Molecular clustering of patients with *Mycobacterium tuberculosis* strains cultured from the diabetic and non-diabetic newly diagnosed TB positive cases

Josephine W. Mburua, Leonard Kingwara, Magiri Esther, Nyerere Andrew

**ARTICLE INFO**

**Keywords:**
- Tuberculosis
- Diabetes
- Molecular clustering

**ABSTRACT**

**Background:** Social determinants of health, biological, and individual variants have been associated with Pulmonary TB (PTB) case clustering. None of the studies have focused on diabetes mellitus (DM) despite it being one of the co-morbidities affecting TB patients. Minimal data is available and it is not clear whether patients with DM and TB are more likely than TB patients without DM to be grouped into similar molecular clusters thus indicating a bias in transmission among TB/DM co-morbidity patients. 

**Objective:** To determine proportion of TB strains within TB and TB/DM cases that were clustered with their corresponding clinical outcomes and hence could be attributable to active TB transmission in the two urban counties of Nairobi, Kenya.

**Methods:** We carried out a prospective cohort study of non-pregnant patients aged 15 years and above that tested positive for TB in two peri-urban counties in Kenya between February 2014 and August 2015. Clinical and socio-demographic data were obtained from a questionnaire and medical records of the National TB program patient data base at two, three, five, and six months. Spoligotyping data was then obtained and compared from previously identified strains in a data bank from the spolDB4.

**Results:** We identified 7 different TB strains out of which East Asia Beijing, Euro America and Indo oceanic being the most dominant strain within the two counties accounting for 92.4% of the infections. DM was not a significant factor in increasing the likelihood of PTB patients to cluster according to the genotype of the infecting *Mycobacterium tuberculosis* bacillus. TB lineages, DM and County of the patient were found to be independent of the clinical outcomes that were observed in the study.

**Conclusion:** Diabetes mellitus is not a significant factor in increasing the molecular clustering among PTB patients.

**Introduction**

*Mycobacterium tuberculosis* and the other members of the *M. tuberculosis* complex (MTBC) remains one of the main causes of morbidity and mortality in low and medium-income countries, currently experiencing an upsurge in diabetes mellitus (DM) cases [1–3]. Studies indicates an increased risk of TB among patients with DM in addition to poor prognosis of patients suffering from DM/TB co-morbidity [4]. Some studies have further explored the relationship between diabetes mellitus (DM) and tuberculosis infecting *M. tuberculosis* bacillus. Socio-epidemiological data, molecular strain typing together with conventional epidemiologic methods have previously helped to further characterize *M. tuberculosis* strains and understand the dynamics and patterns of transmission of the pathogen in different regions and populations [5–8]. In molecular characterization, patients with identical strains of *M. tuberculosis* are considered to belong to one cluster resulting from recent transmission and rapid progression. Unique patterns results mainly from reactivation of latent infection or recent transmission from patients out of the period or area under study [9,10].

Variation in individual, biological and social determinants have also been associated with Pulmonary TB (PTB) case clustering. Other determinants include being male, young, country of birth, resident of an urban area, alcohol or drug consumption, homelessness, HIV infection or having acid-fast bacilli in sputum smear [11,12]. Similar determinants have been described in medium income countries. Despite numerous studies on the association of HIV and increased risk of *M.
tuberculosis transmission within the sub-Saharan region, fewer studies have investigated if DM increases the likelihood of PTB patients to cluster according to the genotype of the infecting M. tuberculosis bacillus. Furthermore, there is dearth of data addressing DM-TB co-infection and it remains unclear whether patients with DM and TB are more likely than TB patients without DM to be grouped into molecular clusters defined according to the genotype of the infecting M. tuberculosis bacillus. Specifically, the question as to whether there is convincing molecular epidemiological evidence for TB transmission among DM patients or the impact of lineage in the clinical outcome of a patient on treatment is yet to be addressed.

Information derived from DM-TB co-infection will contribute to developing effective preventive strategies of TB transmission. Alternatively, those not clustered might represent patients latently infected in the remote past presenting with reactivated active TB disease or might be recently infected but with unique strains to that population [7,8,13,14]. In any case, this important information can be used for outbreak investigation or contact tracing, active case-finding or instituting other control measures to mitigate further TB transmission in at risk population (i.e., those identified as clustered or un-clustered). Here, we sought to identify the most frequently isolated TB strains among patients with TB/DM and also to determine proportion of TB/DM-TB co-infection within the sub-Saharan region. Fewer studies have investigated if DM increases the likelihood of PTB patients to cluster according to the genotype of the infecting M. tuberculosis bacillus. Furthermore, there is dearth of data addressing DM-TB co-infection and it remains unclear whether patients with DM and TB are more likely than TB patients without DM to be grouped into molecular clusters defined according to the genotype of the infecting M. tuberculosis bacillus. Specifically, the question as to whether there is convincing molecular epidemiological evidence for TB transmission among DM patients or the impact of lineage in the clinical outcome of a patient on treatment is yet to be addressed.

Table 1
Difference in various clinical outcomes among different TB lineages: C-Completed, D- Died, F – Failure, NC- Not Complete, OOC – Out of Control, TC– Treatment Complete, TO- Transfer Out. The results presented in tables using frequencies with corresponding percentages within parenthesis. Treatment failures were reported in East Asia Beijing and Indo oceanic.

| Lineage            | Outcome | C   | D   | F   | NC  | OOC | TC  | TO  | Total |
|--------------------|---------|-----|-----|-----|-----|-----|-----|-----|-------|
| East Asia Beijing  | 55 (90.2) | 0 (0.0) | 1 (1.6) | 0 (0.0) | 2 (3.3) | 2 (3.3) | 1 (1.6) | 61 (24.5) |
| Euro America       | 84 (85.7) | 0 (0.0) | 0 (0.0) | 4 (4.1) | 2 (2.0) | 6 (6.1) | 2 (2.0) | 98 (39.4) |
| Indo oceanic       | 63 (88.7) | 1 (1.4) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 3 (4.2) | 2 (2.8) | 71 (28.5) |
| M Africanum        | 1 (100.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (0.4) |
| M bovis            | 7 (100.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 7 (10.0) |
| West Africa 1      | 3 (75.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (25.0) | 0 (0.0) | 0 (0.0) | 4 (1.6) |
| West Africa 2      | 4 (57.1) | 1 (14.3) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 1 (14.3) | 1 (14.3) | 7 (2.8) |
| Total              | 217     | 2   | 2   | 4   | 6   | 12  | 6   | 249 |

Table 2
Difference in TB lineage among diabetics and non-diabetics.

| Lineage          | Diabetes condition | Total |
|------------------|--------------------|-------|
|                  | Yes                | No    |
| East Asia Beijing| 46 (75.4)          | 15 (24.6) | 61 (24.5) |
| Euro America     | 62 (63.3)          | 36 (36.7) | 98 (39.4) |
| Indo oceanic     | 54 (76.1)          | 17 (23.9) | 71 (28.5) |
| M Africanum      | 0 (0.0)            | 1 (100.0) | 1 (0.4) |
| M bovis          | 4 (57.1)           | 3 (42.9) | 7 (2.8) |
| West Africa 1    | 3 (75.0)           | 1 (25.0) | 4 (1.6) |
| West Africa 2    | 4 (57.1)           | 3 (42.9) | 7 (2.8) |
| Total            | 173                | 76    | 249 |

Table 3
Difference in TB lineages within Nairobi and Kiambu counties.

| Lineage          | Country | Total |
|------------------|---------|-------|
|                  | Kiambu  | Nairobi |
| East Asia Beijing| 13 (21.3) | 48 (78.7) | 61 (24.5) |
| Euro America     | 10 (10.2) | 88 (89.8) | 98 (39.4) |
| Indo oceanic     | 12 (16.9) | 59 (83.1) | 71 (28.5) |
| M Africanum      | 0 (0.0) | 1 (100.0) | 1 (0.4) |
| M bovis          | 1 (14.3) | 6 (85.7) | 7 (2.8) |
| West Africa 1    | 2 (50.0) | 2 (50.0) | 4 (1.6) |
| West Africa 2    | 2 (28.6) | 5 (71.4) | 7 (2.8) |
| Total            | 40      | 209   | 249 |
were con
sputum smear microscopy were registered at the TB clinic once they
Signi
Table 5
comes and hence could be attributable to active TB transmission in the
DM cases that were clustered with their corresponding clinical outcomes: hence could be attributable to active TB transmission in the two urban counties of Nairobi, Kenya.

Materials and methods

Study design

A prospective cohort study was carried out in Kiambu and Nairobi counties of, Kenya between February 2014 and August 2015. Patients aged above 15 years who tested positive for M. tuberculosis complex on sputum smear microscopy were registered at the TB clinic once they were confirmed as smear positive. The diagnosis was based on the

Table 4
Strain Characterization and the various clinical outcomes: The pr value of 0.328 is > 0.05. This indicate no relationship between the lineage and the clinical outcome.

DM = < 7 (Normal)

| Study Design | Materials and methods |

| Table 5 | Significance of the county, diabetes status and TB lineage on the clinical outcomes: There is no relationship between the county, Tb lineage and the diabetes status on the clinical outcomes of the patients. Similar information is corroborated in Fig. 2 below. |

| Variable | Level | Total (N = 340) (%) | Not done n = 58 (%) | DNA fingerprints |

| Outcomes | Favorable | 308 (90) | 50 (86) | 221 (92) | 37 (10) | 0.761 |

| Time-to-negative smears* | Baseline (positive) | 335/336 (99) | 58 (100) | 236/237 (100) | 41 (100) | 0.811 |

| Diabetes Mellitus (DM) | Non-DM | 258 (76) | 42 (72) | 183 (76) | 33 (80) | 0.998 |

| HIV test result | Positive | 77 (23) | 12 (21) | 56 (23) | 9 (22) | 0.568 |

| TB regimen | 2HRZE/4HR | 308 (91) | 52 (90) | 217 (90) | 39 (95) | 0.615 |

| County | K | 56 (17) | 7 (12) | 39 (16) | 10 (24) | 0.037 |

| Gender | Female | 96 (28) | 16 (29) | 79 (23) | 9 (22) | 0.010 |

| Ever smoked | Yes | 244 (72) | 50 (86) | 162 (67) | 32 (78) | 0.838 |

| Age (years) | (mean [SD]) | 32.1 (8.80) | 32.15 (9.13) | 32.18 (9.04) | 31.60 (8.81) | 0.979 |

| Weight (KG) | (mean [SD]) | 54.61 (10.32) | 54.61 (11.47) | 54.86 (10.34) | 53.14 (8.39) | 0.615 |

| BMI (Kg/m²) | (mean [SD]) | 19.55 (3.91) | 19.37 (3.56) | 19.71 (4.11) | 18.93 (3.23) | 0.456 |

| Blood glucose | (mean [SD]) | 3.61 (1.19) | 3.81 (1.23) | 3.59 (1.18) | 3.39 (1.20) | 0.220 |

| Hba1c (%) | (mean [SD]) | 5.75 (2.22) | 5.64 (2.42) | 5.82 (2.18) | 5.54 (2.21) | 0.699 |

| BUN | (mean [SD]) | 3.76 (1.17) | 4.1 (1.36) | 3.68 (1.08) | 3.73 (0.32) | 0.172 |

Laboratory analysis

Laboratory sample processing was done at the National Reference Tuberculosis laboratory (NTRL) Kenya. Staining was done with ZN and FM, while sputa were decontaminated with Sodium hydroxide (NaOH). Ten microliters of the processed sputa was inoculated on 2 slopes of LJ and incubated at 37 °C for 8 weeks and read weekly. Positive LJ was confirmed by use of ZN, conventional biochemical tests to discriminate MTB complex mycobacteria other than Tuberculosis (MOTT).
DNA extraction and spoligotyping analyses

Extraction of genomic DNA from TB positive culture slopes and spoligotyping were done at the South Africa Medical Research Council (SAMRC, Pretoria) as described by Kamerbeek et al. Spoligotyping was performed using the spoligotype kit (Ocimum bio solutions company, Netherlands) according to the manufacturer's instructions. The direct repeats (DR) were amplified by ABI 9700 thermocycler using oligonucleotides primers (DRa: 5′ GGT TTT GGG TCT GAC GAC 3′ and DRb: 5′ CCG AGA GGG GAC GGA AAC 3′) derived from the DR sequence. The DRa primer is biotinylated at the 5′-end. Biotinylation exploits polymorphisms in the spacer sequences found in the direct repeats (DR) locus in the chromosomes of MTBC Strains. Currently the 94 different spacer sequences were identified of which 43 are used for MTBC strain differentiation which determine the absence or presence of the 43 defined spacer sequences.

Definition of terms

A clustered case was defined as any TB case from the study populations whose strain type, based on the standard spoligotyping assay, was not indistinguishable from that of at least one other case, while a non-clustered cases shared a unique strain not found in the study population. Since the TB cases were from adjacent and proximate counties and all cases were notified within the same calendar year, the specific geographic and temporal timing of cases were not used in the definition of a cluster. This means that recent transmission was assumed to have occurred between TB cases who shared the same strain, either directly or indirectly via another identified or missed case, within the same population.

Data analysis

We defined diabetes using a cut-off of the variable HB1AC such that a person is considered to be diabetic if HB1AC > 6.5, else non-diabetic. TB treatment outcome was grouped into seven categories (C, D, F, NC, OOC, TC, and TO), diabetic conditions grouped into two groups (diabetic and non-diabetic), and two counties considered (Nairobi and Kiambu). Cross tabulation of TB lineage was done with Outcome, diabetic conditions, and counties to compare the proportion of each lineage across each of the categories of the other variables. The results presented in tables using frequencies with corresponding percentages within parenthesis.

All genotyping data were entered into a Microsoft Excel sheet. Spoligotyping molecular analysis patterns was converted into binary and octal codes designations for easier analysis and interpretation. This data was compared from previously identified strains in a database from the spolDB4 and for binary spoligotype was entered in the SITVIT2 web database from which contains more than 75,000 MTB isolates from different countries as opposed to orphans which designates patterns reported as a single isolates while the lineage of the Mycobacteria was obtained by SPOTCLUST online software (http://www.tbinsight.cs.rpi.edu). MLVA Compare V1.03 software (Genoscreen; Lille, France) was used to draw the minimum spanning trees (MSTs). The SIT or orphan spoligotype number appeared inside each node, and the distance (number of spacers of difference) between two nodes was shown on the edge linking these nodes. These phylogenetic trees were colored in function of various characteristics such as the MTBC lineages described in SITVIT and county of isolation. The MST is a graph which is undirected and connected. The MST links all isolates together with the fewest possible linkages between nearest neighbors. Furthermore, a spoligoforest was drawn as a “hierarchical layout” using the SpolTools software (available through http://www. emi.unsw.edu.au/spolTools). As opposed to the MST, the spoligoforest is a directed and not necessarily connected phylogenetic tree illustrating the parent to descendant relationships between spoligotype (considering the fact that spoligotype rather evolve by loss of spacers). TBVis tool (available at http://tbinsight.cs.rpi.edu/; [29, 30]) was used to visualize and map the spoligotype shared between different lineages and split by county of isolation. Maps were reproduced and designed...
accordance to terms described in the Creative Commons 3.0 Attribution License (http://creativecommons.org/licenses/by/3.0/).

**Results**

Newly diagnosed TB patients (374) were enrolled to participate in the study. We collected 347 isolates from the 5 health facilities in both Kiambu and Nairobi counties between August 2014 and August 2015 and analyzed in terms of sample size and within the Health facilities or tuberculosis lineages by genetic analysis using Spacers oligonucleotide typing direct repeat (spoligotype) originating from the population based on the cross sectional study. This was done through strains detected by spoligotyping and matching it with the Spoligotype43 space. The spoligotype analysis builds upon previous research that has classified MTBC strain isolates into 9 major spoligotype based families: *Mycobacterium africanum*, *M. bovis*, East African Indian (EAI), Beijing, Haarlem, Latin American and Mediterranean (LAM), Central and Middle Eastern Asia (CAS), a European family X and a default family T. These 9 families are further subdivided into 36 more subfamilies in the global database SpolDB. Groups of the related spoligotype are interchangeably called (sub) families, (sub) clades and (sub) classes. Spoligotype that have no match in the SpolDB database are defined as orphans or unclustered. Distinguishing both the clustered and the orphan clusters are more likely to be directly or indirectly involved in the same chain of Transmission and the unique genotype is difficult to get plausible origin.

The dendrogram in Fig. 1 indicated high clustering at 33% among the dominant strains with a divergent evolutionary trend among the orphan clusters. The results below in Tables 1–5 and Fig. 2 shows there was no difference in the clinical outcomes depending on TB lineage, county or diabetes status.

**Discussion**

In Countries like Kenya with high burden of TB, studies determining the population structure of strains in different geographical areas are important to monitor transmission. Information regarding MTB strains circulating in Nairobi and Kiambu counties, situated at the center of Kenya with huge influx of people across the country, is lacking. The only available data are from a previous study including few isolates thus not representative of MTB strains in Nairobi and Kiambu counties. The study documents 7 different lineages with East Asia Beijing, Euro America and Indo oceanic being the most dominant within the two counties. DM and county were not significant factors in increasing the likelihood of PTB patients to cluster according to the genotype of the infecting *M. tuberculosis* bacillus. Difference in TB lineages, DM and County of the patient were also found to be independent of the various clinical outcomes that were observed in the study.

Within the two counties, the most dominant strains were Euro America (39.4%), Indo oceanic (28.5%) and East Asia Beijing (24.5%). This indicates that the modern strains are the most circulating in the two counties and not ancestral lineages. The three clades were also the most prevalent genotypes in East Africa [16–18], The East Asia Beijing, Euro American and Indo oceanic lineages occurred in all the countries within the region, but the highest frequencies of the Indo oceanic lineage were in Kenya, depicting it as the dominant strain within the two counties. Our clustering rate of 33% was similar to that observed in the region, but the highest frequencies of the Indo oceanic lineage were in Kenya, depicting it as the dominant strain within the two counties. DM and county were not significant factors in increasing the likelihood of PTB patients to cluster according to the genotype of the infecting *M. tuberculosis* bacillus. This may indicate non-clonal transmission with diverse strains contributing to TB dynamics [25–27]. In addition there remains a need to type current isolates to get a clear understanding on the genotypic population structure of TB strains and its transmission dynamics. The heavy clustering among the three strains might also be due to recent transmission and rapid progression [21, 26, 27]. The few Unique genetic patterns observed are likely due to low reactivation of latent infection or almost non-existence recent transmission from patients out of the period or area under study [26–28]. Being that there was no relationship between DM and clustering of the cases also points to being no convincing molecular epidemiological evidence of higher TB transmission among DM patients similarly as observed in other settings [15, 18, 21]. These settings have observed no association between DM and molecular clustering of PTB in the community at a pooled estimate of 0.84 (CI 95%: 0.40 ± 1.72) [19, 21].

Our study did not indicate any relationship between the TB lineages, DM and County of the patient to the various clinical outcomes observed. Other studies show poorer outcomes among patients with TB and DM [29, 30]. Public health impact of the comorbidity is greater in regions of low and middle income, where 84% of patients with DM live, many of whom are unaware of their condition [28, 30]. Lack of association between the DM and poor clinical outcomes in our study could be attributed to high economic status of the two counties. It also indicates that the two counties have almost similar health systems and economic status.

**Limitations**

The study is limited by the fact that it is a single center study, the results of which cannot be generalized. However this provide a window of opportunity for clinicians and researchers to carry out further study and follow the patients on early screening among Newly diagnosed *M. tuberculosis* for DM as we may have underestimated the prevalence of DM and also cross transmission through contacts.

**Conclusion**

East Asia Beijing, Euro America and Indo oceanic are the most dominant strain within the two counties of Kiambu and Nairobi Counties. DM and county were not significant factors in increasing the likelihood of PTB patients to cluster according to the genotype of the infecting *M. tuberculosis* bacillus indicating no convincing molecular epidemiological evidence of higher TB transmission among DM patients similarly as observed in other settings.

**References**

[1] Pan SC, et al. Effect of diabetes on tuberculosis control in 13 countries with high tuberculosis: a modelling study. Lancet Diabetes Endocrinol 2015;3:323–30.

[2] Bailey SL, Grant P. The tubercular diabetic: The impact of diabetes mellitus on tuberculosis and its threat to global tuberculosis control. Clin Med J R Coll Physicians London 2011;11:344–7.

[3] Rodriguez-Campos S, Smith NH, Boniotti MB, Aranaz A. Overview and phylogeny of Mycobacterium tuberculosis complex organisms: Implications for diagnostics and legislation of bovine tuberculosis. Res Vet Sci 2014;97:55–19.

[4] De La Rua-Domech R. Human Mycobacterium bovis infection in the United Kingdom: Incidence, risks, control measures and review of the zoonotic aspects of bovine tuberculosis. Tuberculosis 2006;86:77–109.

[5] Mathema B, Kurepina NE, Bitari P, Kreiswirth BN, Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev 2006;19:658–85.

[6] Kremer K, et al. Discriminatory power and reproducibility of novel DNA typing methods for Mycobacterium tuberculosis complex strains. J Clin Microbiol 2005;43:5626–38.

[7] Jagielski T, et al. Current methods in the molecular typing of mycobacterium tuberculosis and other Mycobacteria. BioMed Res Int 2014.

[8] Nicol MP, Wilkinson RJ. The clinical consequences of strain diversity in Mycobacterium tuberculosis. Trans R Soc Trop Med Hyg 2008;102:955–65.

[9] Kamerbeek J, et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 1997:59:907–14.

[10] Kremer K, et al. Comparison of methods based on different molecular
epidemiological markers for typing of Mycobacterium tuberculosis complex strains: Interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.

[11] Dickman KR, et al. Detection of multiple strains of Mycobacterium tuberculosis using MIRU-VNTR in patients with pulmonary tuberculosis in Kampala, Uganda. BMC Infect Dis 2010;10:349.

[12] Malik ANJ, Godfrey-Faussett P. Effects of genetic variability of Mycobacterium tuberculosis strains on the presentation of disease 1. Lancet Infect Dis 2005;5:174–83.

[13] Schirch AC, van Soolingen D. DNA fingerprinting of Mycobacterium tuberculosis: from phage typing to whole-genome sequencing. Infect Genet Evol 2012;12:602–9.

[14] Kanduma E, McHugh TD, Gillespie SH. Molecular methods for Mycobacterium tuberculosis strain typing: a users guide. J Appl Microbiol 2003;94:781–91.

[15] Van der Zanden AGM, et al. Improvement of differentiation and interpretability of spoligotyping for Mycobacterium tuberculosis complex isolates by introduction of new spacer oligonucleotides. J Clin Microbiol 2002;40:4628–39.

[16] Ahmed N, Ehtesham NZ, Hasnain SE. Ancestral Mycobacterium tuberculosis genotypes in India: implications for TB control programmes. Infect Genet Evol 2009;9:142–6.

[17] Thomas SK, et al. Modern and ancestral genotypes of mycobacterium tuberculosis from Andhra Pradesh, India. PLoS One 2011;6.

[18] Truuer JM, Denholm JT, McBryde ES. Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific. J Theor Biol 2014;358:74–84.

[19] Barniol J, et al. Transmission dynamics of pulmonary tuberculosis between autochthonous and immigrant sub-populations. Clin Develop Immunol 2011.

[20] Magee MJ, et al. Clinical characteristics, drug resistance, and treatment outcomes among tuberculosis patients with diabetes in Peru. Int J Infect Dis 2013;17:e404–12.