niemann, & Supply, 2017). Contact tracing is a primary epidemiological method for investigating transmission networks of TB, mainly based on patient interviews (Fox, Barry, Britton, & Marks, 2012). Although this method is often seen as a gold standard of transmission linking, it does not always match the true transmission patterns, even in low incidence settings (Behr et al., 1998; Diel et al., 2002; Roetzer et al., 2011, 2013; Small et al., 1994) and misses many connections (Bjorn-Mortensen et al., 2017; Vluggen et al., 2017). The implementation of molecular epidemiological approaches has overcome these limitations and is often used as the main approach for cluster analysis. Classical genotyping has involved IS*6110* DNA fingerprinting (Thierry et al., 1990; van Embden et al., 1993), spoligotyping (Goguet de la Salmonière et al., 1997; Guernier, Sola, Brudey, Guégan, & Rastogi, 2008; Kamerbeek et al., 1997), and variable-number tandem repeats of mycobacterial interspersed repetitive units (MIRU-VNTR) (Supply, Magdalena, Himpens, & Locht, 1997) which is the most common method at the moment (Merker et al., 2017). The latter method is based on copy numbers of a sequence in tandem repeat patterns derived from 24 distinct loci within the genome (Supply et al., 2006). If two patients have the same classical genotyping pattern such as a 24-loci MIRU-VNTR pattern (or up to one locus

72 difference (Jonsson et al., 2014; Supply et al., 2006)) they are considered to be within a local
73 transmission chain. The combination of spoligotyping and MIRU-VNTR-typing, where patterns
74 must match in both methods to be considered a transmission link, is often considered the
75 molecular gold standard for transmission linking and genotyping (Supply et al., 2006).
76 However, examples of unlinked patients with identical patterns have been observed,
77 suggesting that this threshold covers too broad a genetic diversity and timespan between
78 infections (Gardy et al., 2011; Roetzer et al., 2013).
79
80 The application of (whole genome) sequence-based approaches for similarity analysis of Mtbc
81 isolates and cluster determination is known to have high discriminatory power when
82 assessing transmission dynamics (Comas, 2017; Roetzer et al., 2013; Walker et al., 2013, 2018;
83 Wyllie et al., 2018). Single nucleotide polymorphisms (SNPs) in the *pncA* gene are associated
84 with resistance to pyrazinamide (PZA) and can be used to improve the discriminatory power
85 of spoligotyping in a method referred to as SpoNC (Said et al., 2016). However, this is limited
86 by the low occurrence of PZA resistance, even in MDR-TB isolates (Kurbatova, Cavanaugh,
87 Dalton, S. Click, & Cegielski, 2013; Ngabonziza et al., 2017; Xu et al., 2016; Zignol et al., 2016).
88 The advent of widespread whole genome sequencing (WGS) capabilities has allowed for
89 highly discriminatory analyses of Mtbc strains either using core genome multi-locus sequence
90 typing (cgMLST) (Thomas A. Kohl et al., 2014) or SNP distances (Bjorn-Mortensen et al., 2016;
91 Gardy et al., 2011; Roetzer et al., 2013; Walker et al., 2013, 2018). WGS-based approaches
92 compare the genetic relatedness of the genomes of the clinical strains under consideration,
93 albeit usually excluding large repetitive portions of the genome, with the assumption that
94 highly similar strains are linked by a recent transmission event (Roetzer et al., 2013; Walker
95 et al., 2013). Although many SNP cut-offs for linking isolates have been proposed (Hatherell

et al., 2016), the most commonly employed is based on the finding that a 5 SNP cut-off will cluster the genomes of strains from the majority of epidemiologically linked TB patients, with an upper bound of 12 SNPs between any two linked isolates (Walker et al., 2013). The widespread use of WGS has quickly pushed these cut-offs to be considered the new molecular gold standard of recent transmission linking, although SNP distances may vary for technical reasons (e.g. assembly pipelines or filter criteria (Guthrie & Gardy, 2017)) and between study populations e.g. high and low incidence settings (Bjorn-Mortensen et al., 2016).

In addition to cluster detection, uncovering the timing of transmission events within a given cluster is highly useful information for TB control e.g. for assessing the impact of interventions on the spread of an outbreak. Accordingly, knowledge of the rate change associated with different genotyping methods is essential for correct timing. The whole genome mutation rate of Mtbc strains has been estimated by several studies as between 10^{-7} and 10^{-8} substitutions per site per year or ~0.3-0.5 SNPs per genome per year (Bryant et al., 2013; Duchêne et al., 2016; Eldholm et al., 2015; Eldholm & Balloux, 2016; Roetzer et al., 2013; Walker et al., 2013) while the rate of change in the MIRU-VNTR loci specifically is known to be quicker (~10^{-3}) (Ragheb et al., 2013; Wirth et al., 2008). Since these mutation rates have been shown to also vary by lineage (Duchêne et al., 2016; Ford et al., 2013) and over short periods of time (Bryant et al., 2013), such variation needs to be accounted for, e.g. in Bayesian phylogenetic dating techniques (Bryant et al., 2013; Merker et al., 2015; Wirth et al., 2008).

Considering the multiple genotyping methods currently available, many of them proposed as a “gold standard”, there is an urgent need to precisely define the individual capacity of each method to accurately detect recent transmission events and perform timing of outbreaks. To

provide this essential information, this study harnesses the power of WGS-based phylogenetic dating methods to assign timespans onto Mtbc transmission chains encompassed by the different genotypic clustering methods commonly used in TB transmission studies.

Results

In this study, we assessed 20 different approaches for generating putative *M. tuberculosis* transmission clusters (see methods for approaches and naming schemes) using a dataset of 324 phenotypically rifampicin resistant isolates collected 2005-2010 from retreatment cases in Kinshasa, Democratic Republic of Congo (DRC). These 20 sets of clustering patterns were then characterised using whole genome sequence data and the propensity for convergence of clustering patterns was estimated (see methods). Bayesian phylodynamic approaches implemented in BEAST-2 (Bouckaert et al., 2014) were then utilised to assign timespans to the transmission events estimated by each genotyping method.

As expected, both the genome- and membrane-based spoligotyping approaches (named Gen-Spo and Mem-Spo respectively), clustered the most strains, with the lowest resolution (i.e. highest clustering rate) (Figure 1, Table 1). Convergent evolution (defined as the same pattern observed in unrelated strains; see methods) was found to affect 39% (12) of Mem-Spo clusters and 25% (7) of Gen-Spo clusters. Additionally, some discrepancies between the Mem-Spo and Gen-Spo patterns of each isolate were observed, with 291 isolates (90%) having the same pattern in both Mem-Spo and Gen-Spo approaches with 1 mismatch allowed (Supplementary table 1). The remaining 33 isolates mismatched with 2 to 17 spacers (average of 5 spacers). Although MIRU-VNTR performed far better than spoligotyping, 16% (6) of

144 clustering patterns were influenced by convergence in this study (see methods) (Table 1,
145 Figure 1). Mixed MIRU-VNTR patterns were observed in 18 isolates although this mixing was
146 not observed in the WGS data.
147
148 WGS-based methods had by far the highest discriminatory power and low SNP cut-offs
149 grouped isolates into smaller clusters (e.g. 2-10 isolates per cluster for a 5 SNP cut-off) (Table
150 1, Figure 1). When the clusters were expanded to better represent transmission chains using
151 the novel phylogenetic inclusion method implemented here (see methods), the resulting SNP
152 clusters often did not increase dramatically in size (Table 1). Discriminatory power and cluster
153 sizes based on cgMLST alleles were similar to the SNP-based clusters (Table 1, Figure 1).
154

Table 1: Clustering method overview.

For each clustering method, the general features are outlined in the tables. a) All clusters for each method affected by convergence. b) Clusters derived only from non-convergent patterns. c) SNP- and cgMLST-based methods Mean ages and 95% HPD ranges are based upon the BEAST2 estimates of clade mean heights.

a)

Method	**Strains in clusters**	**Number of clusters**	**Percent of strains in clusters**	**Cluster sizes**	**Maximum SNP distances**	**Clustering rate**
Gen-Spo	293	29	90.43	2-42	1-653	0.8148
GenSpo-MIRU	190	39	58.64	2-27	0-48	0.466
Gen-SpoNC	76	23	23.46	2-10	0-195	0.1636
Mem-Spo	276	33	85.19	2-39	1-685	0.75
MemSpo-MIRU	174	36	53.7	2-25	0-611	0.4259
Mem-SpoNC	64	18	19.75	2-10	0-21	0.142
MIRU-VNTR	207	38	63.89	2-30	0-611	0.5216
MIRU-NC	59	17	18.21	2-9	0-21	0.1296

b)

Method	Strains in clusters	Number of clusters	Percent of strains in clusters	Cluster sizes	Maximum SNP distances	Clustering rate	Mean Timespan	Timespan 95% HPD
Gen-Spo	191	22	58.95	2-37	1-322	0.5216	382.8101	0.96 - 1893.15
GenSpo-MIRU	77	22	23.77	2-10	0-48	0.1698	63.91188	0 - 278.77
Gen-SpoNC	34	11	10.49	2-6	0-14	0.071	21.52556	0.16 - 94.95
Mem-Spo	118	21	36.42	2-28	0-189	0.2994	141.1556	0.81 - 823.21
MemSpo-MIRU	50	12	15.43	2-10	2-48	0.1173	48.80688	0.8 - 216.31
Mem-SpoNC	15	5	4.63	2-4	0-14	0.0309	21.38239	1.03 - 97.91
MIRU-VNTR	121	32	37.35	2-11	0-48	0.2747	37.97812	0 - 162.27
MIRU-NC	25	9	7.72	2-3	1-11	0.0494	15.45935	0.77 - 58.38

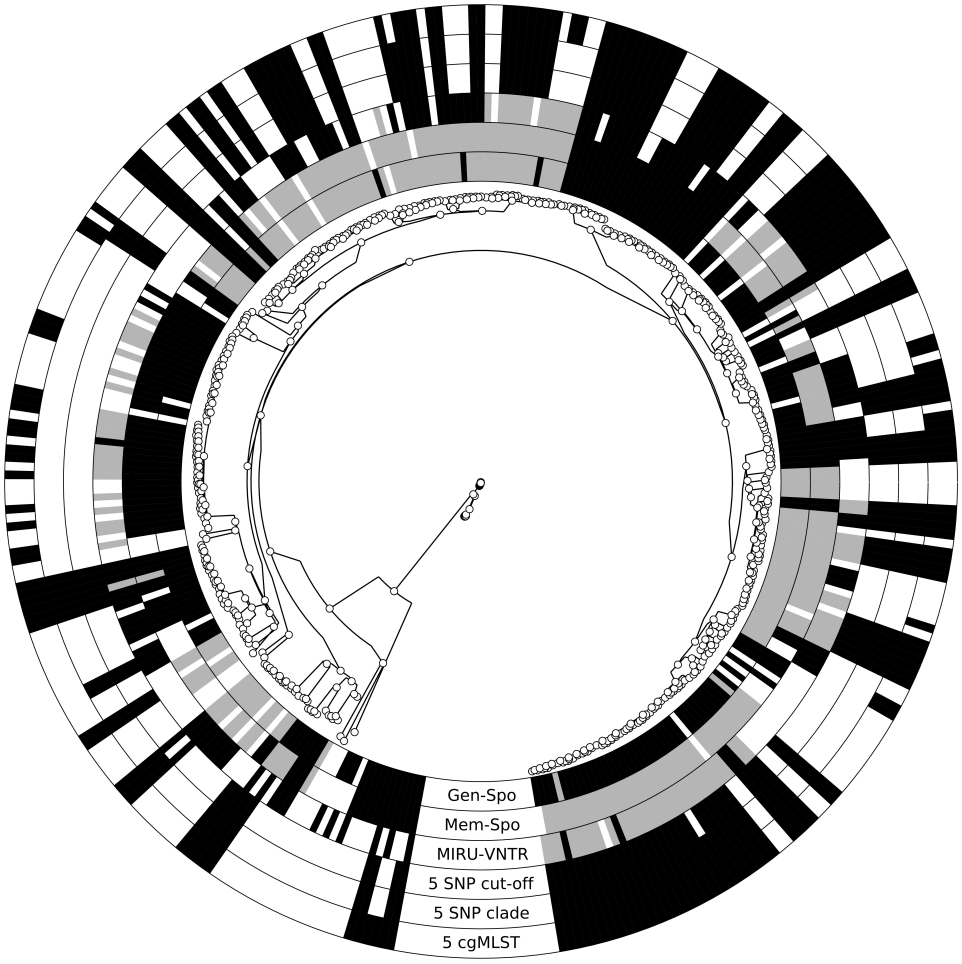
c)

Method	Strains in clusters	Number of clusters	Percent of strains in clusters	Cluster sizes	Maximum SNP distances	Clustering rate	Mean Timespan	Timespan 95% HPD
0 SNP cluster	54	25	16.67	2-4	0	0.0895	4.309937	0 - 15.9
1 SNP cluster	74	29	22.84	2-6	0-2	0.1389	5.698197	0 - 23.54
5 SNP cluster	147	40	45.37	2-27	0-10	0.3302	13.4115	0 - 47.07
12 SNP cluster	242	47	74.69	2-34	0-23	0.6019	28.95219	0 - 102.58
0 SNP clade	66	21	20.37	2-9	0-9	0.1389	5.746077	0 - 23.96
1 SNP clade	80	27	24.69	2-9	0-9	0.1636	6.104103	0 - 25.74
5 SNP clade	149	40	45.99	2-28	0-11	0.3364	13.48716	0 - 47.41
12 SNP clade	253	45	78.09	2-39	0-27	0.642	29.73941	0 - 104.64
0 allele cgMLST	51	24	15.74	2-4	0-1	0.0833	4.231405	0.03 - 15.48
1 allele cgMLST	80	31	24.69	2-6	0-4	0.1512	6.371668	0 - 24.65
5 allele cgMLST	173	42	53.4	2-28	0-22	0.4043	17.54352	0 - 68.53
12 allele cgMLST	254	45	78.4	2-39	0-51	0.6451	30.08732	0 - 112.25

159

160 Figure 1: Clustering of *M. tuberculosis* isolates.

161 For a representative approach of each of the main methods (Mem-Spo, Gen-Spo, MIRU-VNTR,
162 5 SNP cut-off, 5 SNP clade and 5 cgMLST) the inclusion of an isolate into a cluster is outlined
163 in the surrounding circles using GraPhlAn(Asnicar, Weingart, Tickle, Huttenhower, & Segata,
164 2015). If an isolate is in a cluster not affected by convergence, it is highlighted in black for the
165 given method. If an isolate is in a cluster affected by convergence, it is shown in grey. The
166 clustering based on all 20 approaches is shown in Supplementary Figure 1.



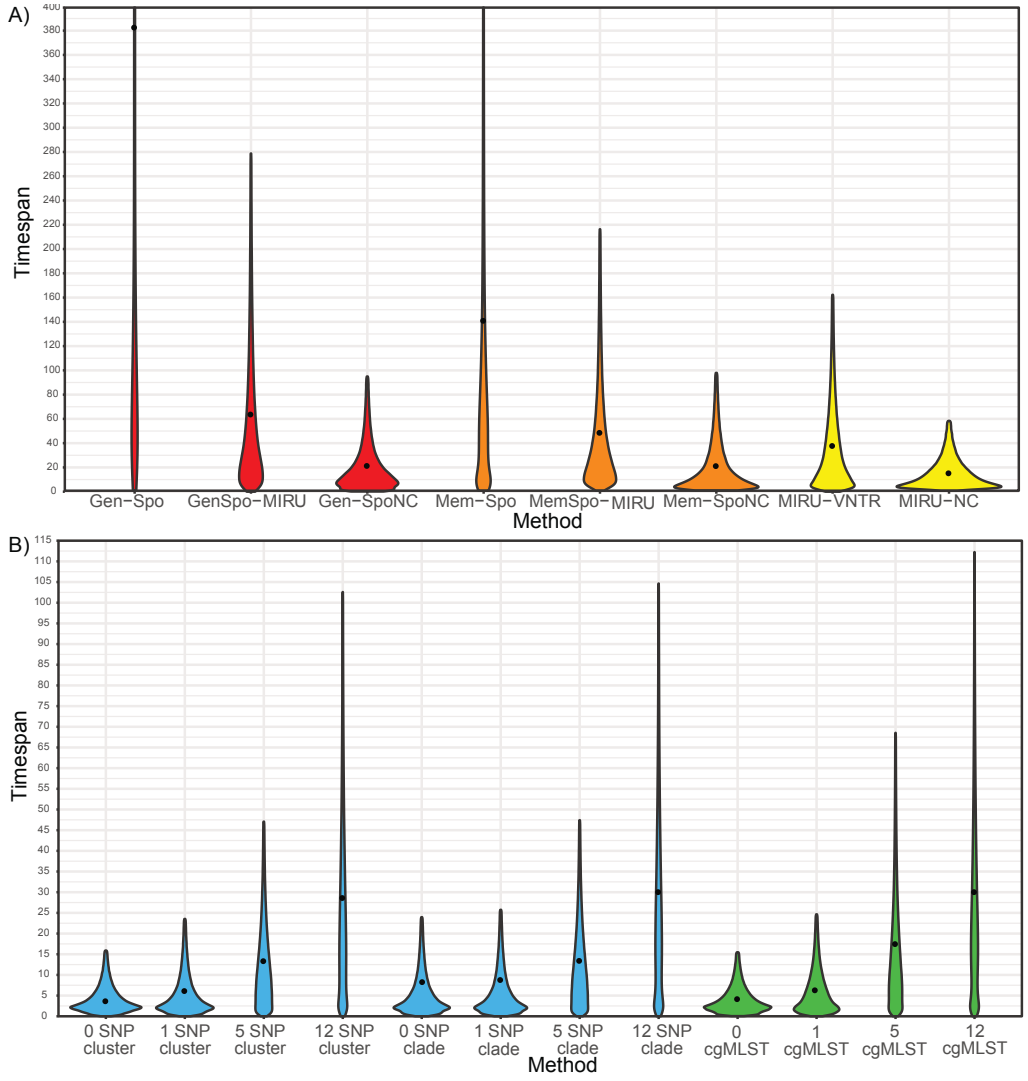
167

168 Statistical estimation of the timeframe associated with particular transmission chains showed
169 large differences in estimated cluster ages between the genotyping approaches used (Table
170 1, Figure 2), correlating well with the difference in discriminatory power. Cluster ages are
171 defined here as the most ancient transmission event that links any two isolates within a
172 specific cluster. Thus, in phylogenetic terms, the cluster age is the difference in time between
173 when the most recent common ancestor (MCRA) of the entire cluster existed and the date of
174 isolation of the furthest isolate from this ancestor. The aggregate mean ages of clusters
175 derived from spoligotyping approaches were found to often be several hundreds of years old
176 (Gen-Spo: 383 years ago (95% HPD: 1-1893); Mem-Spo: 141 years ago (95% HPD: 1-823))
177 (Table 1b, Figure 2a). The addition of MIRU-VNTR or *pncA* mutation data to spoligotyping
178 resulted in clusters that, on average, originated less than 100 years ago (Table 1b, Figure 2a).
179 MIRU-VNTR alone gave similar cluster ages as to when combined with spoligotyping (MIRU-
180 VNTR: 38 (0-162); GenSpo-MIRU: 64 (0-279); MemSpo-MIRU: 49 (1-216)) (Table 1b, Figure
181 2a).
182
183 Clusters based on SNP cut-offs correlated to 4 years of transmission using a 0 SNP cut-off (95%
184 HPD: 0-16), 6 years using a 1 SNP cut-off (95% HPD: 0-24), 13 years using a 5 SNP cut-off (95%
185 HPD: 0-47), and 29 years using a 12 SNP cut-off (95% HPD: 0-103) (Table 1c, Figure 2b).
186 Extension on the tree using the phylogenetic inclusion approach to form SNP clades did not
187 greatly increase the lengths of transmissions encompassed by clusters (one year increase, on
188 average) (Table 1c). Similar findings were obtained when clusters were based on allele
189 differences in the cgMLST method: 4 years of transmission using a 0 cgMLST cut-off (95% HPD:
190 0-15), 6 years using a 1 cgMLST cut-off (95% HPD: 0-25), 18 years using a 5 cgMLST cut-off

191 (95% HPD: 0-68), and 30 years using a 12 cgMLST cut-off (95% HPD: 0-112) (Table 1c, Figure
192 2b)

193 Figure 2: Timespans associated with transmission clusters

194 For each clustering method, the timespan associated with a cluster was estimated using
195 BEAST-2. The ages of each cluster (Y-axis) was aggregated per clustering method (X-axis).
196 Violin plots show the mean (black dot) for timespans along with the proportion of clusters
197 with a given age (coloured kernel plots). Methods are split as follows: A) Spoligotype-based
198 (Gen-Spo-based (red), Mem-Spo-based (orange)) and MIRU-VNTR-based (yellow), B) SNP-
199 based (blue) and cgMLST-based (green). Note the y-axis is different for each and panel A) is
200 cut at 400 years.



201

Discussion

The term 'recent transmission' is often applied to gain a better understanding of the current transmission dynamics of pathogens in a given population. However, little data is available on how recent a likely transmission event occurred when measured with different genotyping methods. To get a better understanding of the discriminatory power of different classical genotyping techniques and WGS-based approaches in relation to outbreak timing, this study has performed an in-depth comparison of clustering rates and dated phylogenies obtained in a collection of 324 Mtbc strains from a high incidence setting (Kinshasa, DRC). With a whole genome phylodynamic approach employed as a gold standard, our study demonstrates that each genotyping method was associated with a specific discriminatory power resulting in clusters representing vastly different time periods of transmission events (Table 1 and Figure 2). This has significant implications for data interpretations e.g. when selecting and utilising different genotyping methods/clustering approaches for epidemiological studies and assessing the effectiveness of public health intervention strategies.

As the most extreme example, spoligotyping-derived clusters were associated with transmission events that can be hundreds of years old. This low discriminatory power coupled with the high rate of convergent evolution (the same spoligotype pattern found in phylogenetically distant isolates) in both Mem-Spo and Gen-Spo add weight to the previous suggestion that these techniques are not suitable for recent transmission studies (Comas, Homolka, Niemann, & Gagneux, 2009), although Mem-Spo may be of use as a low-cost method of sorting Mtbc strains into the seven primary lineages (Filliol et al., 2006; Kato-Maeda et al., 2011). Differences between Mem-Spo and Gen-Spo patterns from the same isolate were observed for 10% of isolates in this study, even after rechecking of patterns,

226 requiring more investigation into which method is closer to the 'true' spoligotyping pattern
227 within a genome (Coll et al., 2012; Mokrousov et al., 2016; Warren et al., 2002; Xia, Teo, &
228 Ong, 2016).
229
230 In line with previous findings (Comas et al., 2009; Scott et al., 2005), convergent evolution of
231 24-loci MIRU-VNTR patterns was rarer than observed for spoligotyping, but did occur in 16%
232 of MIRU-VNTR-based clusters. Additionally, the transmission times encompassed by MIRU-
233 VNTR clusters spanned several decades (Table 1b, Figure 2a), confirming previous studies
234 showing over-estimation of recent transmission with this method (Bjorn-Mortensen et al.,
235 2016; Roetzer et al., 2013; Stucki et al., 2016; Wyllie et al., 2018).
236
237 The combination of MIRU-VNTR or spoligotyping with *pncA* mutations (MIRU-NC and Gen-
238 SpoNC/Mem-SpoNC) appeared to reflect true clusters of PZA resistance transmission based
239 on the relatively young ages of such transmission clusters (Table 1b). Thus, as discussed
240 before (den Hertog, Sengstake, & Anthony, 2015; Sengstake et al., 2017), although
241 transmission of *pncA* mutations seems to occur, further investigation is needed to find out
242 whether *pncA* mutants are less transmissible than those with a wildtype gene.
243
244 For defining transmission events that occurred in more recent time frames before sampling,
245 WGS-based methods (SNP or cgMLST) were found to be better suited than classical
246 genotyping methods (Table 1, Figure 2). The 12 SNP cut-off, currently the recommended
247 upper bound for clustering isolates, likely defines transmission events that occurred on
248 average three decades prior to sampling, similar in age to clusters estimated by MIRU-VNTR.
249 This suggests that the 12 SNP cluster method may be a good replacement for MIRU-VNTR as

it detects larger transmission networks spanning similar transmission time periods but is less affected by convergent evolution. Isolates clustered at identical (0 SNP) or nearly identical (1 SNP) cut-offs were found to represent transmission events occurring four to six years previous. These findings correlate well with previous studies where confirmed contact tracing-based epidemiological links were found between patients that were two (Walker et al., 2014), three (Roetzer et al., 2013) or five (Walker et al., 2013) SNPs apart. Indeed, a recent study of a cross-country MDR-TB outbreak found only a maximum of two SNP differences between all 29 isolates involved in the origin of the outbreak (Walker et al., 2018). Although this supports their use for detection or exclusion of very recent transmission, this low variability between isolates makes robust identification of transmission direction impossible, especially during short timespans.

Comparisons between the SNP-based (using almost all genomic differences) and the cgMLST-based cluster detection (using a defined core set of genes) demonstrated that the latter approach gives similar estimations to full SNP approaches. However, as current SNP assembly pipelines for Illumina data exclude repetitive region such as PE/PPE genes, larger differences between cgMLST and full SNP estimation may be seen once all aspects of the genome can be utilised.

Different clustering approaches can be applied when grouping isolates by SNP distance. Two partitional clustering methods are primarily utilised: either the creation of tight clusters (where the maximum pairwise distance between isolates in a cluster is less than the SNP cut-off; e.g. (Thomas A. Kohl et al., 2014)) or loose clusters (where each isolate is less than the SNP cut-off distance from at least one other isolate in the cluster; e.g. (Walker et al., 2014)).

Tight clusters ensure high connectivity within clusters, but may result in isolates belonging to multiple groups, making interpretation and delineation of transmission events difficult. Loose clusters (the definition used in this study), separate isolates into non-overlapping clusters, but may result in low connectivity within clusters. Here we present an extension of the loose cluster, termed the phylogenetic inclusion method, which adds all other isolates with the same phylogenetically defined common ancestor to the cluster, potentially identifying larger circulating genotypes. Tight, loose and phylogenetic inclusion clusters each aim to define different levels of connectivity through time, an aspect that should be considered when selecting the appropriate clustering approach.

The mutation rate of *M. tuberculosis* has been estimated to be between 10^{-7} and 10^{-8} substitutions per site per year (Duchêne et al., 2016; Merker et al., 2015; Roetzer et al., 2013). Within the Bayesian analysis employed here, the mutation rate was free to vary between these values but was found to strongly favour ~$3x10^{-8}$ (ESS > 1000 for all runs), translating to approximately 0.3 SNPs per genome per year. While the mutation rate used here is primarily applicable for lineage 4 (which most of this dataset is comprised of) and in line with previous estimates for this lineage (Duchêne et al., 2016), it may be similar in other lineages, although this has only been shown for lineage 2 (Duchêne et al., 2016; Merker et al., 2015). Thus, per-lineage estimates are required for all seven lineages to ensure similar transmission times are linked to genotyping methods across the whole population diversity of the Mtbc.

While this study has many advantages due to its five year population based design in an endemic setting coupled with the application of three different genotyping methods (membrane based spoligotyping analysis, 24-locus MIRU-VNTR and WGS), future

298 confirmatory studies could address the following drawbacks that are inherent to genomic
299 epidemiology (Comas, 2017; Guthrie & Gardy, 2017): 1) studies employing contact tracing
300 and/or digital epidemiology (Salathé et al., 2012) in conjunction with these genotyping
301 methods can help confirm transmission times associated with different clusters; 2) as outlined
302 above, strains of other lineages of the Mtbc should be analysed in a similar fashion to ensure
303 transferability of findings across the entire complex; 3) a broad range of drug resistance
304 profiles should be included to fully assess the impact of such mutations on transmission
305 estimates; 4) improved WGS methods, such as directly from clinical samples to help reduce
306 culture biases (Sanoussi, Affolabi, Rigouts, Anagonou, & de Jong, 2017) and longer reads (e.g.
307 PacBio SMRT or Nanopore MinION) to capture the entire genome, including repetitive regions
308 such as PE/PPE genes known to impact genome remodelling (Ates et al., 2018; Phelan et al.,
309 2016), will ensure that the maximum diversity between isolates is captured and 5)
310 standardised SNP calling pipelines appropriate across all lineages, with high true positive/low
311 false negative rates, will ensure that Mtbc molecular epidemiology can be uniformly
312 implemented and comparable across studies.
313
314 In conclusion, since each method was found to represent different timespans and clustering
315 definitions, they can be used in a stratified manner in an integrated epidemiological and
316 public health investigation addressing the transmission of Mtbc strains. For instance,
317 although spoligotyping clusters represented potentially very old transmission events, the low
318 associated cost and its ability to be applied directly on sputum helps reduce culture bias and
319 thus robustly assign lineages. Thus, spoligotyping and/or MIRU-VNTR would serve well as
320 first-line surveillance of potential transmission events in the population, guiding further
321 investigations and resource allocations.

These potential transmission hotspots could be further investigated with contact tracing and/or WGS. Employment of different cut-offs and clustering approaches to WGS data can then address several questions. The 12 SNP cluster/clade or 12 allele cgMLST approaches serve well for high level surveillance targeting larger (older) transmission networks, akin to what is currently often done using MIRU-VNTR (e.g. (Guthrie et al., 2018; Walker et al., 2018)). Recent transmission events can then be detected through employment of low SNP or cgMLST-based cut-offs (e.g. 5 SNPs for transmission in the past 15 years or 0-1 SNPs for transmission in the past 5 years). These clusters can then be linked to historical isolates or other clusters through employment of the phylogenetic inclusion method to resolve the local circulating genotypes. This is especially useful if bursts of sampling are undertaken such as in drug resistance surveys (WHO, 2015), which are increasingly employing WGS approaches (Cabibbe & Cirillo, 2016; Zignol et al., 2016, 2018). Alternatively, in high incidence/low diversity settings where amalgamation of clusters may inadvertently obscure distinct hotspots of transmission at different time points, subdivision into distinct time-dependant clusters can be undertaken using the algorithm presented in such a study in East Greenland (Bjorn-Mortensen et al., 2016).

Overall, phylodynamic approaches applied to whole genome sequences, as undertaken here, are recommended to fully investigate the specific transmission dynamics within a study population to account for setting-specific conditions, such as low/high TB incidence, low/high pathogen population diversity, sampling fractions and social factors influencing transmission. Thus, each genotyping method can be employed as part of an overall evidence gathering

345 program for transmission, placing molecular epidemiological approaches as an integral part
346 in tracking and stopping the spread of TB.
347
348 **Materials and Methods**
349 Dataset and sequencing
350 A set of 324 isolates from Kinshasa, Democratic Republic of Congo were collected from
351 consecutive retreatment TB patients between 2005 and 2010 at TB clinics, servicing an
352 estimated 30% of the population of Kinshasa. All isolates were phenotypically resistant to
353 rifampicin (RR-TB) and the majority are also isoniazid resistant (i.e. MDR-TB). Use of the
354 stored isolates without any linked personal information was approved by the health
355 authorities of the DRC and the Institutional Review Board of the ITM in Antwerp (ref no
356 945/14). Libraries for whole genome sequencing were prepared from extracted genomic DNA
357 with the Illumina Nextera XT kit, and run on the Illumina NextSeq platform in a 2x151bp run
358 according to manufacturer's instructions. Illumina read sets will be available at ReSeqTB
359 (platform.reseqtb.org) upon publication.
360
361 Genome reconstruction and maximum likelihood phylogeny estimation
362 The MTBseq pipeline(Thomas Andreas Kohl et al., 2018) was used to detect the SNPs for each
363 isolate using the H37Rv reference genome (NCBI accession number NC000962.3) (Lew,
364 Kapopoulou, Jones, & Cole, 2011; Médigue, Cole, Camus, & Pryor, 2002). Sites known to be
365 involved in drug resistance (as outlined in the PhyResSE list of drug mutations v27 (Feuerriegel
366 et al., 2015)) were excluded from the alignment and additional filtering of sites with
367 ambiguous calls in >5% of isolates and those SNPs within a 12bp window of each other was
368 also applied.

The SNP alignment of all isolates was used as the basis for creating a maximum likelihood (ML) phylogeny. RAxML-NG version 0.5.1b (Kozlov, 2017) was used to reconstruct the phylogeny from this alignment using a GTR+GAMMA model of evolution, accounting for ascertainment bias(Lewis, 2001) with the Stamatakis reconstituted DNA approach (Leaché et al., 2015) and site repeat optimisation (Kobert, Stamatakis, & Flouri, 2016) with 20 different starting trees and 100 bootstraps. All subsequent topology visualisation was undertaken using FigTree version 1.4.3 (Rambaut, 2016) and GraPhlAn (Asnicar et al., 2015).

Transmission cluster estimation methods

Several standard transmission clustering approaches were chosen for comparison and analysis. For each method, the total SNP distances were calculated to investigate the range of variability encompassed within each cluster. Maximum SNP distances were derived from pairwise comparisons of isolates within the SNP alignment using custom python scripts. A clustering rate was calculated for each method using the formula $(n_c - c)/n$, where n_c is the total number of isolates clustered by a given method, c is the number of clusters, and n is the total number of isolates in the dataset (n=324).

Spoligotyping

Spoligotype patterns were estimated by 2 methods: membrane-based and genome-based. Membrane-based patterns were obtained following the previously published protocol (Kamerbeek et al., 1997). This method is referred to as Mem-Spo. Genome-based spoligotyping was derived from the Illumina reads of each isolate using SpoTyping v2.1 (Xia et al., 2016). Reads (both forward and reverse) were input to SpoTyping with default

parameters and the 43 spacer values were extracted from the output. This method is referred to as Gen-Spo. For both methods, isolates were said to be clustered if all 43 spacers matched.

MIRU-VNTR

Genotyping by MIRU-VNTR was undertaken as previously described (Supply et al., 2006). 2 µl of DNA was extracted from cultures and amplified using the 24 loci MIRU-VNTR typing kit (Genoscreen, Lille, France). Analysis of patterns was undertaken using the ABI 3500 automatic sequencer (Applied Biosystems, California, USA) and Genemapper software (Applied Biosystems). Isolates were said to be clustered if all 24 loci matched. MIRU-VNTR patterns were also combined with spoligotyping patterns for additional refinement of clusters. Isolates were clustered if both the spoligotyping pattern and the 24 loci MIRU-VNTR pattern matched. These clustering methods are referred to as MemSpo-MIRU and GenSpo-MIRU.

SpoNC

Transmission estimation using spoligotyping has been shown to be improved if combined with *pncA* mutations (Said et al., 2016). This method, referred to as SpoNC, was applied to both Mem-Spo (Mem-SpoNC) and Gen-Spo (Gen-SpoNC). Mutations in *pncA* were extracted from the MTBseq tabular output for each isolate. All mutations were selected, regardless of drug resistance association, as is done in the SpoNC approach. The upstream promoter region of *pncA* did not reveal any mutations in this dataset. Isolates were said to be clustered if all 43 spacers matched and the *pncA* mutation was the same in both isolates. MIRU-VNTR patterns were combined with *pncA* mutations in a similar manner. This is referred to as MIRU-NC.

SNP cut-off clustering

The advent of whole genome reconstruction has allowed for genome-based comparisons for transmission clustering. Previous work has suggested that linked and recent transmission can be estimated by comparison of SNP differences between isolates. The cut-offs proposed by Walker *et al.* (Walker et al., 2013) are the most widely used and have been employed in multiple studies (Casali et al., 2016; Tessema et al., 2017; Witney et al., 2016). In this study, we employed both the 5 SNP (proposed by Walker *et al*. as the likely boundary for linked transmission) and 12 SNP cut-offs (proposed maximum boundary) for cluster definition. Additionally, we employed lower cut-offs of 0 and 1 SNPs to look for clusters of very highly related isolates. Pairwise SNP distances were calculated between all isolates. A loose cluster definition was used, where every isolate in a cluster at most the SNP cut-off from at least 1 other isolate in the cluster.

Phylogenetic information was used to extend these SNP-based clusters to include any other isolates that share the same most recent common ancestor (MRCA). These isolates may exceed the SNP cut-off but should be included as, through sharing an MRCA, they are intrinsically within the same putative transmission chain. The MRCA is defined here as the internal node in a phylogenetic tree that is shared by all the isolates within the putative SNP-based cluster. This extension was achieved by mapping each SNP cluster onto the ML phylogenetic tree and the MRCA (shared internal node) of all isolates was found using DendroPy v4.0.3 (Sukumaran & Holder, 2010). Any additional taxa with the same MRCA were then added to the transmission cluster (Supplemental Figure 2). In other words, all leaf nodes of the MRCA internal node were labelled as being part of the putative transmission cluster. We call this approach the phylogenetic inclusion method and extended clusters are hereafter

440 referred to as extended SNP clades to distinguish them from SNP clusters as created by the
441 standard non-phylogenetic method above. The python script that implements this method
442 can be found at https://github.com/conmeehan/pathophy.
443
444 cgMLST
445 An alternative approach to clustering using WGS data is the concept of core genome MLST
446 (cgMLST) patterns (Thomas A. Kohl et al., 2014). Since SNP detection can be variable between
447 assembly pipelines, SNP clusters between studies may be difficult to compare. The cgMLST
448 approach standardises comparisons by ensuring the same core genes are always compared.
449 BAM files for all isolates are input into Ridom SeqSphere^{+} software (Ridom GmbH, Münster,
450 Germany) to compile an allelic distance matrix based on the cgMLST v2 scheme consisting of
451 2,891 core Mtbc genes. Loose clusters were then defined as above using allelic differences of
452 0, 1, 5 and 12 as cut-offs. These methods are referred to as 0/1/5/12 cgMLST respectively.
453
454 Detection of convergent evolution
455 Convergent evolution towards identical patterns may occur for Spoligotyping, MIRU-VNTR
456 and *pncA* mutations (Driscoll, 2009; Miotto et al., 2014; Scott et al., 2005; Warren et al., 2002).
457 Convergence was detected and cross-checked with two methods. Firstly, Mtbc lineage and
458 sub-lineage numbering (Coll et al., 2014) was applied to all isolates based on the PhyResSE
459 lineage-defining SNP list v27 (Feuerriegel et al., 2015). If the same clustering pattern was
460 observed in two different sub-lineages, with other patterns seen in-between, this was flagged
461 as potential convergence. Additional convergence confirmation was also undertaken using
462 phylogenetic distances, as estimated by DendroPy. If the phylogenetic distance (combined
463 branch lengths that separate 2 isolates) between two isolates with identical clustering

464 patterns was greater than 0.0005, this was flagged as potential convergence. Any isolates
465 flagged by both methods (lineage-based and distance-based) were marked as clustered by
466 convergence. For example, if isolates with the same spoligotyping pattern appeared in lineage
467 4,1 and 4,6 with different patterns in-between and these isolates were distant on the tree
468 (distance greater than 0.0005), this was confirmed as a convergent pattern. Convergence was
469 checked for all approaches except the SNP cut-off clusters/clades, which, by definition, could
470 not be convergent. Clustering methods that combined two other methods (e.g. Gen-SpoNC)
471 were first checked separately for convergence and then combined to create the final clusters.
472
473 Estimation of transmission times
474 To estimate the age and timespan of potential transmission clusters, SNP alignments were
475 created from the convergence-free version of the five primary clustering types: Gen-Spo,
476 Mem-Spo, MIRU-VNTR, extended 12 SNP clades and 12 allele cgMLST. All other methods are
477 sub-clustering methods of at least one of these five methods (e.g. Mem-SpoNC clusters are
478 inherently included in any Mem-Spo clusters, and all SNP-based clusters are sub-clusters of
479 the 12 SNP clades).
480
481 A Bayesian approach to transmission time estimation was then undertaken. The SNP
482 alignments were created as above for the five high-level clustering types. Each cluster method
483 alignment was separately input to BEAST-2 v2.4.7 (Bouckaert et al., 2014) to create a time
484 tree for those isolates. These phylogenies were built using the following priors: GTR+GAMMA
485 substitution model, a log-normal relaxed molecular clock model to account for variation in
486 mutation rates (Alexei J Drummond, Ho, Phillips, Rambaut, & Rambaut, 2006) and coalescent
487 constant size demographic model (A. J. Drummond, Rambaut, Shapiro, & Pybus, 2005), both

488 of which have been found to be suitable for lineage 4 isolates in a previous study (Bjorn-
489 Mortensen et al., 2016). The MCMC chain was run six times independently per alignment with
490 a length of at least 400 million, sampled every 40,000th step (Gen-Spo: 400 million; extended
491 12 SNP & cgMLST: 500 million; MIRU & MemSpo: 600 million). A log normal prior (mean
492 1.5x10^{-7}; variance 1.0) was used for the clock model to reflect the previously estimated
493 mutation rate of *M. tuberculosis* lineage 4 (Bryant et al., 2013; Duchêne et al., 2016; Eldholm
494 et al., 2015; Eldholm & Balloux, 2016; Roetzer et al., 2013; Walker et al., 2013), while allowing
495 for variation as previously suggested (Bryant et al., 2013). A 1/X non-informative prior was
496 selected for the population size parameter of the demographic model. Isolation dates were
497 used as informative heterochronous tip dates and the SNP alignment was augmented with a
498 count of invariant sites for each of the four nucleotide bases to avoid ascertainment
499 bias(Leaché et al., 2015). Tracer v1.6 (Rambaut, Suchard, Xie, & Drummond, 2013) was used
500 to determine adequate mixing and convergence of chains (ESS >150) after a 25% burn-in. The
501 chains were combined via LogCombiner v2.4.8 (Bouckaert et al., 2014) to obtain a single chain
502 for each clustering type with high (>1000) effective sample sizes. The tree samples were
503 combined in the same manner and resampled at a lower frequency to create thinned samples
504 of (minimum) 20,000 trees.
505
506 The timespan of transmission events estimated by each method was then calculated as
507 follows: for each cluster created by the given method, we defined the MRCA node as the
508 internal node that connects all taxa in that cluster. The youngest node was then defined as
509 the tip that is furthest from this MRCA within the clade (i.e. the tip descendant from that node
510 that was sampled closest to the present time). For each retained tree in the MCMC process,
511 the difference in age between the MRCA node and youngest node was calculated. This gave

512 a distribution of likely maximum transmission event times within that cluster. For each
513 method, these per-cluster aggregated ages were then combined across all clusters to give a
514 per-method distribution of transmission event times represented by the clusters. The 95%
515 HPD interval of these distributions was calculated with the LaplacesDemon p.interval function
516 (Statisticat, 2016) in R v3.4.0 (R Core Team, 2017) and the distribution within this interval for
517 each method along with the mean based upon this interval were then visualized in violin plots
518 per clustering method using ggplot2 (Wickham, 2009) in R.

519 **Acknowledgements/Funding**

520 The authors would like to thank Armand Van Deun and Koen Vandelannoote for valuable
521 discussion and input and Cecile Uwizeye for aid with spoligotyping. This work was supported
522 by an ERC grant (INTERRUPTB; no. 311725) to BdJ, FG and CJM; an ERC grant to TS (PhyPD;
523 no. 335529); an FWO PhD fellowship to PM (grant number 1141217N); the German Centre
524 for Infection Research (DZIF) for TAK, MM and SN; a SNF SystemsX grant (TBX) to JP and TS
525 and a Marie Heim-Vögtlin fellowship granted to DK by the Swiss National Science Foundation.
526 The computational resources and services used in this work were provided by the VSC
527 (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and
528 the Flemish Government – department EWI;

529 **Competing interests**

530 The authors declare there are no competing interests attached to this work.

531

532 **References**

533 Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C., & Segata, N. (2015). Compact
534 graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ*, *3*,
535 e1029. https://doi.org/10.7717/peerj.1029

536 Ates, L. S., Dippenaar, A., Ummels, R., Piersma, S. R., van der Woude, A. D., van der Kuij, K.,
537 … Bitter, W. (2018). Mutations in ppe38 block PE_PGRS secretion and increase
538 virulence of Mycobacterium tuberculosis. *Nature Microbiology*, *3*(2), 181–188.
539 https://doi.org/10.1038/s41564-017-0090-6
540 Behr, M. A., Hopewell, P. C., Antonio Paz, E., Masae Kawamura, L., Schecter, G. F., & Small,
541 P. M. (1998). Predictive Value of Contact Investigation for Identifying Recent
542 Transmission of Mycobacterium tuberculosis. *American Journal of Respiratory and*
543 *Critical Care Medicine*, *158*(2), 465–469.
544 https://doi.org/10.1164/ajrccm.158.2.9801062
545 Bjorn-Mortensen, K., Lillebaek, T., Koch, A., Soborg, B., Ladefoged, K., Sørensen, H. C. F., …
546 Andersen, A. B. (2017). Extent of transmission captured by contact tracing in a
547 tuberculosis high endemic setting. *European Respiratory Journal*, *49*(3). Retrieved from
548 http://erj.ersjournals.com/content/49/3/1601851
549 Bjorn-Mortensen, K., Soborg, B., Koch, A., Ladefoged, K., Merker, M., Lillebaek, T., … Kohl, T.
550 A. (2016). Tracing Mycobacterium tuberculosis transmission by whole genome
551 sequencing in a high incidence setting: a retrospective population-based study in East
552 Greenland. *Scientific Reports*, *6*, 33180. https://doi.org/10.1038/srep33180
553 Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., … Drummond, A. J.
554 (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS*
555 *Computational Biology*, *10*(4), e1003537.
556 https://doi.org/10.1371/journal.pcbi.1003537
557 Bryant, J. M., Schürch, A. C., van Deutekom, H., Harris, S. R., de Beer, J. L., de Jager, V., … van
558 Soolingen, D. (2013). Inferring patient to patient transmission of Mycobacterium
559 tuberculosis from whole genome sequencing data. *BMC Infectious Diseases*, *13*(1), 110.

https://doi.org/10.1186/1471-2334-13-110

Cabibbe, A. M., & Cirillo, D. M. (2016). Best approaches to drug-resistance surveillance at the country level. *International Journal of Mycobacteriology*, *5*, S40–S41. https://doi.org/10.1016/j.ijmyco.2016.09.010

Casali, N., Broda, A., Harris, S. R., 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 Medicine*, *13*(10), e1002137. https://doi.org/10.1371/journal.pmed.1002137

Coll, F., Mallard, K., Preston, M. D., Bentley, S., Parkhill, J., McNerney, R., … Clark, T. G. (2012). SpolPred: rapid and accurate prediction of Mycobacterium tuberculosis spoligotypes from short genomic sequences. *Bioinformatics (Oxford, England)*, *28*(22), 2991–3. https://doi.org/10.1093/bioinformatics/bts544

Coll, F., McNerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigão, J., Viveiros, M., … Clark, T. G. (2014). A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. *Nature Communications*, *5*, 4812. https://doi.org/10.1038/ncomms5812

Comas, I. (2017). Genomic Epidemiology of Tuberculosis (pp. 79–93). Springer, Cham. https://doi.org/10.1007/978-3-319-64371-7_4

Comas, I., Homolka, S., Niemann, S., & Gagneux, S. (2009). Genotyping of Genetically Monomorphic Bacteria: DNA Sequencing in Mycobacterium tuberculosis Highlights the Limitations of Current Methodologies. *PLoS ONE*, *4*(11), e7815. https://doi.org/10.1371/journal.pone.0007815

den Hertog, A. L., Sengstake, S., & Anthony, R. M. (2015). Pyrazinamide resistance in Mycobacterium tuberculosis fails to bite? *Pathogens and Disease*, *73*(6), ftv037. https://doi.org/10.1093/femspd/ftv037

584 Diel, R., Schneider, S., Meywald-Walter, K., Ruf, C.-M., Rusch-Gerdes, S., & Niemann, S.
585 (2002). Epidemiology of Tuberculosis in Hamburg, Germany: Long-Term Population-
586 Based Analysis Applying Classical and Molecular Epidemiological Techniques. *Journal of*
587 *Clinical Microbiology*, *40*(2), 532–539. https://doi.org/10.1128/JCM.40.2.532-539.2002
588 Driscoll, J. R. (2009). Spoligotyping for Molecular Epidemiology of the Mycobacterium
589 tuberculosis Complex. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 551, pp. 117–
590 128). https://doi.org/10.1007/978-1-60327-999-4_10
591 Drummond, A. J., Ho, S. Y. W., Phillips, M. J., Rambaut, A., & Rambaut, A. (2006). Relaxed
592 Phylogenetics and Dating with Confidence. *PLoS Biology*, *4*(5), e88.
593 https://doi.org/10.1371/journal.pbio.0040088
594 Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian Coalescent
595 Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology*
596 *and Evolution*, *22*(5), 1185–1192. https://doi.org/10.1093/molbev/msi103
597 Duchêne, S., Holt, K. E., Weill, F.-X., Le Hello, S., Hawkey, J., Edwards, D. J., … Holmes, E. C.
598 (2016). Genome-scale rates of evolutionary change in bacteria. *Microbial Genomics*,
599 *2*(11), e000094. https://doi.org/10.1099/mgen.0.000094
600 Eldholm, V., & Balloux, F. (2016). Antimicrobial Resistance in Mycobacterium tuberculosis:
601 The Odd One Out. *Trends in Microbiology*. https://doi.org/10.1016/j.tim.2016.03.007
602 Eldholm, V., Monteserin, J., Rieux, A., Lopez, B., Sobkowiak, B., Ritacco, V., … Lemey, P.
603 (2015). Four decades of transmission of a multidrug-resistant Mycobacterium
604 tuberculosis outbreak strain. *Nature Communications*, *6*, 7119.
605 https://doi.org/10.1038/ncomms8119
606 Feuerriegel, S., Schleusener, V., Beckert, P., Kohl, T. A., Miotto, P., Cirillo, D. M., …
607 Fellenberg, K. (2015). PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis

608 Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. *Journal of*
609 *Clinical Microbiology*, *53*(6), 1908–1914. https://doi.org/10.1128/JCM.00025-15
610 Filliol, I., Motiwala, A. S., Cavatore, M., Qi, W., Hazbón, M. H., Bobadilla del Valle, M., …
611 Alland, D. (2006). Global phylogeny of Mycobacterium tuberculosis based on single
612 nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution,
613 phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for
614 a minimal standard SNP set. *Journal of Bacteriology*, *188*(2), 759–72.
615 https://doi.org/10.1128/JB.188.2.759-772.2006
616 Ford, C. B., Shah, R. R., Maeda, M. K., Gagneux, S., Murray, M. B., Cohen, T., … Fortune, S. M.
617 (2013). Mycobacterium tuberculosis mutation rate estimates from different lineages
618 predict substantial differences in the emergence of drug-resistant tuberculosis. *Nature*
619 *Genetics*, *45*(7), 784–90. https://doi.org/10.1038/ng.2656
620 Fox, G. J., Barry, S. E., Britton, W. J., & Marks, G. B. (2012). Contact investigation for
621 tuberculosis: a systematic review and meta-analysis. *European Respiratory Journal*,
622 *41*(1). Retrieved from http://erj.ersjournals.com/content/41/1/140.short
623 Gardy, J. L., Johnston, J. C., Sui, S. J. H., Cook, V. J., Shah, L., Brodkin, E., … Tang, P. (2011).
624 Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak.
625 *New England Journal of Medicine*, *364*(8), 730–739.
626 https://doi.org/10.1056/NEJMoa1003176
627 Goguet de la Salmonière, Y. O., Li, H. M., Torrea, G., Bunschoten, A., van Embden, J., &
628 Gicquel, B. (1997). Evaluation of spoligotyping in a study of the transmission of
629 Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, *35*(9), 2210–4. Retrieved
630 from http://www.ncbi.nlm.nih.gov/pubmed/9276389
631 Guernier, V., Sola, C., Brudey, K., Guégan, J.-F., & Rastogi, N. (2008). Use of cluster-graphs

632 from spoligotyping data to study genotype similarities and a comparison of three
633 indices to quantify recent tuberculosis transmission among culture positive cases in
634 French Guiana during a eight year period. *BMC Infectious Diseases*, *8*(1), 46.
635 https://doi.org/10.1186/1471-2334-8-46
636 Guthrie, J. L., & Gardy, J. L. (2017). A brief primer on genomic epidemiology: lessons learned
637 from *Mycobacterium tuberculosis*. *Annals of the New York Academy of Sciences*,
638 *1388*(1), 59–77. https://doi.org/10.1111/nyas.13273
639 Guthrie, J. L., Kong, C., Roth, D., Jorgensen, D., Rodrigues, M., Hoang, L., … Gardy, J. L.
640 (2018). Molecular Epidemiology of Tuberculosis in British Columbia, Canada: A 10-Year
641 Retrospective Study. *Clinical Infectious Diseases*, *66*(6), 849–856.
642 https://doi.org/10.1093/cid/cix906
643 Hatherell, H.-A., Colijn, C., Stagg, H. R., Jackson, C., Winter, J. R., & Abubakar, I. (2016).
644 Interpreting whole genome sequencing for investigating tuberculosis transmission: a
645 systematic review. *BMC Medicine*, *14*(1), 21. https://doi.org/10.1186/s12916-016-
646 0566-x
647 Jonsson, J., Hoffner, S., Berggren, I., Bruchfeld, J., Ghebremichael, S., Pennhag, A., &
648 Groenheit, R. (2014). Comparison between RFLP and MIRU-VNTR Genotyping of
649 Mycobacterium tuberculosis Strains Isolated in Stockholm 2009 to 2011. *PLoS ONE*,
650 *9*(4), e95159. https://doi.org/10.1371/journal.pone.0095159
651 Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., … van
652 Embden, J. (1997). Simultaneous detection and strain differentiation of Mycobacterium
653 tuberculosis for diagnosis and epidemiology. *Journal of Clinical Microbiology*, *35*(4),
654 907–14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9157152
655 Kato-Maeda, M., Gagneux, S., Flores, L. L., Kim, E. Y., Small, P. M., Desmond, E. P., &

Hopewell, P. C. (2011). Strain classification of Mycobacterium tuberculosis: congruence between large sequence polymorphisms and spoligotypes. *The International Journal of Tuberculosis and Lung Disease : The Official Journal of the International Union against Tuberculosis and Lung Disease*, *15*(1), 131–3. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21276309

Kendall, E. A., Fofana, M. O., Dowdy, D. D. W., WHO, Dye, C., Garnett, G., … al., et. (2015). Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis. *The Lancet Respiratory Medicine*, *3*(12), 963–972. https://doi.org/10.1016/S2213-2600(15)00458-0

Klopper, M., Warren, R. M., Hayes, C., Gey van Pittius, N. C., Streicher, E. M., Müller, B., … Trollip, A. P. (2013). Emergence and Spread of Extensively and Totally Drug-Resistant Tuberculosis, South Africa. *Emerging Infectious Diseases*, *19*(3), 449–455. https://doi.org/10.3201/eid1903.120246

Kobert, K., Stamatakis, A., & Flouri, T. (2016). Efficient Detection of Repeating Sites to Accelerate Phylogenetic Likelihood Calculations. *Systematic Biology*, *66*(2), syw075. https://doi.org/10.1093/sysbio/syw075

Kohl, T. A., Diel, R., Harmsen, D., Rothgänger, J., Walter, K. M., Merker, M., … Niemann, S. (2014). Whole-genome-based Mycobacterium tuberculosis surveillance: a standardized, portable, and expandable approach. *Journal of Clinical Microbiology*, *52*(7), 2479–86. https://doi.org/10.1128/JCM.00567-14

Kohl, T. A., Utpatel, C., Schleusener, V., De Filippo, M. R., Beckert, P., Cirillo, D. M., & Niemann, S. (2018). MTBseq: A comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates. *In Preparation*.

Kozlov, A. (2017). RAxML-NG.

https://doi.org/https://zenodo.org/record/888146#.Wm8r2Uso9TY

Kurbatova, E. V., Cavanaugh, J. S., Dalton, T., S. Click, E., & Cegielski, J. P. (2013). Epidemiology of Pyrazinamide-Resistant Tuberculosis in the United States, 1999–2009. *Clinical Infectious Diseases*, *57*(8), 1081–1093. https://doi.org/10.1093/cid/cit452

Leaché, A. D., Banbury, B. L., Felsenstein, J., de Oca, A. nieto-M., Stamatakis, A., K., S., … J., W. (2015). Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic Biology*, *64*(6), 1032–1047. https://doi.org/10.1093/sysbio/syv053

Lew, J. M., Kapopoulou, A., Jones, L. M., & Cole, S. T. (2011). TubercuList – 10 years after. *Tuberculosis*, *91*(1), 1–7. https://doi.org/10.1016/j.tube.2010.09.008

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, *50*(6), 913–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12116640

Médigue, C., Cole, S. T., Camus, J.-C., & Pryor, M. J. (2002). Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. *Microbiology*, *148*(10), 2967–2973. https://doi.org/10.1099/00221287-148-10-2967

Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., … Wirth, T. (2015). Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nature Genetics*, *47*(3), 242–249. https://doi.org/10.1038/ng.3195

Merker, M., Kohl, T. A., Niemann, S., & Supply, P. (2017). The Evolution of Strain Typing in the Mycobacterium tuberculosis Complex. In *Advances in experimental medicine and biology* (Vol. 1019, pp. 43–78). https://doi.org/10.1007/978-3-319-64371-7_3

Miotto, P., Cabibbe, A. M., Feuerriegel, S., Casali, N., Drobniewski, F., Rodionova, Y., … Cirillo, D. M. (2014). Mycobacterium tuberculosis pyrazinamide resistance

704 determinants: a multicenter study. *MBio*, *5*(5), e01819-14.
705 https://doi.org/10.1128/mBio.01819-14
706 Mokrousov, I., Chernyaeva, E., Vyazovaya, A., Sinkov, V., Zhuravlev, V., & Narvskaya, O.
707 (2016). Next-Generation Sequencing of Mycobacterium tuberculosis. *Emerging*
708 *Infectious Diseases*, *22*(6), 1127–9. https://doi.org/10.3201/eid2206.152051
709 Ngabonziza, J. C. S., Diallo, A. B., Tagliani, E., Diarra, B., Kadanga, A. E., Togo, A. C. G., …
710 group", the "Union short M.-T. regimen study. (2017). Half of rifampicin-resistant
711 Mycobacterium tuberculosis complex isolated from tuberculosis patients in Sub-
712 Saharan Africa have concomitant resistance to pyrazinamide. *PLOS ONE*, *12*(10),
713 e0187211. https://doi.org/10.1371/journal.pone.0187211
714 Phelan, J. E., Coll, F., Bergval, I., Anthony, R. M., Warren, R., Sampson, S. L., … Clark, T. G.
715 (2016). Recombination in pe/ppe genes contributes to genetic variation in
716 Mycobacterium tuberculosis lineages. *BMC Genomics*, *17*, 151.
717 https://doi.org/10.1186/s12864-016-2467-y
718 R Core Team. (2017). R: A language and environment for statistical computing. R Foundation
719 for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org/
720 Ragheb, M. N., Ford, C. B., Chase, M. R., Lin, P. L., Flynn, J. L., & Fortune, S. M. (2013). The
721 mutation rate of mycobacterial repetitive unit loci in strains of M. tuberculosis from
722 cynomolgus macaque infection. *BMC Genomics*, *14*(1), 145.
723 https://doi.org/10.1186/1471-2164-14-145
724 Rambaut, A. (2016). Figtree. Retrieved May 3, 2017, from
725 http://tree.bio.ed.ac.uk/software/figtree/
726 Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. (2013). Tracer. Retrieved December
727 11, 2013, from http://beast.bio.ed.ac.uk/Tracer

728 Roetzer, A., Diel, R., Kohl, T. A., Rückert, C., Nübel, U., Blom, J., … Niemann, S. (2013). Whole
729 genome sequencing versus traditional genotyping for investigation of a Mycobacterium
730 tuberculosis outbreak: a longitudinal molecular epidemiological study. *PLoS Medicine*,
731 *10*(2), e1001387. https://doi.org/10.1371/journal.pmed.1001387
732 Roetzer, A., Schuback, S., Diel, R., Gasau, F., Ubben, T., di Nauta, A., … Niemann, S. (2011).
733 Evaluation of Mycobacterium tuberculosis typing methods in a 4-year study in
734 Schleswig-Holstein, Northern Germany. *Journal of Clinical Microbiology*, *49*(12), 4173–
735 8. https://doi.org/10.1128/JCM.05293-11
736 Said, H. M., Kushner, N., Omar, S. V, Dreyer, A. W., Koornhof, H., Erasmus, L., … Ismail, N. A.
737 (2016). A Novel Molecular Strategy for Surveillance of Multidrug Resistant Tuberculosis
738 in High Burden Settings. *PloS One*, *11*(1), e0146106.
739 https://doi.org/10.1371/journal.pone.0146106
740 Salathé, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., …
741 Vespignani, A. (2012). Digital Epidemiology. *PLoS Computational Biology*, *8*(7),
742 e1002616. https://doi.org/10.1371/journal.pcbi.1002616
743 Sanoussi, C. N., Affolabi, D., Rigouts, L., Anagonou, S., & de Jong, B. (2017). Genotypic
744 characterization directly applied to sputum improves the detection of Mycobacterium
745 africanum West African 1, under-represented in positive cultures. *PLOS Neglected*
746 *Tropical Diseases*, *11*(9), e0005900. https://doi.org/10.1371/journal.pntd.0005900
747 Scott, A. N., Menzies, D., Tannenbaum, T.-N., Thibert, L., Kozak, R., Joseph, L., … Behr, M. A.
748 (2005). Sensitivities and specificities of spoligotyping and mycobacterial interspersed
749 repetitive unit-variable-number tandem repeat typing methods for studying molecular
750 epidemiology of tuberculosis. *Journal of Clinical Microbiology*, *43*(1), 89–94.
751 https://doi.org/10.1128/JCM.43.1.89-94.2005

752 Sengstake, S., Bergval, I. L., Schuitema, A. R., Beer, J. L., Phelan, J., Zwaan, R., … Anthony, R.
753 M. (2017). Pyrazinamide resistance-conferring mutations in pncA and the transmission
754 of multidrug resistant TB in Georgia. *BMC Infectious Diseases 2017 17:1*, *17*(1), 491.
755 https://doi.org/10.1186/s12879-017-2594-3
756 Shah, N. S., Auld, S. C., Brust, J. C. M., Mathema, B., Ismail, N., Moodley, P., … Gandhi, N. R.
757 (2017). Transmission of Extensively Drug-Resistant Tuberculosis in South Africa. *New*
758 *England Journal of Medicine*, *376*(3), 243–253.
759 https://doi.org/10.1056/NEJMoa1604544
760 Small, P. M., Hopewell, P. C., Singh, S. P., Paz, A., Parsonnet, J., Ruston, D. C., … Schoolnik, G.
761 K. (1994). The Epidemiology of Tuberculosis in San Francisco -- A Population-Based
762 Study Using Conventional and Molecular Methods. *New England Journal of Medicine*,
763 *330*(24), 1703–1709. https://doi.org/10.1056/NEJM199406163302402
764 Statisticat. (2016). LaplacesDemon: Complete Environment for Bayesian Inference.
765 Bayesian-Inference.com. R package version 16.0.1. Retrieved from
766 https://web.archive.org/web/20150206004624/http://www.bayesian-
767 inference.com/software
768 Stucki, D., Ballif, M., Egger, M., Furrer, H., Altpeter, E., Battegay, M., … Fenner, L. (2016).
769 Standard Genotyping Overestimates Transmission of Mycobacterium tuberculosis
770 among Immigrants in a Low-Incidence Country. *Journal of Clinical Microbiology*, *54*(7),
771 1862–70. https://doi.org/10.1128/JCM.00126-16
772 Sukumaran, J., & Holder, M. T. (2010). DendroPy: a Python library for phylogenetic
773 computing. *Bioinformatics*, *26*(12), 1569–1571.
774 https://doi.org/10.1093/bioinformatics/btq228
775 Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rusch-Gerdes, S., Willery, E., … van

776 Soolingen, D. (2006). Proposal for Standardization of Optimized Mycobacterial
777 Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of
778 Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, *44*(12), 4498–4510.
779 https://doi.org/10.1128/JCM.01392-06
780 Supply, P., Magdalena, J., Himpens, S., & Locht, C. (1997). Identification of novel intergenic
781 repetitive units in a mycobacterial two-component system operon. *Molecular*
782 *Microbiology*, *26*(5), 991–1003. Retrieved from
783 http://www.ncbi.nlm.nih.gov/pubmed/9426136
784 Tessema, B., Nabeta, P., Valli, E., Albertini, A., Collantes, J., Lan, N. H., … Dolinger, D. L.
785 (2017). FIND Tuberculosis Strain Bank: a Resource for Researchers and Developers
786 Working on Tests To Detect Mycobacterium tuberculosis and Related Drug Resistance.
787 *Journal of Clinical Microbiology*, *55*(4), 1066–1073.
788 https://doi.org/10.1128/JCM.01662-16
789 Thierry, D., Cave, M. D., Eisenach, K. D., Crawford, J. T., Bates, J. H., Gicquel, B., & Guesdon,
790 J. L. (1990). IS6110, an IS-like element of Mycobacterium tuberculosis complex. *Nucleic*
791 *Acids Research*, *18*(1), 188. Retrieved from
792 http://www.ncbi.nlm.nih.gov/pubmed/2155396
793 Trauer, J. M., Denholm, J. T., & McBryde, E. S. (2014). Construction of a mathematical model
794 for tuberculosis transmission in highly endemic regions of the Asia-pacific. *Journal of*
795 *Theoretical Biology*, *358*, 74–84. https://doi.org/10.1016/j.jtbi.2014.05.023
796 van Embden, J. D., Cave, M. D., Crawford, J. T., Dale, J. W., Eisenach, K. D., Gicquel, B., … al.,
797 et. (1993). Strain identification of Mycobacterium tuberculosis by DNA fingerprinting:
798 recommendations for a standardized methodology. *Journal of Clinical Microbiology*,
799 *31*(2), 406–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8381814

Vluggen, C., Soetaert, K., Groenen, G., Wanlin, M., Spitaels, M., Arrazola de Oñate, W., … Mathys, V. (2017). Molecular epidemiology of Mycobacterium tuberculosis complex in Brussels, 2010–2013. *PLOS ONE*, *12*(2), e0172554. https://doi.org/10.1371/journal.pone.0172554

Walker, T. M., Ip, C. L. C., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., … Peto, T. E. A. (2013). Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. *The Lancet Infectious Diseases*, *13*(2), 137–46. https://doi.org/10.1016/S1473-3099(12)70277-3

Walker, T. M., Lalor, M. K., Broda, A., Ortega, L. S., Morgan, M., Parker, L., … Conlon, C. P. (2014). Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. *The Lancet. Respiratory Medicine*, *2*(4), 285–292. https://doi.org/10.1016/S2213-2600(14)70027-X

Walker, T. M., Merker, M., Knoblauch, A. M., Helbling, P., Schoch, O. D., van der Werf, M. J., … Witschi, M. (2018). A cluster of multidrug-resistant Mycobacterium tuberculosis among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(18)30004-5

Warren, R. M., Streicher, E. M., Sampson, S. L., van der Spuy, G. D., Richardson, M., Nguyen, D., … van Helden, P. D. (2002). Microevolution of the direct repeat region of Mycobacterium tuberculosis: implications for interpretation of spoligotyping data. *Journal of Clinical Microbiology*, *40*(12), 4457–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12454136

WHO. (2015). *Guidelines for surveillance of drug resistance in tuberculosis* (5th ed.). Geneva.

WHO. (2018). *Global tuberculosis report 2017*. Retrieved from http://www.who.int/tb/publications/global_report/archive/en/

824 Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
825 Retrieved from http://ggplot2.org
826 Wirth, T., Hildebrand, F., Allix-Béguec, C., Wölbeling, F., Kubica, T., Kremer, K., … Niemann,
827 S. (2008). Origin, Spread and Demography of the Mycobacterium tuberculosis Complex.
828 *PLoS Pathogens*, *4*(9), e1000160. https://doi.org/10.1371/journal.ppat.1000160
829 Witney, A. A., Cosgrove, C. A., Arnold, A., Hinds, J., Stoker, N. G., & Butcher, P. D. (2016).
830 Clinical use of whole genome sequencing for Mycobacterium tuberculosis. *BMC*
831 *Medicine*, *14*(1), 46. https://doi.org/10.1186/s12916-016-0598-2
832 Wyllie, D., Davidson, J., Walker, T., Rathod, P., Peto, T., Robinson, E., … Campbell, C. (2018).
833 A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for
834 identifying Mycobacterium tuberculosis transmission: A prospective observational
835 cohort study. *BioRxiv*, 252734. https://doi.org/10.1101/252734
836 Xia, E., Teo, Y.-Y., & Ong, R. T.-H. (2016). SpoTyping: fast and accurate in silico
837 Mycobacterium spoligotyping from sequence reads. *Genome Medicine*, *8*(1), 19.
838 https://doi.org/10.1186/s13073-016-0270-7
839 Xu, P., Wu, J., Yang, C., Luo, T., Shen, X., Zhang, Y., … Gao, Q. (2016). Prevalence and
840 transmission of pyrazinamide resistant Mycobacterium tuberculosis in China.
841 *Tuberculosis*, *98*, 56–61. https://doi.org/10.1016/j.tube.2016.02.008
842 Zignol, M., Cabibbe, A. M., Dean, A. S., Glaziou, P., Alikhanova, N., Ama, C., … Raviglione, M.
843 C. (2018). Genetic sequencing for surveillance of drug resistance in tuberculosis in
844 highly endemic countries: a multi-country population-based surveillance study. *The*
845 *Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(18)30073-2
846 Zignol, M., Dean, A. S., Alikhanova, N., Andres, S., Cabibbe, A. M. A., Cirillo, D. M., … Valafar,
847 F. (2016). Population-based resistance of Mycobacterium tuberculosis isolates to

848 pyrazinamide and fluoroquinolones: results from a multicountry surveillance project.

849 *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(16)30190-6

850

851 Supplementary legends

852 Supplementary Table 1: Spoligotyping patterns derived from membrane- and genome-based

853 methods

854 Spoligotyping patterns were estimated from both membrane (Mem-Spo) and genome (Gen-

855 Spo) approaches. Patterns are shown from both methods per isolates with the number of

856 mismatches between patterns recorded.

857

Isolate	Mem-Spoligotype	Gen-Spoligotype	Mismatches
DRC-052577	11111110000011111111110100000000000010001111	11111110000011111111110100000000000110001111	1
DRC-052578	11111110000000000000000001111111000001111111	11111100000000000000000001111111000001111111	0
DRC-052654	11111111111111111110000011111111000001111111	11011111111111111111011111110111000001101111	7
DRC-052667	11111111000111111110000000000000000000001100	11111111111100001110000000000000000000111100	9
DRC-052750	11111110000011111111101100000000000010001111	11111110000011111111101000000000000110001111	3
DRC-052956	11111111111111111111111111111111100001110111	11111111111111111111111111111111000011101111	0
DRC-052958	11111111111111111111111111011110000111011 1	11111111111111111111111111011110000011101111	0
DRC-052959	11111111111111111111111111111111000011111111	11111111111111111111111111111111000001111111	1
DRC-052964	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-060116	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-060118	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-060120	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-060131	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-060237	11111111111111111111111111111111000011111111	11111111111111111111111111111111000011111111	0
DRC-060244	00000000000000000000000000111111000011111111	00000000000000000000000011111111000011101111	3
DRC-060248	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-060253	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-060260	11111100000000000000000001111111000010111111	11111100000000000000000001111111001011111111	2
DRC-060360	11111010000011111111111110000000000010000011	11111110000011111111111110000000000110000011	2
DRC-060361	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-060362	11111110000011111111110110000000000010001111	11111110000011111111110100000000000110001111	2
DRC-060366	11111100000000000000000001111111000010111111	11111100000000000000000001111111000011111111	1
DRC-060375	11000000111111101111000011111111000011111111	11000000111111101111000011111111000011111111	2
DRC-060376	11011110001111111111000011111111000000011111	11011111111111111111000011111111000011111111	5

DRC-060381	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-060382	11111100000000000000000001111111000001111111	11111100000000000000000001111111000001111111	0
DRC-060514	11000000111111110111000011111111000001111111	11000000111111110111000011111111000001111111	0
DRC-060526	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-060527	11111111111100001110000111111111000001011111	11111111111100001110000111111111000001011111	0
DRC-060528	11111100000000000000000001111111000001011111	11111100000000000000000001111111000001111111	1
DRC-060549	10000011111111111110111111111111000001110111	10000011111111111111111111111111000001111111	2
DRC-060562	11111111111111111111111111111111000001000011	11111111111111111111111111111111000001000011	0
DRC-060565	11111111111111111111111111111111000001110111	11111111111111111111111111111111000001110111	0
DRC-060756	11111100000000000000000001111111000001111111	11111100000000000000000001111111000001111111	0
DRC-060757	11111111111111111111111111011110000001110111	11111111111111111111111111011110000001110111	0
DRC-060772	11111110000011111111110100000000000010001111	11111110000011111111110100000000000110001111	1
DRC-060778	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-060856	11111111111100001110000000000000000000111100	11111111111100001110000000000000000000111100	0
DRC-060858	11111111111111111111000011111111000001011111	11111111111111111111111111011110000001110111	7
DRC-060859	11111111111111111111000011111111000001111111	11111111111111111111000011111111000001111111	0
DRC-061069	11111110000011111111111100000000000010001111	11111110000011111111111100000000000110001111	1
DRC-061123	11111111111111111111000011111111000001111111	11111111111111111111000011111111000001111111	0
DRC-061124	11011111111111111111000011111111000001111111	11111111111111111111000011111111000001111111	1
DRC-061127	11111110000000000000000001111111000001111111	11111100000000000000000001111111000001111111	0
DRC-061136	11111100000000000000000001111111000001011111	11111100000000000000000001111111000001111111	1
DRC-061137	11111100000000000000000001111111000001011111	11111100000000000000000001111111000001111111	1
DRC-061138	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-061142	11111111111111111111000011111111000001111111	11111111111111111111000011111111000001111111	0
DRC-061431	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-061493	11111111111111111111111111111111000001110111	11111111101111111111111111111111000001110111	1
DRC-062295	11111111111111000001111111111111000001110111	11111111111111000001111111111111000001110111	0
DRC-062298	11111111111100001110000000000000000000111100	11111111111100001110000000000000000000111100	0
DRC-062303	11000000111111101111000011111111000001111111	11000000111111101111000011111111000001111111	2
DRC-062444	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-062446	11111111111111111111000011110111000001111111	11111111111111111111000011110111000001111111	0
DRC-062447	11000000111111101111000011111111000001111111	11000000111111101111000011111111000001111111	0
DRC-062495	11111111111111111111111111111111000001111111	11111111111111111111111111111111000001111111	0
DRC-062502	11111111111111111111000011111111000001011111	11111111111111111111000011111111000001011111	0
DRC-062503	11111111111111111111111111111111000001111111	11111111111111111111111111111111000001110111	1

DRC-062822	110111111111111111110000111111100001111111	110111111111111111110000111111100001111111	0
DRC-062827	110111111111111111110000111111100001111111	110111111111111111110000111111100001111111	0
DRC-062840	111111111111111111111111111111100001111111	111111111111111111111111111111100001111111	0
DRC-062841	110111111111111111110000011011100001111111	111111111111111111110000011011100001111111	2
DRC-062844	110111111111111111110000111111100001111111	110111111111111111110000111111100001111111	0
DRC-062849	111111111101111111110000111111100001111111	111111111101111111110000111111100001111111	0
DRC-063522	111111111111111111111111111111100001110111	111111111111111111111111111111100001110111	0
DRC-063540	111111100011111111000000000000000000011111	111111100011111111000000000000000000011111	0
DRC-063541	111111000000000000000000111111100001111111	111111000000000000000000111111100001111111	0
DRC-063545	111111000001111111110100000000000010001111	111111000001111111110100000000000110001111	1
DRC-063556	111111100011111111000000000000000000011111	111111100011111111000000000000000000011111	0
DRC-063557	111111000000000000000000111111100001111111	111111000000000000000000111111100001111111	0
DRC-063645	111111111111111111110000111111100001111111	111111111110111111110000011011100001111111	3
DRC-063667	110111111111111111110000111111100001111111	110111111111111111110000111111100001111111	0
DRC-063803	111111111111111111110000011011100001111111	111111111111111111110000011011100001111111	1
DRC-063887	111111111111111111110000111111100001111111	111111111111111111110000111111100001111111	0
DRC-063888	111111111111111111110000111111100001111111	111111111111111111110000111111100001111111	0
DRC-063889	111111111111111111110000111111100001111111	111111111111111111110000111111100001111111	0
DRC-063896	111111111111111111111111111111100001111111	111111111111111111111111111111100001111111	0
DRC-063898	111111111111111111110000111011100001111111	111111111111111111110000111011100001111111	0
DRC-064081	111111111111111111110000111111100001111111	111111111111111111110000111111100001111111	1
DRC-064085	111111111111111111111111111111100001110111	111111111111111111111111111111100001110111	0
DRC-064090	111111000000000000000000111111100001111111	111111000000000000000000111111100001111111	0
DRC-064101	111111111111111111110000111111100001111111	111111111111111111110000111111100001111111	0
DRC-064208	111111111111111111111111111111100001110111	111111111111111111111111111111100001110111	0
DRC-070012	101111111101011111111111111111100001110111	101111111110111111111111111111100001110111	1
DRC-070038	111111111111111111110000111111100001111111	111111111111111111110000111111100001111111	0
DRC-070043	111000011111111111111000000000000111111111	111000011111111111111000000000000111111111	0
DRC-070049	111111100011111111000000000000000000011111	111111100011111111000000000000000000011111	0
DRC-070777	111111111111111111110000111011100001111111	111111111111111111110000111011100001111111	0
DRC-070779	111111000000000000000000111111100001111111	111111000000000000000000111111100001111111	0
DRC-070781	111111111111011111111111111111100001110111	111111111111011111111111111111100001110111	0
DRC-070783	111111111111111111110000000000000000011111	111111111101111111110000000000000000011111	1
DRC-070786	111111000001111111111010000000000010001111	111111000001111111111010000000000110001111	1
DRC-070788	111100001111000011111111111111111100001111	111100001111000011111111111111111100001111	0

DRC-070790	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-070792	11111111111111111111111000000011000011111111	11111111111111111111111000000011000011111111	0
DRC-070794	10000011111111111111111111111111100001111111	10000011111111111111111111111111100001111111	0
DRC-070804	11111111111111111111000011111111000010111111	11111111111111111111000011111111000010111111	0
DRC-070807	11111111111111111111111111011100001110111	11111111111111111111111111011100001110111	0
DRC-070808	11111111111111111111111111111100001110111	11111111111111111111111111111100001110111	0
DRC-070809	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-071052	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-071057	11111111111111111111111111111111000011110111	11111111111111111111111111111111000011110111	0
DRC-071074	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-071075	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-071076	11111110000011111111110100000000000010001111	11111110000011111111110100000000000110001111	1
DRC-071458	11111111111111111111111111111111000011110111	11111111111111111111111111111111000011110111	0
DRC-071460	11111111111111111111000111111111000011110111	11111111111111111111000111111111000011110111	0
DRC-071472	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-071479	11111111111111111111111111111101000011111111	11111111111111111111111111111111000011111111	1
DRC-071686	11111111111111111111111111111111000011110111	11111111111111111111111111111111000011110111	0
DRC-071691	11011111111110111111000011111111000011111111	11111111111101111111000011111111000011111111	3
DRC-071696	11111111111101111111111111111111000011110111	11111111111101111111111111111111000011110111	0
DRC-071697	11111111111111111111000111110100001111111	11111111111111111111000111111111000011111111	1
DRC-071712	11111111111111111111111111111111000011110111	11111111111111111111111111111111000011110111	0
DRC-071960	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-071968	11111111111111111111111111111111000011110111	11111111111111111111111111111111000011110111	0
DRC-071972	11111111111000011100000000000000000000111100	11111111111000011100000000000000000000111100	0
DRC-071973	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-071975	11111111111101111111111111111111000011110111	11111111111101111111111111111111000011110111	0
DRC-071988	11111111111111111111000000110111000000011111	11111111111111111111000000111011000011111111	3
DRC-071990	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-071998	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-072061	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-072379	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-072384	11111010000011111111111010000000010000011	11111110000011111111111000000000110000011	3
DRC-072390	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-072603	11111111111111111011111111111111000011111111	11111111111111111011111111111111000011111111	0
DRC-072604	11110000000000000000000011111111000011110111	11110000000000000000000011111111000011110111	0

DRC-072605	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-072606	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-072608	11111111111101111111111111111111100001110111	11111111111101111111111111111111100001110111	0
DRC-072609	11111111111111111111111111111111100001110111	11011111111101111111000011111111000011111111	7
DRC-072613	11011111110111111111000011111111000011111111	11111111110111111111000011111111000011111111	1
DRC-072615	11111110000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-072618	11111111111100001110000000000000000000111100	11111111111100001110000000000000000000111100	0
DRC-072625	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-072818	11011111111111111111000011111111000011111111	11111111111100001110000000000000000000111100	17
DRC-072819	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-072820	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-072824	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-072825	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-072827	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-072829	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-072830	11111111111100001110000000000000000000111100	11111111111100001110000000000000000000111100	0
DRC-072834	11111111111101111111111111111111100001110111	11111111111101111111111111111111100001110111	0
DRC-072836	11111111111111111111011001111111000010101111	11111111111111111111011001111111000011101111	1
DRC-072837	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-072943	11111111111100001110000000000000000000111100	11111111111100001110000000000000000000111100	0
DRC-072945	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-072957	11011111111101111111000011111111000011111111	11111111111111111111111111111111100001110111	7
DRC-072959	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-072963	11111111111111111011111111101111000011111111	11111111111111111111111111101111000011111111	1
DRC-072974	11111111111111111111000011111111000010111111	11111111111111111111000011111111000010111111	0
DRC-072986	11111111111111111111000011000011000010111111	11111111111111111111000011000011000011111111	1
DRC-072988	11111100000000000000000001111111000011110111	11111100000000000000000001111111000011110111	0
DRC-073093	11111110000011111111110110000000000010001111	11111110000011111111110100000000000110001111	2
DRC-073099	11111010000011111111111110000000000010000011	11111110000011111111111100000000000110000011	3
DRC-073168	11111111111100001110000000000000000000111100	11111111111100001110000000000000000000111100	0
DRC-073169	11111111111111111111000011110111000011110111	11111111111111111111000011110111000011110111	0
DRC-073170	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-073173	11011111111101111111000011111111000011111111	11011111111101111111000011111111000011111111	0
DRC-073175	11011111111111111111111111111111100001110111	11011111111111111111111111111111100001110111	0
DRC-073183	11111111111111111011111111111111100001111111	11111111111111111111111111111111100001111111	1

DRC-073186	11111111111111111111111111011100001110111	11111111111111111111111111011100001110111	0
DRC-073413	11111111111111111110000111011100001111111	11111111111111101111000011101110000111111	1
DRC-073414	00000000000000000000000011111110000111011	00000000000000000000000011111110000111011	0
DRC-073415	00000000000000000000000011111110000111011	00000000000000000000000011111110000111011	0
DRC-073420	11111111111111111110000111111110000111111	11111111111111111110000111111110000111111	0
DRC-073422	11111111111111111110000111111110000111111	11111111111111111110000111111110000111111	0
DRC-073423	11111111111111110111111111111110000100001	11111111111111110111111111111110000100001	0
DRC-073424	11111111111111111110000111111110000111111	11111111101111111110000111111110000111111	1
DRC-073429	11111100000000000000000011111110000111111	11111100000000000000000011111110000111111	0
DRC-073430	11111111111111111111111111111110000111011	11111111111111111111111111111110000111011	0
DRC-073432	11111111111111111111111111111110000111011	11111111101111111111111111111110000111011	1
DRC-073434	11111111111101111101111111111110000111011	11111111111101111101111111111110000111011	0
DRC-073443	11111111111000011100000000000000000111100	11111111111000011100000000000000000111100	0
DRC-073515	11111100000000000000000011111110000111111	11111100000000000000000011111110000111111	0
DRC-073712	11000000111111110111000011111110000111111	11000000111111110111000011111110000111111	0
DRC-073716	11011111111111111111000011111110000111111	11011111111111111111000011111110000111111	0
DRC-080008	11111111000111111110000000000000000011111	11111111000111111110000000000000000011111	0
DRC-080016	11111111111110111111111111111110000111011	11111111111111111111111111111110000111011	1
DRC-080021	11111100000000000000000011111110000111111	11111100000000000000000011111110000111111	0
DRC-080028	11111111000111111110000000000000000011111	11111111000111111110000000000000000011111	0
DRC-080040	11111100000000000000000011111110000111111	11111100000000000000000011111110000111111	0
DRC-080046	00000000000000000000000011111110000111011	00000000000000000000000011111110000111011	0
DRC-080545	11111111111111111111000111111110000111111	11111111111111111111000111111110000111111	0
DRC-080548	11111110000011111111101100000000000010001111	11111110000011111111101000000000000110001111	3
DRC-080552	11111111101111111111000011111110000101111	11111111101111111111000011111110000101111	0
DRC-080553	11111111111111111111111111111110000111011	11111111111111111111111111111110000111011	0
DRC-080561	11111111111111111111000011111110000101111	11111111111111111111000011111110000101111	0
DRC-080567	11111111111111111111111111011100001110111	11111111111111111111111111011100001110111	0
DRC-081290	11111111101111111111111111111110000111111	11111111101111111111000011111110000111111	4
DRC-081527	11111111111111111111000011111110000111111	11111111111111111111000011111110000111111	0
DRC-081585	11111111111111111111111111111110000111011	11111111111111111111111111111110000111011	0
DRC-081588	11000000111111110111000011111110000111111	11000000111111110111000011111110000111111	0
DRC-081589	11101111111000011100000000000000000111100	11101111111000011100000000000000000111100	0
DRC-081592	11111111111111011111000011111110000111111	11111111111111111111000011111110000111111	1
DRC-081593	11111111111111111111111111111110000111011	11111111111111111111111111111110000111011	0

DRC-081595	11111111111111111111111111111100001110111	11111111111111111111111111111100001110111	0
DRC-081600	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-081602	11111111111011111111111111111100001110111	11111111111011111111111111111100001110111	0
DRC-081610	11111111001111111111111111111100001111111	11111111001111111111111111111100001111111	0
DRC-081622	11111111111111111110000110000110000111111	11111111111111111110000110000110000111111	0
DRC-081722	11011111111111111110000111111110000111111	11011111111111111110000111111110000111111	0
DRC-081873	11111111111111111110000111111110000101111	11111111111111111110000111111110000101111	0
DRC-081879	11111111111111110111111111111100001100001	11111111111111110111111111111100001100001	0
DRC-081881	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-081888	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-081985	11111111111111111011111111111100001111111	11111111111111111011111111111100001111111	0
DRC-081986	11111111111010111111111111111100001110111	11111111111011111111111111111100001110111	1
DRC-081989	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-081990	00000000000000000000000000000000111111111	00000000000000000000000000000000111111111	0
DRC-082120	11111111111111111111111111111100001110111	11111111111111111111111111111100001110111	0
DRC-082121	11111111111111111111111111111100001110111	11111111111111111111111111111100001110111	0
DRC-082130	11111111111111111110000110000110000101111	01111111111111111110000110000110000101111	1
DRC-082131	11111111111111110111111111111100001111111	11111111111111110111111111111100001111111	0
DRC-082317	11111111111111111110110011111100001010111	11111111111111111111111111111100001110111	4
DRC-082318	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-082327	11111111111110111111111110111100001110111	11111111111111111111111110111100001110111	1
DRC-082328	11111110001111111000000000000000000011111	11111110001111111000000000000000000011111	0
DRC-082330	11011111111111111110000111111110000111111	11011111111111111110000111111110000111111	0
DRC-082584	11111111111111111111111111111100001110111	11111111111111111111111111111100001110111	0
DRC-082592	11111111111111111110110011111100001010111	11111111111111111110110011111100001110111	1
DRC-082639	11111110001111111000000000000000000011111	11111110001111111000000000000000000011111	0
DRC-082656	11111111111111111110000111101100001111111	11111111111111111110000111101100001111111	0
DRC-082700	11111111111111111110000111111110000111111	11111111111111111110000111111110000111111	0
DRC-082919	11000000111111110111000011111110000111111	11000000111111110111000011111110000111111	0
DRC-082923	11111110000011111111110110000000000010001111	11111110000011111111110100000000000110001111	2
DRC-082939	11111111111111111111000011111110000111111	11111111111111111111000011111110000111111	0
DRC-082940	11111110001111111000000000000000000011111	11111110001111111000000000000000000011111	0
DRC-082945	11111111111111111111000011111110000111111	11111111111111111111000011111110000111111	0
DRC-082946	11111111111111111111111111111100001110111	11111111111111111111111111111100001110111	0
DRC-082948	11111111111111111111000011110111000011101111	11111111111111111111000011110111000011101111	0

DRC-082949	11111111111111011111111111111111100001111111	11111111111111111111111111111100001111111	1
DRC-083047	11111111111111111111000011110111000011111111	11111111111111111111000011110111000011111111	0
DRC-083048	11111111111111111111000000110111000011111111	11111111111111111111000001110111000011111111	1
DRC-083050	11111111111111111111000011110111000010011111	11111111111111111111000011110111000010011111	0
DRC-083051	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-083052	11111111111111111111100011111111000011111111	11111111111111111111100011111111000011111111	0
DRC-083066	11111111111000011100000000000000000000111100	11111111111000011100000000000000000000111100	0
DRC-083067	11111111111011111111111111111111100001110111	11111111111011111111111111111111100001110111	0
DRC-083190	11111111111111111111000011110111000011111111	11111111111111111111000011111111000011111111	1
DRC-083191	11111111111111111111000011110111000011111111	11111111111111111111000011110111000011111111	0
DRC-083200	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-083201	10000011111111111110000011111111000011111111	10000011111111111111111111111111000011111111	5
DRC-083202	11111111111111111111000011110111000010011111	11111111111111111111000011110111000010011111	0
DRC-083204	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-083226	11111101011111111111111110000000000000000000	11111101011111111111111110000000000000000000	1
DRC-083250	11111111111111111111111111111101000011111111	11111111111111111111111111111111000011111111	1
DRC-083736	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-083801	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-084109	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-084204	11011111011111111111000011111111000011110111	11011111011111111111000011111111000011110111	0
DRC-084210	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-090010	11111111111111111111000011110111000011111111	11111111111111111111000011110111000011111111	0
DRC-090012	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-090016	11111111111111111111000011110111000011111111	11111111111111111111000011110111000011111111	0
DRC-090022	11111111111011111111111111111111100001110111	11111111111011111111111111111111100001110111	0
DRC-090037	11111111111111111111000011000011000011111111	11111111111111111111000011000011000011111111	0
DRC-090066	11111110000011111111110110000000000010001111	11111110000011111111110100000000000110001111	2
DRC-090069	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-090077	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-090078	11111111111111111111000011111111000010111111	11111111111111111111000011111111000010111111	0
DRC-090165	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-090172	11111111111000011100000000000000000000111100	11111111111000011100000000000000000000111100	0
DRC-090299	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-090308	11111100000000000000000001111111000011111111	11111100000000000000000001111111000011111111	0
DRC-090309	11111111111101111111000011000011000011111111	11111111111101111111000011000011000011111111	0

DRC-090310	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-090432	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-090451	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-090452	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-090456	11111111111111111111111111011100001110111	11111111111111111111111111011100001110111	0
DRC-090702	11111100000000000000000011111100001111111	11111100000000000000000011111100001111111	0
DRC-090708	11111111111111110111111111111100001100001	11111111111111110111111111111100001100001	0
DRC-090710	11111111111000011100000000000000000000111100	11111111111000011100000000000000000000111100	0
DRC-090909	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-090913	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-090947	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-090976	11111111111111111111000011000011000011111111	11111111111111111111000011000011000011111111	0
DRC-090977	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-090999	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-091003	11111111111111111111111111011100001110111	11111111101111111111111111011100001110111	1
DRC-091004	11111100000000000000000011111100001111111	11111100000000000000000011111100001111111	0
DRC-091005	11111111111010111111111111111100001111111	11111100000000000000000011111111000011111111	16
DRC-091168	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-091171	11111111111111111111000011111111000011111111	11111111111111111111000011110111000011111111	1
DRC-091172	11111111111111111111111111011100001110111	11111111111111111111111111011100001110111	0
DRC-091696	11111111101111111110000111111110000111111	11111111101111111110000111111110000111111	0
DRC-091821	11111111111111111111111111011100001110111	11111111111111111111111111011100001110111	0
DRC-092401	11111111111111111111000111111110000111111	11111111111111111111000111111110000111111	0
DRC-092631	11111111111111111111000011000011000011110000	11111111111111111111000011000011000011110000	0
DRC-092748	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-092750	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-093215	11011111011111111111000011111111000011110111	11011111011111111111000011111111000011110111	0
DRC-093480	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-093482	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-093488	11011111111111111111000011111111000011111111	11011111111111111111000011111111000011111111	0
DRC-093493	11111111111111111111000011111111000011111111	11111111111111111111000011111111000011111111	0
DRC-093495	11111111111111111111111111011100001110111	11111111111111111111000011111111000011111111	6
DRC-093664	11111100000000000000000011111100001111111	11111100000000000000000011111100001111111	0
DRC-093666	11111111111101111111000101111000000011111111	11111111111101111111000010111000000011111111	1
DRC-093674	11111111111111111111111111111111100001111111	11111111111111111111111111111111100001111111	0

DRC-100033	11111111111111111011111111111111100001111111	11111111111111111011111111111111100001111111	0
DRC-100035	11111111111011111111111111111111100001110111	11111111111011111111111111111111100001110111	0
DRC-100037	11011111111011111110000111111111100001111111	11011111111011111110000111111111100001111111	0
DRC-100099	11111100000000000000000001111111100001111111	11111100000000000000000001111111100001111111	0
DRC-100122	11111111111111111111111111111001100001110111	11111111111111111111111111111111100001110111	2
DRC-100299	11111111111111111110000111111111100001011111	11111111111111111110000111111111100001011111	0
DRC-100300	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-100301	11111111111111111110000111111111100001011111	11111111111111111110000111111111100001011111	0
DRC-100307	11111111111111111110000110000110000001111111	11111111111111111110000110000110000001111111	0
DRC-100310	11111111111110000011111111111111100001110111	11111111111110000011111111111111100001110111	0
DRC-100503	11111111111111111111111111111111100001110111	11111111111111111111111111111111100001110111	0
DRC-100740	11111111111111111110000111111111100001111111	11111111111111111110000111111111100001111111	0
DRC-100753	11111111111111111111111111011100001111111	11111111111111111111111111111111100001110111	2
DRC-100759	11011111111111111110000111111111100001111111	11011111111111111110000111111111100001111111	0
DRC-100923	11111111111111111110000111101110000111111	11111111111111111110000111111111100001011111	2
DRC-100929	11111111000111111110000000000000000000011111	11111111000111111110000000000000000000011111	0
DRC-100931	11011111111111111110000111111111100001111111	11011111111111111110000111111111100001111111	0
DRC-101072	11111111111111111110000111111111100001111111	11111111111111111110000111111111100001111111	0
DRC-101306	11111111111111111110000111101110000111111	11111111111111111110000111101110000111111	0
DRC-101308	11011111111111111110000111111111100001111111	11011111111111111110000111111111100001111111	0

858

859

860

861 Supplementary Figure 1: Clustering of *M. tuberculosis* isolates based on 20 methods.

862 For each method, the inclusion of an isolate into a cluster is outlined in the surrounding
863 circles using GraPhlAn (Asnicar et al., 2015). If an isolate is in a cluster not affected by
864 convergence, it is highlighted in black for the given method. If an isolate is in a cluster
865 affected by convergence, it is shown in grey.

866

867

Supplementary Figure 2: Algorithm of the phylogenetic inclusion method

An extension of the SNP-based clustering methods is used to detect complete chains of transmission that may be missed by pairwise SNP comparisons. In this example: A) shows the inferred phylogenetic sub-tree for five taxa and associated SNPs for each branch; B) a SNP cut-off is applied, which places 2 taxa (Taxon 1 and Taxon 4) together into a cluster; the most recent common ancestor (MRCA) of all taxa in this cluster (i.e. Taxon 1 & Taxon 4) is found; D) all descendants of this MRCA are added to the putative transmission chain.