

1 **High proportion of tuberculosis transmission among social contacts**  
2 **in rural China: a 12-year prospective population-based genomic**  
3 **epidemiological study**

4  
5 Meng Li<sup>1\*</sup>, Mingcheng Guo<sup>2\*</sup>, Ying Peng<sup>3\*</sup>, Qi Jiang<sup>1,4</sup>, Lan Xia<sup>5</sup>, Sheng Zhong<sup>6</sup>,  
6 Yong Qiu<sup>2</sup>, Xin Su<sup>6</sup>, Shu Zhang<sup>5</sup>, Chongguang Yang<sup>1,7</sup>, Peierdun Mijiti<sup>1</sup>, Qizhi Mao<sup>1</sup>,  
7 Howard Takiff<sup>8,9</sup>, Fabian Li<sup>3†</sup>, Chuang Chen<sup>5†</sup>, Qian Gao<sup>1†</sup>

8  
9 <sup>1</sup> Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS), School of  
10 Basic Medical Science, Shanghai Medical College, Shanghai Institute of Infectious  
11 Disease and Biosecurity, Fudan University, Shanghai, China

12 <sup>2</sup> Wusheng County Center for Disease Control and Prevention, Guangan, China

13 <sup>3</sup> Heilongjiang Provincial Center for Tuberculosis Prevention and Control, Harbin,  
14 China

15 <sup>4</sup> School of Public Health, Wuhan University, Wuhan, China

16 <sup>5</sup> Institution for Tuberculosis Prevention and Control, Sichuan Provincial Center for  
17 Disease Control and Prevention, Chengdu, China

18 <sup>6</sup> Wuchang City Center for Tuberculosis Control and Prevention, Harbin, China

19 <sup>7</sup> School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University,  
20 Shenzhen, China

21 <sup>8</sup> Unité de Pathogenetique Intégrée Mycobacterienne, Institut Pasteur, Paris, France

22 <sup>9</sup> Laboratorio de Genética Molecular, CMBC, IVIC, Caracas, Venezuela

23  
24 \* These authors contributed equally to this study.

25 † These authors contributed equally to this study.

26  
27 **Word count:** Abstract 250; Text: 3448

28  
29 **Correspondence to:**

30 Dr. Qian Gao, Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS),  
31 School of Basic Medical Science, Shanghai Medical College, Shanghai Institute of  
32 Infectious Disease and Biosecurity, Fudan University, Shanghai, China, 131 Dongan  
33 Road, Shanghai, China 200032, E-mail: [qiangao@fudan.edu.cn](mailto:qiangao@fudan.edu.cn)

34  
35 Chuang Chen, Institution for Tuberculosis Prevention and Control, Sichuan Provincial  
36 Center for Disease Control and Prevention, 6 Zhongxue Road, Chengdu, China

37 610041, E-mail: [chenchuang67@sina.com](mailto:chenchuang67@sina.com)

38

39 Fabin Li, Heilongjiang Provincial Center for Tuberculosis Prevention and Control, 40

40 Youfang Road, Harbin, China 150036, E-mail: [hljlf@163.com](mailto:hljlf@163.com)

41 **Evidence before this study**

42 We searched PubMed for genomic epidemiological studies of *Mycobacterium*  
43 *tuberculosis* published in English before April 2022 employing whole-genome  
44 sequencing, using the search terms “tuberculosis”, “transmission”, “population based”,  
45 and “whole-genome sequencing”. We identified only 11 studies in which  
46 whole-genome sequencing was used to investigate transmission of *M tuberculosis* at  
47 the population level. We also searched the China national knowledge infrastructure  
48 (CNKI) and WANFANG databases with the same search terms for papers published  
49 in Chinese, but did not identify any studies. The duration of most of the 11 studies we  
50 identified was less than 5 years. Seven studies conducted epidemiological  
51 investigations of genomic-clustered cases, but the proportion of cases with confirmed  
52 epidemiological links was very low. Therefore, no studies had sufficient evidence to  
53 identify populations and sites at high risk of TB transmission. Five studies were  
54 conducted in China but all were in urban areas and focused on MDR-TB patients and  
55 internal migrants. The pattern of TB transmission in rural China, where TB is more  
56 prevalent, had not been addressed.

57

58 **Added value of this study**

59 To our knowledge, ours is the first population-based genomic epidemiological study  
60 to delineate TB transmission patterns in rural China. Close contacts have been shown  
61 to be a high-risk group for TB transmission in other countries. In China, however, the  
62 huge number of TB patients, limited resources for TB prevention and control and the  
63 stigma associated with tuberculosis all contribute to a failure to identify and screen  
64 many close contacts. As a consequence, close contacts have been calculated to  
65 contribute only about 2% of the total TB burden. In this study, through the  
66 investigation of genomic-clustered patients, we found at least 41·8% of clustered  
67 patients were close contacts who comprised 9·9% of the total TB patients in the study.  
68 Moreover, more than two-thirds of the close contacts were social contacts rather than  
69 members of the immediate family. The composition of social contacts differed  
70 between the two study sites due to differences in climate and lifestyle habits. The  
71 average time interval between the diagnosis of clustered student contacts was shorter  
72 than for family or community contacts. In addition, transmission of MDR strains was  
73 likely responsible for 81·4% of MDR-TB cases, with minimal acquisition of  
74 additional resistance mutations. Our 12-year study identified patterns of TB  
75 transmission not identified by previous studies, demonstrating the value of long-term  
76 genomic epidemiological studies.

77

78 **Implications of all the available evidence**

79 Our study demonstrates that much of the transmission of TB in rural China was

80 among close contacts, especially social contacts. Therefore, strengthening and  
81 improving proactive screening of close social contacts can identify more TB patients  
82 and shorten the time to patient detection. We believe that this type of vigorous active  
83 case-finding is essential for reducing TB transmission and the considerable TB burden  
84 in China. Long-term prospective genomic epidemiological studies provide a useful  
85 picture of TB transmission patterns that can help guide the design of strategies to  
86 improve TB prevention and control.

87 **Summary**

88 **Background** Tuberculosis (TB) is more prevalent in rural than urban areas in China,  
89 and delineating TB transmission patterns in rural populations could improve TB  
90 control.

91  
92 **Methods** We conducted a prospective population-based study of culture-positive  
93 pulmonary TB patients diagnosed between July 1, 2009 and December 31, 2020 in  
94 two rural counties in China. Genomic clusters were defined with a threshold distance  
95 of 12-single-nucleotide-polymorphisms, based on whole-genome sequencing. Risk  
96 factors for clustering were identified by logistic regression. Transmission links were  
97 sought through epidemiological investigation of genomic-clustered patients.

98  
99 **Findings** Of 1517 and 751 culture-positive pulmonary TB patients in Wusheng and  
100 Wuchang counties, respectively, 1289 and 699 strains were sequenced. Overall, 624  
101 (31.4%, 624/1988) patients were grouped into 225 genomic clusters. Epidemiological  
102 links were confirmed in 41.8% (196/469) of clustered isolates, including family  
103 (32.7%, 64/196) and social contacts (67.3%, 132/196). Social contacts were generally  
104 with relatives, within the community or in shared aggregated settings outside the  
105 community, but the proportion of clustered contacts in each category differed between  
106 the two sites. The time interval between diagnosis of student cases and contacts was  
107 significantly shorter than family and social contacts, probably due to enhanced student  
108 contact screening. Transmission of multidrug-resistant strains was likely responsible  
109 for 81.4% (83/102) of MDR-TB cases, with minimal acquisition of additional  
110 resistance mutations.

111  
112 **Interpretation** A large proportion of TB transmission in rural China occurred among  
113 social contacts, suggesting that active screening and aggressive contact tracing could  
114 benefit TB control, but contact screening should be tailored to local patterns of social  
115 interactions.

116  
117 **Funding** National Science and Technology Major Project of China, Natural Science  
118 Foundation of China, and Science and Technology Major Project of Shanghai

119 **Introduction**

120 Tuberculosis (TB) remains an important threat to global health, with an estimated 10  
121 million people worldwide falling ill and at least 1-4 million deaths in 2019.<sup>1</sup> Although  
122 the global incidence of TB has declined in recent years, there is still a significant gap  
123 between current rates of decline and the goals of the WHO End TB Strategy.<sup>2,3</sup>

124

125 It is estimated that approximately one-third of TB patients worldwide were not  
126 diagnosed,<sup>1</sup> so measures that increase diagnosis, especially early in the disease, are  
127 key to reducing transmission and thereby controlling TB. One of the core strategies to  
128 increase TB diagnosis and achieve global TB control is active case-finding.<sup>4</sup> The  
129 World Health Organization (WHO) recommends systematic TB screening of the  
130 general population in areas with an estimated TB prevalence of 0.5% or higher.<sup>5</sup> For  
131 areas with a lower TB prevalence, a cost-effective strategy involves identifying  
132 populations at high risk for transmission and then implementing targeted screening.

133

134 Genomic-epidemiological studies of tuberculosis can help identify populations at high  
135 risk of transmission, but the limited discriminatory power of previously employed  
136 genotyping methods and the barriers to epidemiological investigations have made it  
137 difficult to define these populations in China.<sup>6</sup> The populations at high risk for TB in  
138 other countries, such as HIV-positive, homeless, and drug abusing individuals, are  
139 less important contributors to the TB burden in China.<sup>5,7</sup> Instead, TB screening in  
140 China is focused principally on the elderly and internal migrants, but because of the  
141 large size of these populations in China, generalized screening is not practical. Most  
142 genomic-epidemiological studies of tuberculosis in China have been conducted in  
143 large cities, but China also has extensive rural areas where the prevalence of TB is up  
144 to three times higher than in urban regions.<sup>8</sup> Therefore, identifying high-risk  
145 populations and delineating transmission patterns in rural areas is important for  
146 improving TB control in China. To define the transmission patterns of the rural TB  
147 burden and thereby provide guidance for improving TB control, we conducted a  
148 12-year prospective population-based genomic-epidemiological study in two rural  
149 counties in China.

150

151 **Methods**

152 **Study design and participants**

153 Wusheng and Wuchang are rural counties in southwestern and northeastern China,  
154 respectively. Wusheng is located in the east of Sichuan Province, with an area of 960  
155 square kilometers and an estimated 556,000 inhabitants in 2020. Wuchang is located  
156 in the south of Heilongjiang Province, with an area of 7,512 square kilometers. The  
157 study included 14 townships (2,299.5 square kilometers) in Wuchang, with a

158 combined estimated 419,000 inhabitants in 2020. The average annual reported TB  
159 incidences were 95.8 (per 100,000 population) in Wusheng and 68.9 in Wuchang.  
160 Lower incidences recorded in 2020 were probably inaccurate and a result of reduced  
161 TB diagnosis due to the COVID-19 pandemic.

162

163 The study population was comprised of all culture-positive pulmonary TB patients 15  
164 years or older who were diagnosed between July 1, 2009 and December 31, 2020 in  
165 the regions included in the study. The study was approved by the institutional review  
166 board of the Institutes of Biomedical Sciences, Fudan University and all enrolled  
167 patients provided written informed consent.

168

### 169 **Diagnostic procedures**

170 The Wusheng County Center for Disease Control and Prevention and the Wuchang  
171 City Center for Tuberculosis Control and Prevention are responsible for the diagnosis  
172 and treatment of TB in their respective areas. Individuals with TB-like symptoms or  
173 abnormal chest radiographs in general hospitals and township health centers are  
174 referred to these designated medical institutions for diagnosis by sputum smear and  
175 culture. Sputum specimens were collected when the patients presented for TB  
176 diagnosis, before starting treatment, and culture-positive isolates were stored in the  
177 site laboratory.

178

### 179 **Whole-genome sequencing**

180 All stored clinical strains isolated during the study period were re-cultured from  
181 frozen stocks. DNA was extracted using the CTAB method and sequenced as  
182 previously described.<sup>9</sup> Raw sequence reads were trimmed with Sickle (version 1.33)  
183 and aligned to the inferred *Mycobacterium tuberculosis* complex ancestor sequence<sup>10</sup>  
184 using BWA-MEM. SAMtools (version 1.3.1) and Varscan (version 2.3.6) were then  
185 used to identify single nucleotide polymorphisms (SNPs). Pairwise SNP distances  
186 were calculated based on fixed SNPs (frequency  $\geq 75\%$ ), excluding those in  
187 drug-resistance associated genes and repetitive regions of the genome (e.g.,  
188 PPE/PE-PGRS family genes, phage sequences, and insertion or mobile genetic  
189 elements). A genomic cluster was defined as strains differing by 12 or fewer SNPs,  
190 consistent with linkage through recent transmission.<sup>11</sup> Sequencing data were deposited  
191 in the Genome Sequence Archive (<https://bigd.big.ac.cn/gsa>) under BioProject  
192 PRJCA008815 and PRJCA008816.

193

194 Based on the identified SNPs, a phylogeny tree was constructed with RAxML-NG  
195 (version 1.0.2) software, using the maximum likelihood method with 100 bootstraps  
196 and visualized with Interactive Tree of Life (<https://itol.embl.de/>). Strains were

197 classified into different lineages according to Coll et al.<sup>12</sup> Strains belonging to lineage  
198 2 (L2), also termed the Beijing family, were divided into L2-3, representing “modern”  
199 Beijing, and other sub-lineages considered “ancient” Beijing.<sup>13</sup>

200

201 Drug-resistance profiles were predicted for 14 anti-TB drugs based on the mutations  
202 reported to be associated with resistance,<sup>14</sup> using Whole-genome sequencing (WGS)  
203 data as described previously.<sup>15</sup> Pan-susceptible TB was defined as susceptibility to the  
204 four first-line drugs (isoniazid, rifampicin, ethambutol and pyrazinamide).

205 Multidrug-resistant TB (MDR-TB) was defined as resistance to at least isoniazid and  
206 rifampicin. Strains with resistance to any of the four first line drugs but not MDR  
207 were termed other drug resistant (DR) TB.

208

### 209 **Epidemiological investigation**

210 We conducted epidemiological investigations twice. First, at the time of TB diagnosis  
211 we collected demographic, clinical and laboratory information on each patient and  
212 also collected data on their close contacts. We then conducted in-depth  
213 epidemiological investigation of patients whose strains belonged to genomic clusters,  
214 including their residences, workplaces, and public community centers and facilities  
215 frequented in the three years prior to their TB diagnosis. Putative transmission  
216 networks were constructed based on the structure of the genomic phylogeny and the  
217 epidemiologic links. The epidemiologic links were defined as confirmed -- patients  
218 who knew each other and had a history of contact before diagnosis, or probable --  
219 patients who did not know each other but lived in the same village.

220

### 221 **Statistical analysis**

222 The clustering rate was calculated by the N method and the cumulative clustering rate  
223 was obtained by calculating the clustering rate in 2009 and then adding the TB  
224 patients from every successive year of the study, as previously described.<sup>16</sup> Changes  
225 in temporal trends of annual clustering were detected using joinpoint regression  
226 analysis. The distribution of continuous and categorical variables between groups was  
227 compared using the Wilcoxon rank sum test or the chi-square test. Logistic regression  
228 analysis was used to calculate the odds ratios (OR) and 95% confidence intervals (CI)  
229 for risk factors associated with genomic clustering. Variables with p-values less than  
230 0.2 in the univariable analysis were included in the multivariable analysis to calculate  
231 the adjusted odds ratios (aOR). Factors with a p-value less than 0.05 in the final  
232 model were considered statistically significant. All analyses were performed in Stata  
233 (version 14.0).

234

### 235 **Results**



236 Between July 1, 2009 and December 31, 2020, 5502 and 3435 pulmonary TB patients  
237 were officially registered in Wusheng and Wuchang, respectively, of whom 1517 and  
238 751 had positive sputum cultures. After excluding 103 strains that failed in re-culture,  
239 1452 and 713 strains were sequenced. We further excluded 123 patients with  
240 non-tuberculous mycobacteria isolates and 54 patients with recurrent TB. The final  
241 analysis thus included 1289 and 699 patients from Wusheng and Wuchang,  
242 respectively (Figure 1), whose characteristics are shown in Table 1. The proportions  
243 of patients less than 25 years or students were higher in Wusheng [15.4% (198/1289)  
244 and 7.6% (98/1289)] than in Wuchang [9.9% (69/699) and 2.9% (20/699)]. Patients in  
245 Wuchang were more likely to have had longer diagnosis delays, chest cavities and  
246 smear positivity.

247  
248 Phylogenetic analysis revealed that most strains belonged to the Beijing family  
249 lineage (69.0%, 1372/1988) (Figure 2), with more belonging to the modern Beijing  
250 sublineage (74.1%, 1016/1372) than ancient Beijing sublineages (25.9%, 356/1372).  
251 All of the non-Beijing strains belonged to L4 (99.8%, 615/616), except for one L1  
252 strain. Analysis of the genome sequences for mutations conferring resistance to 14  
253 anti-TB drugs showed that the drug-resistance profiles were similar between the two  
254 sites (Figure 2; Table 2). In total, there were 1716 (86.3%) pan-susceptible strains and  
255 272 (13.7%) strains resistant to at least one anti-TB drug. The isoniazid-resistant,  
256 rifampicin-resistant and MDR strains accounted for 10.7% (212/1988), 7.6%  
257 (152/1988) and 5.1% (102/1988) respectively, and 29.4% (30/102) of MDR strains  
258 were predicted to also be resistant to fluoroquinolones.

259  
260 To estimate the level of recent transmission, we calculated the clustering rate between  
261 2009 to 2020. A total of 624 (31.4%, 624/1988) strains, 347 (26.9%, 347/1289) from  
262 Wusheng and 277 (39.6%, 277/699) from Wuchang, were grouped into 225 genomic  
263 clusters containing 2 to 13 strains (Figure 2). To understand the dynamics of  
264 clustering trends in the two sites, we calculated the cumulative clustering rate and  
265 found that the rate gradually increased as more strains were analyzed, but the trend  
266 eased between 2015 and 2016 in both sites. As seen in Figure 3, the clustering rate  
267 increased significantly after 2016, especially in Wusheng ( $p=0.004$ , Figure S1, rising  
268 from 21.8% in 2017 to 26.9% in 2020), which may have been the result of strategies  
269 we implemented starting in 2017: active case-finding based on symptoms identified  
270 15-25% more patients; stepwise sputum collection to improve the diagnostic quality  
271 of the sputum<sup>17</sup> and sputum cultures in liquid media increased the percentage of  
272 culture-positive patients from 20-30% to 40-50%.

273

274 In the analysis of risk factors for MDR-TB, the only association was a higher

275 proportion of MDR in retreated patients (16·1%, 24/149) than in new cases (4·2%,  
276 78/1839). To investigate the cause of MDR we analyzed the number of clustered and  
277 non-clustered patients in the 102 MDR patients and found that 34 were clustered.  
278 While retreated MDR-TB cases could have secondary resistance that was acquired  
279 during treatment, new cases were presumably infected with MDR strains and  
280 represent primary resistance. Among the non-clustered MDR patients, 49 were new  
281 patients. Therefore, in total, 81·4% (83/102) of MDR-TB patients were likely caused  
282 by the transmission of MDR strains. We also looked for accumulated additional  
283 drug-resistance mutations along the transmission chain. Among 32 transmission  
284 events of MDR-TB belonging to 11 genomic-clusters, we found just two events (6·3%,  
285 2/32) of accumulation of additional resistance (Figure S2), which is a lower incidence  
286 of acquired resistance than has been reported in urban areas of China.<sup>11</sup>

287

288 To further delineate the transmission links between the genomic-clustered patients we  
289 performed in-depth epidemiological investigation of all patients with clustered  
290 isolates. These investigations were completed for 469 (75·2%, 469/624) patients, 294  
291 in Wusheng and 175 in Wuchang, but the other 155 clustered patients (24·8%) had  
292 either died or were lost to follow-up. Confirmed epidemiological links were identified  
293 in 196 (41·8%, 196/469) of the genomic-clustered patients investigated (110 in  
294 Wusheng and 86 in Wuchang), and probable epidemiological links were identified in  
295 26 (5·5%, 26/469) patients. The characteristics of each cluster with confirmed or  
296 probable epidemiological links are shown in Table S1.

297

298 The 196 patients with confirmed epidemiological links were all close contacts.  
299 Among them (Figure 4A), transmission between family contacts was identified in  
300 32·7% (64/196) and between social contacts in 67·3% (132/196). To further analyze  
301 transmission patterns, we divided the social contacts into three categories: relatives  
302 outside of the immediate family, community contacts, and contacts who shared  
303 aggregated settings such as teahouses located outside the community.  
304 Epidemiological links were found in community contacts in 14·5% (16/110) of the  
305 confirmed transmission in Wusheng but 51·2% (44/86) in Wuchang. Surprisingly, the  
306 proportions were reversed for confirmed epidemiological links in aggregated settings  
307 outside the community, with 43·6% (48/110) in Wusheng and 15·1% (13/86) in  
308 Wuchang. The percentages of confirmed links to relatives beyond the immediate  
309 family were low in both sites, with 4·5% (5/110) in Wusheng and 7·0% (6/86) in  
310 Wuchang.

311

312 We constructed putative transmission networks based on genomic phylogeny and  
313 epidemiological links. Figure 5 shows examples of three of the largest clusters. The

314 inhabitants of the villages in Wuchang live in fairly close proximity, often with daily  
315 contact in village shops that may explain the high level of transmission occurring  
316 within the community (Figure 5A). In Wusheng, by contrast, the majority of  
317 identified epidemiological links were outside the community and mainly in schools  
318 and teahouses (Figure 5B/5C). The teahouses in Wusheng are generally located in  
319 townships outside the villages and are frequented by people from several different  
320 villages.

321

322 We then analyzed the time interval between TB diagnosis in clustered cases in the  
323 different categories of contacts (Figure 4B). We found no statistical difference  
324 ( $p=0.90$ ) in the time interval between diagnosis of family contacts (3.1 years,  
325 Interquartile range [IQR] 0.9-5.9 years) and community contacts (3.0 years, IQR  
326 1.3-4.6 years). Unexpectedly though, the average diagnostic time interval between  
327 contacts who shared the aggregated settings outside the community (1.1 years, IQR  
328 0.4-2.3 years) was significantly shorter than for family ( $p<0.01$ ) or community  
329 contacts ( $p<0.01$ ). This shorter interval was perhaps because most of the contacts who  
330 shared aggregated settings outside the community were students, and the average time  
331 interval between the diagnosis of student contacts (0.9 years, IQR 0.3-1.4 years) was  
332 significantly shorter ( $p<0.01$ ) than for all other contacts (1.9 years, IQR 0.7-3.6  
333 years).

334

335 Finally, we used logistic regression to identify risk factors associated with clustering.  
336 The univariate analysis found that occupation and the Beijing strain were significantly  
337 associated with clustering, and both associations persisted in the multivariate analysis  
338 (Table 3). Patients with a Beijing lineage strain had a greater risk of clustering (aOR  
339 1.46, 95% CI, 1.17-1.82,  $p=0.001$ ) than patients with L4 strains. Also, compared to  
340 any other occupation, farmers had a lower risk (aOR 0.60, 95% CI, 0.44-0.82,  
341  $p=0.001$ ) and students had a higher risk (aOR, 2.11; 95% CI, 1.20-3.71,  $p=0.01$ ) of  
342 clustering.

343

## 344 **Discussion**

345 To our knowledge, this was the longest longitudinal population-based genomic  
346 epidemiological study of TB transmission in rural China. The overall cumulative  
347 clustering rate during the study period was 31.4%. Among 196 genomic-clustered  
348 patients with confirmed epidemiological links, 32.7% of transmission occurred  
349 between family contacts and 67.3% between social contacts. Of all MDR-TB cases  
350 diagnosed during the study, 81.4% were likely from transmission of MDR strains.  
351 The average time interval between the diagnosis of clustered student contacts was  
352 much shorter than for non-student contacts.

353

354 Close contacts of TB patients are at a high-risk for TB transmission. A recent  
355 systematic review suggested that the pooled prevalence of TB in close contacts was  
356 3·6% (95% CI: 3·3-4·0), and that contact investigation could increase TB case  
357 notification and thereby decrease the incidence of TB in the population.<sup>18</sup> In  
358 developed countries, index cases have been found to have 6-12 close contacts<sup>19-22</sup> who  
359 could potentially constitute 10-20%<sup>23,24</sup> of the total TB burden. Accordingly, the  
360 WHO strongly recommends systematic screening for TB disease among close  
361 contacts.<sup>5,25</sup> In China, however, the importance of screening close contacts has been  
362 underappreciated and its role in case finding has been limited. In China, only 2-3  
363 close contacts per index case were identified and most of these were family members,  
364 and it was therefore estimated that close contacts would contribute only about 1·0-3·3%  
365 of the total TB patients.<sup>26,27</sup> A recent community-based study of active case-finding  
366 that included 320,000 people in 10 provinces in China during 2013-2015 concluded  
367 that screening of close contacts contributed less than 2% of TB patients.<sup>7</sup>

368

369 Our study found that 41·8% of clustered patients were close contacts of other patients  
370 and contributed 9·9% of the total TB patients, suggesting a high risk of TB  
371 transmission among close contacts in rural China. However, it was only through  
372 in-depth investigation of these genomic-clustered patients that we could identify the  
373 epidemiological links that are usually overlooked in routine contact investigation. In  
374 contrast to our findings, the routine TB contact investigations that were conducted in  
375 the study sites during the duration of our study registered only 2·4 close contacts per  
376 index case who contributed just 1·8% of the total patients. Without the WGS data that  
377 identified clustered patients and directed our investigations, nearly 80% of the  
378 transmission events between close contacts would have been missed. Most secondary  
379 cases were diagnosed between 1-3 years after the index case, but the long study  
380 duration of 12 years, using WGS to analyze all TB isolates in the study areas, allowed  
381 us to identify clustered isolates that would be missed in studies with a shorter  
382 duration.

383

384 Screening of close contacts in China has been poorly implemented for a number of  
385 reasons. The huge number of TB patients, limited resources for TB prevention and  
386 control and the stigma of tuberculosis all contribute to a failure to identify and screen  
387 many close contacts. When contact screening is poorly implemented, the contribution  
388 of close contacts to the overall TB burden is underestimated and therefore less  
389 attention is paid to the screening, creating a vicious cycle. WHO guidelines defined  
390 close contacts as persons sharing an enclosed space with the index case during the  
391 three months prior to commencement of the current treatment episode.<sup>25</sup> While the

392 identification of family contacts is generally straightforward, the definition and  
393 identification of non-family contacts is difficult. In China, non-family contacts refer  
394 mainly to classmates and colleagues.<sup>27</sup> However, in this study we found that many  
395 patients with confirmed epidemiologic links were contacts either within the  
396 community or contacts who shared the aggregated settings outside the community.  
397 Therefore, in addition to family contacts, we classified others as social contacts in the  
398 hope of understanding where contacts occur and highlighting their relevance for TB  
399 transmission in rural China.

400

401 We found that 67.3% of the transmission occurred among social contacts in rural  
402 China, which was significantly higher than has been found in Vietnam, Zambia and  
403 South Africa (15-50%).<sup>28,29</sup> This suggests that active case-finding in rural China  
404 should go beyond family contacts to include social contacts. In this study we proposed  
405 three categories of social contacts: relatives outside of the immediate family; people  
406 who know each other and live in the same community or village; and people  
407 frequenting aggregated settings outside the community. Largely due to differences in  
408 climate and lifestyle, the importance of the different categories of contacts varied  
409 between the two study sites. The villages in Wuchang are very small, consisting of  
410 perhaps only a dozen families and the houses are generally adjacent to each other.  
411 Wuchang is located in the northeast of China and has an average annual temperature  
412 of only 4°C. Consequently, villagers spend more than half of the year indoors and  
413 frequently interact to socialize and play cards with other villagers in the village shops.  
414 It is therefore not surprising that 51.2% of epidemiological links were attributable to  
415 transmission between social contacts within these communities. Wusheng, in contrast,  
416 is located in the southwest of China with an average annual temperature of 18°C. and  
417 villagers live in homes that are relatively scattered in the region. We found that  
418 transmission between villagers in this setting was less common than transmission at  
419 aggregated settings such as schools and teahouses, where individuals from several  
420 different villages congregate. Therefore, strategies for screening and social contact  
421 tracing must be tailored to the conditions and association habits of the local  
422 populations.

423

424 The purpose of active case-finding is to diagnose patients early and reduce  
425 transmission. Importantly, we found that the time interval between TB diagnosis of  
426 student patients and contacts was 3-4 times shorter than for family or community  
427 contacts (Figure 4B). Students are a special group in China and receive special  
428 attention from the government and society. Once a student patient is diagnosed with  
429 TB, their classmates and even schoolmates will be screened. If a similar enhanced  
430 screening strategy could be extended to other social contacts, patients might be

431 diagnosed earlier in the course of their disease, thereby reducing rural TB  
432 transmission.

433

434 This study have several limitations. The clustering rate we calculated is likely an  
435 underestimation of the extent of local TB transmission. We previously found that the  
436 current passive case-finding strategy in Wusheng identified only about 30% of the  
437 incident TB patients<sup>30</sup>, but we were unable to even estimate the number of TB patients  
438 in the sampling region who were undiagnosed or diagnosed elsewhere. Some of the  
439 non-clustered patients could have been clustered with locally diagnosed TB patients  
440 without cultured isolates or patients diagnosed outside the study period or the  
441 geographic regions sampled. In addition, some clustered patients died or were lost to  
442 follow-up before the epidemiological investigations could be completed.

443

444 In conclusion, this long-term genomic-epidemiological study helped to delineate the  
445 patterns of TB transmission in rural China. Transmission appears to occur principally  
446 among close contacts, but contact investigation should be extended to include the  
447 social interactions that are common in the targeted population. Further improvement  
448 and implementation of screening of close contacts are critical to halt transmission and  
449 improve TB control in rural areas.

450

#### 451 **Contributors**

452 ML, QJ, CY, LX, FL, CC and QG conceived, designed and managed the study. SZ  
453 (Sheng Zhong), XS, YP, SZ (Shu Zhang), MG, and YQ contributed the demographic,  
454 clinical and microbiological data. ML, MG, YP, SZ (Sheng Zhong), XS, and PM  
455 contributed to the epidemiological investigations. ML, QM, QJ, and PM did the data  
456 analyses. ML, QJ, CY, HT, and QG drafted and revised the manuscript. All coauthors  
457 reviewed and approved the final manuscript before submission.

458

#### 459 **Declaration of interests**

460 We declare no competing interests.

461

#### 462 **Acknowledgments**

463 This study was funded by the National Science and Technology Major Project of  
464 China (grants 2017ZX10201302-006). It was also funded by Natural Science  
465 Foundation of China (81661128043) and Science and Technology Major Project of  
466 Shanghai. We thank the tuberculosis public health teams in the Wusheng County  
467 Center for Disease Control and Prevention (Wusheng CDC) and the Wuchang City  
468 Center for Tuberculosis Control and Prevention (Wuchang CTC).

469

470 **Role of the funding source**

471 The funders of the study had no role in study design, data collection, data analysis,  
472 data interpretation, or writing of the report. The corresponding authors had full access  
473 to all the data in the study and had final responsibility for the decision to submit the  
474 manuscript for publication.

475

476 **Data sharing**

477 De-identified participant data from the study will be made available with publication  
478 to medical researchers on a not for profit basis by email request to the corresponding  
479 author for the purposes of propensity matching or meta-analysis. Detailed  
480 supplementary materials will be made available by email request to the corresponding  
481 author.

**Table 1: Characteristics of tuberculosis patients in Wusheng and Wuchang, 2009-2020.**

	Wusheng (n=1289)	Wuchang (n=699)
<b>Sex</b>		
Female	271 (21.0)	202 (28.9)
Male	1018 (79.0)	497 (71.1)
<b>Age</b>		
<25	198 (15.4)	69 (9.9)
25–44	328 (25.4)	193 (27.6)
45–64	519 (40.3)	296 (42.3)
≥65	244 (18.9)	141 (20.2)
<b>Occupation</b>		
Farmer	1096 (85.0)	578 (82.7)
Students	98 (7.6)	20 (2.9)
Others	95 (7.4)	101 (14.4)
<b>History of tuberculosis</b>		
New	1197 (92.9)	642 (91.8)
Retreated	92 (7.1)	57 (8.2)
<b>Diagnostic delay</b>		
<2 weeks	428 (33.2)	103 (14.7)
2–4 weeks	227 (17.6)	162 (23.2)
4–8 weeks	371 (28.8)	208 (29.8)
≥8 weeks	263 (20.4)	226 (32.3)
<b>Chest cavitation</b>		
No	889 (69.0)	407 (58.2)
Yes	400 (31.0)	292 (41.8)
<b>Sputum smear status</b>		
Negative	634 (49.2)	244 (34.9)
Positive	655 (50.8)	455 (65.1)
<b>Drug-resistance profile</b>		
Pan-susceptible	1128 (87.5)	588 (84.1)
Other DR	95 (7.4)	75 (10.7)
MDR	66 (5.1)	36 (5.2)

Data are n (%) unless otherwise specified. DR=drug resistance. MDR=multidrug resistance.



**Table 2: Drug-resistance profile, stratified by new and retreated cases.**

	New cases (n=1839)	Retreated cases (n=149)	Total (n=1988)
Pan-susceptible	1610 (87.5)	106 (71.1)	1716 (86.3)
Resistance to INH	180 (9.8)	32 (21.5)	212 (10.7)
Resistance to RIF	119 (6.5)	33 (22.1)	152 (7.6)
Resistance to EMB	54 (2.9)	22 (14.8)	76 (3.8)
Resistance to PZA	24 (1.3)	9 (6.0)	33 (1.7)
MDR	78 (4.2)	24 (16.1)	102 (5.1)
MDR plus resistance to FQ	24 (1.3)	6 (4.0)	30 (1.5)

Data are n (%) unless otherwise specified. INH=isoniazid. RIF=rifampicin. EMB=ethambutol.

PZA=pyrazinamide. MDR=multidrug resistance. FQ=fluoroquinolone.

**Table 3: Univariate and multivariable logistic regression of risk factors for clustering.**

	Clustered (n=624)	Non-clustered (n=1364)	Univariate regression		Multivariable regression	
			OR (95% CI)	p value	aOR (95% CI)	p value
<b>Sex</b>						
Female	148 (31.3)	325 (68.7)	1.00		..	
Male	476 (31.4)	746 (68.6)	1.01 (0.81, 1.26)	0.958	..	..
<b>Age</b>						
<25	117 (43.8)	150 (56.2)	2.41 (1.73, 3.38)	<0.001	1.35 (0.89, 2.06)	0.160
25–44	164 (31.5)	357 (68.5)	1.42 (1.06, 1.91)	0.020	1.35 (1.00, 1.82)	0.054
45–64	249 (30.6)	566 (69.4)	1.36 (1.03, 1.80)	0.028	1.34 (1.01, 1.77)	0.041
≥65	94 (24.4)	291 (75.6)	1.00		1.00	
<b>Occupation</b>						
Farmer	475 (28.4)	1199 (71.6)	0.56 (0.42, 0.76)	<0.001	0.60 (0.44, 0.82)	0.001
Students	68 (57.6)	50 (42.4)	1.93 (1.22, 3.07)	0.005	2.11 (1.20, 3.71)	0.010
Others	81 (41.3)	115 (58.7)	1.00		1.00	
<b>History of tuberculosis</b>						
New	582 (31.6)	1257 (68.4)	1.00		..	
Retreated	42 (28.2)	107 (71.8)	0.85 (0.59, 1.23)	0.382	..	..
<b>Diagnostic delay</b>						
<2 weeks	157 (29.6)	374 (70.4)	1.00		..	
2–4 weeks	124 (31.9)	265 (68.1)	1.11 (0.84, 1.48)	0.452	..	..
4–8 weeks	188 (32.5)	391 (67.5)	1.15 (0.89, 1.48)	0.297	..	..
≥8 weeks	155 (31.7)	334 (68.3)	1.11 (0.85, 1.44)	0.461	..	..
<b>Chest cavitation</b>						
No	385 (29.7)	911 (70.3)	1.00		1.00	
Yes	239 (34.5)	453 (65.5)	1.25 (1.03, 1.52)	0.027	1.22 (1.00, 1.50)	0.054
<b>Sputum smear status</b>						
Negative	258 (29.4)	620 (70.6)	1.00		1.00	
Positive	366 (33.0)	744 (67.0)	1.18 (0.98, 1.43)	0.087	1.19 (0.98, 1.46)	0.085
<b>Drug-resistance profile</b>						
Pan-susceptible	547 (31.9)	1169 (68.1)	1.00		1.00	
Other DR	43 (25.3)	127 (74.7)	0.72 (0.50, 1.04)	0.079	0.68 (0.47, 0.98)	0.040
MDR	34 (33.3)	68 (66.7)	1.07 (0.70, 1.63)	0.759	1.00 (0.65, 1.54)	1.000
<b>Beijing strain</b>						
No	158 (25.6)	458 (74.4)	1.00		1.00	
Yes	466 (34.0)	906 (66.0)	1.49 (1.21, 1.84)	<0.001	1.46 (1.17, 1.82)	0.001

Data are n (%) unless otherwise specified. OR=odds ratio. aOR=adjusted odds ratio. DR=drug resistance. MDR=multidrug resistance.

**Table S1. Characteristics of genomic clusters with confirmed or probable epidemiological links based on whole-genome sequencing analysis in Wusheng and Wuchang.**

	Sites	Number of cases in clusters	Mean age (years [IQR])	Number of patients living in the same village	Number of new cases	Number of patients receiving investigation	Epidemiologically linked	Nature of epidemiological link [number of patients]*
Cluster01	Wusheng	2	55	2	1	2	Confirmed	family [2]
Cluster02	Wusheng	3	45	2	3	3	Confirmed	family [2]
Cluster04	Wusheng	2	23	2	2	2	Probable	same village [2]
Cluster10	Wusheng	3	47	3	3	3	Confirmed/Probable	family [2]; same village [1]
Cluster14	Wusheng	5	57	0	3	5	Confirmed	teahouse [5]
Cluster16	Wusheng	4	50	2	4	3	Confirmed	family [2]
Cluster20	Wusheng	2	17	0	2	2	Confirmed	classmate [2]
Cluster22	Wusheng	3	60	2	3	3	Confirmed	family [2]
Cluster24	Wusheng	2	41	2	2	2	Confirmed	family [2]
Cluster27	Wusheng	5	47	2	5	5	Confirmed/Probable	workplace [3]; relatives [1]; same village [1]
Cluster32	Wusheng	3	17	0	3	3	Confirmed	classmate [2]
Cluster35	Wusheng	3	18	0	3	3	Confirmed	classmate [3]
Cluster36	Wusheng	2	56	2	2	2	Confirmed	same village and knew each other [2]
Cluster37	Wusheng	2	18	0	2	2	Confirmed	classmate [2]
Cluster39	Wusheng	3	39	2	3	3	Confirmed	family [2]
Cluster42	Wusheng	2	60	2	2	2	Confirmed	same village and knew each other [2]
Cluster44	Wusheng	7	54	3	6	7	Confirmed	family [2]; neighbor [1]
Cluster45	Wusheng	2	57	2	2	2	Confirmed	same village and knew each other [2]
Cluster46	Wusheng	4	41	2	3	3	Confirmed	same village and knew each other [2]
Cluster48	Wusheng	2	44	2	2	2	Confirmed	relatives [2]
Cluster49	Wusheng	3	34	2	2	2	Confirmed	family [2]
Cluster52	Wusheng	10	18	0	10	10	Confirmed	classmate [9]
Cluster53	Wusheng	2	49	2	2	2	Confirmed	family [2]
Cluster55	Wusheng	2	48	2	2	2	Confirmed	family [2]
Cluster57	Wusheng	3	51	2	2	3	Confirmed	family [2]; teacher [1]
Cluster61	Wusheng	3	59	2	3	3	Confirmed	family [2]
Cluster66	Wusheng	2	31	2	2	2	Confirmed	family [2]
Cluster67	Wusheng	9	32	3	8	8	Confirmed	family [3]; classmate [3]; same village and knew each other [1]
Cluster69	Wusheng	6	57	0	6	6	Confirmed	friends [2]
Cluster70	Wusheng	2	51	2	2	2	Probable	same village [2]

Cluster72	Wusheng	3	62	0	3	3	Confirmed	teahouse [2]
Cluster73	Wusheng	3	45	0	3	3	Confirmed	mental hospital [3]
Cluster75	Wusheng	3	64	2	3	3	Confirmed	friends [2]
Cluster78	Wusheng	2	41	2	2	2	Confirmed	friends [2]
Cluster79	Wusheng	5	38	5	5	5	Confirmed/Probable	family [2]; same village [3]
Cluster86	Wusheng	2	58	2	2	2	Confirmed	same village and knew each other [2]
Cluster87	Wusheng	4	24	0	4	4	Confirmed	classmate [3]
Cluster88	Wusheng	3	56	2	3	3	Confirmed	same village and knew each other [2]
Cluster93	Wusheng	3	46	0	3	3	Confirmed	mental hospital [2]
Cluster97	Wusheng	2	55	0	2	2	Confirmed	friends [2]
Cluster98	Wusheng	2	50	2	2	2	Probable	same village [2]
Cluster103	Wusheng	2	30	0	2	2	Confirmed	friends [2]
Cluster107	Wusheng	7	32	6	6	7	Confirmed/Probable	family [2]; same village [4]
Cluster109	Wusheng	2	44	2	2	2	Confirmed	same village and knew each other [2]
Cluster111	Wusheng	2	34	2	2	2	Confirmed	friends [2]
Cluster117	Wusheng	2	59	2	2	2	Confirmed	friends [2]
Cluster119	Wusheng	2	31	2	2	2	Confirmed	relatives [2]
Cluster03	Wuchang	5	52	3	6	4	Confirmed	family [2]; same village and knew each other [1]
Cluster04	Wuchang	6	45	2	6	4	Confirmed	same village and knew each other [2]
Cluster05	Wuchang	4	49	3	3	4	Confirmed	same village and knew each other [1]; relatives [2]
Cluster08	Wuchang	3	71	2	3	3	Confirmed	relatives [2]
Cluster09	Wuchang	2	19	0	2	2	Confirmed	workplace [2]
Cluster11	Wuchang	3	18	0	3	3	Confirmed	classmates [3]
Cluster12	Wuchang	8	43	4	7	7	Confirmed	family [3]; same village and knew each other [2]
Cluster17	Wuchang	2	55	2	2	2	Confirmed	same village and knew each other [2]
Cluster21	Wuchang	2	40	0	2	2	Confirmed	same village and knew each other [2]
Cluster23	Wuchang	2	27	2	2	0	Probable	same village [2]
Cluster25	Wuchang	7	54	6	7	6	Confirmed/probable	same village and knew each other [4]; same village [2]
Cluster27	Wuchang	5	57	2	5	3	Confirmed	family [2]
Cluster30	Wuchang	3	55	0	3	2	Confirmed	family [2]
Cluster37	Wuchang	3	43	2	3	2	Confirmed	same village and knew each other [2]
Cluster39	Wuchang	6	54	0	6	3	Confirmed	Workplace [2]

Cluster40	Wuchang	3	38	2	3	0	Probable	same village [2]
Cluster42	Wuchang	2	32	2	2	2	Confirmed	family [2]
Cluster44	Wuchang	2	45	2	0	2	Confirmed	family [2]
Cluster45	Wuchang	4	45	3	4	4	Confirmed	same village and knew each other [3]
Cluster47	Wuchang	3	50	0	3	3	Confirmed	nursing home [3]
Cluster50	Wuchang	3	41	0	3	3	Confirmed	family [2]
Cluster53	Wuchang	3	40	0	3	3	Confirmed	friends [3]
Cluster54	Wuchang	3	36	0	3	0	Confirmed	family [2]
Cluster55	Wuchang	3	46	3	3	3	Confirmed	same village and knew each other [3]
Cluster56	Wuchang	4	47	3	4	3	Confirmed/Probable	same village and knew each other [2]; same village [1]
Cluster58	Wuchang	3	46	2	3	0	Probable	same village [2]
Cluster60	Wuchang	12	43	6	11	10	Confirmed	relatives [2]; same village and knew each other [6]
Cluster64	Wuchang	2	63	2	2	2	Confirmed	same village and knew each other [2]
Cluster66	Wuchang	2	44	2	2	2	Confirmed	same village and knew each other [2]
Cluster69	Wuchang	2	36	2	2	2	Confirmed	family [2]
Cluster70	Wuchang	2	39	2	2	0	Probable	same village [2]
Cluster73	Wuchang	8	50	0	6	6	Confirmed	family [2]
Cluster77	Wuchang	3	53	2	3	2	Confirmed	same village and knew each other [2]
Cluster80	Wuchang	2	68	2	2	2	Confirmed	same village and knew each other [2]
Cluster82	Wuchang	2	33	2	2	2	Confirmed	same village and knew each other [2]
Cluster89	Wuchang	2	42	2	2	2	Confirmed	family [2]
Cluster94	Wuchang	4	39	2	4	2	Confirmed	same village and knew each other [2]
Cluster95	Wuchang	2	41	2	2	2	Confirmed	same village and knew each other [2]

\*Red indicates household contacts, yellow indicates social contacts, black indicates genomic-clustered cases with probable epidemiological links.

## References

1. WHO. Global tuberculosis report 2020. [https://www.who.int/tb/publications/global\\_report/en/](https://www.who.int/tb/publications/global_report/en/) (accessed Apr 7, 2022).
2. Churchyard G, Kim P, Shah NS, et al. What We Know About Tuberculosis Transmission: An Overview. *J Infect Dis* 2017; **216**(suppl\_6): S629-S35.
3. Floyd K, Glaziou P, Zumla A, Raviglione M. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. *Lancet Respir Med* 2018; **6**(4): 299-314.
4. WHO. The End TB Strategy: global strategy and targets for tuberculosis prevention, care and control after 2015. [https://www.who.int/tb/strategy/End\\_TB\\_Strategy.pdf](https://www.who.int/tb/strategy/End_TB_Strategy.pdf) (accessed Apr 4, 2022).
5. WHO. WHO consolidated guidelines on tuberculosis. Module 2: screening – systematic screening for tuberculosis disease. <https://www.who.int/publications-detail-redirect/9789240022614> (accessed Apr 4, 2022).
6. Yang C, Gao Q. Recent transmission of *Mycobacterium tuberculosis* in China: the implication of molecular epidemiology for tuberculosis control. *Front Med* 2018; **12**(1): 76-83.
7. Zhang H, Cheng J, Yu YL, et al. Evaluation of the effectiveness of community-based pulmonary tuberculosis active case-finding among key populations: a multicenter prospective cohort study. *Chin J Antituberc* 2021; **43**(12): 1248-59 (in Chinese).
8. Wang L, Zhang H, Ruan Y, et al. Tuberculosis prevalence in China, 1990-2010; a longitudinal analysis of national survey data. *Lancet* 2014; **383**(9934): 2057-64.
9. Jiang Q, Liu Q, Ji L, et al. Citywide Transmission of Multidrug-resistant Tuberculosis Under China's Rapid Urbanization: A Retrospective Population-based Genomic Spatial Epidemiological Study. *Clin Infect Dis* 2020; **71**(1): 142-51.
10. Comas I, Coscolla M, Luo T, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013; **45**(10): 1176-82.
11. Yang CG, Luo T, Shen X, et al. Transmission of multidrug-resistant *Mycobacterium tuberculosis* in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. *Lancet Infect Dis* 2017; **17**(3): 275-84.
12. Coll F, McNerney R, Guerra-Assuncao JA, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 2014; **5**(1): 4812.
13. Luo T, Comas I, Luo D, et al. Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. *Proc Natl Acad Sci U S A* 2015; **112**(26): 8136-41.

14. Papaventsis D, Casali N, Kontsevaya I, Drobniowski F, Cirillo DM, Nikolayevskyy V. Whole genome sequencing of *Mycobacterium tuberculosis* for detection of drug resistance: a systematic review. *Clin Microbiol Infect* 2017; **23**(2): 61-8.
15. Yang T, Wang Y, Liu Q, et al. A population-based genomic epidemiological study of the source of tuberculosis infections in an emerging city: Shenzhen, China. *Lancet Reg Health-W* 2021; **8**: 100106.
16. Izumi K, Murase Y, Uchimura K, et al. Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Tokyo, Japan: a population-based molecular epidemiological study. *BMJ Open* 2019; **9**(5): e029295.
17. Jiang Q, Ji L, Qiu Y, et al. A randomised controlled trial of stepwise sputum collection to increase yields of confirmed tuberculosis. *Int J Tuberc Lung Dis* 2019; **23**(6): 685-91.
18. Velen K, Shingde RV, Ho J, Fox GJ. The effectiveness of contact investigation among contacts of tuberculosis patients: a systematic review and meta-analysis. *Eur Respir J* 2021; **58**(6).
19. Denholm JT, Leslie DE, Jenkin GA, et al. Long-term follow-up of contacts exposed to multidrug-resistant tuberculosis in Victoria, Australia, 1995-2010. *Int J Tuberc Lung Dis* 2012; **16**(10): 1320-5.
20. Hirsch-Moverman Y, Cronin WA, Chen B, et al. HIV counseling and testing in tuberculosis contact investigations in the United States and Canada. *Int J Tuberc Lung Dis* 2015; **19**(8): 943-53.
21. Young KH, Ehman M, Reves R, et al. Tuberculosis Contact Investigations--United States, 2003-2012. *MMWR Morb Mortal Wkly Rep* 2016; **64**(50-51): 1369-74.
22. Izumi K, Ohkado A, Uchimura K, et al. Evaluation of tuberculosis contact investigations in Japan. *Int J Tuberc Lung Dis* 2017; **21**(2): 188-95.
23. Reichler MR, Khan A, Sterling TR, et al. Risk and Timing of Tuberculosis Among Close Contacts of Persons with Infectious Tuberculosis. *J Infect Dis* 2018; **218**(6): 1000-8.
24. Martin-Sanchez M, Brugueras S, de Andres A, et al. Tuberculosis incidence among infected contacts detected through contact tracing of smear-positive patients. *PLoS One* 2019; **14**(4): e0215322.
25. WHO. Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income countries. <https://www.who.int/publications/i/item/9789241504492> (accessed Apr 7, 2022).
26. Gao CN, Tan QY, Xu ZW, Luo SX. Analysis of TB case detection through investigation of close contacts of smear positive pulmonary tuberculosis patients. *Chin J Antituberc*. 2011;33(06):328-30 (in Chinese).

27. Jiang Q, Lu L, Wu J, et al. Assessment of tuberculosis contact investigation in Shanghai, China: An 8-year cohort study. *Tuberculosis (Edinb)* 2018; **108**: 10-5.
28. Horby P, Pham QT, Hens N, et al. Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS One* 2011; **6**(2): e16965.
29. Dodd PJ, Looker C, Plumb ID, et al. Age- and Sex-Specific Social Contact Patterns and Incidence of *Mycobacterium tuberculosis* Infection. *Am J Epidemiol* 2016; **183**(2): 156-66.
30. Liao PJ, Qiu Y, Guo MC, et al. Analysis of patients from active case finding in rural areas of Wusheng, Sichuan. *Chin J Antituberc* 2016; **38**(7): 576-81 (in Chinese).



## FIGURE LEGENDS

### **Figure 1: Sample enrollment in Wusheng (A) and Wuchang (B).**

NTM=non-tuberculous mycobacteria.

**Figure 2: Phylogeny, clustering, and resistance profile of 1289 *Mycobacterium tuberculosis* strains isolated in Wusheng (A) and 699 *Mycobacterium tuberculosis* strains isolated in Wuchang (B).** The different colors on the branches indicate different lineages and sublineages. The outer grey circle indicates genomic-clustered strains differing by  $\leq 12$  single-nucleotide polymorphisms. The outer yellow-red circle indicates other drug resistance and multidrug resistance.

### **Figure 3: Cumulative clustering rate in Wusheng (A) and Wuchang (B),**

**2009-2020.** The bar indicates the cumulative number of cases, the line indicates the cumulative clustering rate.

### **Figure 4: Categories of confirmed epidemiological links and their time interval**

**between diagnoses.** The four colors of the circle indicate four categories of contacts with confirmed epidemiological links. The darker the color, the closer the relationship. (A) The yellow border indicates family contacts (Family) and the red border indicates social contacts (Relatives, Community and Aggregated settings outside the community). The number in the ring indicates the number and proportion of each category of contacts. The width of the ring indicates the proportion of each category. (B) The time interval between diagnoses in the four categories of contacts and students. Due to the small number of relatives, they were not compared with the other categories. \*\* $p < 0.01$ , NS refers to no significance (given by Wilcoxon non-parametric rank sum test).

### **Figure 5: Putative transmission network for 3 clusters based on the structure of**

**the genomic phylogeny and the epidemiological links.** The color of the line indicates the categories of contacts in Figure 4. epi-link=epidemiological link.

**Figure S1. Changes in temporal trends in cumulative clustering rate in Wusheng (left) and Wuchang (right), 2009-2020.** \* APC is significantly different from zero at the level of  $\alpha=0.05$ .

**Figure S2. Transmission networks of multidrug-resistant tuberculosis based on genomic phylogeny and drug-resistance mutations.** For each network, the arrow

indicates the root and circles represent *Mycobacterium tuberculosis* isolates. M tuberculosis isolates are separated by lines with length representing genetic distance. Different colors of rectangle indicate resistance to different drugs. Two new MDR-TB patients were clustered with non-MDR-TB patients and were not included in the analysis.

**A**

5502 pulmonary TB cases during Jul. 2009 -  
Dec. 2020 in Wusheng

1776 (32.3%) bacteriologically confirmed cases

1517 (85.4%) cases with culture positive strains

65 strains failed re-culture

1452 (95.7%) cases with sequencing data

119 strains were NTM

1333 (91.8%) cases with available sequencing data

exclude 44 recurrent cases

1289 (96.7%) cases were included in the analysis

**B**

3435 pulmonary TB cases during Jul. 2009 -  
Dec. 2020 in Wuchang

869 (25.3%) bacteriologically confirmed cases

751 (86.4%) cases with culture positive strains

38 strains failed re-culture

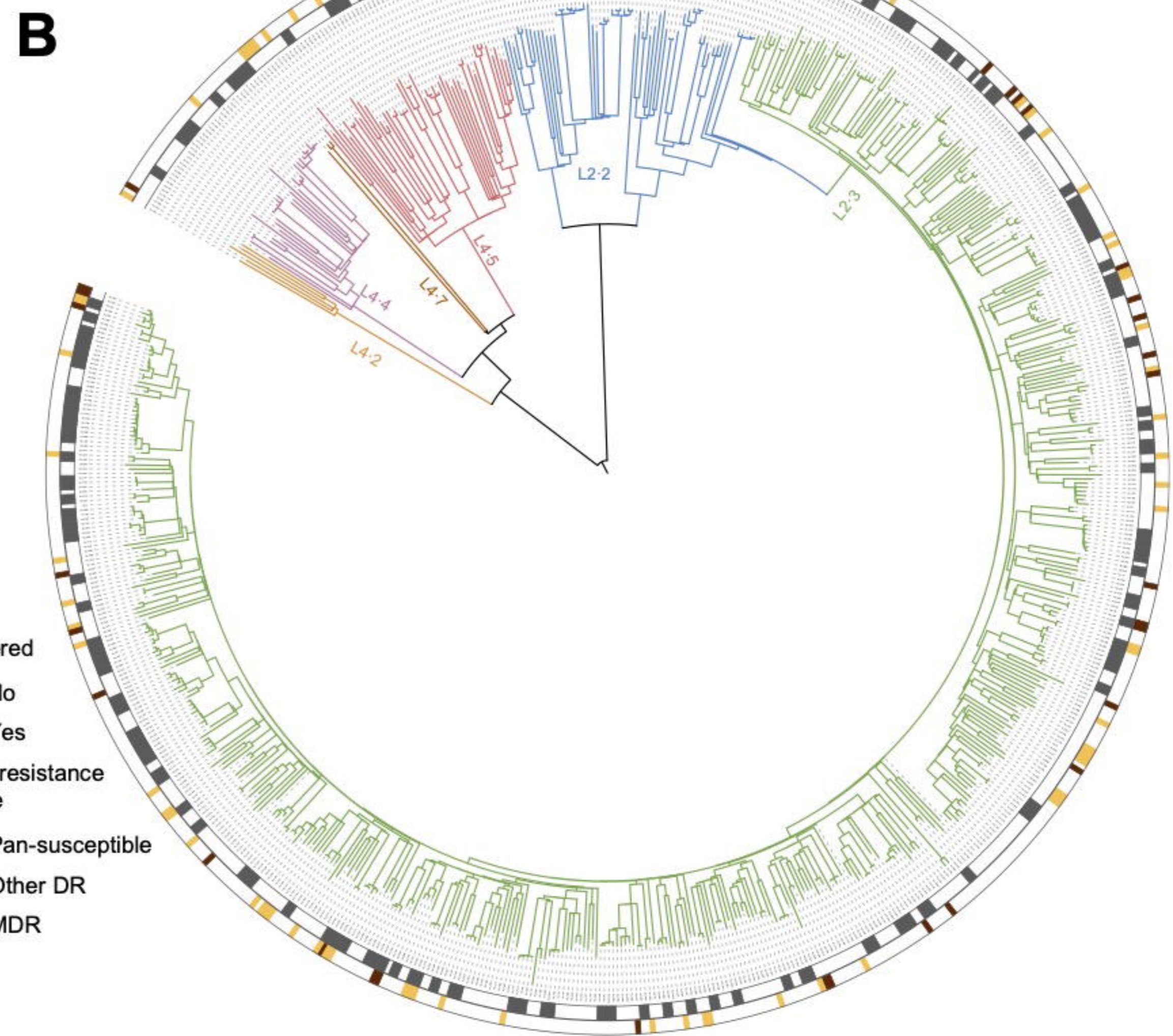
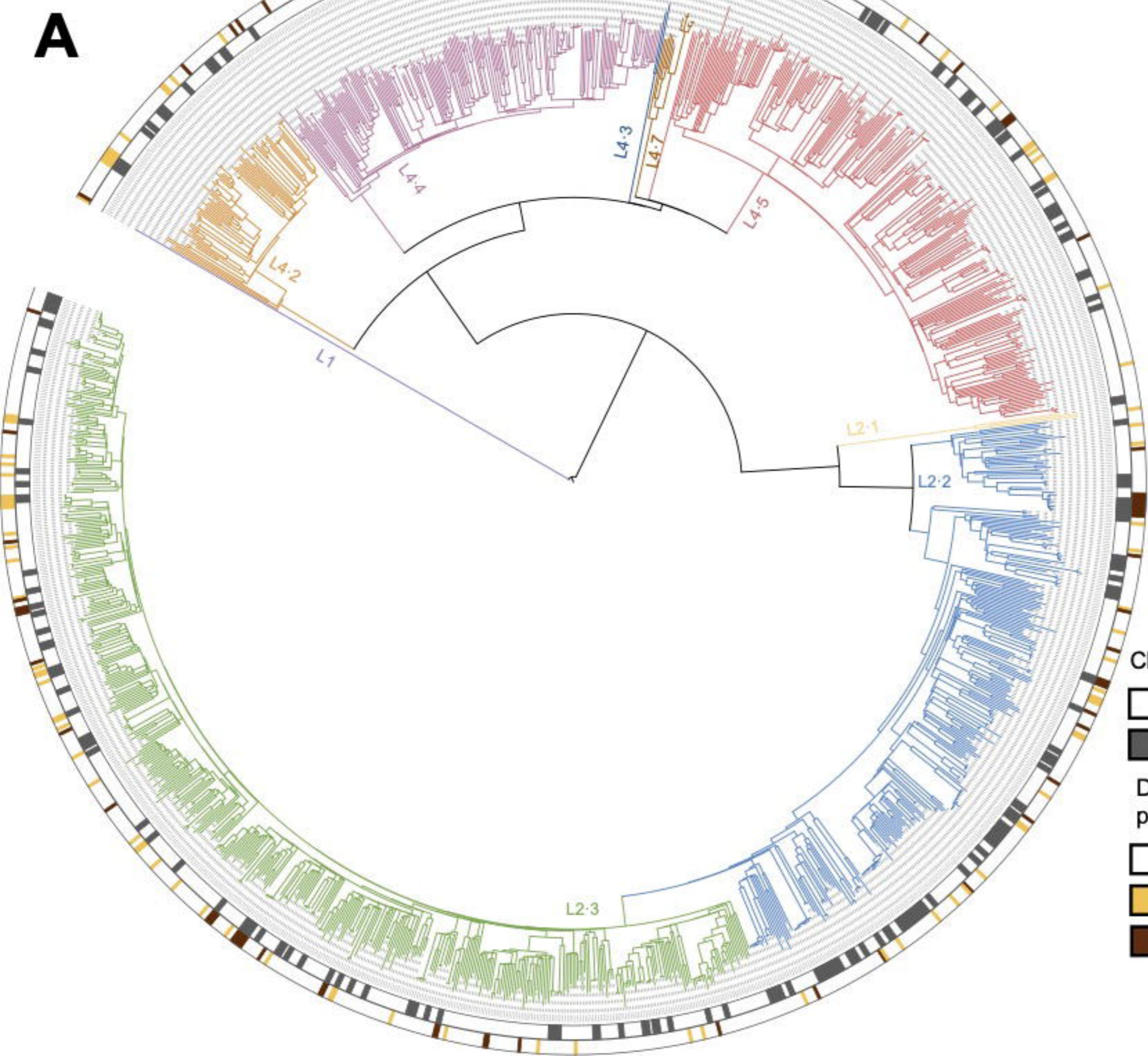
713 (94.9%) cases with sequencing data

4 strains were NTM

709 (99.4%) cases with available sequencing data

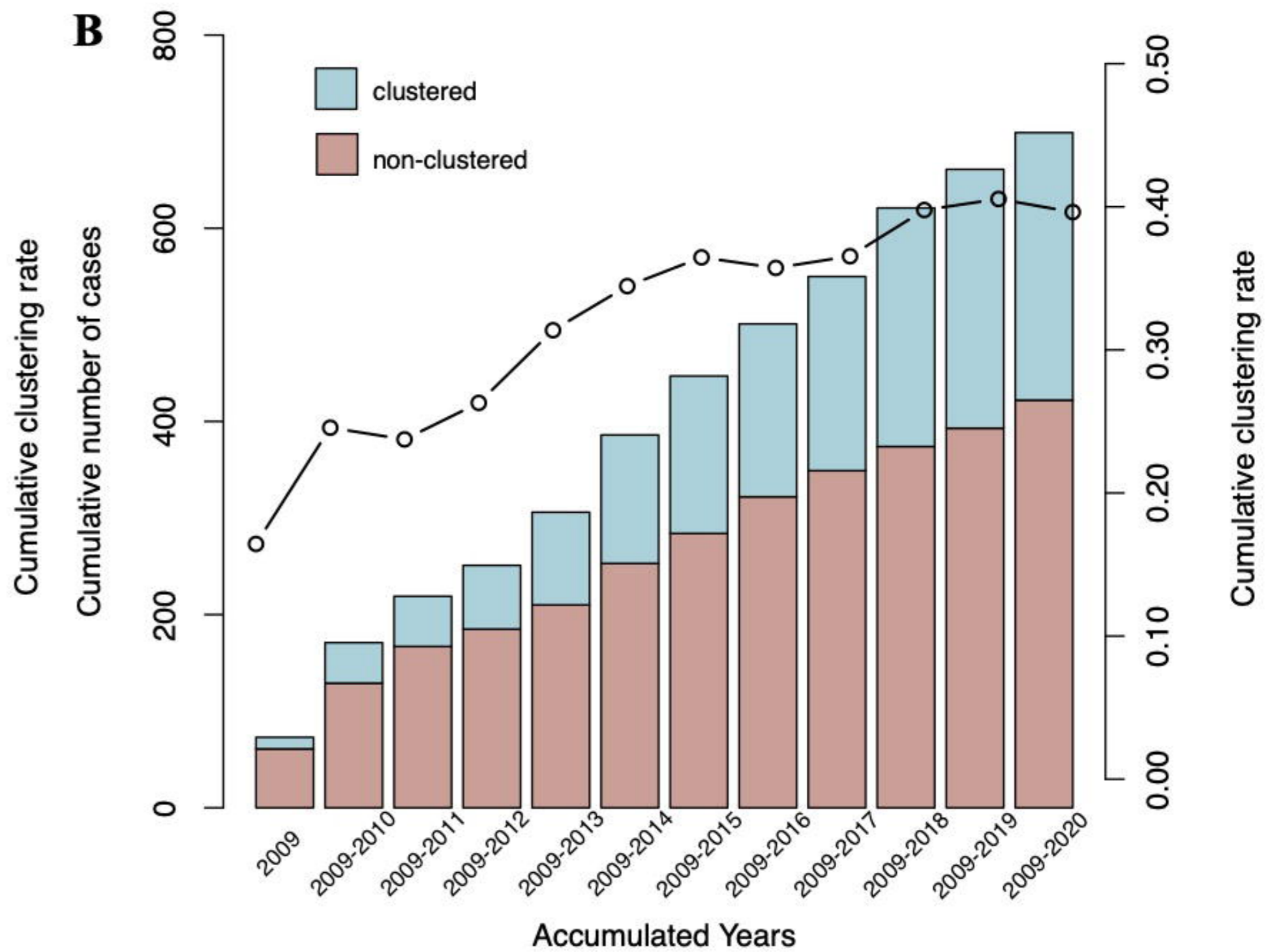
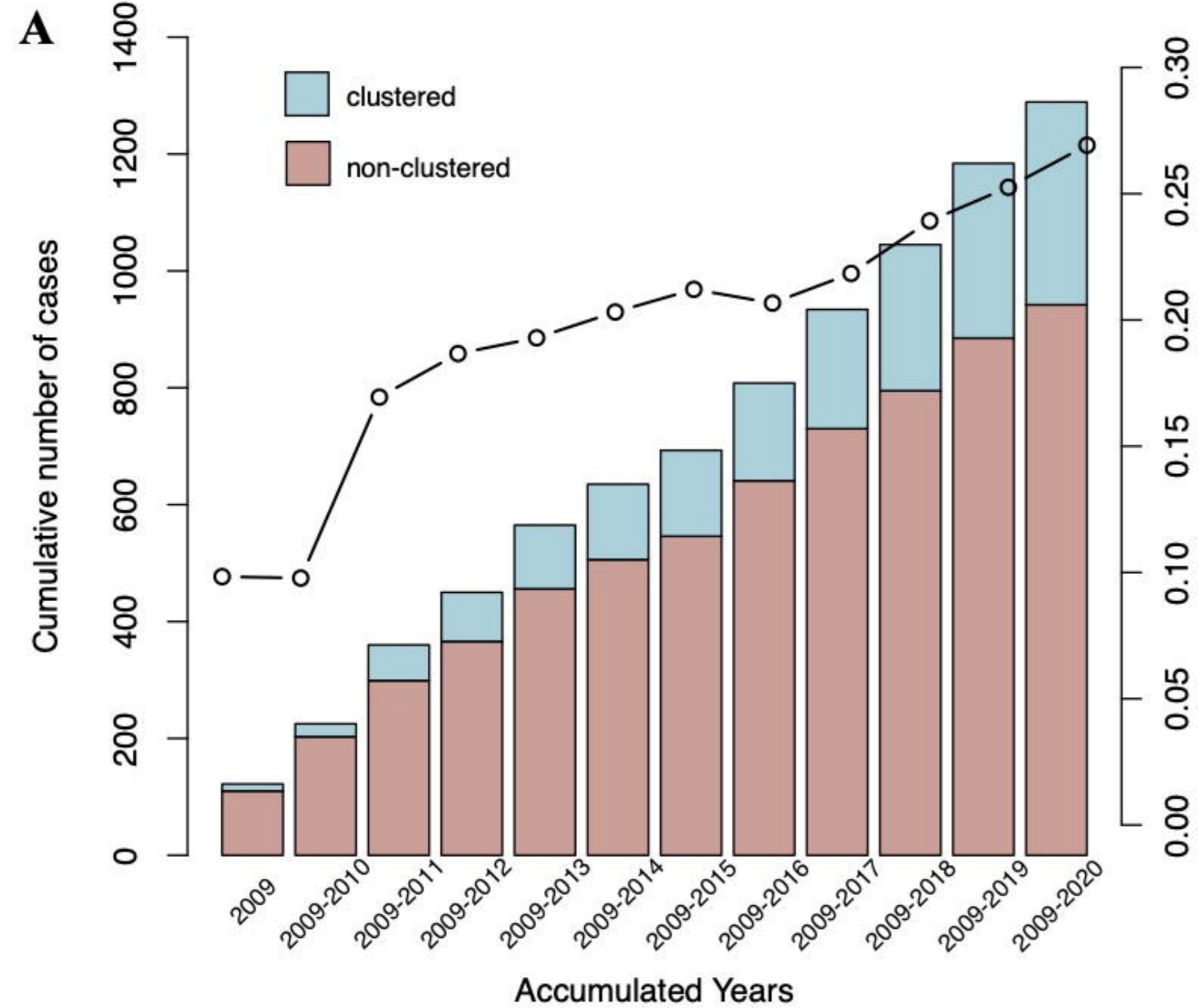
exclude 10 recurrent cases

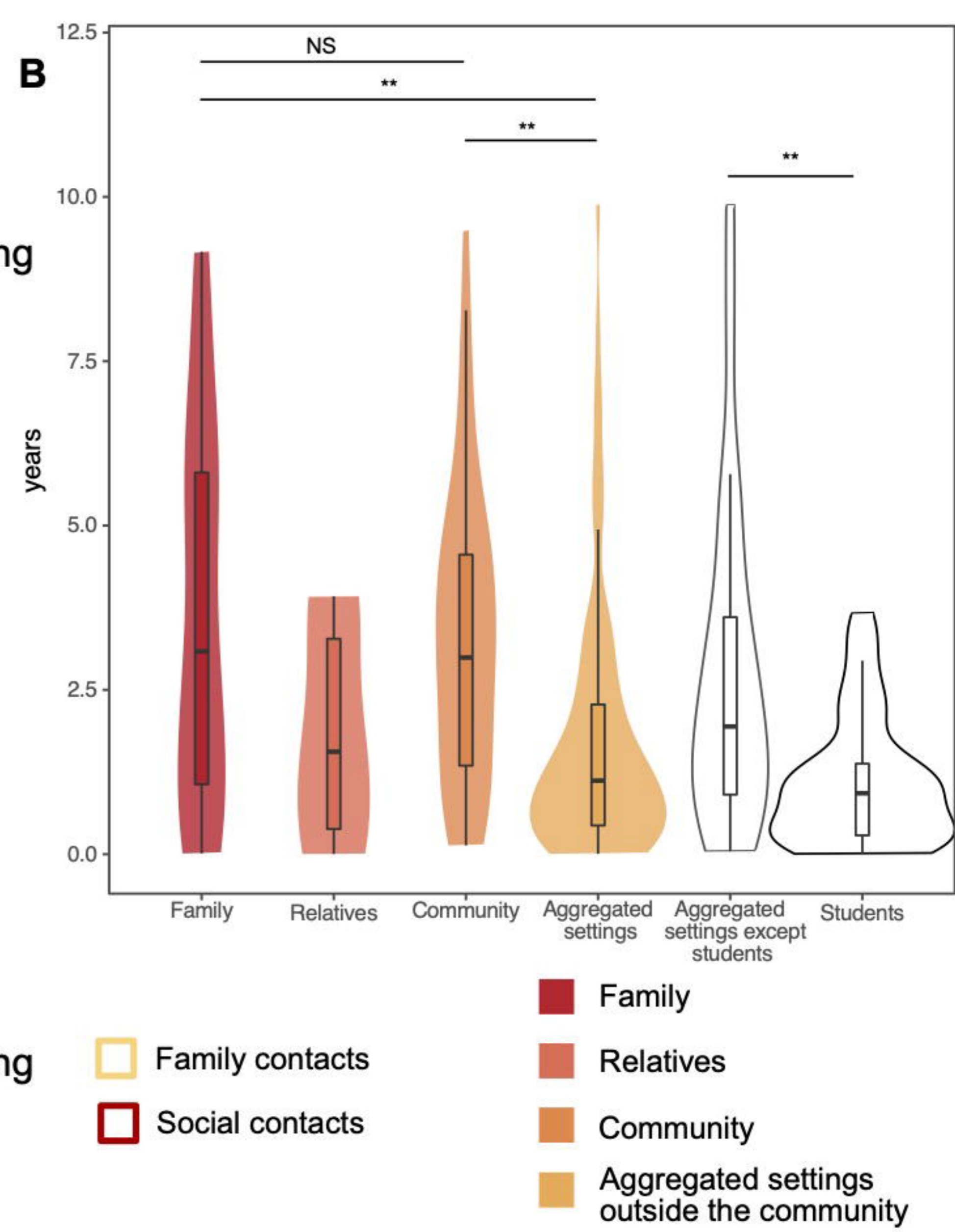
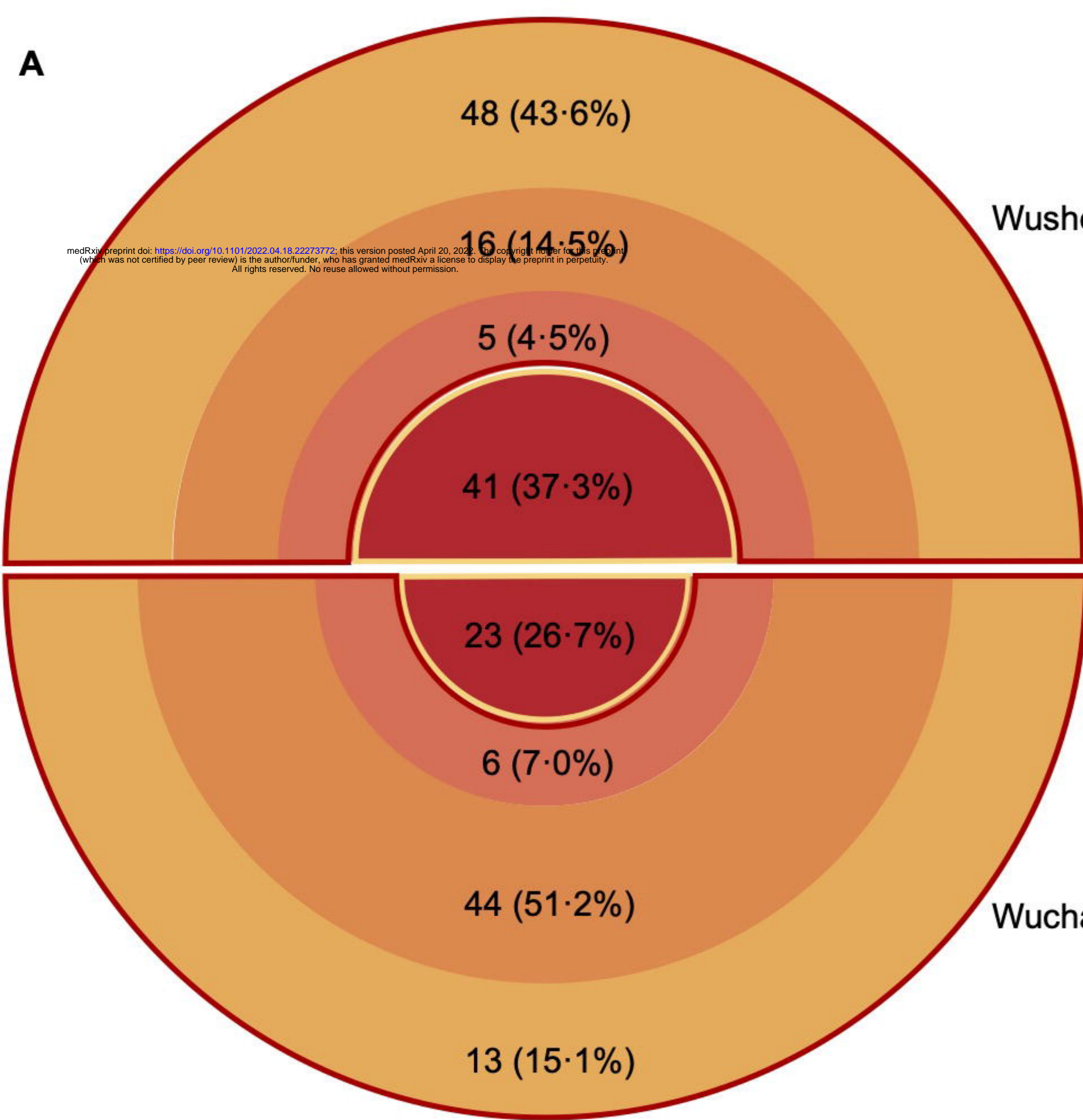
699 (98.6%) cases were included in the analysis

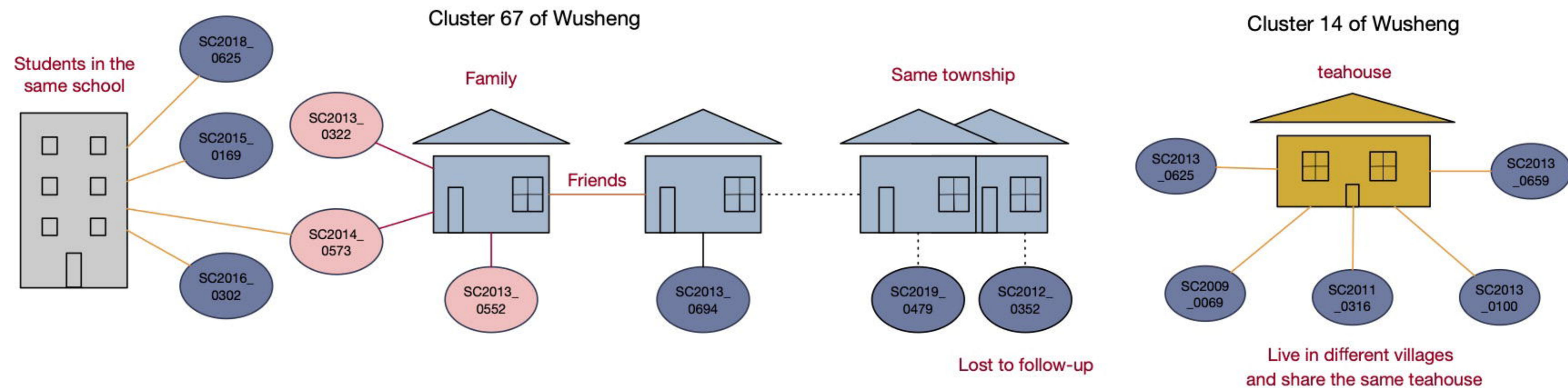
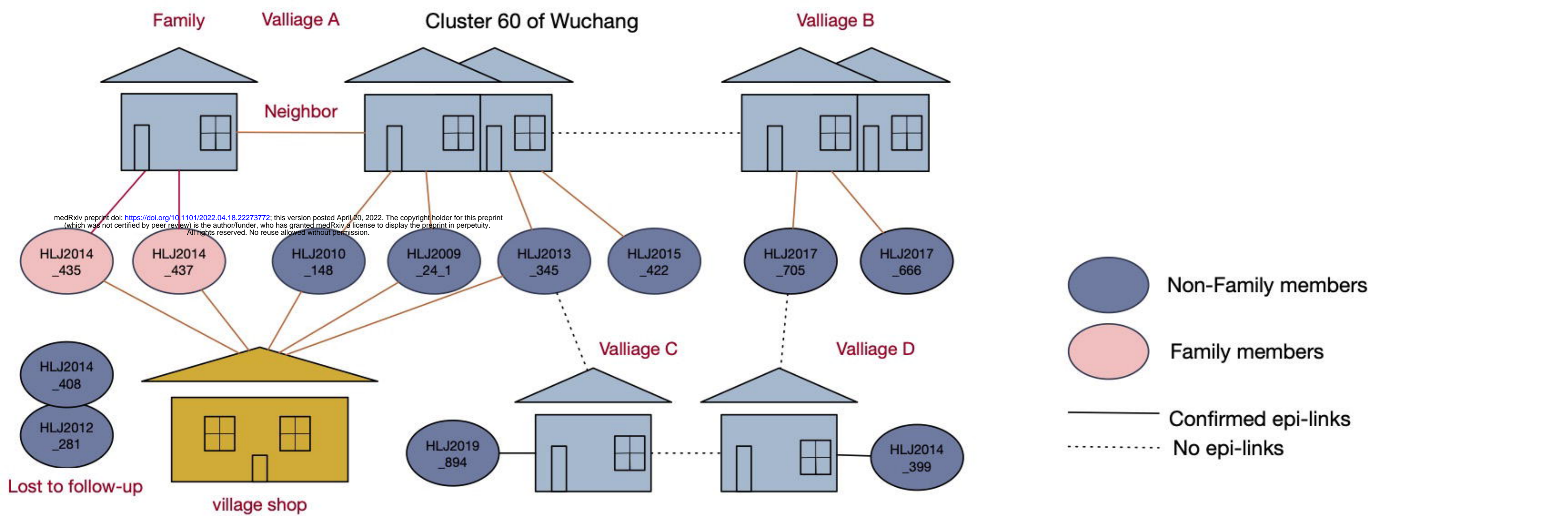


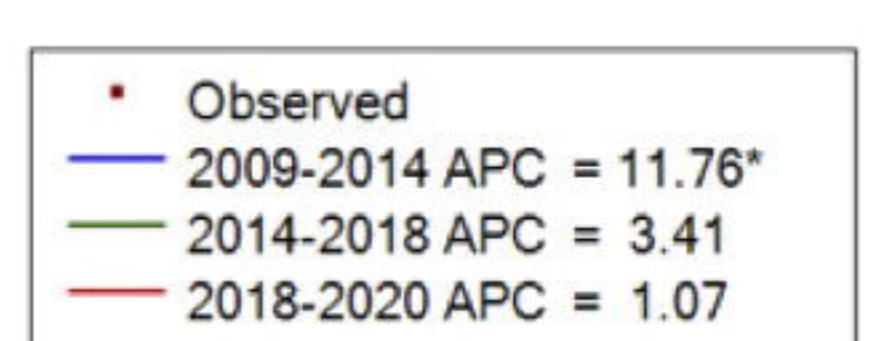
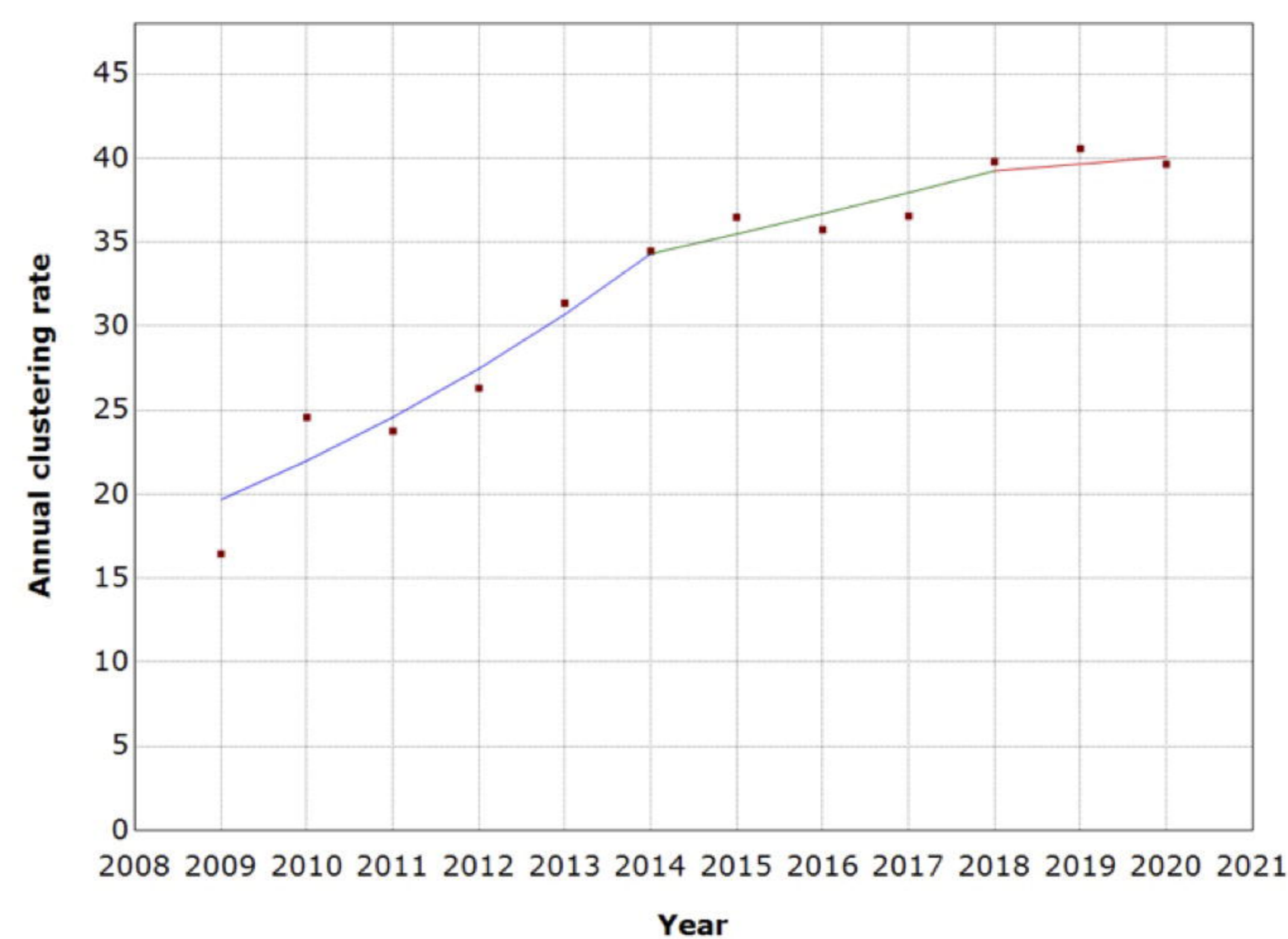
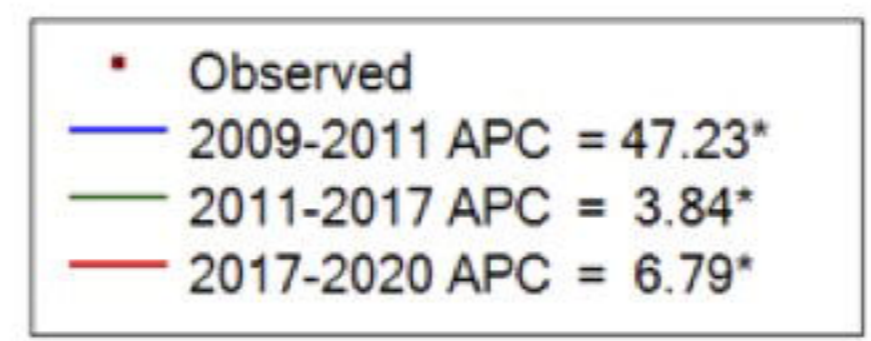
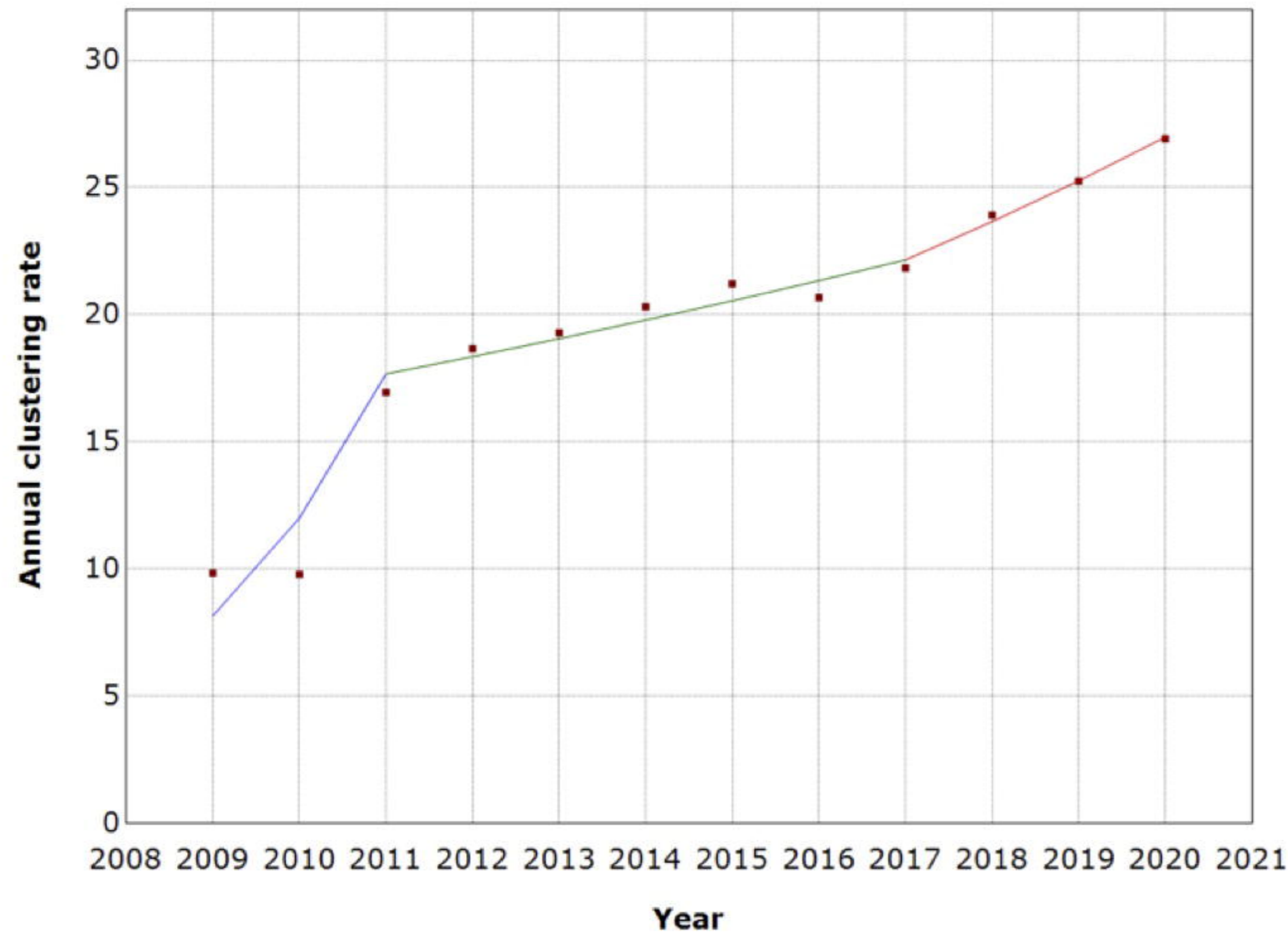
Clustered  
□ No  
■ Yes

Drug-resistance profile  
□ Pan-susceptible  
■ Other DR  
■ MDR

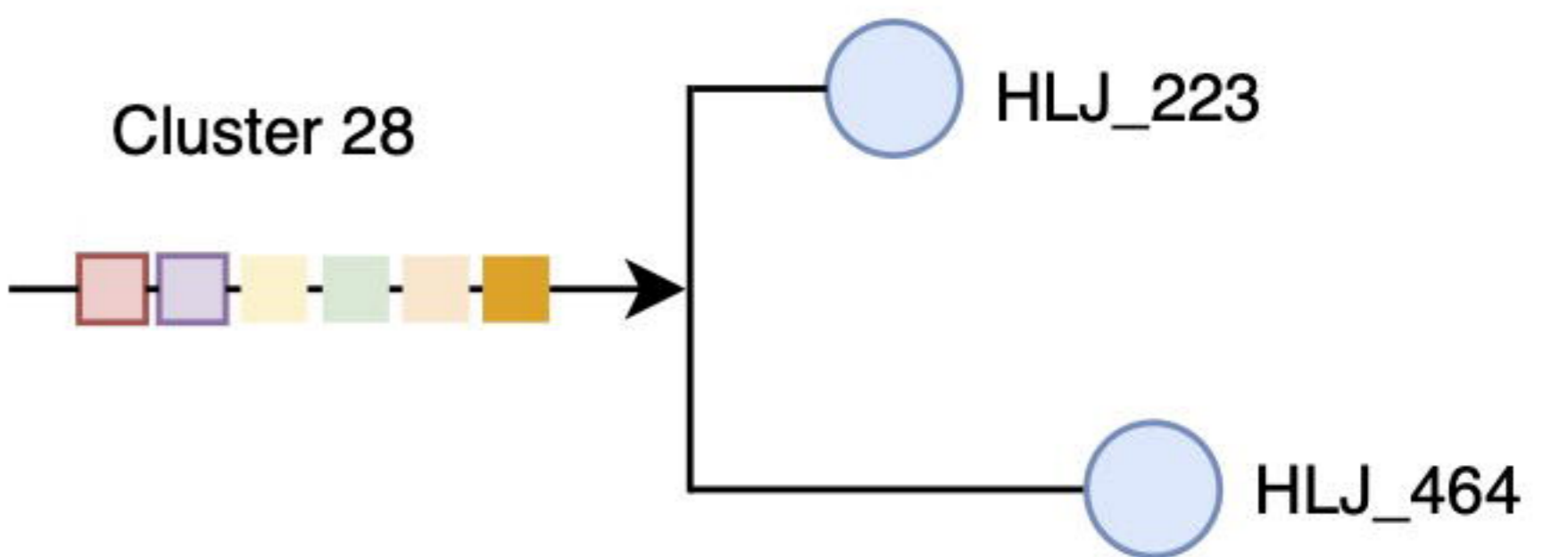
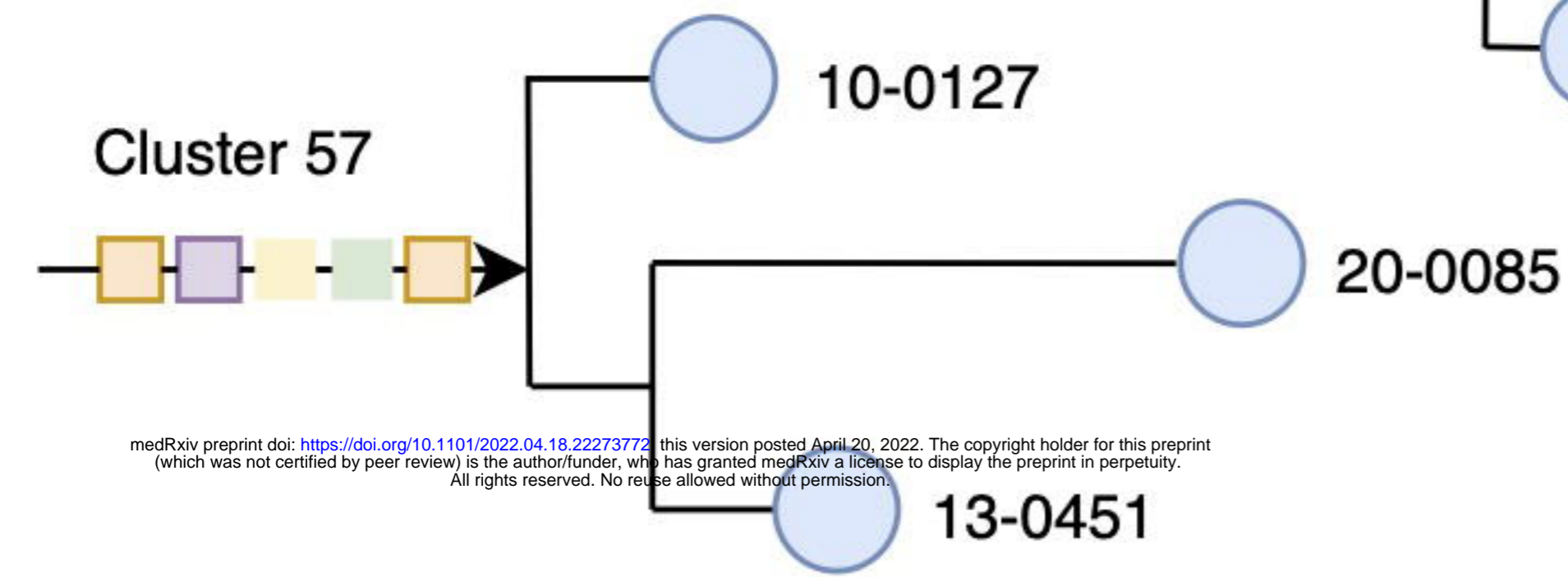
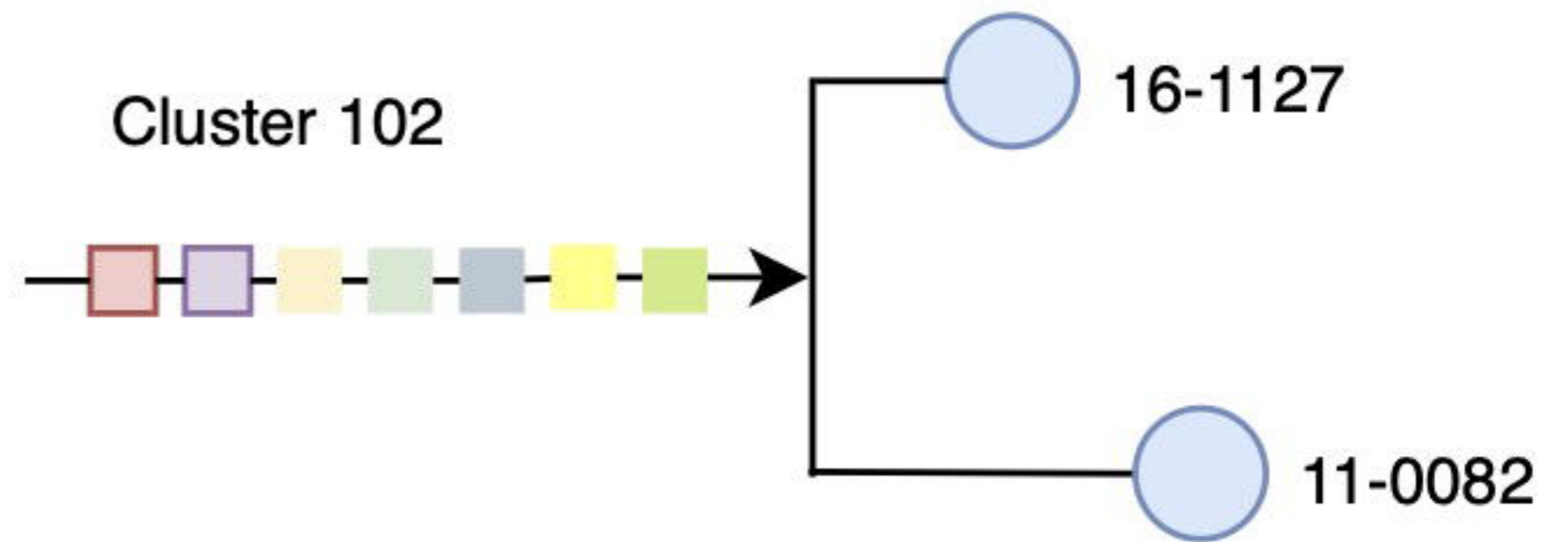
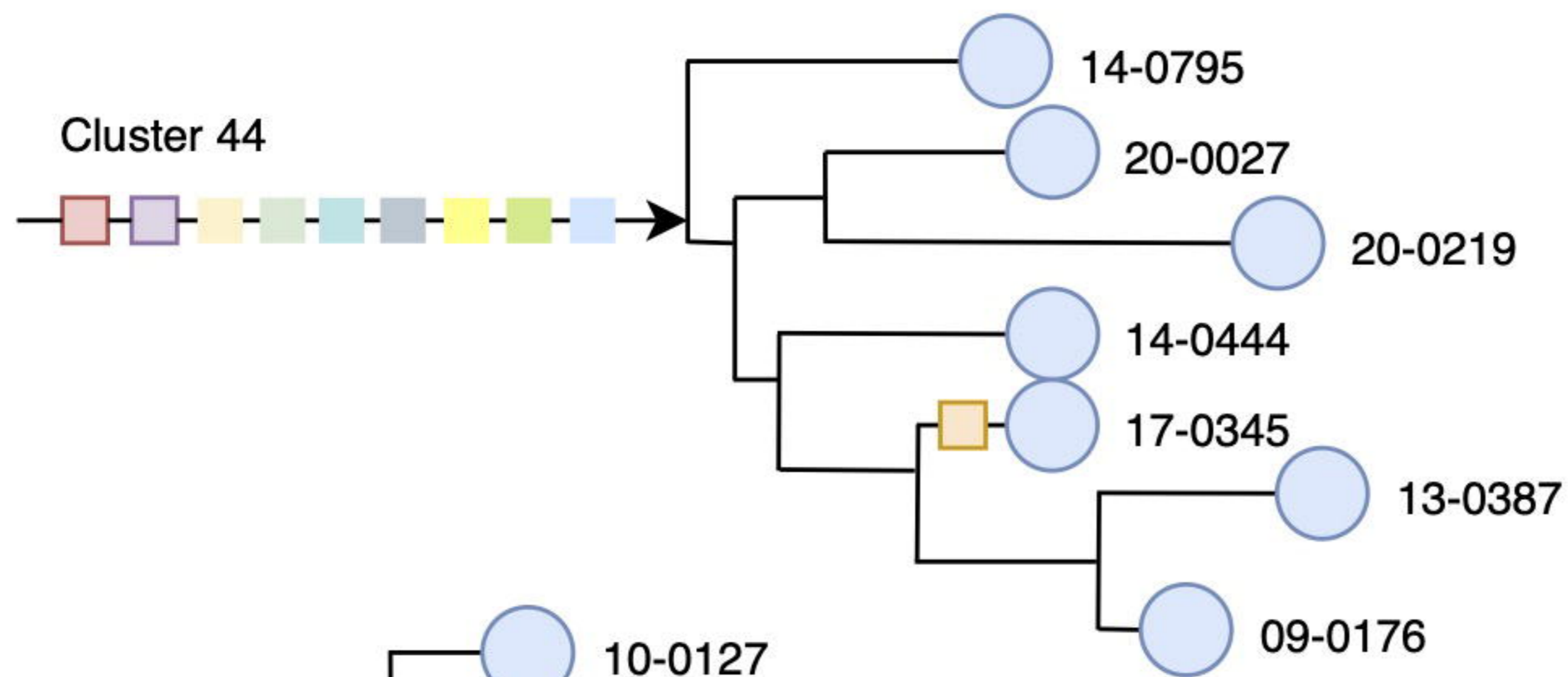
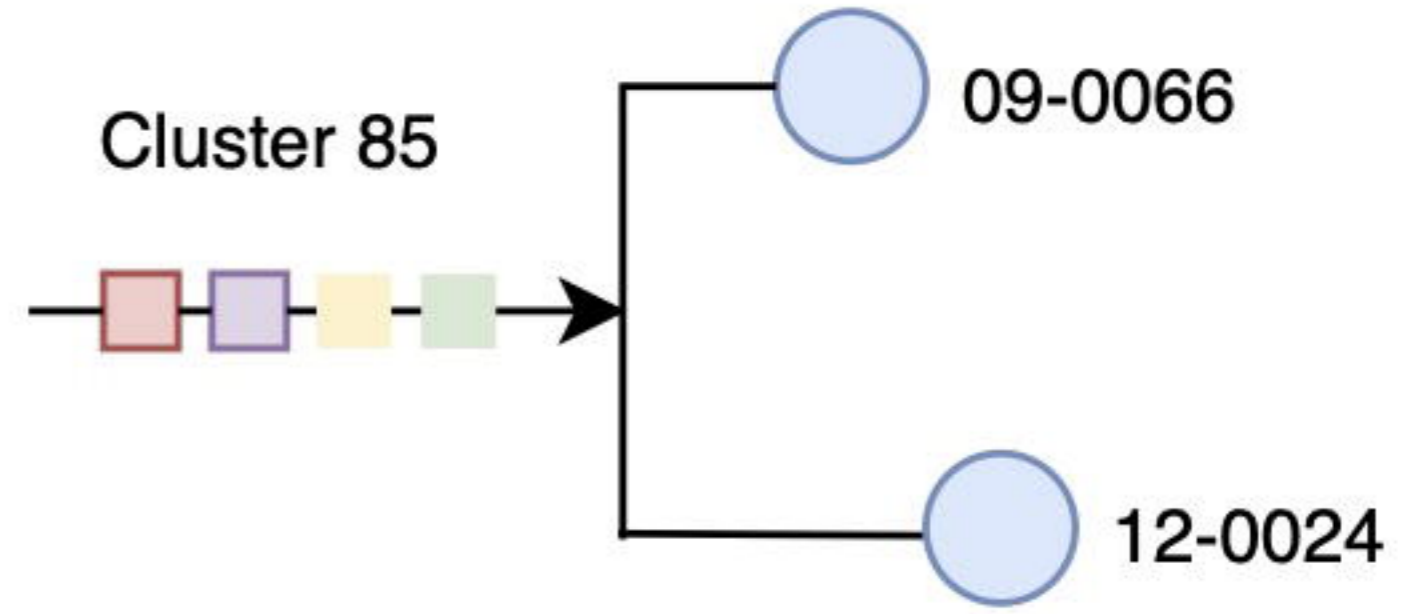
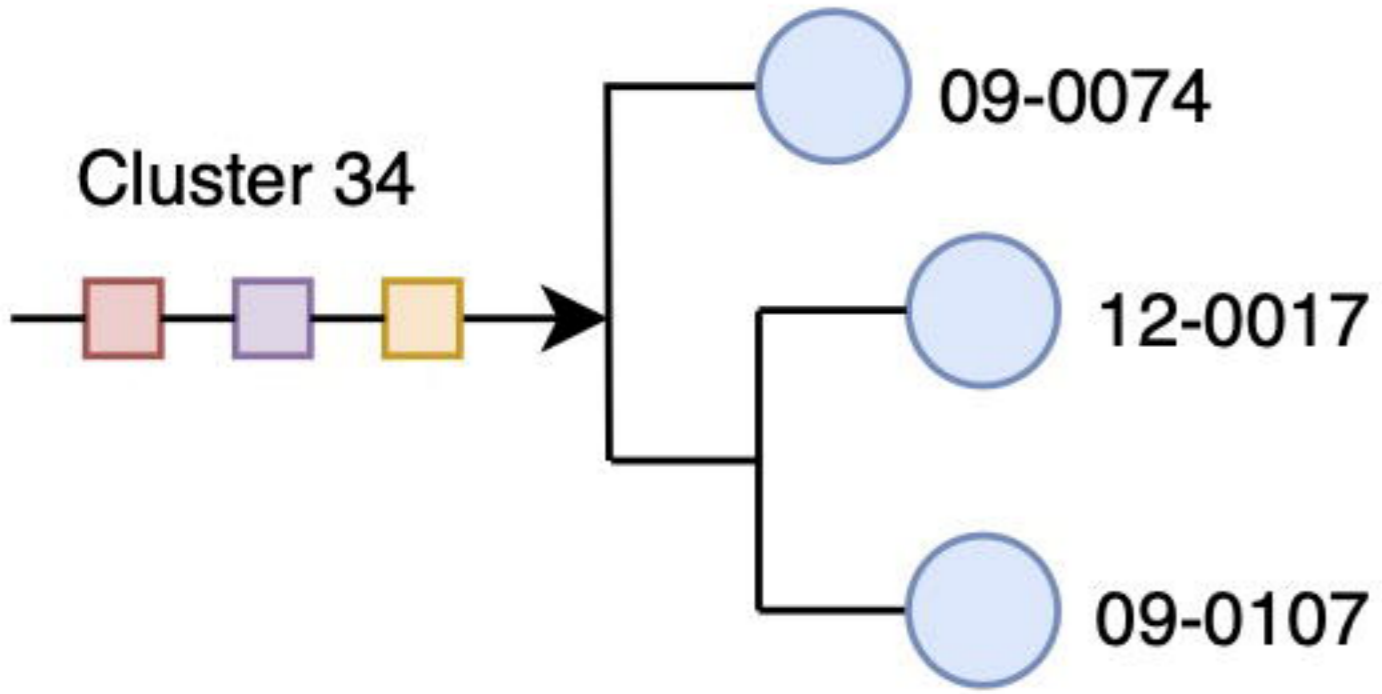
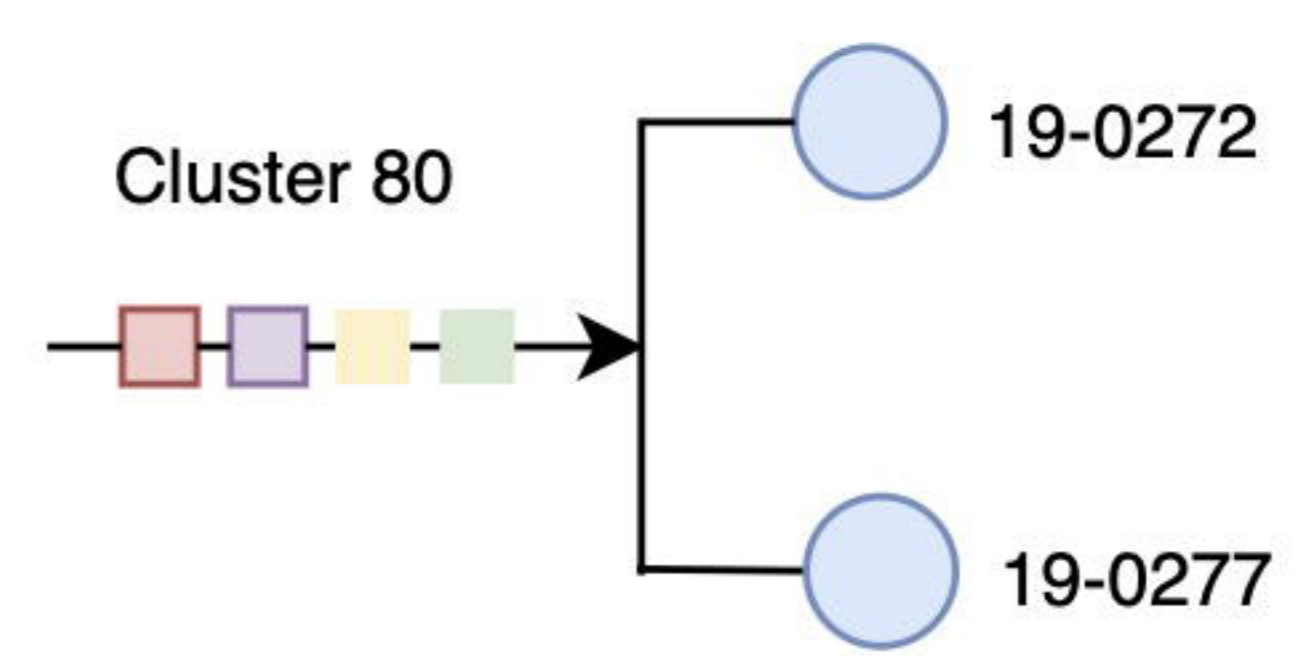
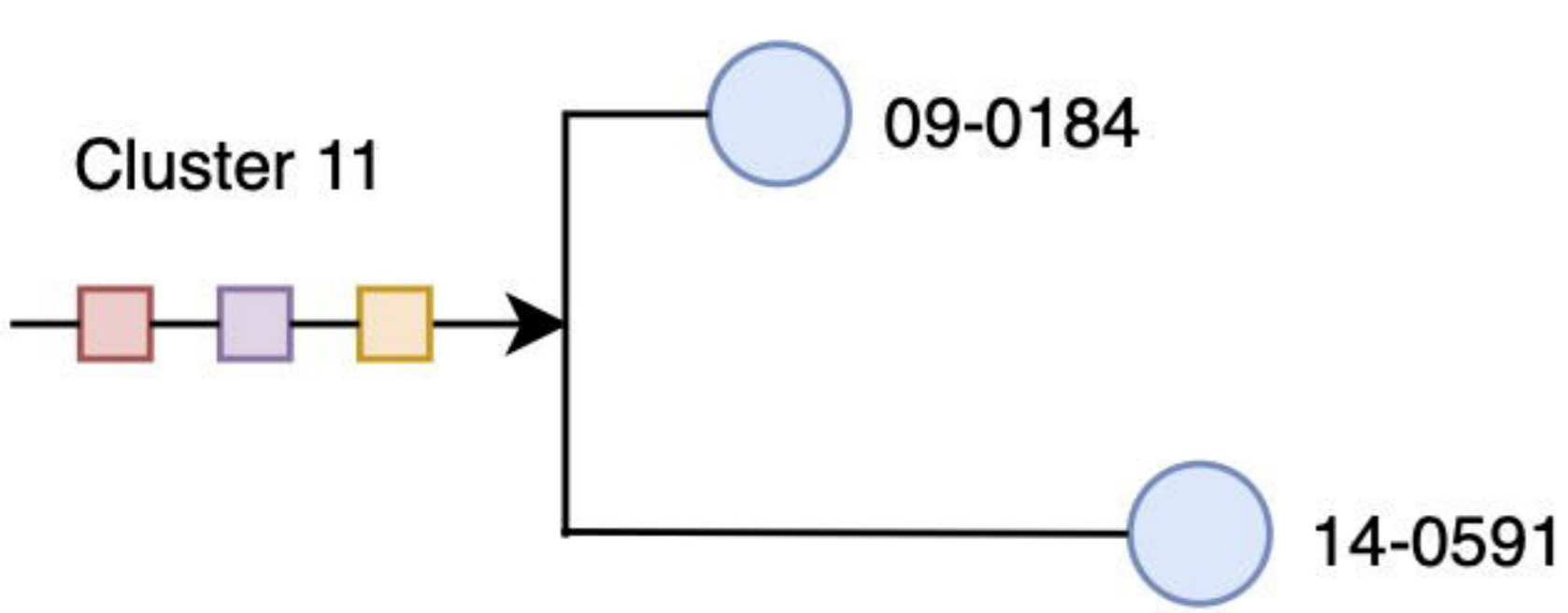












medRxiv preprint doi: <https://doi.org/10.1101/2022.04.18.22273772>; this version posted April 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

