JCM
Journals.ASM.org

Clustering of Tuberculosis Cases Based on Variable-Number Tandem-Repeat Typing in Relation to the Population Structure of *Mycobacterium tuberculosis* in the Netherlands

Rosa Sloot,[a] Martien W. Borgdorff,[b] Jessica L. de Beer,[c] Jakko van Ingen,[d] Philip Supply,[e,f,g,h,i] Dick van Soolingen[c,d,j]

Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands[a]; Department of Infectious Diseases, Public Health Service, Amsterdam, the Netherlands[b]; National Tuberculosis Reference Laboratory, National Institute for Public Health and the Environment, Bilthoven, the Netherlands[c]; Department of Medical Microbiology, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands[d]; Institut National de la Santé et de la Recherche Médicale U1019, Center for Infection and Immunity of Lille, Lille, France[e]; Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8204, Center for Infection and Immunity of Lille, Lille, France[f]; Université Lille Nord de France, Center for Infection and Immunity of Lille, Lille, France[g]; Institut Pasteur de Lille, Center for Infection and Immunity of Lille, Lille, France[h]; GenoScreen, Lille, France[i]; Department of Pulmonary Diseases, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands[j]

The population structure of 3,776 *Mycobacterium tuberculosis* isolates was determined using variable-number tandem-repeat (VNTR) typing. The degree of clonality was so high that a more relaxed definition of clustering cannot be applied. Among recent immigrants with non-Euro-American isolates, transmission is overestimated if based on identical VNTR patterns.

DNA typing is a powerful tool to trace tuberculosis (TB) transmission and outbreaks. Clustering of *Mycobacterium tuberculosis* isolates based on identical DNA fingerprints is commonly used as a proxy for recent transmission (1). However, this assumption is not always correct and depends on many factors, such as circulation of genetically similar strains, evolution of *M. tuberculosis* over time, transmission rate, DNA typing methods applied, duration of the study period, sampling, and effectiveness of TB control (2, 3). Various studies have shown that not all cases in DNA fingerprint clusters have epidemiological links with other cases in the cluster (4, 5). Moreover, epidemiological links have been found between cases caused by bacteria with slightly different DNA fingerprints (6). Clustering results among cases in the immigrant population especially should be interpreted with caution (7, 8), as isolates from these patients often belong to genetically compact strain lineages predominating in the countries of origin (9, 10, 11, 12).

In the Netherlands, more than 70% of all TB cases are found among foreign-born persons, and extensive information on each patient is stored in a national registry. We aimed to investigate the population structure of *M. tuberculosis* isolates among native and immigrant cases and to determine the consequences for the interpretation of recent transmission based on variable-number tandem-repeat (VNTR) typing results.

Culture-confirmed TB cases from October 2003 to December 2008 were included in this study. Patient information was obtained from the Netherlands Tuberculosis Register (NTR), held by the KNCV Tuberculosis Foundation. In total, 3,975 *M. tuberculosis* isolates were typed by IS*6110*/PGRS restriction fragment length polymorphism (RFLP) and standard 24-locus VNTR typing (13, 14) at the RIVM or by Genoscreen (Lille). Molecular data were matched with demographic data using the date of birth, sex, postal area code, and year of diagnosis, resulting in 3,793 (95%) matching cases. After exclusion of 17 foreign-born individuals because of incomplete data for several variables, 3,776 (95%) cases remained eligible.

Genotype information was uploaded to the MIRU-VNTRplus web-application (http://www.miru-vntrplus.org) (15) for phylogenetic lineage prediction, which was performed stepwise as described by Allix-Beguec et al. (16). Isolates that were part of the CAS, Beijing, EAI, *Mycobacterium bovis*, and *Mycobacterium africanum* lineages were categorized as non-Euro-American and the remaining as a Euro-American superlineage (16).

Clonal complexes, defined as groups of at least two isolates differing in not more than 3/24 loci, were identified on a minimum-spanning tree with BioNumerics software (Applied Maths, Kortrijk, Belgium), using MIRU-VNTR data and the categorical distance, which scores the number of alleles shared or different over the 24 markers used.

Multiple imputation was used to account for 184 (5%) and 37 (1%) of 3,776 cases with missing data for the variables "time since immigration at TB diagnosis" and "gender," respectively. All remaining variables were used to create five imputed data sets, and results are based on pooled statistics. Two different cluster definitions were used to investigate the interpretation of recent transmission; identical VNTR patterns and single-locus variants (SLVs). The theory behind this was that allowing SLVs to be clustered might involve genetically closely related strains that in fact share the same transmission chain. The analyses were performed separately for the Euro-American and non-Euro-American lineages and completed with SPSS 18.0 (SPSS, Chicago, IL) and statistical program R version 2.11.0.

A minimum-spanning tree was produced for all 3,776 isolates included in the analysis. In total 3,377 (89%) isolates were distributed over 83 clonal complexes (Fig. 1), whereas the remaining 399 (11%) isolates did not belong to any clonal complex. Within each complex, all the VNTR patterns represented the same lineage type,

Received 21 February 2013 **Returned for modification** 4 April 2013
Accepted 1 May 2013
Published ahead of print 8 May 2013
Address correspondence to Dick van Soolingen, Dick.van.Soolingen@rivm.nl.

doi:10.1128/JCM.00489-13

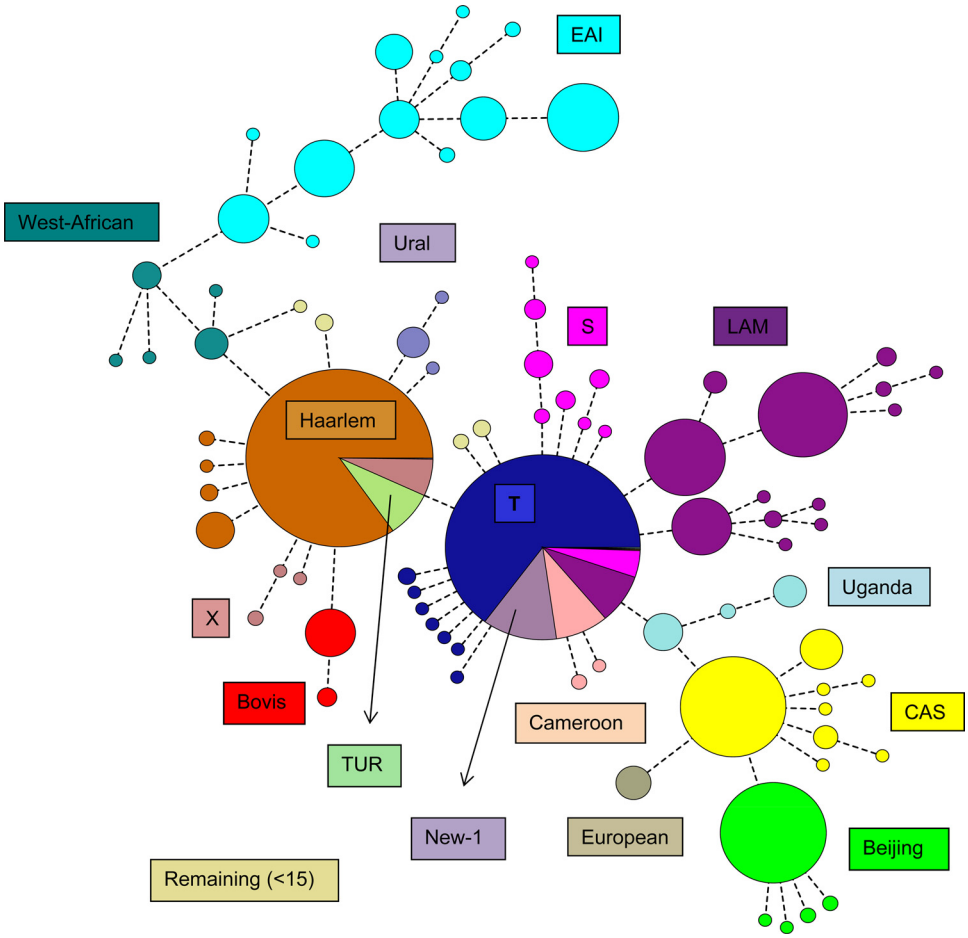

FIG 1 Identification of clonal complexes in the total study population in a minimum-spanning tree.

except the two largest complexes comprising 84% Haarlem strains and 67% T-specific strains.

Of the 3,776 isolates, 1,130 (30%) represented the non-Euro-American lineages (Table 1). We reasoned that recently arrived immigrant cases having nonclustered *M. tuberculosis* isolates most likely represent importation of foreign genotypes. Among the 504 nonclustered recent-immigrant cases, 239 (47%) were caused by isolates of the non-Euro-American lineages, of which EAI (45%), CAS (26%), and Beijing (18%) constituted the majority (Table 1). Cases caused by these non-Euro-American lineages originated from Asia (41%) and Africa (56%). In contrast, recent-immigrant nonclustered cases with Euro-American lineages had a higher diversity in geographical origin. Furthermore, 40 (8%) of the 504 recent-immigrant nonclustered cases originated from European countries, of which the majority (93%) were caused by the Euro-American lineages. Among the 564 native Dutch cases with nonclustered *M. tuberculosis* isolates, 459 (81%) had isolates of the Euro-American lineages, of which the majority were of the Haarlem (36%), T-specific (33%), and LAM (13%) lineages (Table 1).

Patient factors significantly associated with VNTR clustering in the whole study population, using identical profiles as a cluster definition, were analyzed. As observed in previous studies (17, 18), we found male sex, young age, urban residence, having pulmonary tuberculosis, and no previous treatment for tuberculosis as significant risk factors for clustering. All risk factors became less

TABLE 1 Distribution of non-Euro-American and Euro-American lineages over clustered and nonclustered immigrant and native Dutch tuberculosis cases

Lineage	No. of isolates					
	Nonclustered cases			Clustered cases		
	Immigrants					
	Resident <3 yr	Resident ≥3 yr	Natives	Immigrants	Natives	Total study population
Non-Euro-American						
EAI	108	135	19	96	26	384
CAS	63	108	15	117	25	328
Beijing	42	69	37	98	28	274
Bovis	5	10	26	10	11	62
West African (I, II)	16	21	2	18	2	59
With <10 isolates	5	10	6	2	0	23
Total	239	353	105	341	92	1,130
Euro-American						
Haarlem	55	133	166	278	182	814
T specific	64	126	151	152	134	627
LAM	68	148	61	236	102	615
S	13	44	30	23	22	132
Uganda (I, II)	9	28	4	36	4	81
New-1	15	47	8	8	7	85
TUR	9	16	5	32	20	82
X	8	16	16	19	21	80
Cameroon	13	15	10	30	6	74
Ural	7	11	6	5	2	31
With <15 isolates	4	14	2	5	0	25
Total	265	598	459	824	500	2,646

TABLE 2 Risk factors for clustering, with and without allowing single-locus variants to be clustered, according to lineage[a]

Parameter	Non-Euro-American lineages					Euro-American lineages					Total no. of isolates in study population
	No locus variation		Allowing SLVs to be clustered			No locus variation		Allowing SLVs to be clustered			
	No. (%) of VNTR clustered isolates	OR (95% CI)	No. (%) of VNTR clustered isolates	OR (95% CI)	Total no. of isolates	No. (%) of VNTR clustered isolates	OR (95% CI)	No. (%) of VNTR clustered isolates	OR (95% CI)	Total no. of isolates	
Sex											
Male	243 (39)	1.1 (0.9–1.4)	373 (60)	1.0 (0.8–1.3)	619	855 (54)	1.5 (1.3–1.8)	1,023 (65)	1.3 (1.1–1.6)	1,579	2,198
Female	190 (37)	1	306 (60)	1	511	469 (44)	1	616 (58)	1	1,067	1,578
Age (yr)											
<30	157 (40)	1.7 (1.0–2.8)	240 (61)	1.7 (1.1–2.7)	392	429 (58)	5.1 (3.9–6.7)	501 (68)	3.5 (2.7–4.6)	734	1,126
30–55	202 (38)	1.6 (0.9–2.6)	323 (61)	1.6 (1.0–2.6)	534	645 (56)	4.6 (3.5–5.9)	771 (67)	3.3 (2.6–4.2)	1,153	1,687
55–70	49 (43)	1.9 (1.1–3.4)	73 (64)	1.9 (1.1–3.3)	115	162 (46)	3.1 (2.3–4.2)	213 (61)	2.5 (1.9–3.4)	352	467
>70	25 (28)	1	43 (48)	1	89	88 (22)	1	154 (38)	1	407	496
Residence											
Urban	165 (42)	1.3 (0.9–1.6)	237 (60)	1.0 (0.8–1.3)	393	601 (58)	1.7 (1.4–1.9)	721 (69)	1.7 (1.4–2)	1,040	1,433
Rural	268 (36)	1	442 (60)	1	737	723 (45)	1	918 (57)	1	1,606	2,343
Had TB before											
Yes	12 (26)	0.5 (0.3–1.1)	25 (53)	0.7 (0.4–1.3)	47	72 (38)	0.6 (0.4–0.8)	89 (47)	0.5 (0.4–0.7)	190	237
No/unknown	421 (39)	1	654 (60)	1	1,083	1,252 (51)	1	1,550 (63)	1	2,456	3,539
Localization of TB											
Pulmonary	228 (44)	1.8 (1.4–2.4)	332 (65)	1.5 (1.2–2.0)	514	844 (53)	1.6 (1.3–1.9)	1,026 (65)	1.5 (1.3–1.8)	1,589	2,103
Extrapulmonary	141 (31)	1	251 (54)	1	463	303 (41)	1	396 (54)	1	732	1,195
Pulmonary + extrapulmonary	64 (42)	1.6 (1.1–2.4)	96 (63)	1.4 (0.9–2.1)	153	177 (55)	1.7 (1.3–2.2)	217 (67)	1.7 (1.3–2.2)	325	478
RFLP											
Clustered	256 (74)	9.9 (7.4–13.3)	306 (89)	8.7 (6.0–12.4)	345	1,061 (84)	21.9 (17.9–26.7)	1,145 (90)	16.8 (13.5–20.8)	1,267	1,612
Nonclustered	177 (23)	1	373 (48)	1	785	263 (19)	1	494 (36)	1	1,379	2,164
Total	433		679		1,130	1,324		1,639		2,646	3,776

[a] SLV, single-locus variant (maximum threshold of one MIRU locus variation from a central strain or any other strain within the same lineage [Euro-American versus non-Euro-American lineages]); OR, odds ratio; CI, confidence interval.

strongly associated when analyses was restricted to cases caused by non-Euro-American lineages (Table 2). Only age (<30 and 55 to 70 years) and having pulmonary tuberculosis remained significantly associated with clustering.

Allowing SLVs to be clustered increased the clustering proportion by 24% and 57% among cases with Euro-American isolates and non-Euro-American isolates, respectively (Table 2). This resulted in a decreased magnitude of association between the risk factors and clustering, in particular for the cases caused by non-Euro-American lineages (Table 2).

Similarly, the association between RFLP and VNTR clustering in cases caused by non-Euro-American lineages was reduced compared to that in cases caused by Euro-American lineages. Discrepancy between VNTR and RFLP typing (i.e., clustered by either VNTR or RFLP) was in most cases caused by VNTR clustered and RFLP nonclustered isolates (Table 2). Among 1,130 non-Euro-American isolates, 177 (16%) were clustered by VNTR and nonclustered by RFLP. In contrast, among 2,646 Euro-American isolates, 263 (10%) were clustered by VNTR and nonclustered by RFLP (Table 2).

This study shows the lineage-dependent degree of reliability of the inference on transmission. Classification of lineage type was based on the geographical association between patient origin and strain lineage, defined as Euro-American and non-Euro-American. Among nonclustered native Dutch TB cases, Euro-American lineages were most frequently isolated. Domination of these lineages among native TB cases has also been shown in other European populations (19, 20), suggesting that these lineages have been circulating in Europe for centuries (21). In contrast, recent-immigrant cases caused by nonclustered non-Euro-American strains originated from distant geographical areas.

Risk factors for recent transmission, as determined by VNTR clustering, were reduced in the non-Euro-American lineages compared to the Euro-American lineages, indicating the lineage dependence. This was further visible when testing the effect of tolerating single-locus variants in the cluster definition, as the increase in clustered non-Euro-American strains was twice as high as that among Euro-American strains, reflecting the clonality of the former strains in the study population. Furthermore, the magnitude of association between risk factors and clustering decreased after allowing single-locus variants, especially among cases with non-Euro-American isolates, thus increasing overestimation of recent transmission for cases caused by non-Euro-American lineages.

In conclusion, to remain useful in TB control practice, the definition of a cluster on the basis of VNTR typing should be a fully identical 24-locus VNTR typing result. This study further indicated limits in the interpretation of recent transmission based on clustering by VNTR typing in the recent-immigrant population. Our findings are in particular relevant for other European low-incidence countries having similar forms of immigration.

ACKNOWLEDGMENTS

We thank all the Municipal Health Services that reported their cluster investigations to the national surveillance unit and the microbiological laboratories in the Netherlands for their efforts to send *Mycobacterium tuberculosis* isolates to the RIVM.

P.S. is a consultant for GenoScreen, Lille, France. All other authors have declared that no competing interests exist.

REFERENCES

1. **Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, Schecter GF, Daley CL, Schoolnik GK.** 1994. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N. Engl. J. Med. **330**:1703–1709.
2. **Jasmer RM, Hahn JA, Small PM, Daley CL, Behr MA, Moss AR, Creasman JM, Schecter GF, Paz EA, Hopewell PC.** 1999. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991-1997. Ann. Intern. Med. **130**:971–978.
3. **Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, Godfrey-Faussett P, Vynnycky E.** 1999. Interpreting DNA fingerprint clusters of *Mycobacterium tuberculosis*. Int. J. Tuberc. Lung Dis. **3**:1055–1060.
4. **Ellis BA, Crawford JT, Braden CR, McNabb SJ, Moore M, Kammerer S.** 2002. Molecular epidemiology of tuberculosis in a sentinel surveillance population. Emerg. Infect. Dis. **8**:1197–1209.
5. **Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro KG, Moers D, Yang Z, Stead WW, Bates JH.** 1997. Interpretation of restriction fragment length polymorphism analysis of Mycobacterium tuberculosis isolates from a state with a large rural population. J. Infect. Dis. **175**:1446–1452.
6. **Schürch AC, Kremer K, Daviena O, Kiers A, Boeree MJ, Siezen RJ, van Soolingen D.** 2010. High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. J. Clin. Microbiol. **48**: 3403–3406.
7. **Diel R, Rüsch-Gerdes S, Niemann S.** 2004. Molecular epidemiology of tuberculosis among immigrants in Hamburg, Germany. J. Clin. Microbiol. **42**:2952–2960.
8. **Iñigo J, García de Viedma D, Arce A, Palenque E, Alonso Rodríguez N, Rodríguez E, Ruiz Serrano MJ, Andrés S, Bouza E, Chaves F.** 2007. Analysis of changes in recent tuberculosis transmission patterns after a sharp increase in immigration. J. Clin. Microbiol. **45**:63–69.
9. **van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, Qing HZ, Enkhsaikan D, Nymadawa P, van Embden JD.** 1995. Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. J. Clin. Microbiol. **33**:3234–3238.
10. **Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM.** 2004. Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proc. Natl. Acad. Sci. U. S. A. **101**:4871–4876.
11. **Cardoso Oelemann M, Gomes HM, Willery E, Possuelo L, Batista Lima KV, Allix-Béguec C, Locht C, Goguet de la Salmonière YO, Gutierrez MC, Suffys P, Supply P.** 2011. The forest behind the tree: phylogenetic exploration of a dominant Mycobacterium tuberculosisstrain lineage from a high tuberculosis burden country. PLoS One **6**:e18256. doi:10.1371/journal.pone.0018256.
12. **Lillebaek T, Andersen AB, Bauer J, Dirksen A, Glismann S, de Haas P, Kok-Jensen A.** 2001. Risk of Mycobacterium tuberculosis transmission in a low-incidence country due to immigration from high-incidence areas. J. Clin. Microbiol. **39**:855–861.
13. **Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D.** 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J. Clin. Microbiol. **44**:4498–4510.
14. **van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, Small PM.** 1993. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31**:406–409.
15. **Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D.** 2010. MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. Nucleic Acids Res. **38**:W326–W331.
16. **Allix-Béguec C, Fauville-Dufaux M, Supply P.** 2008. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of Mycobacterium tuberculosis. J. Clin. Microbiol. **46**:1398–1406.
17. **Bauer J, Yang Z, Poulsen S, Andersen AB.** 1998. Results from 5 years of nationwide DNA fingerprinting of *Mycobacterium tuberculosis* complex isolates in a country with a low incidence of *M. tuberculosis* infection. J. Clin. Microbiol. **36**:305–308.

18. **van Soolingen D, Borgdorff MW, de Haas PE, Sebek MM, Veen J, Dessens M, Kremer K, van Embden JD.** 1999. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. J. Infect. Dis. **180**:726–736.
19. **Garzelli C, Lari N, Cuccu B, Tortoli E, Rindi L.** 2010. Impact of immigration on tuberculosis in a low-incidence area of Italy: a molecular epidemiological approach. Clin. Microbiol. Infect. **16**:1691–1697.
20. **Kinander W, Bruvik T, Dahle UR.** 2009. Dominant Mycobacterium tuberculosis lineages in elderly patients born in Norway. PLoS One **4**:e8373. doi:10.1371/journal.pone.0008373.
21. **Lillebaek T, Dirksen A, Vynnycky E, Baess I, Thomsen VØ, Andersen AB.** 2003. Stability of DNA patterns and evidence of Mycobacterium tuberculosis reactivation occurring decades after the initial infection. J. Infect. Dis. **188**:1032–1039.