

# A Molecular Epidemiological and Genetic Diversity Study of Tuberculosis in Ibadan, Nnewi and Abuja, Nigeria

Lovett Lawson<sup>1</sup>, Jian Zhang<sup>2</sup>, Michel K. Gomgnimbou<sup>2</sup>, Sadiq T. Abdurrahman<sup>3</sup>, Stéphanie Le Moullec<sup>2</sup>, Fatima Mohamed<sup>2</sup>, Gertrude N. Uzoewulu<sup>4</sup>, Olumide M. Sogaolu<sup>5</sup>, Khye Seng Goh<sup>6</sup>, Nnamdi Emenyonu<sup>1</sup>, Guislaine Refrégier<sup>2</sup>, Luis E. Cuevas<sup>7</sup>, Christophe Sola<sup>2\*</sup>

**1** Zankli Medical Centre, Abuja, Nigeria, **2** Institut de Génétique et Microbiologie UMR8621, CNRS-Université Paris-Sud, Orsay, France, **3** National Tuberculosis and Leprosy Control Programme, Abuja, Nigeria, **4** Nnamdi Azikiwe Teaching Hospital, Nnewi, Nigeria, **5** University College Hospital, Ibadan, Nigeria, **6** Tuberculosis Laboratory Consultant, Les Aymes, Guadeloupe, France, **7** Liverpool School of Tropical Medicine, Liverpool, United Kingdom

## Abstract

**Background:** Nigeria has the tenth highest burden of tuberculosis (TB) among the 22 TB high-burden countries in the world. This study describes the biodiversity and epidemiology of drug-susceptible and drug-resistant TB in Ibadan, Nnewi and Abuja, using 409 DNAs extracted from culture positive TB isolates.

**Methodology/Principal Findings:** DNAs extracted from clinical isolates of *Mycobacterium tuberculosis* complex were studied by spoligotyping and 24 VNTR typing. The Cameroon clade (CAM) was predominant followed by the *M. africanum* (West African 1) and T (mainly T2) clades. By using a smooth definition of clusters, 32 likely epi-linked clusters related to the Cameroon genotype family and 15 likely epi-linked clusters related to other “modern” genotypes were detected. Eight clusters concerned *M. africanum* West African 1. The recent transmission rate of TB was 38%. This large study shows that the recent transmission of TB in Nigeria is high, without major regional differences, with MDR-TB clusters. Improvement in the TB control programme is imperative to address the TB control problem in Nigeria.

**Citation:** Lawson L, Zhang J, Gomgnimbou MK, Abdurrahman ST, Le Moullec S, et al. (2012) A Molecular Epidemiological and Genetic Diversity Study of Tuberculosis in Ibadan, Nnewi and Abuja, Nigeria. PLoS ONE 7(6): e38409. doi:10.1371/journal.pone.0038409

**Editor:** Igor Mokrousov, St. Petersburg Pasteur Institute, Russian Federation

**Received:** February 2, 2012; **Accepted:** May 5, 2012; **Published:** June 18, 2012

**Copyright:** © 2012 Lawson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This project did not receive external grant funding. Costs for laboratory tests and JZ salary costs were provided by internal Zankli Medical Centre resources, the centre coordinating the study. Luminex Corp., Austin, Texas, provided CS with free reagents for the partial testing of specimens for the study. Luminex Corp had no role in the study design, data collection and analysis, decision to publish and a preparation of the manuscript.

**Competing Interests:** All authors, except CS, declare to have no competing interests. CS is an Academic Editor of PLoS ONE. CS also acknowledges receiving travel grants for speaking or participation at Luminex meetings. He does not have stocks or shares, is not paid, employed or consultant, does not belong to the Board membership, does not hold Patent applications (pending or actual) or benefitted from Research grants from Luminex Corp. Austin, TX. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

\* E-mail: christophe.sola@u-psud.fr

## Introduction

Multi-drug-resistant *Mycobacterium tuberculosis* (MDR-TB) has emerged as a major global public health problem [1]. WHO estimates that in 2008, between 390,000 and 510,000 persons developed MDR-TB worldwide with 69,000 cases occurring in Africa and 11,000 in Nigeria [2]. Nigeria has the tenth highest burden of TB among the 22 TB high-burden countries and an estimated TB incidence rate of 320/100000 population (WHO 2011). MDR-TB is an emerging problem in Nigeria with as much as 8% of all cultured specimens being MDR-TB in Ibadan, Nnewi and Abuja [2].

Despite the ever growing importance of TB in Nigeria, available molecular epidemiological studies do not represent an extensive picture of TB epi-links in this country due to non-standard genotyping protocols and restricted sampling areas [3–6]. This is due to molecular diagnostic methods being until now poorly adapted to high TB prevalence due to high costs or suboptimal protocols to ensure epi-links detection. Hence, the African TB molecular epidemiology is poorly described with the exception of South Africa [7,8].

Recent innovations in molecular diagnostics (e.g. Hain MTBDR<sup>plus</sup>®, MTBDR<sup>sl</sup>®, GenXpert®) and genotyping procedures such as the analysis of 24 Mycobacterial Interspersed Repetitive Units-Variable Number of Tandem Repeats (MIRU-VNTR) and high-throughput spoligotyping have made the analysis of TB transmission more efficient and the MDR-TB diagnostics easier [9,10]. Multiplexed high-throughput technologies are also emerging as powerful tools both for molecular diagnostics and public health with whole genome sequencing (WGS) holding promise for this field [11,12]. 24 MIRU-VNTR combined with spoligotyping is a new standard to replace the IS6110-Restriction Fragment Length Polymorphism (RFLP) fingerprinting method. Analyzing hundreds and even thousands of clinical isolates' DNA with limited resources has now become feasible [13,14].

Combining spoligotyping and MIRU-VNTR allows to analyze the genetic diversity and molecular epidemiology of drug susceptible and MDR-TB strains with the aim to identify the population structure of circulating clinical isolates, to estimate the recent TB transmission rate, and to eventually detect the transmission of MDR-TB cases [2].

In this manuscript, we present the characterization of the drug-susceptible and MDR-TB *M. tuberculosis* isolates previously reported by Lawson's *et al.* (2011) in Nigeria using molecular markers and we estimate the recent TB transmission rate in Nigeria [2]. We also identify the main clades of circulating genotypes.

## Materials and Methods

### Biological Samples, Clinical Isolates, Drug Susceptibility Testing, DNA Extraction

Sputum specimens were obtained through a cross-sectional study aiming at describing TB drug resistance in three cities of Nigeria. These are cities situated in three different geopolitical zones of Nigeria, as reported in Lawson *et al.*, 2011 [2]. Five hundred twenty patients were recruited (520), including 433 newly diagnosed patients with pulmonary TB (PTB) and 87 patients with PTB who had failed to respond to first-line TB treatment attending TB diagnostic centres in three different geopolitical zones of Nigeria. Patients over 18 years old with positive smear microscopy attending Directly Observed Treatment Short course (DOTS) treatment centres were enrolled prospectively from August 2009 to July 2010 at 1) Wuse, Maitama, Asokoro and Nyanya General Hospitals in Abuja, 2) the University College Hospital and five DOTS centres in Ibadan and 3) Nnamdi Azikiwe Teaching Hospital and three DOTS centres in Nnewi. Clinical features of these specimens were described in Lawson *et al.* [2]. The protocol for the study received Ethical approval from the Federal Capital Territory Health and Research Ethics Committee and Zankli Ethical Research Review Board. Written consent was obtained from all participants including permission to store samples to conduct further tests for characterising the strains affecting them.

Drug Susceptibility Tests were Performed in Liquid (Bactec<sup>®</sup>-MGIT960 System, Becton Dickinson, NJ) Media for 426 Culture-positive *Mycobacterium Tuberculosis Complex* (MTBC) of the 520 Participants (82%) with Smear-positive Sputum Samples.

DNA was extracted from isolates derived from sputum specimens cultured on BACTEC<sup>®</sup> MGIT960 by a thermolysate method and sent by express carrier on ice to the Institut de Génétique et Microbiologie UMR8621 CNRS-University Paris-Sud.

### Genotyping Methods, Data Analysis

High-throughput spoligotyping was performed on Luminex 200<sup>®</sup> (Luminex Corp., TX) as previously described [15]. Twenty-four VNTR loci analysis was done on agarose gels after simplex PCR, as previously reported [16], or using an updated in house duplex procedure (Refrégier *et al.* unpublished data). The standardized 24 VNTR loci method was used [10]. DNA was available for 423 isolates (353 from new cases and 70 from previously treated patients); complete genotyping defined as spoligotyping plus at least 20 MIRU-VNTR markers was obtained for 404 samples (97%). From these, 154 were obtained from Abuja (38%), 81 from Ibadan (20%) and 169 from Nnewi (42%). All data were entered into Excel<sup>®</sup> files and transferred to Bionumerics<sup>®</sup> (v.6.6 Applied Maths, Sint Martin Latems, Belgium). Cluster analysis was performed in Bionumerics<sup>®</sup> according to instructions manual (Figure 1, 2, 3, 4). Clade/Family assignment was done based on SpolDB4 for available Spoligo-International-Types (up to SIT1939) and based on SITVITWEB for SIT n<sup>o</sup> 2088 and 2550 [17,18]. MIRU-VNTR international type (MIT) designation was done using SITVITWEB [18]. When no SIT n<sup>o</sup> were available in SpolDB4 or SITVITWEB, the designation “new-x”

(lower cases) was given for orphan pattern and the designation “NEW-X” for new intra-study clusters. The recent transmission Index (RTI) was computed using the  $(n-1)$  method in which the number of isolates in clusters minus the number of clusters divided by the total number of isolates represents the recent tuberculosis transmission rate; in our case:  $(RTI_{n-1} = 152-51/410)$  [19]. Statistical analyses (Odds ratio, Chi-square and Student's T test and ANOVA) were performed in R version 2.13.0 ([www.r-project.org/](http://www.r-project.org/)).

## Results

### Population Description

MTBC strains were isolated from 423 patients (163 men, 253 women, and 7 with unregistered gender). Twenty-three patients were <20 years old, 289 between 20 and 40 years, 87 between 40 and 60 years and 17 >60 years old.

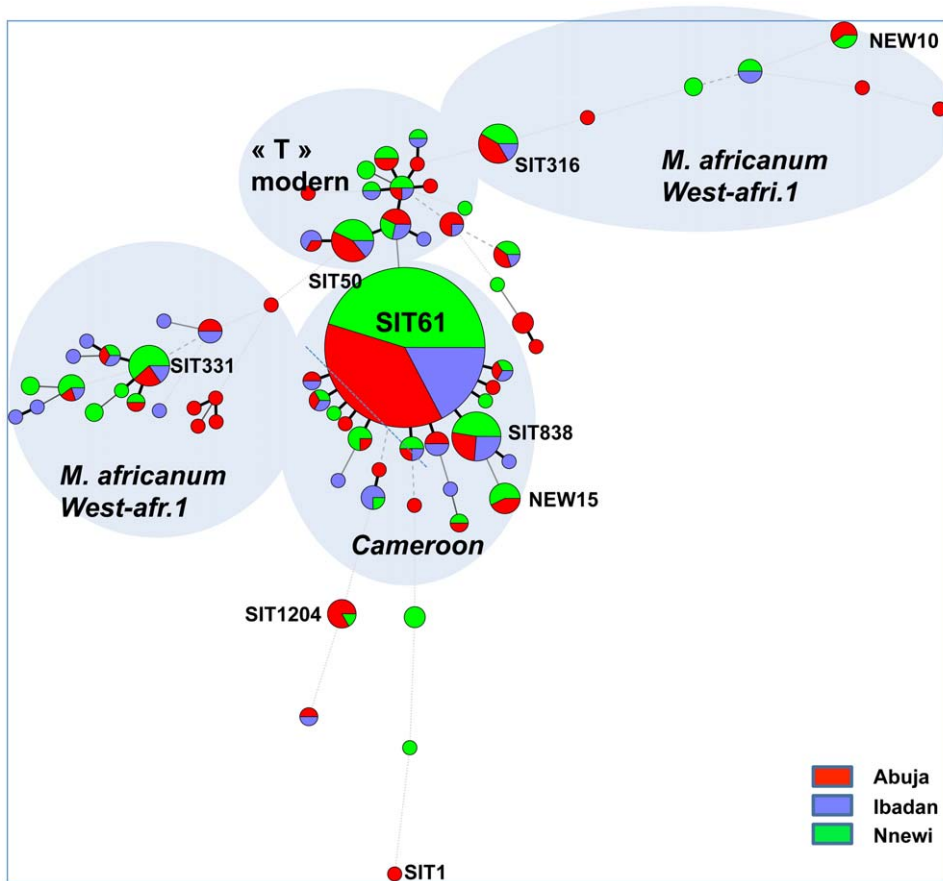
### *M. tuberculosis* Complex Genetic Diversity in Nigeria Studied by Spoligotyping

Spoligotyping results (43 spacers format) were obtained for 412 of 423 (97%) isolates (Fig 1, 2, 3, 4 and Table S1). The global population structure is made up of 4 main clusters: one from the Cameroon clade (CAM), two from the *Mycobacterium africanum* West African 1, and one from other “modern” clinical isolates (Figure 1). Seventy-one different spoligotypes were identified (Figures 2, 3, 4 and Table S1). Thirty-five of these patterns were unique and 36 occurred in clusters totaling 377 isolates, containing a range of 2 to 208 isolates. The CAM family (formerly designated as Latin American and Mediterranean-10-CAM or LAM-10-CAM) predominated, with 269 isolates (65%), of which 261 clustered and 8 were orphans). *M. africanum* West African 1 represented 53 isolates (13%) of the isolates with 42 clustered and 11 orphans. Other genotype families were: 1) “modern” strains including T2 (n = 29; 7%), Haarlem (n = 31; 7%), and Latin-American-Mediterranean (LAM) (n = 2; 0.5%), 2) Beijing (n = 1; 0.25%), 3) *Mycobacterium bovis* (n = 4; 1%) and 4) other unclassified isolates (n = 12; 3%). Most CAM family isolates (n = 208, 77%) had the SIT61; 10 other CAM-related spoligotype profiles were identified (Table 1): SIT838 found in 19 isolates (7%), further 4 SIT types containing from 1 to 4 isolates and 6 undesignated types in the SpolDB4 worldwide database [17] shared by 2 to 6 isolates and 7 orphan patterns. The four spoligotype profiles characteristic of *M. bovis* showed the absence of spacer 30 which is a specific characteristic of the “African 1” *M. bovis* genotype family previously described as predominant in Western Central Africa (Nigeria, Cameroun, Mali, Chad) [20]. By comparison with the *M. bovis* database ([www.mbovis.org](http://www.mbovis.org), see also Table 1), three isolates matched with known profiles SB0944 (SIT1037), SB1026 (new-h), SB1099 (new-d), and one isolate (new-b) had not been previously described.

53 *M. africanum* isolates were found. 52 lacked spacers 8–12 and 37–39 and thus belonged undoubtedly to *M. africanum* West African 1 (MAF1) [21]. The largest cluster within MAF1 was ST331, with 13 isolates. One isolate from Abuja (“new-s”) exhibited a partial signature of MAF1 (absence of spacers 37–39 but presence of spacers 8 and 10–12 [21]). Spoligotyping on an additional 25 spacers set (68 spacers spoligotyping) unfortunately did not much improve the clusterization of all clinical isolates (Table S1).

### 24 VNTR Results; Recent Transmission Index Assessment

Exploitable VNTR results (defined as less than 4 missing values) were obtained for 404 of 423 isolates (Table S1). Some isolates did not provide results for many loci. Among the most problematic



**Figure 1. Minimum Spanning Tree (MST) of all available spoligotypes (n=408; Nnewi n=172 green colour, Abuja n=154 red colour, Ibadan, n=82 blue colour), constructed and drawn using Bionumerics (v.6.6, Applied Maths, Sint-Martens-Latem, Belgium) and the “advanced cluster analysis” method. Some prevalent clades are designated and identified using Spoligotype-International-Types (SIT). Main Clades (Cameroon, *M. africanum* West African 1, modern “T” are also shown.**  
doi:10.1371/journal.pone.0038409.g001

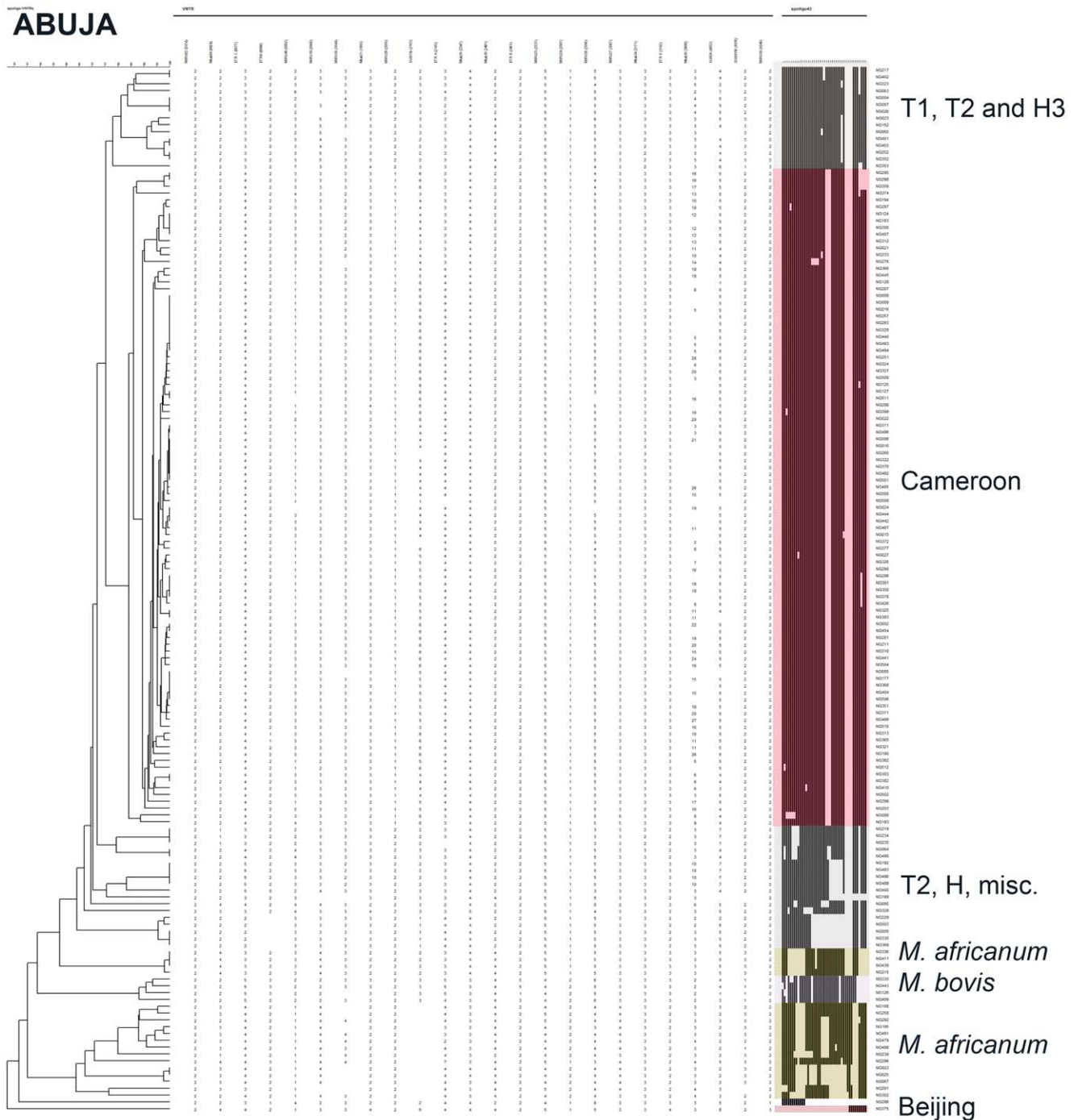
loci, Mtub39 (VNTR3690) characterisation frequently failed (n=78). The failure to amplify Mtub39 was linked to isolates belonging to the CAM family suggesting an amplification problem specific to this family (Odds Ratio = 4.8; IC<sub>95</sub> = [2.5; 10.1]). Recent results in Indian isolates have pointed to the large variability of copy number of VNTR3690 [22]. A similar observation was made for the 57 isolates in which there was a failure to amplify QuB11b (VNTR2163b), among which all *M. africanum*. None of the 53 *M. africanum* clinical isolates provided results on QuB11b whereas only 20 among the 370 other strains did not ( $\chi^2=268$ ). As this feature seems to rely on genetic properties of all *M. africanum* isolates, QuB11b marker was hypothesized to be identical (“X” or not determined “ND” in Table S1) for all isolates from this group. Finally, 45 isolates failed to amplify for MIRU16, 23 of which belong to the *M. africanum* and *M. bovis* clade (OR = 6.4; IC<sub>95</sub> = [3.3; 12.6]).

Sub-populations were tentatively identified among CAM isolates using spoligotyping and 24 VNTR. The identified clusters exhibited 20, 16, 21, 26, 15, 16 and 4 clinical isolates respectively (Table 2). A large proportion of CAM isolates could not be amplified for Mtub39 locus. The remaining isolates had a surprisingly diverse copy numbers. The values ranged from 2 to 28 copies and are noted as “N” (Table 2) (cf. also Table S1). Spoligotyping splits the CAM sub-group G-V isolates into two sub-clusters of 15 and 16 isolates. Among “Unknown” isolates according to SpolDB4, six sharing the spoligotype SIT1204

shared 24 VNTR patterns suggesting a phylogenetic link with CAM family. The SIT1204 genotype was already described in the Cross River State in the South geopolitical zone of Nigeria [4]. These strains differed only on the Mtub39 locus and harboured an intermediate copy number (n=2.5 copies) at the exact tandem repeat D (ETR-D) locus [23]. They may represent an epi-cluster.

If defining clusters by using 100% identity between isolates according to both spoligotyping and VNTR typing, and considering missing values as unique so that any pattern with a missing value cannot belong to a cluster, 134 isolates were grouped in 47 clusters containing 2 to 11 isolates. If these figures are considered a true representation of the epidemiological situation and using the (n-1) method, the recent TB transmission rate would be around 22% in Nigeria [19]. However when single-variants are included, the number of clusters doubles reaching 219 isolates grouped in 65 clusters and a Recent Transmission Index (RTI) of 38% (Table 3). The analysis using VNTR genotyping data alone did not give significantly lower discriminatory power than the composite one, i.e. adding spoligotyping information (Table 3).

Amplification for QuB11b did not work for *M. africanum* isolates and clustering analysis was thus conducted without this marker. A high polymorphism of MLVA (multi-locus VNTR analysis) was observed within *M. africanum* clinical isolates DNA with identical spoligotypes, suggesting that spoligotyping-based clustering represented common ancestors with no clear epidemiological links in most cases. Indeed, among the 49 *M. africanum* isolates for which



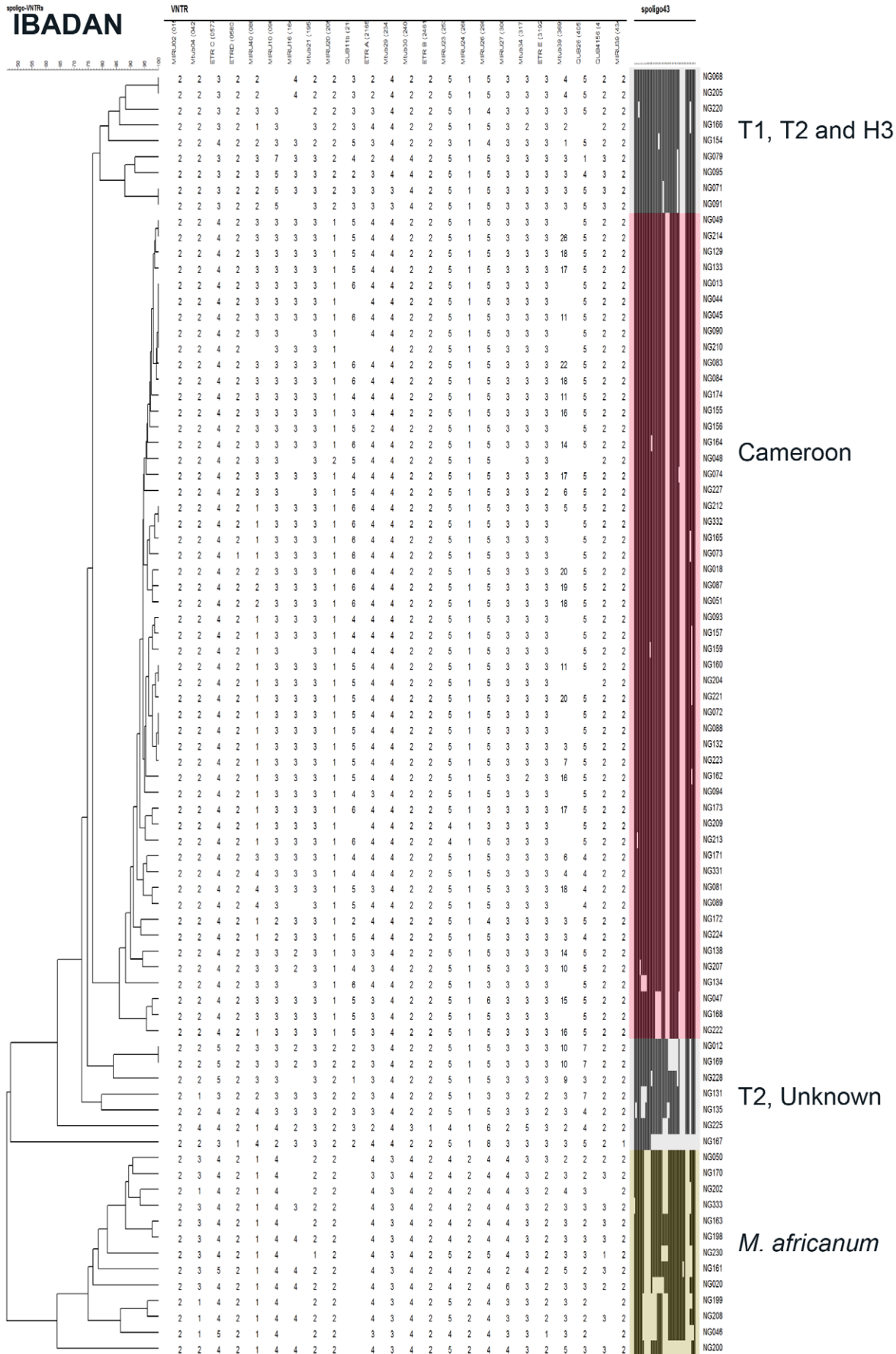
**Figure 2. Unweighted Pair Group Method using Mathematical Averages (UPGMA) dendrogram (first column) built with Bionumerics® on a composite data set (24 VNTR-largest column)-Spoligotyping (coloured column) on the clinical isolates from Abuja patients. (identification number : last column) Main Clades are also annotated right to identification number.**  
doi:10.1371/journal.pone.0038409.g002

MLVA results were available, 4 clusters only of 2 isolates were identified using a strict cluster definition (100% identity). VNTR clusters were also found in the T and H families (Table S1) with six out of eight SIT53 (T1) isolates found in one single cluster. Amongst the 41 SIT52 or derived types, also designated as Ghana family, and using 100% identity, only 17 isolates were found in 4 clusters (4 SIT2088, 4 SIT846, 5 designated as “NEW3” and 4 designated as “NEW5”). 11 of 13 SIT316 isolates (T2-variant)

were found in one cluster, whereas two other isolates differed on only one single VNTR locus, Miru26 [24]. These two isolates are likely to represent a second epidemiologically-linked cluster.

**An Evolutionary Scenario of the “Cameroon” (CAM) Clade**

The CAM clade, was first described in Cameroon and shows a typical SIT61 signature [25,26]. The MLVA analysis of the



**Figure 3. Unweighted Pair Group Method using Mathematical Averages (UPGMA) dendrogram (first column) built with Bionumerics® on a composite data set (24 VNTR-largest column)-Spoligotyping (coloured column) on the clinical isolates from Ibadan patients. (identification number : last column) Main Clades are also annotated right to identification number.**  
doi:10.1371/journal.pone.0038409.g003

CAM isolates in Nigeria, provides evolutionary and epidemiologic information and together with the 43 spacers spoligotyping, describes a global population analysis of at least 7 main clusters. The combination of values obtained on MIRU16 and MIRU40 (greyed out numbers in following text) allows the observation of three main MIRU12 international types (MIT) as described in the SITVITWEB database (see [http://www.pasteur-guadeloupe.fr:8081/SITVIT\\_ONLINE](http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE)): 223315153323, reported as MIRU-international-type 12 (MIT12), 223315153321, reported as MIT266 and 223215153323, reported as MIT264. All three major VNTR 12 types were independently reported in Nigeria in another study [4]. QuB11B provides further epidemiological information within the sub-clades (Table 2). Assuming a molecular evolution by loss of copies on MIRU40, the ancestral character of this marker would be 3 and the ancestral MIT signature would be MIT12, which would have independently evolved in MIT264 and MIT266. The larger diversity observed in Mtub39 for MIT12 and MIT266 (from 5 to 28 copies, see also Table S1) than for MIT264 (from 10 to 12 copies) reinforces this hypothesis.

#### Distribution of Multi-Drug Resistance Isolates

Twenty-nine (29 *i.e.* 7%) of 407 isolates with phenotypic Drug Susceptibility Testing (DST) in BACTEC-MGIT® (Becton Dickinson, NJ, USA) were MDR-TB isolates as previously described [2]. Among these, 23 belonged to the CAM family, of which 17 were SIT61 (data not shown). The proportion of MDR-TB within the CAM family is statistically not different from the percentage of the CAM family in the whole population (Student's test  $T = 0.1$ ;  $df = 260$ ;  $p\text{-value} = 0.9$ ). Three MDR-TB isolates belonged to T, one to LAM, one to *M. africanum* and one to U (Unknown). Nine MDR-TB isolates belonged to the subgroup G-III of the CAM family which contains 21 isolates (*i.e.* 43% MDR in this subgroup). Nine additional isolates were resistant to at least one of the drugs tested (altogether 85% of resistance). The CAM and T clades exhibited a high resistance level with respectively 53% and 54% being resistant to at least one drug. Among the 53 *M. africanum* isolates studied, one only was MDR and 17 (32%) were resistant to one of the drugs tested. The proportion of MDR was higher among the H family (28%, 9 out of 32).

#### Spatial and Phylogenetical Analysis of Diversity and Transmission

To detect if specific MTBC clusters were circulating in specific geographical areas, cluster analyses were performed independently for each collecting center (Figures 2, 3 and 4 for Abuja, Ibadan, and Nnewi, respectively). The number of clustered isolates of the 3 centers was reduced to 72 as compared to 134 in the complete study (54% when considering 100% identity, and 144 as compared to 219 (66%) when allowing for inclusion of SLVs. The prevalence of the main clades was similar in the three cities ( $p = 0.59$ ). These results confirm that Nigeria can be considered as homogeneous in the three settings investigated regarding the origin of isolates. A linear model was searched for to identify possible differences in transmission depending on the city (Abuja, Ibadan, Nnewi) or large isolate families (CAM, other modern isolates, other isolates namely *M. africanum* and *M. bovis*). The clade was found to be significantly linked to the transmission frequency as assessed by clustering, with higher

transmission for T isolates (ANOVA,  $p = 0.012$ ; effect sizes: “*M. africanum* and *bovis*” family = -0.25; “CAM” family = -0.01; other modern isolates [T] = +0.32). Indeed T isolates exhibited the lower proportion of orphans (59% as compared to 94% for *M. africanum* and *M. bovis* cluster and 87% for CAM). No significant statistical differences were detected regarding transmission in the different centers although a tendency for higher transmission in Abuja and Nnewi was detected (Table 3).

#### Discussion

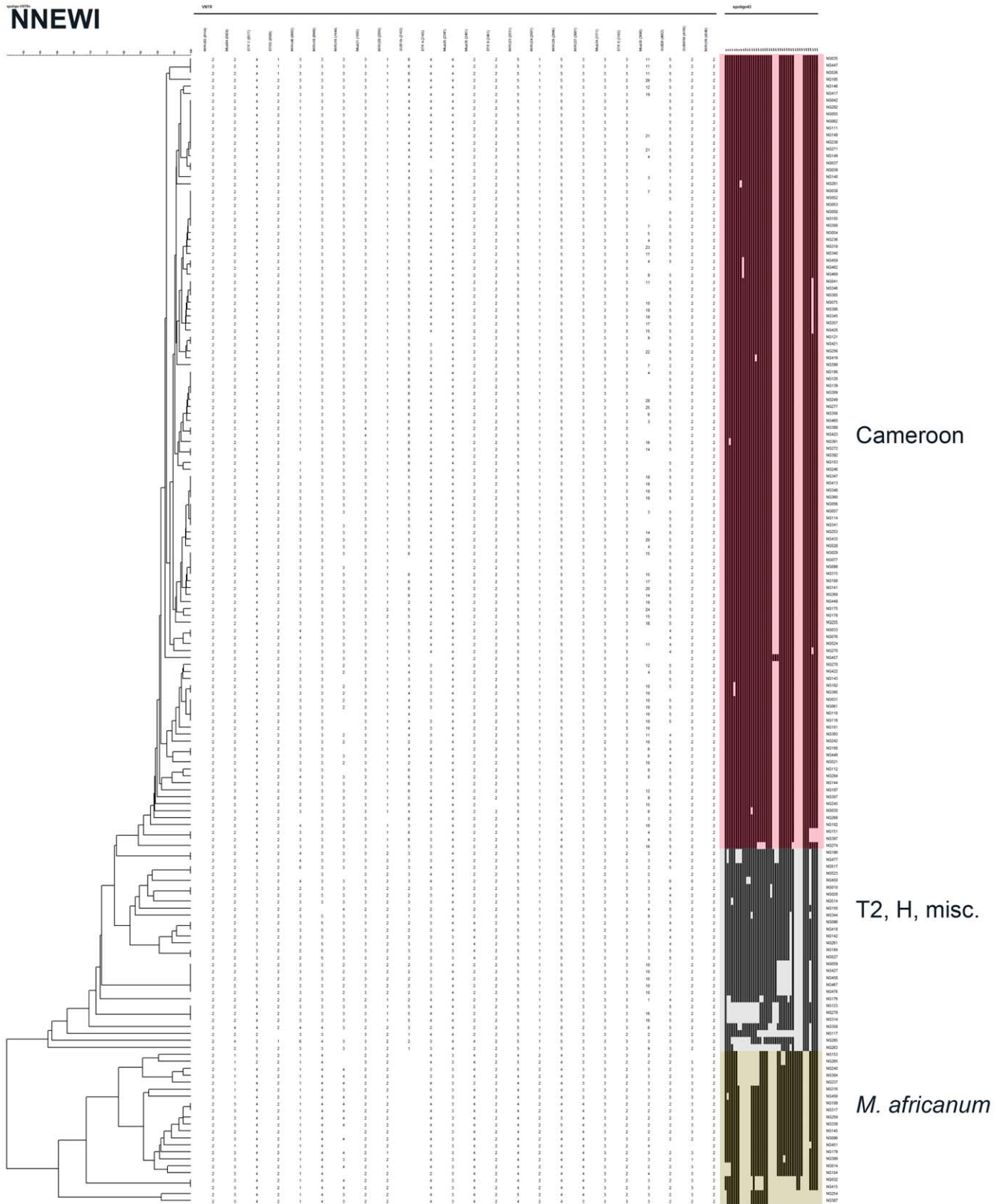
This is the largest and most detailed genetic characterisation on MTBC clinical isolates of patients suffering from TB in Nigeria relying on the analysis of isolates from three main cities [2]. The genetic diversity of MTBC was characterised by spoligotyping (43 and 68 spacers) and by 24 VNTR loci [10,27].

Spoligotyping is a genotyping method that studies the genetic diversity of the Clustered Regularly Interspersed Palindromic Repeats (CRISPR) within the MTBC (for a review on CRISPR see [28]). It enables reliable subspecies identification [17,29,30]. Its recent transfer from a membrane-based to a microbead-based format resulted in a second *youth* to this method, and a similar “CRISPOL” method has recently been developed to track outbreaks for another pathogen, *Salmonella enterica* ser. typhimurium [9,15,31].

Increasing the number of spacers to be analysed in some settings can also improve clustering and reduce the costs of systematic Spoligo+VNTR typing as recently shown in Cambodia where the number of VNTR locus to be analysed was reduced to 8 without loss of discriminatory power [32].

We have shown in this study that the recent tuberculosis transmission rate could be between 22 and 38% using either a strict (100% identity on 24 VNTR) or smooth (including SLV) definition of clusters. We detected an active transmission of TB especially in Abuja and Nnewi, although these data need to be interpreted with caution given the short (one year) recruitment period [33]. However, looking into the social network of patients found in clusters could not be done here and is a clear limitation of this study.

The CAM genotypes were the most prevalent circulating genotypes (66%). This clade was first described in Cameroon, where it represented 34% of the *M. tuberculosis* isolates in 2003 [25]. This group of strains was assumed to have emerged recently and homogeneously in the West province of Cameroon. It is characterized by the SIT61 signature (spacers 23–25 and 33–36 missing) and a homogenous 6 bands-Ligation-Mediated PCR pattern [25]. The CAM clade belongs to the principal genetic group 2 (*i.e.* *modern* strains) and is lacking the TbD1 region [26]. Several CAM spoligotype variants (SIT852, SIT808, SIT403) have been reported in Cameroon and in Nigeria [26]. The 12 MIRU-VNTR signatures differ in 4 out of 12 loci (MIRU16, MIRU26, MIRU27, MIRU40) and they have similar IS6110-RFLP patterns (10 to 15 copies) with seven common bands [26]. This group was also shown not to have an IS6110 copy in the DR locus and four IS6110 copies in open reading frames coding for adenylate cyclase, phospholipase C, *moelI*, and ATP-binding proteins [26]. In rare cases, strains with identical IS6110-RFLP patterns had spoligotypes differing by as much as 15 spacers [26].



**Figure 4. Unweighted Pair Group Method using Mathematical Averages (UPGMA) dendrogram (first column) built with Bionumerics® on a composite data set (24 VNTR-largest column)-Spoligotyping (coloured column) on the clinical isolates from Nnewi patients. (identification number : last column) Main Clades are also annotated right to identification number.**  
 doi:10.1371/journal.pone.0038409.g004





**Table 2.** Main sub-clusters according to Multi Locus VNTR and spoligotyping within the «Cameroon» (CAM) clade.

| Sub   | SIT | Nb | MIRU2 | Mtub04 | ETRC | ETRD | MIRU40 | MIRU10 | MIRU16 | Mtub21 | MIRU20 | Qub11b | ETRA | Mtub29 | Mtub30 | ETRB | MIRU23 | MIRU24 | MIRU26 | MIRU27 | Mtub34 | ETRE | Mtub39 | Qub26 | Qub4156 | MIRU39 | MIRU Int Type |
|-------|-----|----|-------|--------|------|------|--------|--------|--------|--------|--------|--------|------|--------|--------|------|--------|--------|--------|--------|--------|------|--------|-------|---------|--------|---------------|
| G-I   | 61  | 20 | 2     | 2      | 4    | 2    | 3      | 3      | 3      | 3      | 1      | 5      | 4    | 4      | 2      | 2    | 5      | 1      | 5      | 3      | 3      | 3    | N      | 5     | 2       | 2      | MIT12         |
| G-II  | 61  | 16 | 2     | 2      | 4    | 2    | 3      | 3      | 3      | 3      | 1      | 6      | 4    | 4      | 2      | 2    | 5      | 1      | 5      | 3      | 3      | 3    | N      | 5     | 2       | 2      | MIT12         |
| G-III | 61  | 21 | 2     | 2      | 4    | 2    | 1      | 3      | 3      | 3      | 1      | 4      | 4    | 4      | 2      | 2    | 5      | 1      | 5      | 3      | 3      | 3    | N      | 5     | 2       | 2      | MIT266        |
| G-IV  | 61  | 26 | 2     | 2      | 4    | 2    | 1      | 3      | 3      | 3      | 1      | 6      | 4    | 4      | 2      | 2    | 5      | 1      | 5      | 3      | 3      | 3    | N      | 5     | 2       | 2      | MIT266        |
| G-V-1 | 61  | 15 | 2     | 2      | 4    | 2    | 1      | 3      | 3      | 3      | 1      | 5      | 4    | 4      | 2      | 2    | 5      | 1      | 5      | 3      | 3      | 3    | N      | 5     | 2       | 2      | MIT266        |
| G-V-2 | 838 | 16 | 2     | 2      | 4    | 2    | 1      | 3      | 3      | 3      | 1      | 5      | 4    | 4      | 2      | 2    | 5      | 1      | 5      | 3      | 3      | 3    | N      | 5     | 2       | 2      | MIT266        |
| G-VI  | 61  | 4  | 2     | 2      | 4    | 2    | 3      | 3      | 2      | 3      | 1      | ≤4     | 3    | 4      | 2      | 2    | 5      | 1      | 5      | 3      | 3      | 3    | N      | ≤5    | 2       | 2      | MIT264        |

Sub = Subgroup name, SIT = spoligo-International type, Nb = Number of strains in that group. MIT = MIRU-International Type. (MIT266 = 36%; MIT12 = 27%; MIT264 = 7%); Mtub 39, N = high copy number and variable, MIRU16, 40 and Qub11b : poorly variable, all other markers : no variation. Alias designation of MIRU-VNTR loci (with genomic positions) : VNTR0154 = MIRU2, VNTR0424 = Mtub4, VNTR0577 = ETRC, VNTR0580 = ETRD or MIRU4, VNTR0802 = MIRU40, VNTR0960 = MIRU10, VNTR1644 = MIRU16, VNTR1955 = Mtub21, VNTR2059 = MIRU20, VNTR2163 = QUB11B, VNTR2165 = ETRA, VNTR2347 = Mtub29, VNTR2401 = Mtub30, VNTR2461 = ETRB, VNTR2531 = MIRU23, VNTR2687 = MIRU24, VNTR2996 = MIRU26, VNTR3007 = MIRU27, VNTR3171 = Mtub34, VNTR3192 = ETRE or MIRU31, VNTR3690 = Mtub39, VNTR4052 = QUB26, VNTR4156 = QUB4156, VNTR4348 = MIRU39. doi:10.1371/journal.pone.0038409.t002

**Table 3.** Clustering Results (global or by location) and Recent Transmission Index (RTI) computed using the (n–1) method and various cluster definitions.

| Method                           | Total isolates | Number of Clusters | Number of Clustered Isolates* | Number of Orphan isolates* | RTI   |
|----------------------------------|----------------|--------------------|-------------------------------|----------------------------|-------|
| composite* (100% identity)       | 404            | 47                 | 134                           | 270                        | 21.5% |
| composite* (SLV permitted)       | 404            | 65                 | 219                           | 185                        | 38.1% |
| VNTR (100% identity)             | 404            | 51                 | 152                           | 257                        | 25%   |
| Spoligotyping (100% identity)    | 404            | 36                 | 379                           | 30                         | NA**  |
| Abuja-Spoligotyping              | 154            | 15                 | 125                           | 29                         | NA**  |
| Abuja-VNTR                       | 154            | 24                 | 77                            | 77                         | 34%   |
| Abuja-composite (SLV-permitted)  | 154            | 21                 | 70                            | 84                         | 32%   |
| Nnewi-Spoligotyping              | 169            | 19                 | 153                           | 16                         | NA**  |
| Nnewi-VNTR                       | 169            | 29                 | 90                            | 79                         | 36%   |
| Nnewi-composite (SLV-permitted)  | 169            | 28                 | 83                            | 86                         | 33%   |
| Ibadan-Spoligotyping             | 81             | 15                 | 58                            | 23                         | NA**  |
| Ibadan-VNTR                      | 81             | 9                  | 25                            | 56                         | 19%   |
| Ibadan-composite (SLV-permitted) | 81             | 8                  | 21                            | 60                         | 16%   |

SLV permitted = authorizing VNTR single locus variants (SLV) or one « missing data » locus.

composite = spoligo+VNTR dataset; patterns with missing data are treated as unique patterns, \*\* Not applicable.

doi:10.1371/journal.pone.0038409.t003

3R genes are major players of molecular adaptation and evolution [40,41]. Alternatively, the main parameter responsible for the fitness of the CAM clade may be the demographic changes of the Nigerian population with a population of around 150 million in 2012 and a projected 250 million population in 2035.

Even though we did not observe strong geographical differences in the prevalence of the clades in the three cities, further analysis of data stratified by language and ethnic/tribe group and a more thorough spatial analysis may allow to better investigate bacterial genotypes-human hosts associations. Further work to characterise the phenotypic/genotyping links within *M. tuberculosis* strains circulating in Nigeria is also needed, especially on the critical issue of MDR-XDR-TB control. In this sense, new studies that will use integrated molecular methods, aimed at both MDR-TB prevention, outbreak surveillance and patient care should be implemented in Nigeria in a near future.

## References

- Wright A, Zignol M, Van Deun A, Falzon D, Gerdes SR, et al. (2009) Epidemiology of antituberculosis drug resistance 2002–07: an updated analysis of the Global Project on Anti-Tuberculosis Drug Resistance Surveillance. *Lancet* 373: 1861–1873.
- Lawson L, Yassin MA, Abdurrahman ST, Parry CM, Dacombe R, et al. (2011) Resistance to first-line tuberculosis drugs in three cities of Nigeria. *Trop Med Int Health*.
- Ani A, Bruvik T, Okoh Y, Agaba P, Agbaji O, et al. (2010) Genetic diversity of Mycobacterium tuberculosis Complex in Jos, Nigeria. *BMC Infect Dis* 10: 189.
- Thumamo BP, Asuquo AE, Abia-Bassey LN, Lawson L, Hill V, et al. (2011) Molecular epidemiology and genetic diversity of Mycobacterium tuberculosis complex in the Cross River State, Nigeria. *Infect Genet Evol*.
- Jenkins AO, Cadmus SI, Venter EH, Pourcel C, Hauk Y, et al. (2011) Molecular epidemiology of human and animal tuberculosis in Ibadan, Southwestern Nigeria. *Vet Microbiol* 151: 139–147.
- Cadmus S, Hill V, van Soolingen D, Rastogi N (2011) Spoligotype profile of Mycobacterium tuberculosis complex strains from HIV-positive and -negative patients in Nigeria: a comparative analysis. *J Clin Microbiol* 49: 220–226.
- Mlambo CK, Warren RM, Poswa X, Victor TC, Duse AG, et al. (2008) Genotypic diversity of extensively drug-resistant tuberculosis (XDR-TB) in South Africa. *Int J Tuberc Lung Dis* 12: 99–104.
- Streicher EM, Muller B, Chihota V, Mlambo C, Tait M, et al. (2011) Emergence and treatment of multidrug resistant (MDR) and extensively drug-resistant (XDR) tuberculosis in South Africa. *Infect Genet Evol*.
- Cowan LS, Diem L, Brake MC, Crawford JT (2004) Transfer of a Mycobacterium tuberculosis genotyping method, Spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. *J Clin Microbiol* 42: 474–477.
- Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, et al. (2006) Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis. *J Clin Microbiol* 44: 4498–4510.
- Schurch AC, Kremer K, Daviena O, Kiers A, Boeree MJ, et al. (2010) High resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *J Clin Microbiol* 48: 3403–3406.

## Supporting Information

**Table S1 Excel File with Full Experimental Results (Sheet 1: data), Cameroon clade results (Sheet 2: CAM) and Statistics (Sheet 3: analyses).**  
(XLSX)

## Acknowledgments

M. François Topin, Luminex BV, The Netherlands, is acknowledged for technical support. M.K.G is a PhD fellow of the IGEPE Team supported by the CNRS and the Fondation Mérieux, (Lyon, France).

## Author Contributions

Conceived and designed the experiments: GR CS MKG JZ. Performed the experiments: JZ MKG SLM FM KSG NE. Analyzed the data: JZ MKG GR CS. Contributed reagents/materials/analysis tools: LL STA GNU OMS. Wrote the paper: CS LEC LL. performed statistical analysis: GR. provided TB consultancy services: KSG.

12. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 364: 730–739.
13. Cardoso Oelemann M, Gomes HM, Willery E, Possuelo L, Batista Lima KV, et al. (2011) The forest behind the tree: phylogenetic exploration of a dominant *Mycobacterium tuberculosis* strain lineage from a high tuberculosis burden country. *PLoS One* 6: e18256.
14. Yeboah-Manu D, Asante-Poku A, Bodmer T, Stucki D, Koram K, et al. (2011) Genotypic diversity and drug susceptibility patterns among *M. tuberculosis* complex isolates from South-Western Ghana. *PLoS One* 6: e21906.
15. Zhang J, Abadia E, Refregier G, Tafaj S, Boschiroli ML, et al. (2010) *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping' with new spacers and a microbead-based hybridization assay. *J Med Microbiol* 59: 285–294.
16. Tafaj S, Zhang J, Hauck Y, Pourcel C, Hafizi H, et al. (2009) First insight into genetic diversity of the *Mycobacterium tuberculosis* complex in Albania obtained by multilocus variable-number tandem-repeat analysis and spoligotyping reveals the presence of Beijing multidrug-resistant isolates. *J Clin Microbiol* 47: 1581–1584.
17. Brudey K, Driscoll J, Rigouts L, Prodinger WM, Gori A, et al. (2006) *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, Population Genetics, and Epidemiology. *BMC Microbiol* 6: 23.
18. Demay C, Liens B, Burguière T, Hill V, Couvin D, et al. (2012) SITVITWEB – A publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* Feb. 17 [Epub ahead of print].
19. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, et al. (1994) The Epidemiology of Tuberculosis in San Francisco. *N Engl J Med* 330: 1703–1709.
20. Muller B, Hilty M, Berg S, Garcia-Pelayo MC, Dale J, et al. (2009) African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. *J Bacteriol* 191: 1951–1960.
21. Gagneux S, Small PM (2007) Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 7: 328–337.
22. Akhtar P, Singh S, Bifani P, Kaur S, Srivastava BS, et al. (2009) Variable-number tandem repeat 3690 polymorphism in Indian clinical isolates of *Mycobacterium tuberculosis* and its influence on transcription. *J Med Microbiol* 58: 798–805.
23. Frothingham R, Meeker-O'Connell WA (1998) Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* 144 (Pt 5): 1189–1196.
24. Abadia E, Zhang J, Vultros TD, Ritacco V, Kremer K, et al. (2010) Resolving lineage assignment on *Mycobacterium tuberculosis* clinical isolates classified by spoligotyping with a new high-throughput 3R SNPs based method. *Infect Genet Evol* 10: 1066–1074.
25. Niobe-Eyangoh SN, Kuaban C, Sorlin P, Cunin P, Thonnon J, et al. (2003) Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. *J Clin Microbiol* 41: 2547–2553.
26. Niobe-Eyangoh SN, Kuaban C, Sorlin P, Thonnon J, Vincent V, et al. (2004) Molecular characteristics of strains of the Cameroon family, the major group of *Mycobacterium tuberculosis* in a country with a high prevalence of tuberculosis. *J Clin Microbiol* 42: 5029–5035.
27. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 35: 907–914.
28. Sorek R, Kunin V, Hugenholtz P (2008) CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* 6: 181–186.
29. Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One* 4: e7815.
30. Kato-Maeda M, Gagneux S, Flores LL, Kim EY, Small PM, et al. (2011) Strain classification of *Mycobacterium tuberculosis*: congruence between large sequence polymorphisms and spoligotyping. *Int J Tuberc Lung Dis* 15: 131–133.
31. Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, et al. (2012) CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PLoS One* in Press.
32. Zhang J, Heng S, Le Moullec S, Refregier G, Gicquel B, et al. (2011) A first assessment of the genetic diversity of *Mycobacterium tuberculosis* complex in Cambodia. *BMC Infect Dis* 11: 42.
33. Murray M (2002) Determinants of cluster distribution in the molecular epidemiology of tuberculosis. *Proc Natl Acad Sci U S A* 99: 1538–1543.
34. Dos Vultros T, Mestre O, Rauzier J, Golec M, Rastogi N, et al. (2008) Evolution and Diversity of Clonal Bacteria: The Paradigm of *Mycobacterium tuberculosis*. *PLoS ONE* 3: e1538.
35. Diguimbaye C, Hilty M, Ngandolo R, Mahamat HH, Pfyffer GE, et al. (2006) Molecular characterization and drug resistance testing of *Mycobacterium tuberculosis* isolates from Chad. *J Clin Microbiol* 44: 1575–1577.
36. Godreuil S, Renaud F, Van de Perre P, Carriere C, Torrea G, et al. (2007) Genetic diversity and population structure of *Mycobacterium tuberculosis* in HIV-1-infected compared with uninfected individuals in Burkina Faso. *Aids* 21: 248–250.
37. Gombnimbou MK, Refregier G, Diabougba SP, Adama S, Kabore A, et al. (2012) Spoligotyping of *Mycobacterium africanum*, Burkina Faso. *Emerg Infect Dis* 18: 117–119.
38. de Jong BC, Antonio M, Awine T, Ogungbemi K, de Jong YP, et al. (2009) Use of spoligotyping and large-sequence polymorphisms to study the population structure of the *Mycobacterium tuberculosis* complex in a cohort study of consecutive smear positive tuberculosis cases in the Gambia. *J Clin Microbiol*.
39. Cadmus S, Palmer S, Okker M, Dale J, Gover K, et al. (2006) Molecular analysis of human and bovine tubercle bacilli from a local setting in Nigeria. *J Clin Microbiol* 44: 29–34.
40. Dos Vultros T, Mestre O, Tonjum T, Gicquel B (2009) DNA repair in *Mycobacterium tuberculosis* revisited. *FEMS Microbiol Rev* 33: 471–487.
41. Mestre O, Luo T, Dos Vultros T, Kremer K, Murray A, et al. (2011) Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS One* 6: e16020.