

1 **TITLE PAGE:** *J Clin Microbiol, full report – Revision 1*

2

3 **Standard genotyping overestimates transmission of *Mycobacterium***  
4 ***tuberculosis* among immigrants in a low incidence country**

5

6 David Stucki <sup>1,2,3</sup>, Marie Ballif <sup>1,2,3</sup>, Matthias Egger <sup>3</sup>, Hansjakob Furrer <sup>4</sup>, Ekkehardt  
7 Altpeter <sup>5</sup>, Manuel Battegay <sup>6</sup>, Sara Droz <sup>7</sup>, Thomas Bruderer <sup>8</sup>, Mireia Coscolla <sup>1,2</sup>,  
8 Sonia Borrell <sup>1,2</sup>, Kathrin Zürcher <sup>3</sup>, Jean-Paul Janssens <sup>9</sup>, Alexandra Calmy <sup>10</sup>,  
9 Jesica Mazza Stalder <sup>11</sup>, Katia Jatton <sup>12</sup>, Hans L. Rieder <sup>13</sup>, Gaby E. Pfyffer <sup>14</sup>, Hans  
10 H. Siegrist <sup>15</sup>, Matthias Hoffmann <sup>16</sup>, Jan Fehr <sup>17</sup>, Marisa Dolina <sup>18</sup>, Reno Frei <sup>19</sup>,  
11 Jacques Schrenzel <sup>20</sup>, Erik C. Böttger <sup>21</sup>, Sebastien Gagneux <sup>1,2</sup>, Lukas Fenner  
12 <sup>1,2,3, #</sup>, for the Swiss HIV Cohort and Molecular Epidemiology of Tuberculosis Study  
13 Groups \*

14

15 **1** Swiss Tropical and Public Health Institute, Basel, Switzerland  
16 **2** University of Basel, Basel, Switzerland  
17 **3** Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland  
18 **4** Clinic for Infectious Diseases, Bern University Hospital and University of Bern, Bern,  
19 Switzerland  
20 **5** Division of Communicable Diseases, Federal Office of Public Health, Bern, Switzerland  
21 **6** Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel,  
22 Basel, Switzerland  
23 **7** Mycobacteriology Unit, Institute for Infectious Diseases, University of Bern, Bern,  
24 Switzerland  
25 **8** Centre for Laboratory Medicine, St.Gallen, Switzerland  
26 **9** Division of Pneumology, University Hospital Geneva, Geneva, Switzerland  
27 **10** Division of Infectious Diseases, University Hospital Geneva, Geneva, Switzerland  
28 **11** Division of Pneumology, University Hospital Lausanne, Lausanne, Switzerland  
29 **12** Institute of Microbiology, University Hospital of Lausanne, Lausanne, Switzerland  
30 **13** Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Switzerland  
31 **14** Department of Medical Microbiology, Luzerner Kantonsspital, Lucerne, Switzerland  
32 **15** ADMed Microbiology, La Chaux-de-Fonds, Switzerland  
33 **16** Division of Infectious Diseases, Kantonsspital St. Gallen, St. Gallen, Switzerland  
34 **17** University Hospital Zurich, and University of Zurich, Zurich, Switzerland  
35 **18** Clinical Microbiology, EOLAB, Bellinzona, Switzerland  
36 **19** Division of Clinical Microbiology, University Hospital Basel, Basel, Switzerland  
37 **20** Laboratory of Bacteriology, University Hospitals of Geneva, Geneva, Switzerland  
38 **21** Institute of Medical Microbiology, National Center for Mycobacteria, University of Zurich,  
39 Zurich, Switzerland  
40

\* All members of the study groups are listed in the Acknowledgements

41 **Running title:**

42 Tuberculosis transmission among immigrants

43

44 **Word count and inserts:**

45 250 words (abstract), 2945 words (main text), 2 tables, 3 figures, 47 references,

46 supplementary information (2 figures, 1 table)

47

48 **Keywords:**

49 Tuberculosis; transmission; immigrant; MIRU-VNTR; whole genome sequencing;

50 molecular cluster; Swiss HIV Cohort; Switzerland; low TB incidence

51

52 **# Corresponding author:**

53 Lukas Fenner, Swiss Tropical and Public Health Institute, Basel; and Institute of

54 Social and Preventive Medicine, University of Bern, Switzerland

55 Email: lukas.fenner@unibas.ch; lukas.fenner@ispm.unibe.ch

56

57 **ABSTRACT**

58

59 Immigrants from high tuberculosis (TB) incidence regions are a risk group for  
60 TB in low-incidence countries such as Switzerland. In a previous analysis of a  
61 nationwide collection of 520 *Mycobacterium tuberculosis* isolates from 2000-2008, we  
62 identified 35 clusters comprising 90 patients based on standard genotyping (24-loci  
63 MIRU-VNTR and spoligotyping). Here, we used whole genome sequencing (WGS) to  
64 revisit these transmission clusters. Genome-based transmission clusters were  
65 defined as isolate pairs separated by  $\leq 12$  single nucleotide polymorphisms (SNPs).  
66 WGS confirmed 17/35 (49%) MIRU-VNTR clusters; the other 18 clusters contained  
67 pairs separated by  $>12$  SNPs. Most transmission clusters (3/4) of Swiss-born  
68 patients were confirmed by WGS, as opposed to 25% (4/16) of clusters involving only  
69 foreign-born patients. The overall clustering proportion using standard genotyping  
70 was 17% (90 patients, 95% confidence interval [CI]: 14-21%), but only 8% (43  
71 patients, 95% CI: 6-11%) using WGS. The clustering proportion was 17% (67/401,  
72 95% CI: 13-21%) using standard genotyping and 7% (26/401, 95% CI: 4-9%) using  
73 WGS among foreign-born patients, and 19% (23/119, 95% CI: 13-28%) and 14%  
74 (17/119, 95% CI: 9-22%), respectively, among Swiss-born patients. Using weighted  
75 logistic regression, we found weak evidence for an association between birth origin  
76 and transmission (aOR 2.2, 95% CI: 0.9-5.5, comparing Swiss-born patients to  
77 others). In conclusion, standard genotyping overestimated recent TB transmission in  
78 Switzerland when compared to WGS, particularly among immigrants from high TB  
79 incidence regions, where genetically closely related strains often predominate. We  
80 recommend the use of WGS to identify transmission clusters in low TB incidence  
81 settings.

82

83 **INTRODUCTION**

84

85       Tuberculosis (TB) remains an important public health concern in European  
86 countries (1–3). Immigrants from high TB-incidence countries and HIV-infected  
87 populations are common risk groups in Switzerland, as in other European countries  
88 (4–8). We and others have previously shown that transmission of *Mycobacterium*  
89 *tuberculosis* occurs, but is not more common among immigrants than in the native  
90 population (9–12).

91       Mycobacterial interspersed repetitive-unit–variable-number tandem-repeat  
92 (MIRU-VNTR), combined with spoligotyping (13, 14), remains the most commonly  
93 used genotyping method in molecular epidemiology of TB (15). However, MIRU-  
94 VNTR may not distinguish between genetically closely related strains despite the  
95 absence of close epidemiological links between patients (16–18). MIRU-VNTR may  
96 be suboptimal to study transmission among immigrants from high TB incidence  
97 countries, where genetically closely related strains circulate over extended periods of  
98 time (19). Hence, recent transmission among immigrants is potentially overestimated,  
99 but the extent of this phenomenon is largely unknown. In contrast to standard  
100 genotyping methods, whole genome sequencing (WGS) provides an increased  
101 resolution and has been used to study *M. tuberculosis* transmission (20–24). In this  
102 study, we re-analyzed transmission clusters previously defined by MIRU-VNTR,  
103 using WGS to assess transmission of *M. tuberculosis* among Swiss- and foreign-born  
104 TB patients (9).

105

106

## 107 MATERIALS AND METHODS

108

### 109 Study setting and study population

110 In 2012, we conducted a nationwide study of the molecular epidemiology of TB  
111 in Switzerland as a collaborative project between the Swiss HIV Cohort Study  
112 (SHCS), the National Center for Mycobacteria, diagnostic microbiology laboratories,  
113 departments of respiratory medicine and public health, and the Federal Office of  
114 Public Health ([www.tb-network.ch](http://www.tb-network.ch)) (9, 25–28). The study setting was previously  
115 described in detail (9). Briefly, all patients in the SHCS diagnosed with TB between  
116 2000 and 2008 were enrolled (n=93). In addition, we included a random sample of  
117 288 TB cases from the 4,221 culture-confirmed TB patients reported to the National  
118 TB Surveillance Registry, and all drug-resistant TB cases reported in Switzerland  
119 (n=167) during the same period (categories not mutually exclusive). 24-loci MIRU-  
120 VNTR and spoligotyping were used for the molecular detection of transmission  
121 clusters (9). In this follow-up study, we performed WGS on the 90 *M. tuberculosis*  
122 isolates belonging to one of the 35 MIRU-VNTR/spoligotyping clusters.

123

### 124 Clinical data collection and definitions

125 The clinical data collection was previously described in detail (9). The clustering  
126 proportion was determined by the “n” method expressed as the number of patients in  
127 clusters divided by the total number of individuals (29). MIRU-VNTR clusters were  
128 defined as a group of isolates with identical MIRU-VNTR and spoligotyping patterns  
129 (9). In addition, we used IS6110-Restriction Fragment Length Polymorphism (RFLP)  
130 patterns, when available from the National Center for Mycobacteria (30). RFLP has a  
131 higher resolution than MIRU-VNTR, particularly in strains of the “Beijing” genotype.

132 Isolates with identical MIRU-VNTR and spoligotype patterns, but different IS6110  
133 patterns, were considered as non-clustered. “Mixed” molecular clusters were defined  
134 as clusters with Swiss-born and foreign-born individuals, or foreign-born individuals  
135 from different continents.

136

### 137 **Whole genome sequencing and phylogenetic analyses**

138 We generated whole genome sequences for all 90 patient isolates identified as  
139 part of MIRU-VNTR clusters (9). We used Illumina Nextera XT or TruSeq library  
140 preparation kits and Illumina HiSeq, MiSeq or NextSeq devices (Illumina, San Diego,  
141 CA) for WGS according to manufacturer’s instructions. Isolates were re-sequenced  
142 when the mean read depth was below 20x. FastQ files from multiple sequencing runs  
143 of the same isolate were merged. We used KvarQ for initial quality check,  
144 determination of *M. tuberculosis* phylogenetic lineages and *in silico* spoligotyping  
145 pattern, as previously described (31). We then mapped short sequencing reads to a  
146 hypothetical *M. tuberculosis* ancestral genome (identical with H37Rv in structure, but  
147 with maximum likelihood-inferred ancestral bases (32), with BWA 0.6.2 (33)).  
148 Samtools 0.1.19 was used to call variants (SNPs). We only retained positions with a  
149 read depth of  $\geq 10\%$  and  $\leq 200\%$  of the average read depth for the whole genome,  
150 and a phred-scaled quality score of  $\geq 30$ . We excluded positions in known repetitive  
151 regions (23), as well as SNPs in genes in which we have previously identified 50 bp  
152 sequences with homologous sequences elsewhere on the genome (Supplementary  
153 Table 1). We also excluded positions associated with drug resistance (31). For the  
154 analyses of read/allele counts at particular genomic loci, we extracted the number of  
155 high quality bases from Variant Call Format (VCF)-files with SNPeff/SNPst (34).

156

157 **Transmission networks based on SNP distances**

158 We generated an alignment of all variable positions across the 90 isolates. We  
159 then calculated the raw genetic distances (number of SNPs) for each isolate pair in  
160 each MIRU-VNTR cluster with the “Compute Pairwise Distances” function (using the  
161 “Pairwise deletion” option) in MEGA 5.2.2 (35). We defined a MIRU-VNTR cluster as  
162 a “true” transmission cluster if all isolate pairs in the cluster were separated by  $\leq 12$   
163 SNPs. In a sensitivity analysis, we opted for a stricter definition, whereby a MIRU-  
164 VNTR transmission cluster was considered as confirmed if at least one of its isolate  
165 pairs was separated by  $\leq 5$  SNPs. These thresholds of 12 and 5 SNPs were  
166 previously established by Walker *et al.* (17). We imported an alignment of the  
167 variable genomic positions into popart (<http://popart.otago.ac.nz>) to generate median  
168 joining networks. Networks were generated for all 35 transmission clusters identified  
169 by standard genotyping (MIRU-VNTR and spoligotyping) (9).

170

171 **Statistical analysis**

172 We re-analyzed risk factors for transmission using the WGS-based (“true”)   
173 cluster definition. We used weighted logistic regression models to obtain age- and  
174 sex-adjusted odds ratios (aOR) for the probability of belonging to a true molecular  
175 cluster. We used the Kruskal-Wallis rank sum test to assess differences between  
176 mean genetic distances of Swiss-born, foreign-born and mixed clusters. As our study  
177 sample included, by design, more HIV-infected and patients with drug-resistant TB  
178 (9), we calculated weights to take sampling proportions into account. As a sensitivity  
179 analysis, we restricted the analysis of clustering proportion to the patients in the  
180 random sample from the TB registry (n=288) (9). All statistical analyses were  
181 performed in Stata version 14 (Stata Corporation, College Station, TX) and R 3.1.2  
182 (36).

183 In addition, we plotted the mean genetic distances (in SNPs) versus the mean  
184 geographical distances (in km) of all patient pairs in a molecular cluster (distance  
185 between the birth countries' capital cities). Plots were generated with the ggplot2  
186 library in R (37).

187

#### 188 **Ethics approval**

189 The study was approved by the Ethics Committee of the Canton of Bern,  
190 Switzerland (9). Informed consent was obtained from all patients enrolled in the  
191 SHCS. For patients outside the SHCS, informed consent was obtained by the  
192 treating physicians. In some cases, informed consent could not be obtained from  
193 the patient, because he or she could not be located or was known to have died.  
194 For these cases, we obtained permission from the Federal Expert Commission on  
195 Confidentiality in Medical Research to use the data provided by the treating  
196 physician.

197

198



## 199 RESULTS

200

### 201 Study population

202 The study population consisted of 520 TB patients from the nationwide study in  
203 Switzerland (9, 25, 27, 28, 38). The patient characteristics are described in [Table 1](#). A  
204 total of 119 (22.9%) patients were born in Switzerland and 401 (77.1%) abroad.  
205 Median age was 36.5 years (Interquartile Range 28–51). Overall, 113 (21.7%)  
206 patients were HIV positive. Pulmonary TB accounted for 387 (74.4%) of all cases and  
207 extrapulmonary TB for 133 (25.6%) ([Table 1](#)).

208

### 209 Transmission clusters

#### 210 *Whole genome sequencing*

211 Isolates were sequenced with a median sequencing depth of 130x (range 22-  
212 274x). For quality assurance, we compared laboratory-assay-based phylogenetic  
213 lineage classification and spoligotyping pattern with the results generated from the  
214 WGS data using KvarQ (9). We found 100% agreement between the two methods for  
215 lineage identification and up to two discordant spacers in the spoligotyping patterns.

#### 216 *Identification of molecular clusters based on WGS*

217 In the 35 previously defined MIRU-VNTR clusters, we found pairwise genetic  
218 distances of 0 to 224 SNPs (median: 21.5 SNPs) ([Figure 1](#) and [Supplementary](#)  
219 [Figure 1](#)). In the largest cluster (eight isolates), genetic distances were  $\geq 54$  SNPs.  
220 Seventeen of 35 (48.6%) MIRU-VNTR clusters consisted of pairs separated by  $\leq 12$   
221 SNPs, i.e. were confirmed as true transmission clusters, corresponding to 43 of 90  
222 patients (47.8%) ([Table 2](#)). The remaining 18 clusters harbored isolate pairs  
223 separated by  $>12$  SNPs (47 patients).

224 The overall clustering proportion decreased from 17.3% (95% confidence  
225 interval [95% CI]: 14.2-20.8%) based on standard genotyping (spoligotyping and  
226 MIRU-VNTR), to 8.3% (95% CI: 6.0-11.0%) based on WGS. When restricting the  
227 analysis to the 288 randomly selected patients, we found 27 clustered patients in 11  
228 genome-based clusters, resulting in a clustering proportion of 9.4% (95% CI: 6.3-  
229 13.3%). When using a more stringent WGS definition for transmission clusters (at  
230 least one isolate pair in a cluster  $\leq 5$  SNPs distance), 13 of 35 (37.1%, CI: 21.5-  
231 55.1%) MIRU-VNTR clusters were confirmed. These 13 transmission clusters  
232 included 35/520 patients, corresponding to a clustering proportion of 6.7% (95% CI:  
233 4.7-9.2%).

#### 234 *Infection with multiple M. tuberculosis strains*

235 In five isolates that were part of transmission clusters defined by MIRU-VNTR,  
236 we detected multiple alleles at several MIRU-VNTR loci, potentially indicating  
237 infection with multiple *M. tuberculosis* strains. We therefore conducted an allele  
238 frequency analysis based on sequencing reads for each SNP call. Despite the  
239 presence of multi-allelic variant calls in all isolates (potential microevolutionary  
240 events), none of the five isolates with multiple MIRU-VNTR bands showed evidence  
241 of lineage- or sublineage-specific markers with multiple alleles in the sequencing  
242 reads.

243

#### 244 **Molecular clustering in Swiss-born, foreign-born and HIV-positive patients**

245 Four MIRU-VNTR clusters involved Swiss-born patients only, 16 clusters  
246 foreign-born only and 15 clusters were of mixed birth group origin. Three of four  
247 clusters (75.0%) involving only Swiss-born patients were confirmed by WGS as true  
248 clusters (8/10 [80.0%] clustered Swiss-born patients). On the other hand, only 4/16

(25.0%) immigrant clusters (born on the same continent) were true clusters (8/37 [21.6%] patients). Of the 15 mixed clusters, 10 (66.7%) were true clusters (27/43 [62.8%] clustered patients) ([Table 2](#)). We assessed whether foreign-born patients were overrepresented in MIRU-VNTR clusters not confirmed by WGS. Among all 90 patients from the MIRU-VNTR clusters, foreign-born patients were more likely in clusters not confirmed by WGS compared to true clusters (aOR 4.5, CI: 1.5-13.6  $p=0.008$ ) ([Table 2](#)).

We then calculated the true (genome-based) clustering proportion for both Swiss- and foreign-born patients. The clustering proportion among Swiss-born cases decreased only slightly, from 19.3% (23/119, 95% CI: 12.7-27.6%) using MIRU-VNTR to 14.3% (17/119, 95% CI: 8.5-21.9) using WGS data. In contrast, the clustering proportion among immigrants was more than halved, from 16.7% (67/401, 95% CI: 13.2-20.7) to 6.5% (26/401, 95% CI: 4.3-9.4). [Figure 2](#) summarizes the possible factors leading to an overestimation of *M. tuberculosis* transmission based on standard genotyping, among foreign-born and native TB patients in low TB incidence settings.

The median genetic distance differed significantly between the three groups: 9 SNPs (range 8-15) in clusters with Swiss-born individuals only, 2 SNPs (range 0-162) in mixed clusters, and 24 SNPs (2-224) in clusters with foreign-born individuals only ( $p=0.030$ , [Figure 3](#)).

#### **Geographical and genetic distances within molecular clusters**

Plotting the mean genetic distance (in SNPs) versus the mean geographical distance between patient origins (capital cities) for each molecular cluster further supported the different patterns of clustering by birth origin of the patients ([Supplementary Figure 2](#)). Among foreign-born patients, a majority of MIRU-VNTR

275 clusters (11/16, 68.8%) consisted of patients born within 3,500 km of geographic  
276 distance, but harboring genetic distances of >12 SNPs. This indicates that genetically  
277 closely related strains, circulating in a geographically restricted area in the region of  
278 origin, were imported to Switzerland. Among the mixed MIRU-VNTR clusters, for  
279 which transmission is expected to have happened in Switzerland, a majority of  
280 clusters (9/15, 60%) had mean geographic distances above 3,500 km, but genetic  
281 distances of  $\leq 12$  SNPs, indicating recent transmission in Switzerland (Supplementary  
282 Figure 2).

283

#### 284 **Risk factors for transmission**

285 Female (aOR 0.39, CI: 0.18-0.85) and HIV-positive patients (aOR 0.39, CI: 0.16-  
286 0.93) were significantly less likely to be involved in true transmission clusters (Table  
287 1). In contrast, patients with cavitory disease were more likely to be associated with  
288 transmission (aOR 2.31, CI: 1.05-5.10). There was a weak evidence for an  
289 association between being born in Switzerland and being involved in a true  
290 transmission cluster (aOR 2.21, 95% CI: 0.88-5.52 for Swiss-born compared to  
291 foreign-born patients). Overall, the risk factors for transmission remained similar to  
292 those previously reported when using standard genotyping methods (9).

293

294 **DISCUSSION**

295

296 In this nationwide study on 520 TB patients, 18 of 35 transmission clusters  
297 identified by standard molecular genotyping (spoligotyping and MIRU-VNTR) were  
298 refuted by WGS. This suggests that transmission of *M. tuberculosis* is generally  
299 overestimated in low TB incidence countries such as Switzerland. Furthermore, we  
300 found a striking difference between transmission clusters involving Swiss-born cases  
301 and clusters involving foreign-born patients. WGS confirmed three quarters of the  
302 clusters involving Swiss-born individuals only, but only one quarter of clusters  
303 involving foreign-born patients only, hence indicating that transmission was especially  
304 overestimated among the immigrant population.

305 *M. tuberculosis* strains from immigrants, which were defined as clustered by  
306 MIRU-VNTR, but not by WGS, are likely genetically closely related genotypes,  
307 imported independently from a high-incidence region, where they are highly prevalent  
308 (19). Such strains accumulate genetic mutations over time (often leading to pairwise  
309 SNP distances of >12 SNPs), but the MIRU-VNTR pattern may not change. In such a  
310 situation, identical MIRU-VNTR will be wrongly interpreted as recent transmission in  
311 the country of immigration (9). Similar observations were made in the UK, where  
312 immigrant TB cases were identified in transmission clusters based on standard  
313 MIRU-VNTR genotyping, although no epidemiological link could be found during  
314 contact investigations (16).

315 The clustering proportion (indicating recent transmission) among Swiss-born  
316 individuals was similar using standard genotyping and WGS, but more than two-fold  
317 lower among foreign-born individuals when using WGS. In reality, the clustering  
318 proportion among foreign-born individuals might even be lower, as we cannot  
319 exclude that WGS-confirmed clusters ( $\leq 12$  SNPs) involving immigrants might partly

320 represent transmission that happened in the country of origin and not in Switzerland.  
321 Only social contact tracing could provide further insights into transmission dynamics,  
322 but such investigations are notoriously difficult to perform, particularly among  
323 immigrants (23, 39). The low proportion of true transmission clusters among  
324 immigrants in our study was further supported by the weighted analysis of predictors  
325 for transmission, which showed that foreign-born TB cases tended to be less likely  
326 involved in true clusters compared to Swiss-born cases. Of note, the clustering  
327 proportion among immigrants is remarkably similar to previous observations among  
328 immigrant MDR cases diagnosed in Switzerland, which showed a clustering  
329 proportion of 8% (compared to 7% in our study) based on standard genotyping and  
330 contact tracing (30).

331       The majority of mixed molecular clusters as defined by MIRU-VNTR (i.e.  
332 involving Swiss- and foreign-born individuals) showed small SNP distances ( $\leq 12$   
333 SNPs), confirming the intuitive explanation that transmission between Swiss-born  
334 and foreign-born likely occurred in Switzerland. This was further supported by the  
335 analysis of geographical distances between patient birth countries, which indicated  
336 that most isolate pairs in the mixed clusters were from patients born far away from  
337 each other, despite small genetic distances. The five mixed clusters harboring larger  
338 SNP distances may reflect TB cases due to infections by circulating global or  
339 European *M. tuberculosis* genotypes, such as the recently described large cluster in  
340 Eastern Europe (40).

341       We found no evidence of infection with multiple strains among clustered TB  
342 cases using the WGS data, despite the presence of double alleles in the MIRU-VNTR  
343 patterns of five clustered isolates. Infections with multiple strains could potentially  
344 also influence the identification of molecular clusters, as individual strains in an  
345 infection with multiple strains cannot be resolved by MIRU-VNTR typing. The

346 prevalence and relevance of such multiple infections need to be further studied (41–  
347 43).

348 A potential limitation of our study is the definition of transmission clusters by  
349 WGS. The threshold of 12 SNPs, which we used to exclude transmission, has been  
350 established by Walker *et al.* (17) and is in the range of what other studies have used  
351 (21, 44, 45). However, an adequate cluster definition may be adapted according to  
352 the setting (low versus high TB incidence), the study population and the technical  
353 specifications of the WGS analysis pipeline (i.e. whether particular genomic regions  
354 such as the PE/PGRS genes are excluded from the analysis, which was the case  
355 here). For comparison, we repeated the analyses with a stricter definition of  
356 transmission clusters also proposed by Walker *et al.* (MIRU-VNTR clusters in which  
357 at least one pairwise distance was  $\leq 5$  SNPs (17)), which further reduced the number  
358 of true clusters to 13, but did not change the overall clustering proportion significantly.  
359 A further limitation may be our sample size: we included 12.3% of all TB cases  
360 diagnosed between 2000 and 2008 in Switzerland, which potentially underestimates  
361 the overall clustering proportion (29). Indeed, SNP distances could in fact become  
362 shorter upon inclusion of additional patient isolates with intermediate genotypes,  
363 hence increasing the proportion of true clusters.

364 In conclusion, only one quarter of foreign-born transmission clusters previously  
365 identified by MIRU-VNTR were confirmed as true transmission clusters by WGS. We  
366 therefore recommend the use of WGS for a more accurate identification of recent  
367 transmission of *M. tuberculosis* among immigrants in low TB incidence countries, but  
368 also in high TB incidence countries, where genetically closely related strains  
369 circulate. Although analyzing WGS remains resource-intensive, the strategy adopted  
370 in the UK documents that implementing WGS in the routine public laboratory  
371 surveillance system is feasible (21, 46), and allows the prompt identification of

372 transmission clusters as well as information about the drug resistance genotype (46,  
373 47). Our results also indicate that the native population in Switzerland may also play  
374 a role in spreading TB, particularly individuals belonging to high risk populations (22,  
375 23). Additional prospective studies using WGS are needed, possibly complemented  
376 with social network analyses (20), to evaluate the usefulness of real-time analyses of  
377 TB transmission dynamics in low TB incidence countries.

378

379



380 **ACKNOWLEDGEMENTS**

381 We thank all tuberculosis patients for participating in this study, the treating  
382 physicians for providing clinical information and the microbiology laboratories for  
383 providing strains. We are grateful to the National TB Surveillance Registry at the  
384 Federal Office of Public Health.

385 This work was supported by the Swiss National Science Foundation (grant no.  
386 324730-12544 and PP00P3\_150750), the Swiss HIV Cohort Study (grant no. 588  
387 and 740), the Federal Office of Public Health (grant no. 09.007368), the National  
388 Center for Mycobacteria, and the European Research Council (grant no. 309540-  
389 EVODRTB). The Swiss HIV Cohort Study is supported by the Swiss National Science  
390 Foundation (grant no. 33CS30-134277). Sequencing calculations were performed at  
391 sciCORE (<http://scicore.unibas.ch/>) scientific computing core facility at the University  
392 of Basel. The funders had no role in study design, data collection and analysis,  
393 decision to publish, or preparation of the manuscript.

394  
395 **The members of the Molecular Epidemiology of Tuberculosis Study Group**

396 Central coordinating team:

397 Lukas Fenner and Matthias Egger, Institute of Social and Preventive Medicine,  
398 Bern; Sebastien Gagneux and Marcel Tanner, Swiss Tropical and Public Health  
399 Institute, Basel; Hansjakob Furrer, Inselspital Bern.

400 National Center for Mycobacteria:

401 Erik C. Böttger, Institute of Medical Microbiology, University of Zurich, Switzerland.

402 Microbiology laboratories:

403 Reno Frei, Clinical Microbiology, University Hospital Basel; Thomas Bodmer and  
404 Sara Droz, Institute for Infectious Diseases, University of Bern; Jacques Schrenzel,  
405 Laboratory of Bacteriology, University Hospitals of Geneva; Katia Jaton, Institute of

406 Microbiology, University Hospital of Lausanne; Hans Siegrist, ADMed Microbiology,  
407 La Chaux-de-Fonds; Gaby E. Pfyffer, Department of Medical Microbiology, Luzerner  
408 Kantonsspital, Luzern; Thomas Bruderer, Detlev Schultze, Centre for Laboratory  
409 Medicine, St.Gallen; Marisa Dolina, Clinical Microbiology, EOLAB, Bellinzona,  
410 Switzerland; Olivier Dubuis, Viollier AG Switzerland, Allschwil.

411 Swiss HIV Cohort Study:

412 Manuel Battegay, University Hospital Basel; Enos Bernasconi, Andrea Parini  
413 Lugano; Matthias Hoffmann, St.Gallen; Hansjakob Furrer, Inselspital Bern; Matthias  
414 Cavassini, University Hospital of Lausanne; Bernard Hirschel, Alexandra Calmy,  
415 University Hospital of Geneva; Jan Fehr, University Hospital of Zurich.

416 Respiratory clinics:

417 Jean-Paul Janssens, University Hospital of Geneva; Jesica Mazza Stalder,  
418 University Hospital of Lausanne.

419 Federal Office of Public Health:

420 Peter Helbling and Ekkehardt Altpeter, Division of Communicable Diseases.

421 Other:

422 Hans L. Rieder, Epidemiology, Biostatistics and Prevention Institute, University of  
423 Zürich, Switzerland.

424 **The members of the Swiss HIV Cohort Study**

425 Barth J, Battegay M, Bernasconi E, Böni J, Bucher HC, Burton-Jeangros C, Calmy  
426 A, Cavassini M, Celleraï C, Egger M, Elzi L, Fehr J, Fellay J, Flepp M, Francioli P  
427 (President of the SHCS), Furrer H (Chairman of the Clinical and Laboratory  
428 Committee), Fux CA, Gorgievski M, Günthard H (Chairman of the Scientific Board),  
429 Haerry D (deputy of "Positive Council"), Hasse B, Hirsch HH, Hirschel B, Hösli I,  
430 Kahlert C, Kaiser L, Keiser O, Kind C, Klimkait T, Kovari H, Ledergerber B,  
431 Martinetti G, Martinez de Tejada B, Metzner K, Müller N, Nadal D, Pantaleo G,

432 Rauch A, Regenass S, Rickenbach M (Head of Data Center), Rudin C (Chairman of  
433 the Mother & Child Substudy), Schmid P, Schultze D, Schöni-Affolter F, Schüpbach  
434 J, Speck R, Taffé P, Tarr P, Telenti A, Trkola A, Vernazza P, Weber R, Yerly S.

435

436 **ACCESSION NUMBERS**

437 Raw sequencing data were submitted to the European Nucleotide Archive (ENA)  
438 under project accession number PRJEB12179.

439

440 **CONFLICT OF INTEREST**

441 None to declare.

442

443

## 444 REFERENCES

- 445 1. **World Health Organization**. 2014. Global tuberculosis report 2014. WHO,  
446 Geneva, Switzerland.
- 447 2. **Falzon D, Kudjauw Y, Desenclos JC, Fernandez de la Hoz K, Dadu A,**  
448 **Zaleskis R**. 2008. Stopping TB in Europe: some progress but still not there.  
449 *Euro Surveill* **13**.
- 450 3. **Zellweger JP**. 2013. Current issues in the management of tuberculosis in  
451 Europe. *Panminerva Med* **55**:145–55.
- 452 4. **Federal Office of Public Health**. 2011. Tuberkulose in der Schweiz 2005-  
453 2009. *Bull BAG* 205–213.
- 454 5. **Abgrall S, del Giudice P, Melica G, Costagliola D, on behalf of FHDH-**  
455 **ANRS CO4 1**. 2010. HIV-associated tuberculosis and immigration in a high-  
456 income country: incidence trends and risk factors in recent years. *AIDS*  
457 **24**:763–771.
- 458 6. **Dahle UR, Eldholm V, Winje BA, Mannsåker T, Heldal E**. 2007. Impact of  
459 immigration on the molecular epidemiology of *Mycobacterium tuberculosis* in a  
460 low-incidence country. *Am J Respir Crit Care Med* **176**:930–935.
- 461 7. **Wolff H, Janssens JP, Bodenmann P, Meynard A, Delhumeau C, Rochat T,**  
462 **Sudre P, Costanza MC, Gaspoz JM, Morabia A**. 2010. Undocumented  
463 migrants in Switzerland: geographical origin versus legal status as risk factor  
464 for tuberculosis. *J Immigr Minor Health* **12**:18–23.
- 465 8. **Lillebaek T, Andersen AB, Bauer J, Dirksen A, Glismann S, de Haas P,**  
466 **Kok-Jensen A**. 2001. Risk of *Mycobacterium tuberculosis* transmission in a  
467 low-incidence country due to immigration from high-incidence areas. *J Clin*  
468 *Microbiol* **39**:855–61.
- 469 9. **Fenner L, Gagneux S, Helbling P, Battegay M, Rieder HL, Pfyffer GE,**  
470 **Zwahlen M, Furrer H, Siegrist HH, Fehr J, Dolina M, Calmy A, Stucki D,**  
471 **Jaton K, Janssens J-P, Stalder JM, Bodmer T, Ninet B, Böttger EC, Egger**  
472 **M**. 2012. *Mycobacterium tuberculosis* transmission in a country with low  
473 tuberculosis incidence: role of immigration and HIV infection. *J Clin Microbiol*  
474 **50**:388–395.
- 475 10. **Barniol J, Niemann S, Louis VR, Brodhun B, Dreweck C, Richter E,**  
476 **Becher H, Haas W, Junghanss T**. 2009. Transmission dynamics of pulmonary  
477 tuberculosis between autochthonous and immigrant sub-populations. *BMC*  
478 *Infect Dis* **9**:197.
- 479 11. **Kamper-Jorgensen Z, Andersen AB, Kok-Jensen A, Bygbjerg IC,**  
480 **Andersen PH, Thomsen VO, Kamper-Jorgensen M, Lillebaek T**. 2012.  
481 Clustered tuberculosis in a low-burden country: nationwide genotyping through  
482 15 years. *J Clin Microbiol* **50**:2660–2667.
- 483 12. **Guzman Herrador BR, Rønning K, Borgen K, Mannsåker T, Dahle UR**.  
484 2015. Description of the largest cluster of tuberculosis notified in Norway 1997-  
485 2011: is the Norwegian tuberculosis control programme serving its purpose for  
486 high risk groups? *BMC Public Health* **15**:367.
- 487 13. **Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C**. 2000.

- 488 Variable human minisatellite-like regions in the *Mycobacterium tuberculosis*  
489 genome. *Mol Microbiol* **36**:762–771.
- 490 14. **Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D,**  
491 **Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J.**  
492 1997. Simultaneous detection and strain differentiation of *Mycobacterium*  
493 *tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* **35**:907–914.
- 494 15. **Kato-Maeda M, Metcalfe JZ, Flores L.** 2011. Genotyping of *Mycobacterium*  
495 *tuberculosis*: application in epidemiologic studies. *Future Microbiol* **6**:203–216.
- 496 16. **Munang ML, Browne C, Khanom S, Evans JT, Smith EG, Hawkey PM,**  
497 **Kunst H, Welch SB, Dedicoat MJ.** 2015. Tuberculosis microepidemics among  
498 dispersed migrants, Birmingham, UK, 2004–2013. *Emerg Infect Dis* **21**:524–7.
- 499 17. **Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre**  
500 **DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS,**  
501 **Bowden R, Monk P, Smith EG, Peto TEA.** 2013. Whole-genome sequencing  
502 to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective  
503 observational study. *Lancet Infect Dis* **13**:137–146.
- 504 18. **Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PEM, Godfrey-**  
505 **Faussett P, Vynnycky E, European Concerted Action on Molecular**  
506 **Epidemiology and Control of Tuberculosis.** 1999. Interpreting DNA  
507 fingerprint clusters of *Mycobacterium tuberculosis*. *Int J Tuberc Lung Dis*  
508 **3**:1055–1060(6).
- 509 19. **Wampande EM, Mupere E, Debanne SM, Asiimwe BB, Nsereko M,**  
510 **Mayanja H, Eisenach K, Kaplan G, Boom HW, Sebastien G, Joloba ML.**  
511 2013. Long-term dominance of *Mycobacterium tuberculosis* Uganda family in  
512 peri-urban Kampala-Uganda is not associated with cavitary disease. *BMC*  
513 *Infect Dis* **13**:484.
- 514 20. **Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, Rempel**  
515 **S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood**  
516 **K, Jones SJM, Brinkman FSL, Brunham RC, Tang P.** 2011. Whole-genome  
517 sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J*  
518 *Med* **364**:730–739.
- 519 21. **Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, Parker L, Churchill**  
520 **S, Bennett K, Golubchik T, Giess AP, Del Ojo Elias C, Jeffery KJ, Bowler**  
521 **ICJW, Laurenson IF, Barrett A, Drobniewski F, McCarthy ND, Anderson**  
522 **LF, Abubakar I, Thomas HL, Monk P, Smith EG, Walker AS, Crook DW,**  
523 **Peto TEA, Conlon CP.** 2014. Assessment of *Mycobacterium tuberculosis*  
524 transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome  
525 sequences: an observational study. *Lancet Respir Med* **2**:285–292.
- 526 22. **Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, Wirth T, Jaenicke**  
527 **S, Schuback S, Rüsche-Gerdes S, Supply P, Kalinowski J, Niemann S.**  
528 2013. Whole genome sequencing versus traditional genotyping for  
529 investigation of a *Mycobacterium tuberculosis* outbreak: A longitudinal  
530 molecular epidemiological study. *PLoS Med* **10**:e1001387.
- 531 23. **Stucki D, Ballif M, Bodmer T, Coscolla M, Maurer A-M, Droz S, Butz C,**  
532 **Borrell S, Längle C, Feldmann J, Furrer H, Mordasini C, Helbling P, Rieder**  
533 **HL, Egger M, Gagneux S, Fenner L.** 2015. Tracking a tuberculosis outbreak

- 534 over 21 years: strain-specific single-nucleotide polymorphism typing combined  
535 with targeted whole-genome sequencing. *J Infect Dis* **211**:1306–1316.
- 536 24. **Coscolla M, Barry PM, Oeltmann JE, Koshinsky H, Shaw T, Cilnis M,**  
537 **Posey J, Rose J, Weber T, Fofanov VY, Gagneux S, Kato-Maeda M,**  
538 **Metcalf JZ.** 2015. Genomic epidemiology of multidrug-resistant  
539 *Mycobacterium tuberculosis* during transcontinental spread. *J Infect Dis*  
540 **212**:302–310.
- 541 25. **Fenner L, Egger M, Bodmer T, Altpeter E, Zwahlen M, Jatton K, Pfyffer GE,**  
542 **Borrell S, Dubuis O, Bruderer T, Siegrist HH, Furrer H, Calmy A, Fehr J,**  
543 **Stalder JM, Ninet B, Böttger EC, Gagneux S.** 2012. Effect of mutation and  
544 genetic background on drug resistance in *Mycobacterium tuberculosis*.  
545 *Antimicrob Agents Chemother* **56**:3047–53.
- 546 26. **Fenner L, Gagneux S, Janssens J-P, Fehr J, Cavassini M, Hoffmann M,**  
547 **Bernasconi E, Schrenzel J, Bodmer T, Böttger EC, Helbling P, Egger M,**  
548 **for the Swiss HIV Cohort and Molecular Epidemiology of Tuberculosis**  
549 **Study Groups.** 2012. Tuberculosis in HIV-negative and HIV-infected patients  
550 in a low-incidence country: clinical characteristics and treatment outcomes.  
551 *PLoS One* **7**:e34186.
- 552 27. **Fenner L, Malla B, Ninet B, Dubuis O, Stucki D, Borrell S, Huna T, Bodmer**  
553 **T, Egger M, Gagneux S.** 2011. “Pseudo-Beijing”: Evidence for convergent  
554 evolution in the direct repeat region of *Mycobacterium tuberculosis*. *PLoS One*  
555 **6**:e24737.
- 556 28. **Fenner L, Egger M, Bodmer T, Furrer H, Ballif M, Battegay M, Helbling P,**  
557 **Fehr J, Gsponer T, Rieder HL, Zwahlen M, Hoffmann M, Bernasconi E,**  
558 **Cavassini M, Calmy A, Dolina M, Frei R, Janssens J-P, Borrell S, Stucki D,**  
559 **Schrenzel J, Böttger EC, Gagneux S, for the Swiss HIV Cohort and**  
560 **Molecular Epidemiology of Tuberculosis Study Groups.** 2013. HIV infection  
561 disrupts the sympatric host–pathogen relationship in human tuberculosis. *PLoS*  
562 *Genet* **9**:e1003318.
- 563 29. **Glynn JR, Vynnycky E, Fine PE.** 1999. Influence of sampling on estimates of  
564 clustering and recent transmission of *Mycobacterium tuberculosis* derived from  
565 DNA fingerprinting techniques. *Am J Epidemiol* **149**:366–71.
- 566 30. **Somoskovi A, Helbling P, Deggim V, Hömke R, Ritter C, Böttger EC.** 2014.  
567 Transmission of multidrug-resistant tuberculosis in a low-incidence setting,  
568 Switzerland, 2006 to 2012. *Eurosurveillance* **19**.
- 569 31. **Steiner A, Stucki D, Coscolla M, Borrell S, Gagneux S.** 2014. KvarQ:  
570 targeted and direct variant calling from fastq reads of bacterial genomes. *BMC*  
571 *Genomics* **15**:881.
- 572 32. **Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, Kremer K,**  
573 **Ernst JD, Gagneux S.** 2010. Human T cell epitopes of *Mycobacterium*  
574 *tuberculosis* are evolutionarily hyperconserved. *Nat Genet* **42**:498–503.
- 575 33. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-  
576 Wheeler transform. *Bioinformatics* **25**:1754–1760.
- 577 34. **Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X.**  
578 2012. Using *Drosophila melanogaster* as a model for genotoxic chemical  
579 mutational studies with a new program, SnpSift. *Front Genet* **3**.



- 580 35. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011.  
581 MEGA5: molecular evolutionary genetics analysis using maximum likelihood,  
582 evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*  
583 **28**:2731–2739.
- 584 36. **R Core Team.** 2014. R: a language and environment for statistical computing.  
585 Vienna, Austria.
- 586 37. **Wickham H.** 2009. ggplot2: elegant graphics for data analysis. Springer-Verlag  
587 New York.
- 588 38. **Fenner L, Gagneux S, Janssens J-P, Fehr J, Cavassini M, Hoffmann M,**  
589 **Bernasconi E, Schrenzel J, Bodmer T, Böttger EC, Helbling P, Egger M.**  
590 2012. Tuberculosis in HIV-negative and HIV-infected patients in a low-  
591 incidence country: clinical characteristics and treatment outcomes. *PLoS One*  
592 **7**:e34186.
- 593 39. **Cook VJ, Shah L, Gardy J, Bourgeois A-C.** 2012. Recommendations on  
594 modern contact investigation methods for enhancing tuberculosis control. *Int J*  
595 *Tuberc Lung Dis* **16**:297–305.
- 596 40. **De Beer JL, Kodmon C, van der Werf MJ, van Ingen J, van Soolingen D,**  
597 **ECDC MDR-TB Molecular Surveillance Project Participants.** 2014.  
598 Molecular surveillance of multi- and extensively drug-resistant tuberculosis  
599 transmission in the European Union from 2003 to 2011. *Euro Surveill Bull Eur*  
600 *sur les Mal Transm = Eur Commun Dis Bull* **19**.
- 601 41. **Huyen MNT, Kremer K, Lan NTN, Cobelens FGJ, Buu TN, Dung NH, Caws**  
602 **M, Tiemersma EW, van Soolingen D.** 2012. Mixed tuberculosis infections in  
603 rural South Vietnam. *J Clin Microbiol* **50**:1586–92.
- 604 42. **Cohen T, Wilson D, Wallengren K, Samuel EY, Murray M.** 2011. Mixed-  
605 strain *Mycobacterium tuberculosis* infections among patients dying in a hospital  
606 in KwaZulu-Natal, South Africa. *J Clin Microbiol* **49**:385–8.
- 607 43. **Mallard K, McNerney R, Crampin AC, Houben R, Ndlovu R, Munthali L,**  
608 **Warren RM, French N, Glynn JR.** 2010. Molecular detection of mixed  
609 infections of *Mycobacterium tuberculosis* strains in sputum samples from  
610 patients in Karonga District, Malawi. *J Clin Microbiol* **48**:4512–8.
- 611 44. **Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea**  
612 **TP, Walker BJ, Bantubani N, Almeida D V, Alvarado L, Chapman SB,**  
613 **Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S,**  
614 **Howarth C, Larimer JD, Maharaj K, Pearson MD, Priest ME, Zeng Q,**  
615 **Padayatchi N, Grosset J, Young SK, Wortman J, Mlisana KP, O'Donnell**  
616 **MR, Birren BW, Bishai WR, Pym AS, Earl AM.** 2015. Evolution of extensively  
617 drug-resistant tuberculosis over four decades: whole genome sequencing and  
618 dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal.  
619 *PLoS Med* **12**:e1001880.
- 620 45. **Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, van Helden P,**  
621 **Pym A, Mahayiddin AA, Chuchottaworn C, Sanne IM, Louw C, Boeree MJ,**  
622 **Hoelscher M, McHugh TD, Bateson ALC, Hunt RD, Mwaigwisya S, Wright**  
623 **L, Gillespie SH, Bentley SD.** 2013. Whole-genome sequencing to establish  
624 relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective  
625 observational study. *Lancet Respir Med* **1**:786–792.

- 626 46. Pankhurst LJ, del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies  
627 J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N,  
628 Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P,  
629 Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW. 2015. Rapid,  
630 comprehensive, and affordable mycobacterial diagnosis with whole-genome  
631 sequencing: a prospective study. *Lancet Respir Med*.
- 632 47. Walker TM, Kohl TA, Omar S V, Hedge J, Del Ojo Elias C, Bradley P, Iqbal  
633 Z, Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CLC,  
634 Bowden R, Drobniewski FA, Allix-Béguec C, Gaudin C, Parkhill J, Diel R,  
635 Supply P, Crook DW, Smith EG, Walker AS, Ismail N, Niemann S, Peto  
636 TEA. 2015. Whole-genome sequencing for prediction of *Mycobacterium*  
637 *tuberculosis* drug susceptibility and resistance: a retrospective cohort study.  
638 *Lancet Infect Dis* **15**:1193–1202.
- 639



640 **TABLES**

641 **Table 1.** Patient characteristics of tuberculosis (TB) cases diagnosed in Switzerland  
 642 between 2000 and 2008, overall and comparing clustered and unclustered TB  
 643 cases (unweighted), as well as risk factors for transmission (weighted analysis) as  
 644 defined by genome-based molecular clustering.

Characteristic	Unweighted (n=520)						Weighted (n=4,221)	
	All cases n=520		Clustered n=43		Unclustered n=477		Adjusted Odds Ratio (95% CI) *	P-value
	n	(%)	n	(%)	n	(%)		
<b>Age at TB diagnosis, years</b>								0.46
16-29	154	(29.6)	14	(32.6)	140	(29.4)	1	
30-49	226	(43.5)	16	(37.2)	210	(44.0)	0.56 (0.22-1.41)	
≥50	140	(26.9)	13	(30.2)	127	(26.6)	0.71 (0.29-1.73)	
<b>Sex</b>								0.018
Male	254	(48.8)	26	(60.5)	228	(47.8)	1	
Female	266	(51.2)	17	(39.5)	249	(52.2)	0.4 (0.18-0.85)	
<b>Country of birth</b>								0.026
Switzerland	119	(22.9)	17	(39.5)	102	(21.4)	1	
Europe, without Switzerland	106	(20.4)	2	(4.7)	104	(21.8)	0.16 (0.03-0.83)	
Asia	120	(23.1)	3	(7.0)	117	(24.5)	0.24 (0.05-1.08)	
Sub-Saharan Africa	132	(25.4)	16	(37.2)	116	(24.3)	1.11 (0.36-3.45)	
Central and South America	20	(3.8)	3	(7.0)	17	(3.6)	1.99 (0.40-9.85)	
Other regions/Unknown	23	(4.4)	2	(4.7)	21	(4.4)	0.80 (0.13-5.02)	
<b>Born in Switzerland</b>								0.091
No	401	(77.1)	26	(60.5)	375	(78.6)	1	
Yes	119	(22.9)	17	(39.5)	102	(21.4)	2.21 (0.88-5.52)	
<b>Residence status</b>								0.41
Swiss nationals	130	(25.0)	14	(32.6)	116	(24.3)	1	
Foreigner/Residents	199	(38.3)	11	(25.6)	188	(39.4)	0.45 (0.17-1.23)	
Asylum seeker/Refugee	101	(19.4)	8	(18.6)	93	(19.5)	0.42 (0.11-1.56)	
Unknown	90	(17.3)	10	(23.3)	80	(16.8)	0.80 (0.22-2.81)	
<b>HIV status</b>								0.033
Negative	407	(78.3)	34	(79.1)	373	(78.2)	1	
Positive	113	(21.7)	9	(20.9)	104	(21.8)	0.39 (0.16-0.93)	
<b>Site of disease</b>								0.054
Extra-pulmonary TB	133	(25.6)	5	(11.6)	128	(26.8)	1	
Pulmonary TB	387	(74.4)	38	(88.4)	349	(73.2)	3.14 (0.98-10.07)	
<b>Cavitary disease</b>								0.039
No	407	(78.3)	28	(65.1)	379	(79.5)	1	
Yes	113	(21.7)	15	(34.9)	98	(20.5)	2.31 (1.05-5.10)	
<b>Positive smear microscopy</b>								0.049
No	352	(67.7)	21	(48.8)	331	(69.4)	1	
Yes	168	(32.3)	22	(51.2)	146	(30.6)	2.18 (1.00-4.73)	
<b>TB in family or social surroundings in last 2 years</b>								0.068
No	482	(92.7)	38	(88.4)	444	(93.1)	1	
Yes	38	(7.3)	5	(11.6)	33	(6.9)	2.95 (0.92-9.41)	

645 \* Adjusted for age and sex, and weighted for sampling proportions

646 95% CI, 95% confidence interval; TB, tuberculosis

647 **Table 2. Number of MIRU-VNTR clusters and number of patients in MIRU-**  
 648 **VNTR clusters confirmed by whole genome sequencing.** MIRU-VNTR clusters  
 649 not confirmed by WGS ("false positive" clusters) and clusters confirmed by WGS  
 650 ("true" clusters) are presented, according to the countries of birth of cases involved:  
 651 clusters involving Swiss-born TB cases only, foreign-born cases only, or mixed  
 652 clusters (e.g., involving both Swiss-born and foreign-born TB cases, or foreign-born  
 653 cases from different continents).

	MIRU-VNTR clusters confirmed by WGS n (%)	MIRU-VNTR clusters not confirmed by WGS n (%)	Total number of clusters n
<b>Molecular clusters *</b>			
Swiss-born only	3 (75.0)	1 (25.0)	4
Mixed	10 (66.7)	5 (33.3)	15
Foreign-born only	4 (25.0)	12 (75.0)	16
Total	17 (48.6)	18 (51.4)	<b>35</b>
<b>Patients in molecular clusters **</b>			
Swiss-born only	8 (80.0)	2 (20.0)	10
Mixed	27 (62.8)	16 (37.2)	43
Foreign-born only	8 (21.6)	29 (78.4)	37
Total	43 (47.8)	47 (52.2)	<b>90</b>

654 MIRU-VNTR, mycobacterial interspersed repetitive-unit-variable-number tandem-repeat;  
 655 WGS, whole genome sequencing

656  
 657 \* Fisher's exact test: p-value=0.031

658 \*\* Fisher's exact test: p-value <0.0001

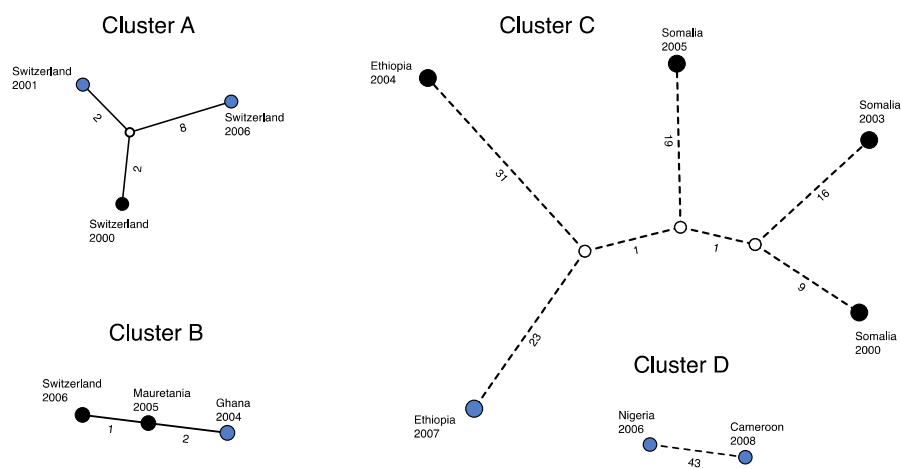
659

660

661 **FIGURES**

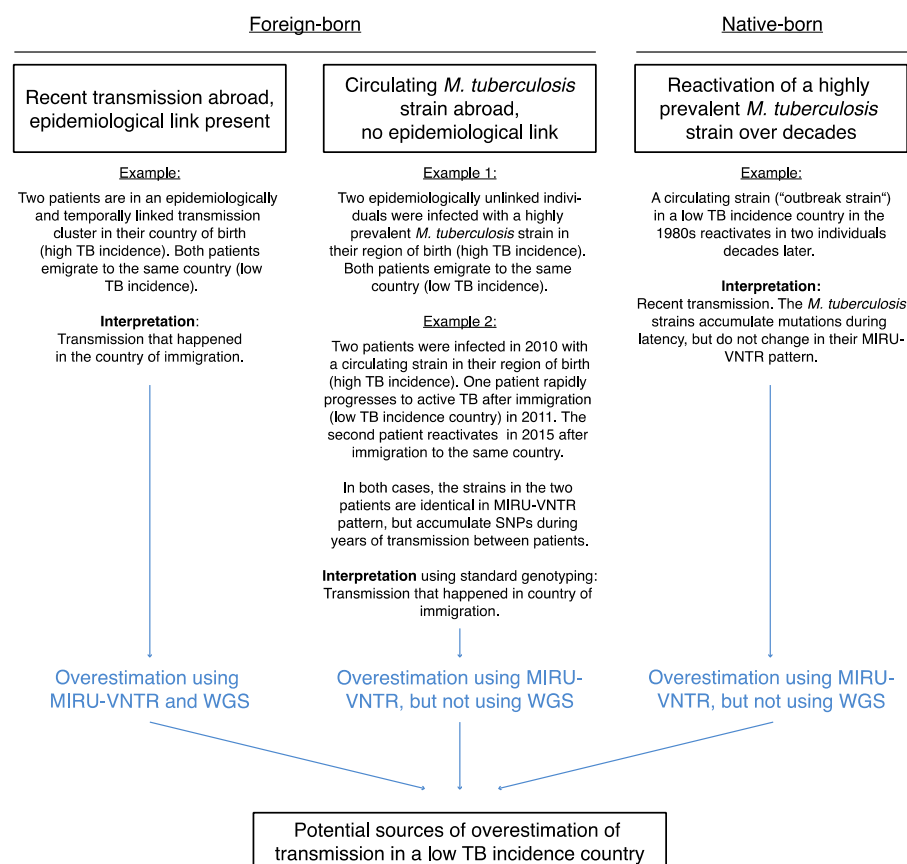
662

663 **Figure 1. *M. tuberculosis* transmission network using SNP data from whole**  
664 **genome sequencing.** Representative examples of different types of transmission  
665 clusters that were previously identified by MIRU-VNTR: cluster A (all patients from  
666 Switzerland) was confirmed as a true transmission cluster by WGS, with distances  
667 of  $\leq 12$  SNPs between all isolates. Cluster B is a mixed cluster (one patient from  
668 Switzerland and two patients from West Africa), which was confirmed to be a true  
669 transmission cluster with one and two SNPs between patient isolates. Clusters C  
670 and D were identified by MIRU-VNTR as transmission clusters, but WGS did not  
671 confirm these clusters as true clusters (genetic distances of  $> 12$  SNPs). Filled  
672 circles represent patient isolates, white circles „median vectors“, i.e. hypothetical  
673 isolates inferred from the sequencing data. Blue circles indicate HIV-positive  
674 patients. Numbers next to lines indicate SNP distances. Countries of birth and years  
675 of tuberculosis diagnosis are indicated next to circles. Clusters with solid lines are  
676 “true” clusters, i.e. clusters confirmed by WGS ( $\leq 12$  SNPs), whereas clusters with  
677 dashed lines are clusters not confirmed by WGS ( $> 12$  SNPs). All other transmission  
678 clusters are shown in Supplementary Figure 1.



679  
680

681 **Figure 2. Possible situations leading to an overestimation of *M. tuberculosis***  
 682 **transmission in a low tuberculosis incidence country.**



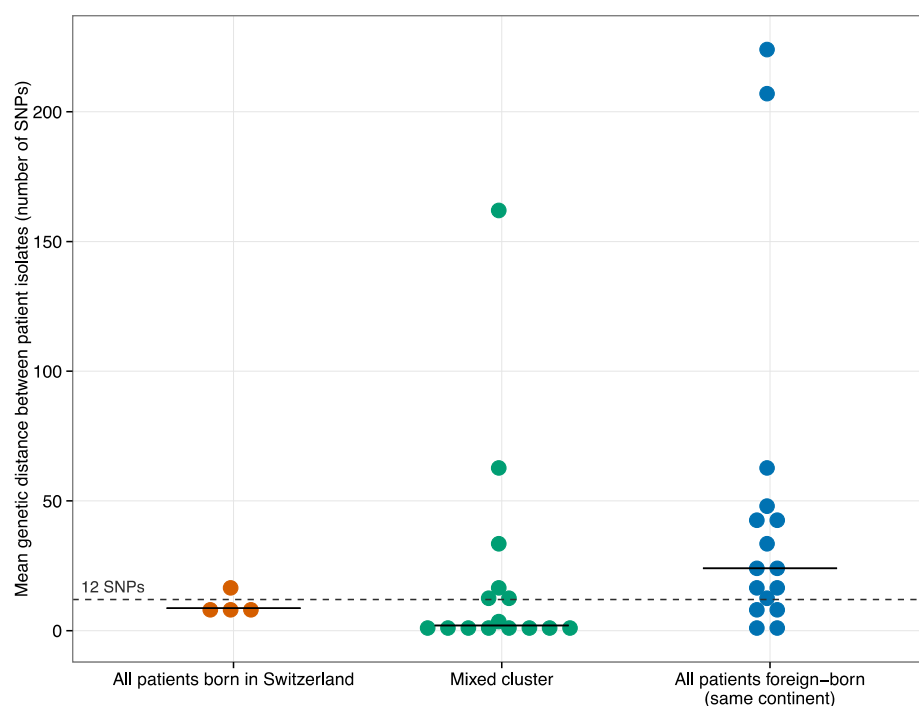
683

684 MIRU-VNTR, mycobacterial interspersed repetitive-unit–variable-number tandem-

685 repeat; TB, tuberculosis; WGS, whole genome sequencing

686

687 **Figure 3. Median genetic distance in MIRU-VNTR-defined transmission**  
688 **clusters.** Each data point shows the mean genetic distance (as number of single  
689 nucleotide polymorphisms [SNPs]) in one of the 35 MIRU-VNTR-defined  
690 transmission clusters. Solid black lines indicate median values of mean pairwise  
691 distances within molecular clusters. The distribution of clusters with patients born in  
692 Switzerland, mixed clusters and immigrant patients born on the same continent  
693 (except Switzerland) was significantly different (Kruskal-Wallis non-parametric test,  
694  $p=0.030$ ). The dashed line represents the cut-off for the definition of genome-based  
695 molecular clusters ( $\geq 12$  SNPs).



696

697