

Mapping of Mycobacterium tuberculosis Complex Genetic Diversity Profiles in Tanzania and Other African Countries

Erasto V Mbugi, Bugwesa Z. Katale, Elizabeth M. Streicher, Julius D. Keyyu, Sharon L. Kendall, Hazel M. Dockrell, Anita L. Michel, Mark M. Rweyemamu, Robin M. Warren, Mecky I. Matee, et al.

▶ **To cite this version:**

Erasto V Mbugi, Bugwesa Z. Katale, Elizabeth M. Streicher, Julius D. Keyyu, Sharon L. Kendall, et al.. Mapping of Mycobacterium tuberculosis Complex Genetic Diversity Profiles in Tanzania and Other African Countries. PLoS ONE, 2016, 11 (5), pp.e0154571. 10.1371/journal.pone.0154571 . pasteur-01652118

HAL Id: pasteur-01652118
https://riip.hal.science/pasteur-01652118

Submitted on 29 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

PLOS ONE

RESEARCH ARTICLE

Mapping of *Mycobacterium tuberculosis* Complex Genetic Diversity Profiles in Tanzania and Other African Countries

Erasto V. Mbugi[1,2]*, Bugwesa Z. Katale[2,4], Elizabeth M. Streicher[3], Julius D. Keyyu[4], Sharon L. Kendall[5], Hazel M. Dockrell[6], Anita L. Michel[7], Mark M. Rweyemamu[8], Robin M. Warren[3], Mecky I. Matee[2], Paul D. van Helden[3], David Couvin[9], Nalin Rastogi[9]

1 Department of Biochemistry, Muhimbili University of Health and Allied Sciences, P. O. Box 65001, Dar es Salaam, Tanzania, **2** Department of Microbiology and Immunology, Muhimbili University of Health and Allied Sciences, P.O. Box 65001, Dar es Salaam, Tanzania, **3** DST/NRF Centre of Excellence for Biomedical Tuberculosis Research/ South African Medical Research Council (MRC) Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Stellenbosch University, P. O. Box 241, Cape Town, 8000, South Africa, **4** Tanzania Wildlife Research Institute (TAWIRI), P.O. Box 661, Arusha, Tanzania, **5** Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, **6** The Royal Veterinary College, Royal College Street, London, NW1 0TU, United Kingdom, **7** Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa, **8** Southern African Centre for Infectious Disease Surveillance, Sokoine University of Agriculture, Morogoro, Tanzania, **9** WHO Supranational TB Reference Laboratory, Tuberculosis & Mycobacteria Unit, Institut Pasteur de la Guadeloupe, Morne Joliviere, BP 484, 97183, Abymes, Guadeloupe

* erasto.mbugi@sacids.org; rerasto@yahoo.com

CrossMark
click for updates

OPEN ACCESS

Citation: Mbugi EV, Katale BZ, Streicher EM, Keyyu JD, Kendall SL, Dockrell HM, et al. (2016) Mapping of *Mycobacterium tuberculosis* Complex Genetic Diversity Profiles in Tanzania and Other African Countries. PLoS ONE 11(5): e0154571. doi:10.1371/journal.pone.0154571

Editor: Srinand Sreevatsan, University of Minnesota, UNITED STATES

Received: January 28, 2016

Accepted: April 15, 2016

Published: May 5, 2016

Copyright: © 2016 Mbugi 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.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Wellcome Trust Grant [WT087546MA] to EVM, MMR and MIM and by MUHAS Sida Sarec Small Grant [000/3177] to EVM, MIM and BZK. The Southern African Centre for Infectious Disease Surveillance (SACIDS) provided a Postdoctoral Research Fellowship to EVM and PhD candidacy for BZK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The aim of this study was to assess and characterize *Mycobacterium tuberculosis* complex (MTBC) genotypic diversity in Tanzania, as well as in neighbouring East and other several African countries. We used spoligotyping to identify a total of 293 *M*. *tuberculosis* clinical isolates (one isolate per patient) collected in the Bunda, Dar es Salaam, Ngorongoro and Serengeti areas in Tanzania. The results were compared with results in the SITVIT2 international database of the Pasteur Institute of Guadeloupe. Genotyping and phylogeographical analyses highlighted the predominance of the CAS, T, EAI, and LAM MTBC lineages in Tanzania. The three most frequent Spoligotype International Types (SITs) were: SIT21/CAS1-Kili (n = 76; 25.94%), SIT59/LAM11-ZWE (n = 22; 7.51%), and SIT126/EAI5 tentatively reclassified as EAI3-TZA (n = 18; 6.14%). Furthermore, three SITs were newly created in this study (SIT4056/EAI5 n = 2, SIT4057/T1 n = 1, and SIT4058/EAI5 n = 1). We noted that the East-African-Indian (EAI) lineage was more predominant in Bunda, the Manu lineage was more common among strains isolated in Ngorongoro, and the Central-Asian (CAS) lineage was more predominant in Dar es Salaam (p-value<0.0001). No statistically significant differences were noted when comparing HIV status of patients vs. major lineages (p-value = 0.103). However, when grouping lineages as Principal Genetic Groups (PGG), we noticed that PGG2/3 group (Haarlem, LAM, S, T, and X) was more associated with HIV-positive patients as compared to PGG1 group (Beijing, CAS, EAI, and Manu) (p-value = 0.03). This study provided mapping of MTBC genetic diversity in Tanzania (containing

Competing Interests: The authors have declared that no competing interests exist.

information on isolates from different cities) and neighbouring East African and other several African countries highlighting differences as regards to MTBC genotypic distribution between Tanzania and other African countries. This work also allowed underlining of spoligotyping patterns tentatively grouped within the newly designated EAI3-TZA lineage (remarkable by absence of spacers 2 and 3, and represented by SIT126) which seems to be specific to Tanzania. However, further genotyping information would be needed to confirm this specificity.

Introduction

Human tuberculosis remains a global leading devastating, often severe and contagious chronic respiratory disease. People with immune-suppression are more susceptible to tuberculosis (TB). With the Global Tuberculosis Report indicating 13% of new TB cases being HIV-positive worldwide and up to 78% of TB cases among people living with HIV worldwide found in African region [1], the disease impact becomes further magnified especially in developing countries. In Tanzania various studies have been conducted using both conventional and non-conventional means to characterize the disease and its spread. Some of these studies focus on epidemiology [2–5], diagnostics [6–9], treatment and control [2, 10–12], strain molecular characterization [13–16], TB-HIV co-infections [17–20], challenges and resource limitation [21], and more recently, TB cross-species transmission at the human-animal interface [22–24]. In the recent study by our group in Serengeti ecosystem [15], a strain unassigned to neither of known spoligotypes, but resembling the CAS strain family was identified and tentatively named 'Serengeti strain'. Although various TB studies have been conducted in Tanzania so far, none of them has given detailed information on all the major phylogenetic lineages of tubercle bacilli and their distribution. In addition, none of the studies compared Tanzanian strain patterns with those prevailing in neighbouring countries and sub-regions to underline differences relating to the presence of specific lineages. All this information is important in establishing phylogenetical relatedness of the strains causing disease within and between countries as well as establishing transmission links between individuals. This study aimed to specifically describe the genotypic diversity of *Mycobacterium tuberculosis* complex (MTBC) in Tanzania, as well as neighbouring and other countries in Africa (having spoligotyping data deposited in SITVIT database), with reference to their global distribution. The study also assessed the association between HIV serological status and *M. tuberculosis* lineages for people who apart from TB had HIV. To better highlight region-specificity or similarity and get a larger view on MTBC distribution in Tanzania, *Mycobacterium tuberculosis* complex (MTBC) genotypic diversity was characterized and compared to diverse genotypes available in other African countries. Basing on genotyping information, this paper also tentatively proposes a sublineage which seems to be phylogeographically specific for Tanzania.

Materials and Methods

Study areas

The Tanzanian studies were conducted in two geographically different areas. First is the Serengeti ecosystem comprised of Bunda, Serengeti and Ngorongoro districts in northern Tanzania [15]. The second is the Dar es Salaam region, along the coast in the eastern zone (newly reported). Dar es Salaam is located at 6°48' South, 39°17' East (−6.8000, 39.2833), along the

natural harbour on the eastern coast of Africa, with sandy beaches in some areas (Population Distribution by Administrative Units, United Republic of Tanzania, 2013). Population density in Dar es Salaam is 3,133 persons per square kilometer. The increased birth and immigration rates as well as transient populations largely influence the relatively high population rate. The population densities and annual growth rates (in brackets) for Bunda, Serengeti and Ngorongoro in persons per square kilometers, are 70 (1.8%), 22.4 (3.5%) and 11.2 (3.0%), respectively.

Neighbouring countries

The main neighboring countries of Tanzania are Kenya and Uganda in the north; Rwanda, Burundi, and the Democratic Republic of the Congo in the west; and Zambia, Malawi, and Mozambique in the south (http://en.wikipedia.org/wiki/Tanzania). The description of *M. tuberculosis* strain profile used the findings from the current study in Tanzania by Mbugi et al [15] and already published data of existing patterns from other countries in the international genotyping database SITVIT2.

Preliminary Laboratory sample processing and Ethics Approval

All preliminary samples processing post field collection for Tanzanian study was performed at the Infectious Disease Centre (IDC) and Muhimbili University of Health and Allied Sciences (MUHAS) in Dar es Salaam. Spoligotyping was done at the DST/NRF Centre of Excellence for Biomedical Tuberculosis Research/South African Medical Research Council (MRC) Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Stellenbosch University, South Africa. Details, as regards the whole study (including location and study design) for the Serengeti ecosystem, and lab analysis for all samples, have been previously reported [15]. Ethical approval for this study was obtained from the Muhimbili University of Health and Allied Sciences (MUHAS) Ethics Review Committee (Ref.MU/PGS/PhD/R/Vol.1) and The Tanzania National Institute for Medical Research (Ref. No. NIMR/HQ/R.8a/Vol. IX/1299). Participants consented to enrol in the study after completing informed consent forms. When patients were found to have tuberculosis, treatment was offered as per the Tanzanian National Guidelines for management of tuberculosis. Information on other infections than tuberculosis was also collected from patients and of particular interest, was HIV serology.

Spoligotypes and new names assignment

Spoligotyping was performed using a commercially available spoligotyping kit (Isogen, Bioscience BV, Maarssen, The Netherlands) as previously described by Kamerbeek et al [25]. The resulting spoligotypes were reported in octal and binary formats (**Table 1**) and compared to existing patterns in an international genotyping database SITVIT2, which is an updated version of previously released SITVITWEB database [26] available at: http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/. Spoligotype patterns were grouped as spoligotype international types (SITs) if they shared identical spoligotype patterns with patterns present in the existing database. Spoligotypes that had no match with previously reported patterns, represented by unique exemplars, were considered as orphans. The database provided tools for queries and conversion into binary or octal values of spoligotypes, as well as geographical distribution maps for worldwide comparisons. Spoligotyping data obtained from Tanzania were compared with patterns from neighbouring countries and sub-regions in order to underline differences relating to the presence or absence of specific lineages. This includes the 'Serengeti strains' that were recently reported in Serengeti ecosystem [15].

Table 1. Distribution of lineages/sublineages of isolates from Tanzania (n = 293).

Lineage/Sublineage	Number	%
Beijing	**12**	**4,10**
CAS	**105**	**35,84**
CAS1-Delhi	19	6,48
CAS1-Kili	78	26,62
CAS2	2	0,68
CAS (or CAS_Like)	6	2,05
EAI	**49**	**16,72**
EAI1-SOM	3	1,02
EAI5	39	13,31
EAI6-BGD1	2	0,68
EAI8-MDG	2	0,68
EAI (or EAI_Like)	3	1,02
Haarlem	**7**	**2,39**
H1	2	0,68
H3	5	1,71
LAM	**43**	**14,68**
LAM6	2	0,68
LAM9	9	3,07
LAM11-ZWE	32	10,92
Manu	**10**	**3,41**
Manu_ancestor	1	0,34
Manu1	2	0,68
Manu2	7	2,39
S	**1**	**0,34**
T	**64**	**21,84**
T1	29	9,90
T2-uganda	19	6,48
T3	4	1,37
T3-ETH	11	3,75
T (or T_Like)	1	0,34
X	**1**	**0,34**
X2	1	0,34
Unknown	**1**	**0,34**

doi:10.1371/journal.pone.0154571.t001

Data analysis

Data were entered using Microsoft Excel and later transferred to relevant software for analysis. STATA version 12 and MegaStat softwares were used for this statistical analysis and a p-value <0.05 was considered significant. Standard deviation (SD) and means were calculated for age of patients. Odds ratio (OR) and 95% confidence interval (CI) were calculated for comparison between HIV information versus PGG1 and PGG2/3 lineage groups. The spoligotype patterns from Dar es Salaam and the Serengeti ecosystem study were pooled together and reanalysed for comparison with already known patterns from neighbouring countries (Kenya, Uganda, Zambia, Malawi, and Mozambique), other African countries (Sudan, Ethiopia, Nigeria, Cameroon, Namibia, South Africa, Zimbabwe, and Madagascar) and those available globally in SITVIT2 database.

Phylogenetic analysis and mapping

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 coloured in function of various characteristics such as the MTBC lineages described in SITVIT, the cities of isolation, and the HIV serological data of patients. The MST is a graph which is undirected and connected. The MST links all isolates together with the fewest possible linkages between nearest neighbours. Furthermore, a spoligoforest was drawn as a "hierarchical layout" using the SpolTools software (available through http://www.emi.unsw.edu.au/spolTools; [27, 28]). As opposed to the MST, the spoligoforest is a directed and not necessarily connected phylogenetic tree illustrating the parent to descendant relationships between spoligotypes (considering the fact that spoligotypes rather evolve by loss of spacers). TBVis tool (available at http://tbinsight.cs.rpi.edu/; [29, 30]) was used to visualize and map the spoligotypes shared between different lineages and split by city of isolation and the information of HIV serology of patients. Maps were reproduced and designed according to terms described in the Creative Commons 3.0 Attribution License (http://creativecommons.org/licenses/by/3.0/).

Results

Distribution of cities of isolation and lineages in Tanzanian study

From a total of 293 Tanzanian isolates, 93 (31.74%) were collected from Ngorongoro; 79 (26.96%) were collected from Dar es Salaam; 77 (26.28%) were collected from Bunda; and 44 (15.02%) were collected from Serengeti. Strains belonging to CAS lineage were found in highest proportion in this study (n = 105 or 35.8% of isolates), followed by ill-defined T lineage strains (n = 64 or 21.8% of isolates). The proportion of EAI lineage strains (n = 49 or 16.72% of isolates) was not negligible (**Table 1**). Among the EAI lineage, EAI5 sublineage was found in relatively higher proportion (n = 39 or 13.31% of all isolates). However, our findings suggest that a high proportion of strains belonging to the newly defined EAI3-TZA sublineage (which seems to be specific to Tanzania) were misclassified as EAI5 sublineage by SITVIT2.

HIV Status, age, and Sex ratio

HIV serology was known for 214 patients: HIV-positive patients accounted for 21 cases, and HIV-negative accounted for 193 cases. No significant statistical differences were noted when comparing lineage distribution vs. HIV serological status (p-value = 0.103), however a significant proportion of serologically HIV-positive patients was associated with the LAM and T lineage strains (n = 6/38 or 15.79% and 7/57 or 12.28%, respectively) based on known HIV status for each lineage. Furthermore, we noticed that PGG2/3 group also defined as the Euro-American family (Haarlem, LAM, S, T, and X) was more associated with HIV-seropositive patients as compared to PGG1 group (Beijing, CAS, EAI, and Manu). Among PGG2/3, 15 patients were HIV-seropositive vs. 89 HIV-seronegative; as compared to 6 HIV-seropositive patients vs. 103 HIV-seronegative patients among PGG1 group (p-value = 0.03; OR = 2.88, 95%CI [1.003–9.4544]).

The mean age of the patients was 35.1 years old, and median age was 35 years old (**Table 2**). The male/female sex ratio was 156/137 or 1.14. HIV serology distribution was not significantly different between female and male patients (p-value = 0.931); neither between the various cities of isolation (p-value = 0.921).

Table 2. Descriptive statistics on age of patients in this study.

Descriptive statistics on	*Age*
Count	293
Mean	35,11
Sample variance	101,56
Sample standard deviation	10,08
Minimum	18
Maximum	75
Range	57
1st quartile	27,00
Median	35,00
3rd quartile	41,00
Interquartile range	14,00
Mode	39,00

No significant statistical differences were noted when comparing lineage distribution vs. gender/sex of the patients (p-value = 0.653).

doi:10.1371/journal.pone.0154571.t002

Distribution of Spoligotype International Types (SITs) in this study

A total of 54/57 SITs containing 264 isolates matched a pre-existing shared-type in the SITVIT2 database, whereas 3/57 SITs (n = 4 isolates) were newly created. A total of 27/57 SITs containing 238 isolates were clustered within this study (2 to 76 isolates per cluster) while 30/57 SITs containing 30 isolates were unique (in addition 25 orphan strains were identified, which brings the number of unclustered isolates in this study to 55/293 or 18.8%, and clustered isolates to 238/293 or 81.2%). The distribution of Spoligotype International Types (SITs) in this study is shown in **Table 3**.

Clusters containing 3 or more isolates and their worldwide distribution in the SITVIT2 database

The results from Dar es Salaam and the Serengeti ecosystem studies were collectively analyzed and proportions of various lineages estimated to represent TB lineages potentially found in Tanzania. There were no major differences from what was reported in a separate Serengeti ecosystem and the combined results, which included samples from Dar es Salaam study [14]. A very high proportion of SIT21/CAS1-Kili was recorded in this pooled study (n = 76 or 25.9% of isolates) (**Table 3**). The description of SIT/lineages and worldwide distribution, as compared to the SITVIT2 database, of clusters containing 3 or more isolates is presented in Table 4. Apart from SIT21/CAS1-Kili, other SITs/lineages were observed in the order from high to low proportion as SIT59/LAM11-ZWE, SIT126/EAI5, SIT53/T1, SIT26/CAS1-Delhi, SIT1/Beijing, SIT8/EAI5, SIT345/T3-ETH, SIT42/LAM9, SIT420/T2-Uganda, SIT2196/LAM11-ZWE, SIT244/T1, SIT37/T3, SIT135/T2-Uganda, SIT149/T3-ETH, SIT25/CAS1-Delhi, SIT50/H3 and SIT128/T2-Uganda.

Phylogenetic lineages vs. demographic characteristics

The distribution of phylogenetic lineages by gender showed no significant difference (**Table 5**). Strains belonging to CAS lineage were more visible with highest proportion in Dar es Salaam followed by Ngorongoro with Bunda and Serengeti having nearly equal proportions. The proportion of strains belonging to the ill-defined T lineage was second highest, Bunda having the

Table 3. Distribution of Spoligotype International Types (SIT) in this study.

SIT*	Spoligotype pattern (Octal format)	Number in study (%)	% in study vs. Database	Lineage**	Unique or Clustered SIT***
1	000000000003771	12 (4.1)	0.11	Beijing	Clustered
4	000000007760771	1 (0.34)	0.26	Unknown	Unique
8	400037777413771	12 (4.1)	6	EAI5	Clustered
10	477777277413771	1 (0.34)	1.12	EAI8-MDG	Unique
21	703377400001771	76 (25.94)	15.97	CAS1-Kili	Clustered
22	703777400001771	1 (0.34)	1.08	CAS1-Delhi	Unique
25	703777740003171	3 (1.02)	0.45	CAS1-Delhi	Clustered
26	703777740003771	14 (4.78)	0.8	CAS1-Delhi	Clustered
34	776377777760771	1 (0.34)	0.11	S	Unique
36	777737777720771	1 (0.34)	0.63	H3	Unique
37	777737777760771	4 (1.37)	0.73	T3	Clustered
42	777777607760771	6 (2.05)	0.17	LAM9	Clustered
48	777777777413731	1 (0.34)	0.22	EAI1-SOM	Unique
50	777777777720771	3 (1.02)	0.07	H3	Clustered
53	777777777760771	15 (5.12)	0.23	T1	Clustered
59	777777606060771	22 (7.51)	4.54	LAM11-ZWE	Clustered
62	777777774020731	1 (0.34)	0.17	H1	Unique
64	777777607560771	2 (0.68)	0.48	LAM6	Clustered
73	777737777760731	1 (0.34)	0.34	T	Unique
126	477777777413771	18 (6.14)	12.08	EAI5	Clustered
128	637775777760730	3 (1.02)	8.57	T2-uganda	Clustered
135	777777777760730	4 (1.37)	11.43	T2-uganda	Clustered
137	777776777760601	1 (0.34)	0.1	X2	Unique
149	777000377760771	4 (1.37)	0.82	T3-ETH	Clustered
244	777777777760601	5 (1.71)	4.03	T1	Clustered
245	777777777760671	1 (0.34)	4.55	T1	Unique
281	777775777760771	1 (0.34)	3.13	T1	Unique
288	700377740003771	2 (0.68)	1.27	CAS2	Clustered
345	777000377760731	7 (2.39)	33.33	T3-ETH	Clustered
420	637774777760730	6 (2.05)	24	T2-uganda	Clustered
458	777777777403771	1 (0.34)	2.7	EAI5	Unique
486	703777740000371	1 (0.34)	1.75	CAS	Unique
522	777777777760770	1 (0.34)	6.25	T1	Unique
523	777777777777771	1 (0.34)	1.69	Manu_ancestor	Unique
702	700775747413771	1 (0.34)	2.86	EAI6-BGD1	Unique
727	777737774020731	1 (0.34)	2.44	H1	Unique
815	777777606060731	2 (0.68)	1.34	LAM11-ZWE	Clustered
891	777777607660771	2 (0.68)	6.67	LAM9	Clustered
1090	077777777413771	2 (0.68)	22.22	EAI5	Clustered
1192	777777677763771	1 (0.34)	10	Manu2	Unique
1369	477777777413671	1 (0.34)	12.5	EAI5	Unique
1549	775777606060731	2 (0.68)	18.18	LAM11-ZWE	Clustered
1765	703337400001771	1 (0.34)	14.29	CAS1-Kili	Unique
1801	777777777413730	1 (0.34)	25	EAI1-SOM	Unique
1824	760377777760701	2 (0.68)	25	T1	Clustered
1957	477777777013771	1 (0.34)	6.67	EAI5	Unique
2196	777775606060731	6 (2.05)	30	LAM11-ZWE	Clustered

(*Continued*)

Table 3. (*Continued*)

SIT*	Spoligotype pattern (Octal format)	Number in study (%)	% in study vs. Database	Lineage**	Unique or Clustered SIT***
2267	701377400001771	1 (0.34)	25	CAS	Unique
2269	703377400000771	1 (0.34)	20	CAS	Unique
2364	703601740003771	1 (0.34)	25	CAS1-Delhi	Unique
2391	703377400003771	1 (0.34)	11.11	CAS1-Kili	Unique
2484	737377677760771	1 (0.34)	25	T1	Unique
2736	777777777760741	1 (0.34)	25	T1	Unique
3880	437774777760730	2 (0.68)	50	T2-uganda	Clustered
4056*	477777777413411	2 (0.68)	100	EAI5	Clustered
4057*	777017777760771	1 (0.34)	50	T1	Unique
4058*	477777777413371	1 (0.34)	50	EAI5	Unique

* A total of 54/57 SITs containing 264 isolates matched a preexisting shared-type in the database, whereas 3/57 SITs (n = 4 isolates) were newly created. A total of 27/57 SITs containing 238 isolates were clustered within this study (2 to 76 isolates per cluster) while 30/57 SITs containing 30 isolates were unique (for total unique strains, one should add to this number the 25 orphan strains, which brings the number of unclustered isolates in this study to 55/293 or 18.77%, and clustered isolates to 238/293 or 81.23%). Note that SITs followed by an asterisk indicates "newly created" SITs due to 2 or more strains belonging to an identical new pattern within this study or after a match with an orphan in the database; SIT designations followed by number of strains: 4056* this study n = 2; 4057* this study n = 1, SDN n = 1; 4058* this study n = 1, NLD n = 1.

** Lineage designations according to SITVIT2; "Unknown" designates patterns with signatures that do not belong to any of the major lineages described in the database.

*** Clustered strains correspond to a similar spoligotype pattern shared by 2 or more strains "within this study"; as opposed to unique strains harboring a spoligotype pattern that does not match with another strain from this study. Unique strains matching a preexisting pattern in the SITVIT2 database are classified as SITs, whereas in case of no match, they are designated as "orphan".

doi:10.1371/journal.pone.0154571.t003

highest proportion, followed by Serengeti, Ngorongoro and Dar es Salaam. The EAI lineage was another lineage found in relatively high proportion predominantly in Bunda, Dar es Salaam and Ngorongoro, Serengeti having the least proportion. The LAM lineage strains were found in high proportions particularly in Ngorongoro and Bunda. For patients who had HIV in association with TB, the predominant strains seemed to be belonging to LAM and T lineages. Other strains were found in small proportions within specific demographic groups (**Table 5**).

Distribution map of *Mycobacterium tuberculosis* lineages in African countries

Comparison of phylogenetic lineages between our study and those isolated in neighboring countries and other countries in Africa (representing a total of 9922 isolates) revealed the predominance of the CAS family especially in the north (Sudan) and Eastern region covering Ethiopia, Kenya, Tanzania and the Madagascar (Fig 1). Small proportions of CAS lineage were also recovered in Uganda, Zambia, Malawi and Mozambique. The EAI lineage was found to be more predominant in Mozambique, Malawi, Sudan, Tanzania and Madagascar with traces of this strain lineage found in Kenya, Uganda, Zambia, Zimbabwe and Ethiopia. The predominance of LAM lineage was notable in Namibia, Zimbabwe, Zambia, Malawi, Mozambique, South Africa, Kenya, Tanzania and Madagascar. Traces of this lineage were found sporadically in Sudan, Ethiopia, Cameroon, Nigeria and Uganda. The T family spread was notable covering all areas where data on TB in Africa were available; however, the smallest proportion of this lineage was noted in Sudan (Fig 1). The Beijing strain spreads out in the Eastern part of Africa (particularly Kenya, Tanzania, and Mozambique) all the way to South Africa including

Table 4. Description of clusters containing 3 or more isolates in this study, and their worldwide distribution in the SITVIT2 database.

SIT/Lineage	Number in study	% in study	% in study vs. database	Distribution in Regions with > = 3% of a given SITs *	Distribution in countries with > = 3% of a given SITs **
21/CAS1-Kili	76	25.94	15.97	AFRI-E 65.97, EURO-N 10.08, EURO-W 6.3, ASIA-W 5.25, AMER-N 5.25, AFRI-S 4.83	TZA 34.66, ETH 9.87, MDG 9.66, USA 5.25, ZAF 4.83, ZMB 4.62, SWE 3.99, KEN 3.57
59/ LAM11-ZWE	22	7.51	4.54	AFRI-E 58.35, AFRI-S 15.26, EURO-W 14.64, AMER-N 3.3	ZMB 21.86, ZWE 15.88, ZAF 15.26, BEL 13.4, TZA 11.13, USA 3.3, MWI 3.3
126/EAI5	18	6.14	12.08	ASIA-S 37.58, ASIA-W 16.78, AFRI-E 15.44, EURO-N 13.42, EURO-W 5.37, ASIA-SE 4.03, AMER-N 4.03	IND 30.2, TZA 14.77, OMN 11.41, GBR 10.07, SAU 5.37, NLD 4.7, LKA 4.7, USA 4.03
53/T1	15	5.12	0.23	AMER-S 14.97, EURO-W 14.82, AMER-N 12.77, EURO-S 8.91, EURO-N 7.09, ASIA-W 6.92, AFRI-S 4.7, AFRI-E 4.63, ASIA-E 4.04, AFRI-N 3.33, EURO-E 3.09, CARI 3.06, AMER-C 3.06	USA 12.5, FXX 7.46, BRA 5.55, ITA 5.05, ZAF 4.6, PER 3.69, TUR 3.29, AUT 3.24
26/ CAS1-Delhi	14	4.78	0.8	ASIA-S 55.97, AMER-N 12.46, ASIA-W 12.12, AFRI-E 5.06, EURO-W 4.84, EURO-N 3.75	IND 38.91, USA 12.46, PAK 8.19, OMN 6.2, BGD 4.72, SAU 4.44, IRN 3.53
1/Beijing	12	4.1	0.11	ASIA-E 32.35, AMER-N 19.02, ASIA-SE 10.11, AFRI-S 7.85, ASIA-N 6.59, ASIA-S 5.64, EURO-N 3.46, AMER-S 3.44	CHN 19.22, USA 18.78, JPN 10.91, ZAF 7.85, RUS 6.59, VNM 3.69, IND 3.26, PER 3.0
8/EAI5	12	4.1	6	AFRI-E 21.5, ASIA-W 18.5, AMER-C 17.0, EURO-N 15.0, AMER-N 14.5, ASIA-S 4.0, AFRI-S 3.0	MEX 17.0, USA 14.5, OMN 14.5, TZA 11.5, DNK 8.0, MOZ 6.0, GBR 5.0, ZMB 4.0, IND 4.0, ZAF 3.0, SAU 3.0
345/T3-ETH	7	2.39	33.33	AFRI-E 57.14, EURO-N 33.33, AMER-N 9.52	TZA 33.33, GBR 19.05, ETH 14.29, USA 9.52, KEN 9.52, FIN 9.52, NOR 4.76
42/LAM9	6	2.05	0.17	AMER-S 31.94, AMER-N 11.17, EURO-S 10.66, EURO-W 8.92, AFRI-N 8.07, EURO-N 4.62, CARI 4.0, AFRI-E 3.52, AMER-C 3.35	BRA 12.89, USA 11.17, COL 7.18, MAR 6.64, ITA 6.16, FXX 4.76, PER 3.49, ESP 3.15, VEN 3.12
420/ T2-uganda	6	2.05	24	AFRI-E 40.0, AMER-N 32.0, EURO-W 24.0, EURO-N 4.0	USA 32.0, TZA 32.0, BEL 20.0, ZMB 4.0, UGA 4.0, SWE 4.0, DEU 4.0
2196/ LAM11-ZWE	6	2.05	30	AFRI-E 70.0, AFRI-S 20.0, EURO-W 5.0, EURO-N 5.0	TZA 35.0, MOZ 25.0, ZAF 20.0, ZMB 10.0, FIN 5.0, BEL 5.0
244/T1	5	1.71	4.03	AFRI-E 16.13, AFRI-S 15.32, EURO-S 14.52, AMER-S 14.52, ASIA-S 12.9, EURO-W 12.1, AFRI-W 6.45, AMER-N 4.03	ZAF 15.32, PRT 12.9, BRA 12.9, BGD 11.29, FXX 10.48, ZMB 6.45, GNB 5.65, TZA 4.84, USA 4.03
37/T3	4	1.37	0.73	AFRI-E 18.98, AMER-S 12.96, EURO-N 11.31, EURO-W 10.95, ASIA-W 10.04, AMER-N 8.21, ASIA-E 5.84, EURO-S 5.47, EURO-E 4.2, ASIA-S 3.47, AFRI-S 3.1	ETH 15.88, USA 7.3, CHL 5.47, SWE 4.93, SAU 4.75, CHN 4.75, FXX 4.01, ITA 3.83, BRA 3.47, ZAF 3.1, DNK 3.1
135/ T2-uganda	4	1.37	11.43	EURO-W 34.29, AMER-N 17.14, AFRI-S 17.14, AFRI-E 17.14, ASIA-W 5.71	ZAF 17.14, USA 17.14, TZA 11.43, DEU 11.43, NLD 8.57, AUT 8.57, SAU 5.71, FXX 5.71
149/T3-ETH	4	1.37	0.82	AFRI-E 58.4, EURO-N 21.52, ASIA-W 6.97, EURO-W 6.35, AMER-N 4.71	ETH 57.58, DNK 11.68, SAU 6.56, USA 4.71, SWE 4.51, NLD 3.48
25/ CAS1-Delhi	3	1.02	0.45	ASIA-W 25.26, AFRI-E 23.76, ASIA-S 17.59, AFRI-N 12.63, AMER-N 7.82, EURO-N 5.41, EURO-W 4.06	ETH 22.26, SAU 14.14, IND 13.38, SDN 10.38, USA 7.82, OMN 6.32, IRQ 4.36
50/H3	3	1.02	0.07	AMER-S 26.56, EURO-W 14.86, AMER-N 14.86, EURO-S 9.77, CARI 4.92, EURO-E 4.67, EURO-N 4.62, AFRI-N 3.6, AFRI-S 3.42, AFRI-M 3.2	USA 14.84, PER 13.54, BRA 7.1, FXX 5.82, AUT 5.15, ITA 4.6, ESP 4.6, ZAF 3.42, CMR 3.15, CZE 3.1
128/ T2-uganda	3	1.02	8.57	EURO-W 40.0, AFRI-S 25.71, AFRI-E 25.71	ZAF 25.71, TZA 17.14, FXX 14.29, DEU 11.43, ZMB 8.57, NLD 8.57, BEL 5.71

* Worldwide distribution is reported for regions with more than 3% of a given SITs as compared to their total number in the SITVIT2 database. The definition of macro-geographical regions and sub-regions (http://unstats.un.org/unsd/methods/m49/m49regin.htm) is according to the United Nations; Regions: AFRI (Africa), AMER (Americas), ASIA (Asia), EURO (Europe), and OCE (Oceania), subdivided in: E (Eastern), M (Middle), C (Central), N (Northern), S (Southern), SE (South-Eastern), and W (Western). Furthermore, CARIB (Caribbean) belongs to Americas, while Oceania is subdivided in 4 sub-regions, AUST (Australasia), MEL (Melanesia), MIC (Micronesia), and POLY (Polynesia). Note that in our classification scheme, Russia has been attributed a new sub-region by itself (Northern Asia) instead of including it among rest of the Eastern Europe. It reflects its geographical localization as well as due to the similarity of specific TB genotypes circulating in Russia (a majority of Beijing genotypes) with those prevalent in Central, Eastern and South-Eastern Asia.

** The 3 letter country codes are according to http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3; countrywide distribution is only shown for SITs with ≥3% of a given SITs as compared to their total number in the SITVIT2 database.

doi:10.1371/journal.pone.0154571.t004

Table 5. Distribution of phylogenetical lineages vs. demographic characteristics of patients and cities of isolation.

Lineage	Female	Male	Mean age [Min—Max]	SD	HIV–	HIV+	Bunda	Dar es Salaam	Ngoron-goro	Serengeti
Beijing	3	9	34.42 [19–51]	9.64	7	1	3	4	1	4
CAS	49	56	36.62 [21–75]	11.15	54	3	11	48	34	12
EAI	25	24	31.27 [19–54]	8.85	32	2	20	15	10	4
Haarlem	3	4	38.14 [30–45]	4.78	6	1	1	0	4	2
LAM	22	21	34.07 [18–53]	8.63	32	6	14	5	18	6
Manu	4	6	34.00 [22–53]	9.49	10	0	1	0	9	0
S	1	0	41.00 [41–41]	_	0	1	1	0	0	0
T	28	36	36.28 [20–63]	10.20	50	7	26	7	15	16
X	1	0	37.00 [37–37]	_	1	0	0	0	1	0
Unknown	1	0	25.00 [25–25]	_	1	0	0	0	1	0

SD = Standard deviation

doi:10.1371/journal.pone.0154571.t005

Madagascar Islands. The Haarlem family which was also detected in our study prevailed in Sudan, Ethiopia, Kenya, Madagascar, Mozambique, Zimbabwe, South Africa and Cameroon. The Manu family seemed to be more prevalent in Sudan and Uganda. The X family was also detected in studies in Tanzania with limited distribution though but it was better visible in South Africa and Kenya, and sporadically present in Mozambique, Malawi, Zambia, Zimbabwe, and Cameroon. Limited distribution of the S family was mostly found in South Africa, Madagascar, Mozambique and Kenya. Unknown lineage TB strains were found in most of the studied countries of Africa. The other lineages, CAM, TUR, AFRI and BOV which were reported in other African countries, were not detected in our Tanzanian study.

Evolutionary relationships between M. tuberculosis spoligotypes

Minimum spanning trees (MSTs) shown in Fig 2 were based on all spoligotypes of this study (n = 293 isolates). Fig 2A illustrated evolutionary relationships of spoligotypes in function of main MTBC lineages. The most visible lineages were CAS, T, EAI, and LAM. Among CAS lineage, the most predominant SITs were SIT21/CAS1-Kili (which is particularly endemic in Tanzania) and SIT26/CAS1-Delhi. Among EAI lineage, we can note the high prevalence of SIT126/EAI5 (tentatively relabeled EAI3-TZA) and SIT8/EAI5. SIT53/T1 was the better representative of the T lineage, and SIT59/LAM11-ZWE was predominant among LAM lineage. One may notice that the phylogenetic lineages were rather well organized in the MST (Fig 2A). Furthermore, the spoligoforest provided in S1 Fig allows a better visualization of main SITs and their tentative parent descendant relationships. As is for Fig 2A, Fig 2B also illustrated relationships among spoligotypes, but in function of the cities of isolation. This figure (Fig 2B) showed the distribution of spoligotypes as compared to the various cities of isolation present in this study (Bunda, Dar es Salaam, Ngorongoro, and Serengeti). Briefly, SIT21/CAS1-Kili was more common in Dar es Salaam followed by Ngorongoro; SIT26/CAS1-Delhi in Serengeti, SIT53/T1 in Bunda, SIT8/EAI5 in Dar es Salaam, SIT126/EAI3-TZA in Bunda, and SIT59/LAM11-ZWE both in Ngorongoro and Bunda (Fig 2B). Fig 2C illustrated phylogenetic relationships of spoligotypes in function of information on HIV serology of patients. We show that among the 82 distinct spoligotype patterns represented in this study, 17 patterns contained strains isolated from HIV-positive patients: i.e. SIT1/Beijing, SIT8/EAI5, SIT21/CAS1-Kili, SIT25/CAS1-Delhi, SIT34/S, SIT42/LAM9, SIT59/LAM11-ZWE, SIT135/T2-Uganda, SIT149/T3-ETH, SIT420/T2-Uganda, SIT1549/LAM11-ZWE, SIT1801/EAI1-SOM, SIT2196/LAM11-ZWE, SIT2484/T1, SIT3880/T2-uganda, including two orphan patterns Or13 and

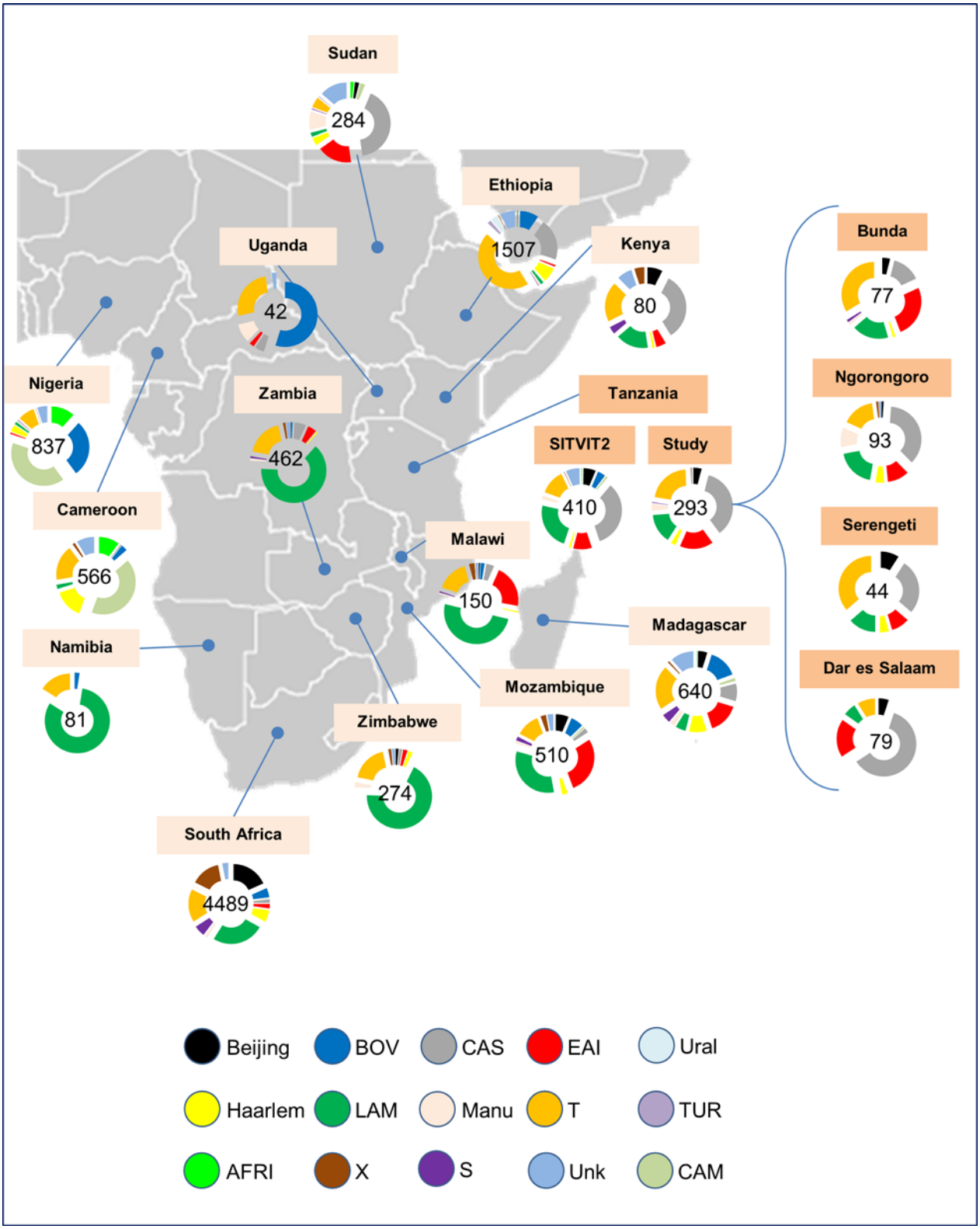


Fig 1. Distribution map of *Mycobacterium tuberculosis* lineages in several countries of Africa according to the SITVIT2 database.

doi:10.1371/journal.pone.0154571.g001

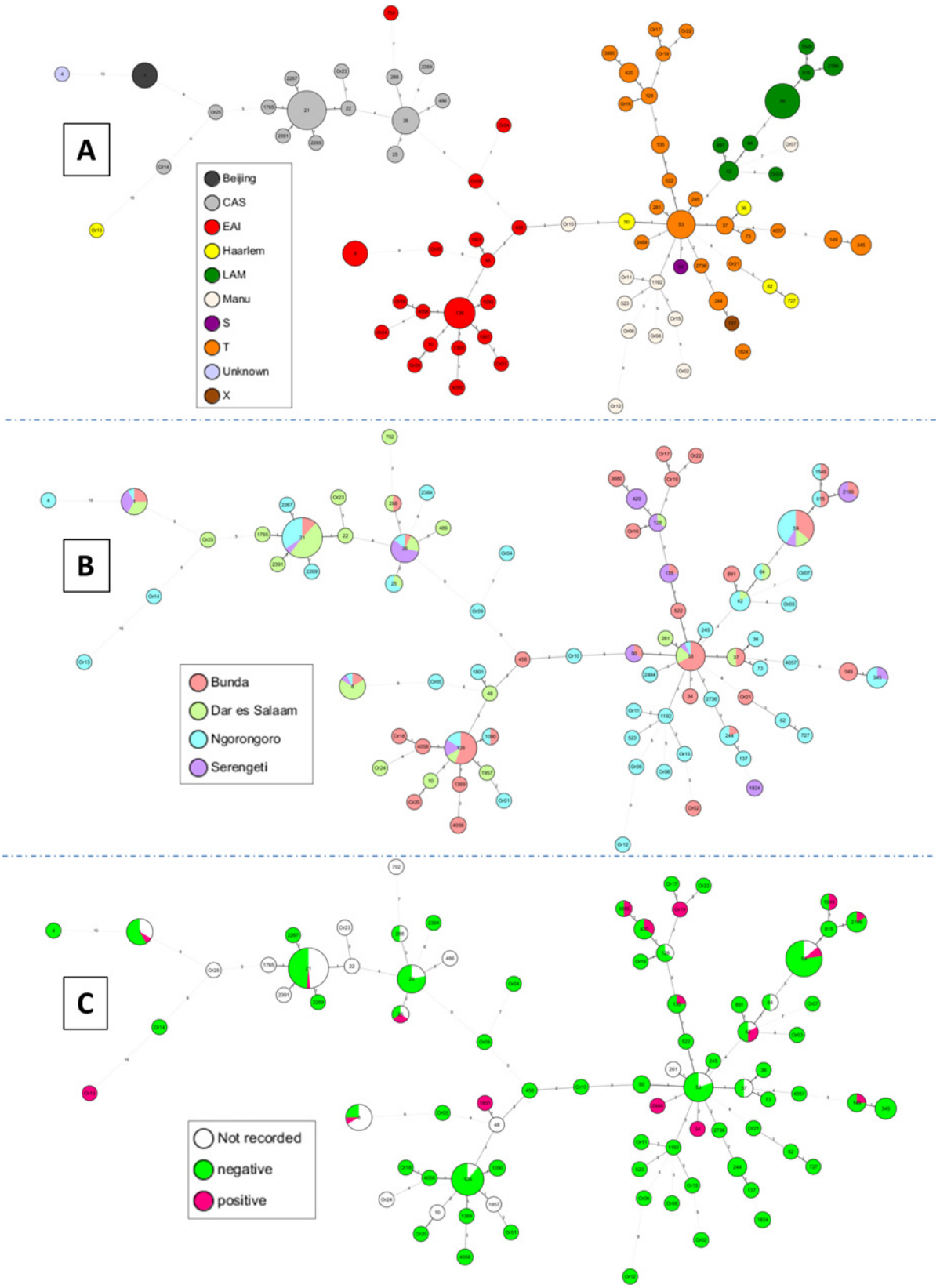

A
Beijing
CAS
EAI
Haarlem
LAM
Manu
S
T
Unknown
X
B
Bunda
Dar es Salaam
Ngorongoro
Serengeti
C
Not recorded
negative
positive

Fig 2. A minimum spanning tree (MST) illustrating evolutionary relationships between *M. tuberculosis* spoligotypes. MST constructed on all isolates including the orphan patterns (n = 293) in function of various characteristics. (**A**)MST based on phylogenetical lineages; (**B**) MST based on cities of isolation; and (**C**) MST based on HIV serology. The phylogenetic tree connects each genotype based on degree of changes required to go from one allele to another. The structure of the tree is represented by branches (continuous vs. dashed and dotted lines) and circles representing each individual pattern. Note that the length of the branches represents the distance between patterns while the complexity of the lines (continuous, gray dashed and gray dotted) denotes the number of allele/spacer changes between two patterns: solid lines, 1 or 2 or more changes (thicker ones indicate a single change, while the thinner one indicate 2 changes); gray dashed lines represent 3 changes; and gray dotted lines represent 4 or more changes. The size of the circle is proportional to the total number of isolates in our study, illustrating unique isolates (smaller nodes) versus clustered isolates (bigger nodes). The color of the circles indicates the phylogenetic lineage to which the specific pattern belongs.

doi:10.1371/journal.pone.0154571.g002

Or19. A TBVis graph representation is provided in S1 Fig, to illustrate the distribution of spoligotyping patterns and lineages as compared to city of isolation and HIV serological information of patients.

Spoligotype geographic distribution and migration routes

The analysis of spoligotype patterns from Tanzania based on the absence of specific spacers is shown in Fig 3A. Absence of spacers 2 and 3 indicated an EAI sublineage that seemed to be specific to Tanzania (pending further investigations). In total, there were 18 isolates belonging to SIT126 and 12 isolates belonging to SIT8. The peculiarity of SIT126 led to tentative re-labeling for the lineage as "EAI3-TZA" to indicate its specificity to Tanzania. This strain seemed to dominate in areas south of Asia particularly the Indian sub-continent (India and Sri-Lanka), Middle East (Oman) and Tanzania. The SIT8 lineage seems to be prevalent in Northern Europe (particularly in Denmark and Norway), Middle East (particularly in Oman), Eastern and Central Africa (particularly in Mozambique and Zambia) as well as southern part of North America particularly in the Mexican area. The geographic distribution maps/Intensity maps of SIT126 and SIT8 (by percentage in country, Fig 3B) has been established according to data recorded in the SITVIT2 database.

Discussion

General spoligotyping results

The present study reports for the first time, the results of mapping of MTBC in Tanzania and other African countries. The study combined spoligotyping data from Dar es Salaam city along the coast and the Serengeti ecosystem that comprises of Bunda, Serengeti and Ngorongoro with heterogeneous population variably involved in different activities as source of their income. Dar es Salaam and Bunda are the busiest cities due to their location that connects several other parts of Tanzania. The distribution of lineages/sublineages of isolates from Tanzania is shown in Table 1. Generally, our findings reveal similar spoligotyping results to those previously reported [13, 14, 16] with the exception that, a new EAI3 strain that seems to be specific to Tanzania was identified. We have tentatively re-labeled this sublineage as "EAI3-TZA" with reference to an international genotyping database SITVIT2, which is an updated version of previously released SITVITWEB database [26].

TB and HIV Status

HIV infection has long been associated with tuberculosis, and this co-infection exacerbates the outcome of co-diseases. As regards TB patients whose HIV serology was known (214 strains), the findings that HIV-positive accounted for 21 (9.8%) cases calls for a joint consideration in designing control strategies. Although there were no statistically significant differences when comparing individual lineages versus HIV serology of the patients ($p = 0.103$), the proportion of HIV-positive patients isolates associated with LAM lineage strains (n = 6/38 or 15.79%)

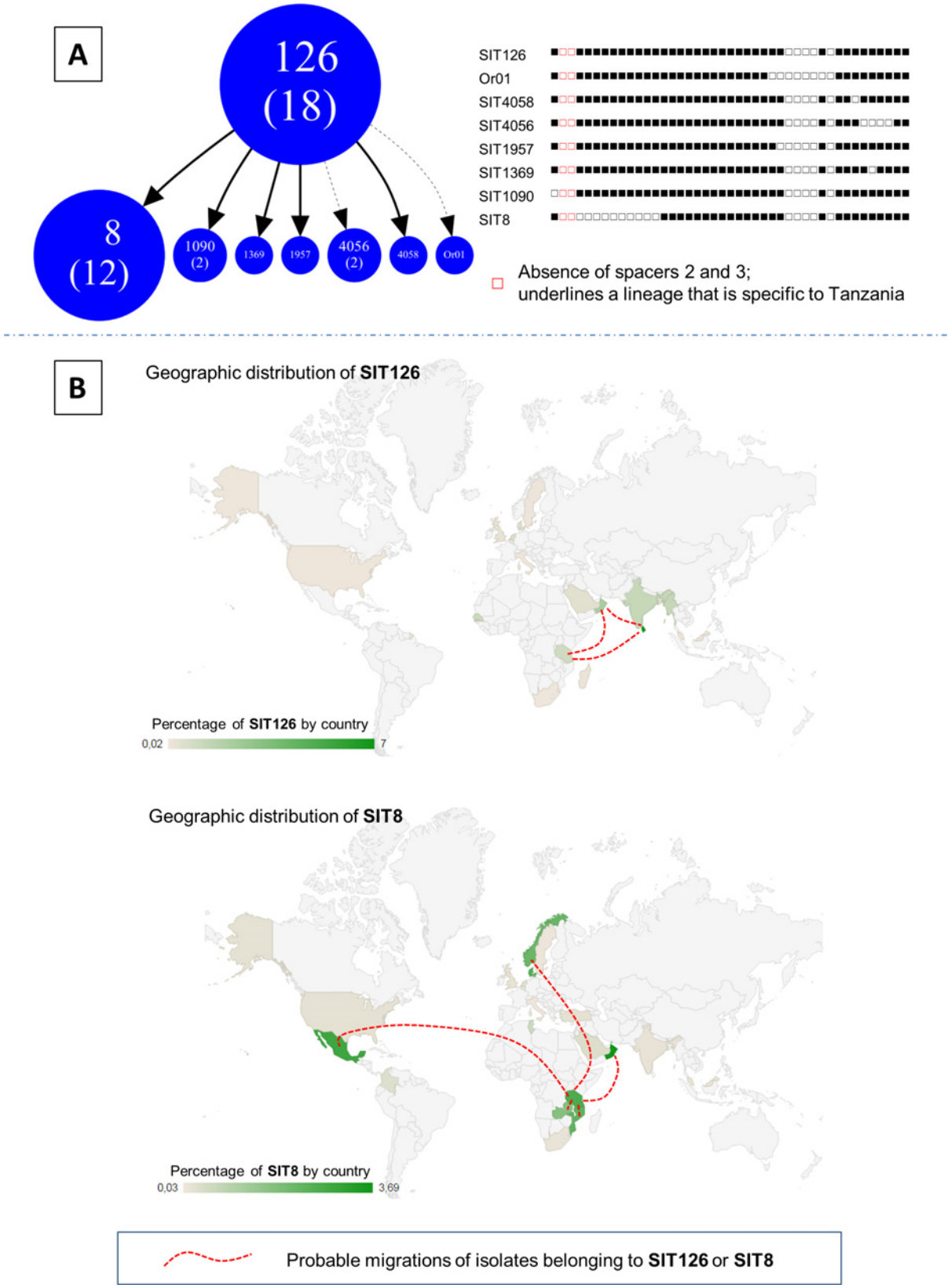

A
126
(18)
8
(12)
1090
(2)
1369
1957
4056
(2)
4058
Or01
SIT126
Or01
SIT4058
SIT4056
SIT1957
SIT1369
SIT1090
SIT8
Absence of spacers 2 and 3;
underlines a lineage that is specific to Tanzania
B
Geographic distribution of **SIT126**
Percentage of **SIT126** by country
0,02
7
Geographic distribution of **SIT8**
Percentage of **SIT8** by country
0,03
3,69
Probable migrations of isolates belonging to **SIT126** or **SIT8**

Fig 3. (**A**) Zoomed part of Spoligoforest tree showing SIT126 tentatively relabeled "EAI3-TZA" and its spoligotype descendants, along with their binary spoligotyping descriptions highlighting the specific absence of spacers 37, 38 and 40; and (**B**) Geographic distribution maps/Intensity maps of SIT126 and SIT8 (by percentage in country) according to SITVIT2 database before entering this study.

doi:10.1371/journal.pone.0154571.g003

provides clues of HIV-TB strain specificity with patients. Some reports have in principle indicated similar associations [31] while others not [14]. With the reported multi-drug resistance in some areas by the LAM lineage [32], the finding is a warning for planning TB control strategy in HIV co-infection. It is important to note that significant number of T lineage (n = 7/57 or 12.28%) was also associated with HIV-positive patients. Despite variable reports on association between TB lineages and HIV [33–37], it seems most likely that there exists specific strain association with HIV. This association might be dominant strain dependent and probably region-specific. These all might be determined by transmission chains of infection. Furthermore, when comparing modern Euro-American PGG2/3 strains and ancestral PGG1 strains versus HIV serology, we found a significant difference. PGG2/3 strains were significantly more associated with HIV-positive patients as compared to PGG1 strains (p-value = 0.03; OR = 2.88, 95%CI [1.003–9.454]). The significant association found between PGG2/3 (modern) lineages group and HIV-positive patients is an observation that should provide some clues regarding treatment of co-infected TB/HIV cases that is mostly anticipated scenario presently. It could also mean that patients infected by evolutionary recent *M*. *tuberculosis* strains have a greater chance of reactivation from latent infection to disease than those infected with ancestral strains such as *M*. *africanum*.

The descriptive statistics on population characteristics reflected HIV serology distribution to not be significantly different between female and male patients (p-value = 0.931). The distribution was too, neither significant between the various cities of isolation (p-value = 0.921). The findings could mean that exposure to infection is more relevant and critical in establishing causal-impact relationship than disease-population characteristics interactions. Previous reports [1] have indicated the age range of 15–40 years, which is the most active and productive age, to be at a higher risk of TB infection than the rest. It is worth noting TB-HIV co-infection because either of the diseases exacerbate the impact or rather progression of the other. For example, it is said that the risk of progressing from latent to active TB in people living with HIV is 12 and 20 times greater than in those without HIV infection (www.who.int/tb/publications/global_report/). Despite the reported challenges in integrating TB and HIV control services (TB Facts.org, http://www.tbfacts.org/tb-hiv.html), it is ideal to have a conscious disease control approach whose focus is on TB-HIV co-infection in endemic areas. This is particularly critical when HIV infection impacts on the epidemiology of drug resistant TB. The idea is to reduce a high risk of development to a situation where MDR TB/HIV could become a co-epidemic [38] and worsening with drug resistant HIV infection. A combined TB/HIV vaccine that could include a BCG vaccine carrying combinations of both mycobacterial and HIV antigens [39, 40] has been suggested, however integrating the knowledge in the mechanistic interactive immune response in tuberculosis-HIV co-infection remains a challenge [40]. Yet, the current BCG vaccine does not confer effective and reliable protection against the prevalent pulmonary TB form in adults in countries near the equator.

Distribution patterns of phylogenetical lineages

Interestingly, this study revealed a low prevalence (around 3%) of isolates belonging to the Manu family. Furthermore, a visible cleavage was noted for the lineage distribution in the studied cities (Fig 1). Bunda is a city located in the north-western part of the country near the Lake

Victoria, as opposed to Dar es Salaam which is located in the eastern part of Tanzania. We noted that EAI lineage was more predominant in Bunda, Manu lineage was more associated to strains isolated in Ngorongoro, and CAS lineage was more predominant in Dar es Salaam (p-value<0.001); this observation may underline an agglomeration of EAI surrounding the Lake Victoria, but further investigations will be needed to clarify this fact as reflected by phylogeographical data.

The EAI strain is a highly polymorphic lineage [41] with considerably low transmission [42] that seem to emerge over 13,000 years ago (most recent common ancestor) having a markedly different genetic structure as compared to other human tuberculosis strains [43]. The strain belongs to the lineage 1 of the Indo-Oceanic family of TB strain [44] that include EAI-5, EAI1-SOM, EAI2-Manila, EAI2-Nonthaburi, EAI3-IND, EAI4-VNM, EAI6-BGD1, EAI7-BGD2, EAI8-MDG [45] in ancient classification [46] which is evolutionary primitive [47, 48]. Supposedly and proposed to have spread to East Africa from Asian continent (Southern India), predominance of this strain around the Lake zone might be due to its closeness to the Serengeti National Park. The CAS strain predominance in Dar es Salaam could be due to the city being very close to the Indian Ocean that links our country with The Asian continent. The CAS strain is said to emerge over 9,000 years ago [43] spreading globally. Using available genetic tools, spoligotyping and MIRU-VNTR, classification, surveillance on TB transmission trends and epidemiology have been made possible [30]. The CAS strain for example has been classified as belonging to the East-African Indian TB strain, lineage 3 according to other researchers [44]. In the SITVIT database, this lineage (labeled CAS) includes the CAS1-Delhi, CAS1-Kili, CAS2 and CAS (or CAS-like) [45]. Originally, while the EAI lineage is more ancestral, predominating in the southern part of India, the CAS lineage is predominant in the north of India and more modern [49]. The Manu lineage is a newly-described ancient clade, closely related to an Indo-Oceanic lineage 1 of an ancestral EAI strain of TB [44]. The presence of the Manu lineage in Ngorongoro could be a new introduction or a descendant of the EAI strain. The two lineages are closely related and have been proposed to share a common ancestor [49] or to represent an intermediate lineage between ancestral and evolutionary modern *M. tuberculosis* [50]. The presence of Manu lineage strains in the area could also be a warning sign as it has been reported to be associated with HIV infection [33].

Tentative designation of EAI3-TZA as a new Tanzania-specific EAI sublineage

East Africa has been prone to nearly all TB strains from all over the world. Similar to other sub-Saharan Africa where HIV/AIDS is endemic, TB caseload increase 5x or more a decade in eastern and southern African countries [51]. This is partly attributed to the historical movements to and from Africa since colonial times particularly because movements were not strictly restricted. Studies done in different places in Asian continent have shown presence of confined local [52, 53] and predominant ancestral TB strains [54] whose worldwide spread might indicate their origins. It might signify that despite ancestral origins of TB strains there are micro-evolutions that are taking place within the MTBC lineages that give rise to new locally evolving strains that share specific signatures [53]. Evolution of *M. tuberculosis* strains analyzed using spoligotyping is generally through loss of some spacers. We understand that evolution is part of adaptation but it is still not known whether this evolution contributes to the 'host-pathogen compatibility' that was previously reported by Gagneux et al [44]. Despite the limited diversity in *M. tuberculosis* compared to other infectious bacteria [55], evolution in *M. tuberculosis* complex strains may result into new largely strain and lineage-specific characteristics that can influence the host interaction and pathogen virulence, consequently, determining the pathological potential and sequelae of infection [56].

Furthermore, a great proportion of SIT126/EAI5 was collected for the first time. The particular spoligotyping signature of this profile (absence of spacers 2 and 3) could indicate that it is endemic to East Africa (and particularly Tanzania). This SIT126 could also be a precursor of EAI3-IND and EAI8-MDG sublineages. We could hereby suggest a new EAI sublineage, tentatively designated as EAI3-TZA, which would be specific to Tanzania, and being mostly represented by SIT126 and its descendant SITs (Fig 3). This lineage which was initially proposed to be 'Serengeti strains' [15], seems to be not only spread in Serengeti ecosystem but also in other areas of Tanzania and possibly nearby countries. The lineage appeared to have spread in South India, Sri Lanka, and other Southeast Asian countries like Malaysia. This spread could be explained by ancestral Afro-Asian trade networks that occurred a long time ago (http://castinet.castilleja.org/users/pmckee/africaweb/swahilistates.html).

This specific family may have spread to other countries, bordering the Indian Ocean, due to historical trade reason or common history, like Oman which has historical links with Tanzania via Zanzibar (http://fr.wikipedia.org/wiki/Oman). However, the relevant proportion of SIT8 detected in Mexico could raise other questions. Several potential migration routes such as the route from East Africa, Middle East, South Asia, East Asia, North Asia, to the Americas by the Bering Strait, could explain the presence of SIT8 in Mexico. Nonetheless, more investigations are obviously needed to assert this hypothesis. Moreover, the large absence of spacers 2–13 in SIT8 does not allow us to draw significant conclusions on this profile. There might also be a parallel evolution that is taking place within *M. tuberculosis* complex providing another arm of consideration for the varying TB strain profiles with regions.

Varying distribution patterns of Mycobacterium tuberculosis lineages in African countries

Variation in predominance of phylogenetic lineages in neighboring and other countries in Africa (Fig 1) largely reflect different ancestral origins of TB strains in the continent. It may also indicate diversity in global TB strains [57]. We noticed that the CAS family was predominant in the north (Sudan) and Eastern region covering Ethiopia, Kenya, and Tanzania, and to a lesser proportion in Madagascar. This lineage is said to descent from Central and Middle Eastern Asia consisting of relatively newly defined sub-lineages CAS1 and CAS2 [58]. Similarly, the EAI lineage was mostly present in Mozambique, Malawi, Sudan, Tanzania and Madagascar, and to a lesser extent in Kenya, Uganda, Zambia, Zimbabwe and Ethiopia. These strains are originally Asian particularly, Central Asia [49] and the region covered nears the Eastern route to and from Asia (Fig 3B). On the other hand, the predominance of LAM lineage was notable in Namibia, Zimbabwe, Zambia, Malawi, Mozambique, South Africa, Kenya, Tanzania and Madagascar. This strain seemed to largely predominate in the south and western part of Africa (Fig 1) that nears the South American-African migratory routes (Fig 3B). This might indicate that these strains originally had focal points of landing in Africa and later spreading within Africa due to movements within the continent. We can suggest that while strains like LAM are spreading from the South towards north along the eastern region, strains such as CAS, EAI, and Beijing counter spread (Fig 1). The T family spread is tremendously notable in most of the areas where TB data in Africa are available. This strain might largely signal evolutional changes that are occurring in the MTBC [53].

The varying predominance of other minor TB strains in Africa could be a result of intermitted migration in and out of Africa due to continued previous and recent movements. For example, the Manu family, which was also detected and reported in our study, could be originating from the north as it seems to prevail in Sudan and Uganda. The dynamics of lineage composition may depend on different selection pressures [59]. The X and S family seemed to be in

limited distribution across African countries. As it was noted in Fig 1, unknown TB strains were detected in the majority of studied countries. These strains might be newly introduced or evolving strains. The other strains, CAM, TUR, AFRI and BOV which were reported in other African countries, were not detected in our Tanzanian study further signifying the localization and originality of various TB strains. It could be possible that, evolutionally, these various pathogens have been adapted to specific human populations [60]. The Beijing strain for instance is said to be very successful, virulent and potentially posing high risk for development of anti-TB drug resistance and multidrug resistance. This strain has an Asian ancestry with a geographical origin centered in northern China, Korea and Japan spreading out in waves [61] to other areas around the world. The worldwide spread of the Beijing lineage, raise concern on spread of drug resistance due to reported association with drug as well as multi-drug resistance [62–64]. With evolution, new strains can be expected. The newly emerging strains are usually more favoured than their ancestors making evolution a useful adaptation tool for the pathogen. New strains of EAI lineage have been reported elsewhere [65–67] indicating a continuum of evolution within this lineage. Similarly, in this study, we found a new EAI strain that does not look like other reported strains and it was tentatively renamed EAI-TZA.

In conclusion, this study provided mapping of MTBC genetic diversity in Tanzania (containing information on isolates from different cities) and several neighbouring African countries allowing us to underline potentially new spoligotyping patterns. However further genotyping information, such as MIRU-VNTRs typing, would be needed to pinpoint epidemiologically important clusters, which is the next important step in devising appropriate TB control strategy in Tanzania.

Supporting Information

S1 Fig. Spoligoforest tree drawn using the SpolTools software (available through http://www.emi.unsw.edu.au/spolTools; Reyes et al. [27], Tang et al. [26]), and shown as a Hierarchical Layout. The Figure was drawn on all patterns including orphan patterns (n = 293). Each spoligotype pattern from the study is represented by a node with area size being proportional to the total number of isolates with that specific pattern. Changes (loss of spacers) are represented by directed edges between nodes, with the arrowheads pointing to descendant spoligotypes. In this representation, the heuristic used selects a single inbound edge with a maximum weight using a Zipf model. Solid black lines link patterns that are very similar, i.e., loss of one spacer only (maximum weight being 1.0), while dashed lines represent links of weight comprised between 0.5 and 1, and dotted lines a weight less than 0.5.
(PDF)

S1 Table. Detailed demographic, epidemiologic and genotyping information on Tanzanian M. tuberculosis isolates. Note that all strains were pansusceptible, and were isolated from newly diagnosed, sputum smear/culture positive pulmonary TB patients. NEW SITs are followed by an asterisk (*) and highlighted in yellow. Orphan spoligotypes are highlighted in blue.
(PDF)

Acknowledgments

The Southern African Centre for Infectious Disease Surveillance (SACIDS) is acknowledged for providing a Postdoctoral Research Fellowship to EVM and PhD candidacy for BZK. Financial support was received from the Wellcome Trust Grant [WT087546MA]; MUHAS Sida Sarec Small Grant [000/3177]. Prof Mark Rweyemamu, The Executive Director of SACIDS is acknowledged for stimulating the idea and advice. The London International Development Centre in the

UK and staff are acknowledged for collaboration. We thank the participants' authorities in Tanzania for allowing us to conduct our study in Dar es Salaam and The Srengeti ecosystem. The WHO Supranational TB Reference Laboratory,Tuberculosis & Mycobacteria Unit, Institut Pasteur de la Guadeloupe is acknowledged for stimulation of ideas to mine up data from TB database for evaluation of TB strains in Tanzania as related to other African Countries.

Author Contributions

Conceived and designed the experiments: EVM EMS BZK DC NR. Performed the experiments: EVM BZK DC. Analyzed the data: EVM DC NR. Contributed reagents/materials/analysis tools: EMS RMW DC NR PDvH MIM. Wrote the paper: EVM BZK EMS JDK SLK HMD ALM MMR RMW MIM PDvH DC NR. Did data mining and established links: DC.

References

1. WHO Global Tuberculosis Report: World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland; 2014.
2. Chonde TM, Basra D, Mfinanga SGM, Range N, Lwilla F, Shirima RP, et al. National anti-tuberculosis drug resistance study in Tanzania. The International Journal of Tuberculosis and Lung Disease. 2010; 14(8):967–72. PMID: 20626940
3. Chonde TM, Doulla B, van Leth F, Mfinanga SGM, Range N, Lwilla F, et al. Implementation of a national anti-tuberculosis drug resistance survey in Tanzania. BMC Public Health. 2008; 8:427–. doi: 10.1186/1471-2458-8-427 PMID: 19116022
4. Mfinanga GS, Kimaro GD, Ngadaya E, Massawe S, Mtandu R, Shayo EH, et al. Health facility-based Active Management of the Third Stage of Labor: findings from a national survey in Tanzania. Health Research Policy and Systems. 2009; 7:6–. doi: 10.1186/1478-4505-7-6 PMID: 19371418
5. Ngadaya ES, Mfinanga GS, Wandwalo ER, Morkve O. Pulmonary tuberculosis among women with cough attending clinics for family planning and maternal and child health in Dar Es Salaam, Tanzania. BMC Public Health. 2009; 9:278–. doi: 10.1186/1471-2458-9-278 PMID: 19646288
6. Holm LL, Rose MV, Kimaro G, Bygbjerg IC, Mfinanga SG, Ravn P, et al. A Comparison of Interferon- γ and IP-10 for the Diagnosis of Tuberculosis. Pediatrics. 2014 December 1, 2014; 134(6):e1568–e75. doi: 10.1542/peds.2014-1570 PMID: 25422019
7. Honeyborne I, Mtafya B, Phillips PPJ, Hoelscher M, Ntinginya EN, Kohlenberg A, et al. The Molecular Bacterial Load Assay Replaces Solid Culture for Measuring Early Bactericidal Response to Antituberculosis Treatment. Journal of Clinical Microbiology. 2014 August 1, 2014; 52(8):3064–7. doi: 10.1128/JCM.01128-14 PMID: 24871215
8. Mfinanga GS, Ngadaya E, Mtandu R, Mutayoba B, Basra D, Kimaro G, et al. The quality of sputum smear microscopy diagnosis of pulmonary tuberculosis in Dar es Salaam, Tanzania. Tanzan Health Res Bull. 2007; 9(3):164–8. PMID: 18087893
9. Reddy EA, Njau BN, Morpeth SC, Lancaster KE, Tribble AC, Maro VP, et al. A randomized controlled trial of standard versus intensified tuberculosis diagnostics on treatment decisions by physicians in Northern Tanzania. BMC Infectious Diseases. 2014; 14:89–. doi: 10.1186/1471-2334-14-89 PMID: 24552306
10. Matee M, Mfinanga S, Holm-Hansen C. Anti-TB drug resistance levels and patterns among Mycobacterium tuberculosis isolated from newly diagnosed cases of pulmonary tuberculosis in Dar es Salaam, Tanzania. APMIS. 2009; 117(4):263–7. doi: 10.1111/j.1600-0463.2008.02429.x PMID: 19338514
11. Mpagama SG, Ndusilo N, Stroup S, Kumburu H, Peloquin CA, Gratz J, et al. Plasma Drug Activity in Patients on Treatment for Multidrug-Resistant Tuberculosis. Antimicrobial Agents and Chemotherapy. 2014 February 1, 2014; 58(2):782–8. doi: 10.1128/AAC.01549-13 PMID: 24247125
12. Pholwat S, Stroup S, Gratz J, Trangan V, Foongladda S, Kumburu H, et al. Pyrazinamide susceptibility testing of Mycobacterium tuberculosis by high resolution melt analysis. Tuberculosis. 2014 2015/01/26; 94(1):20–5. doi: 10.1016/j.tube.2013.10.006 PMID: 24246607
13. Eldholm V, Matee M, Mfinanga SGM, Heun M, Dahle UR. A first insight into the genetic diversity of Mycobacterium tuberculosis in Dar es Salaam, Tanzania, assessed by spoligotyping. BMC Microbiology. 2006; 6:76–. PMID: 16970826
14. Kibiki G, Mulder B, Dolmans W, de Beer J, Boeree M, Sam N, et al. M. tuberculosis genotypic diversity and drug susceptibility pattern in HIV- infected and non-HIV-infected patients in northern Tanzania. BMC microbiology. 2007; 7(1):51.

15. Mbugi EV, Katale BZ, Siame KK, Keyyu JD, Kendall SL, Dockrell HM, et al. Genetic diversity of Mycobacterium tuberculosis isolated from tuberculosis patients in the Serengeti ecosystem in Tanzania. Tuberculosis. 2015; 95(2):170–8. doi: 10.1016/j.tube.2014.11.006 PMID: 25522841

16. Mfinanga SGM, Warren RM, Kazwala R, Kahwa A, Kazimoto T, Kimaro G, et al. Genetic Profile of M. Tuberculosis in relation to Drug Resistance and Treatment Outcomes in human pulmonary tuberculosis in Tanzania. Tanzania J Health Res. 2014; 16(2):13 pages.

17. Kabali C, Mtei L, Brooks DR, Waddell R, Bakari M, Matee M, et al. Increased mortality associated with treated active tuberculosis in HIV-infected adults in Tanzania. Tuberculosis. 2013 2015/01/26; 93 (4):461–6. doi: 10.1016/j.tube.2013.01.001 PMID: 23523641

18. Lahey T, Czechura T, Crabtree S, Arbeit RD, Matee M, Horsburgh CR, et al. Greater Preexisting Interferon-γ Responses to Mycobacterial Antigens and Lower Bacillary Load During HIV-Associated Tuberculosis. Journal of Infectious Diseases. 2013 November 15, 2013; 208(10):1629–33.

19. Lahey T, MacKenzie T, Arbeit RD, Bakari M, Mtei L, Matee M, et al. Recurrent Tuberculosis Risk Among HIV-Infected Adults in Tanzania With Prior Active Tuberculosis. Clinical Infectious Diseases. 2013 January 1, 2013; 56(1):151–8. doi: 10.1093/cid/cis798 PMID: 22972862

20. Range N, Ipuge Y A, Obrien R J, Egwaga S M, Mfinanga S G, Chonde T M, et al. Trend in HIV prevalence among tuberculosis patients in Tanzania, 1991–1998. The International Journal of Tuberculosis and Lung Disease. 2001; 5(5):405–12. PMID: 11336270

21. Semvua HH, Kibiki GS, Kisanga ER, Boeree MJ, Burger DM, Aarnoutse R. Pharmacological interactions between rifampicin and antiretroviral drugs: challenges and research priorities for resource-limited settings. Therapeutic Drug Monitoring. 2015; 37(1):22–32. doi: 10.1097/FTD.0000000000000108 PMID: 24943062

22. Cleaveland S, Shaw DJ, Mfinanga SG, Shirima G, Kazwala RR, Eblate E, et al. Mycobacterium bovis in rural Tanzania: Risk factors for infection in human and cattle populations. Tuberculosis. 2007 2015/01/27; 87(1):30–43. PMID: 16618553

23. Durnez L, Katakweba A, Sadiki H, Katholi CR, Kazwala RR, Machang'u RR, et al. Mycobacteria in Terrestrial Small Mammals on Cattle Farms in Tanzania. Vet Med International. 2011; 2011:1–12.

24. Mbugi EV, Katale BZ, Kendall S, Good L, Kibiki GS, Keyyu JD, et al. Tuberculosis cross-species transmission in Tanzania: Towards a One-Health concept. Onderstepoort J Vet Res. 2012; 79(2):E1–6.

25. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. Journal of Clinical Microbiology. 1997 April 1, 1997; 35(4):907–14. PMID: 9157152

26. Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, et al. SITVITWEB–A publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. Infection, Genetics and Evolution. 2012; 12(4):755–66. doi: 10.1016/j.meegid.2012.02.004 PMID: 22365971

27. Tang C, Reyes JF, Luciani F, Francis AR, Tanaka MM. spolTools: online utilities for analyzing spoligotypes of the Mycobacterium tuberculosis complex. Bioinformatics. 2008 October 15, 2008; 24 (20):2414–5. doi: 10.1093/bioinformatics/btn434 PMID: 18710872

28. Reyes JF, Francis AR, Tanaka MM. Models of deletion for visualizing bacterial variation: an application to tuberculosis spoligotypes. BMC Bioinformatics. 2008; 9:496–. doi: 10.1186/1471-2105-9-496 PMID: 19036166

29. Bennett KP, Ozcaglar C, Ranganathan J, Raghavan S, Katz J, Croft D, et al., editors. Visualization of tuberculosis patient and Mycobacterium tuberculosis complex genotype data via host-pathogen maps. Bioinformatics and Biomedicine Workshops (BIBMW), IEEE International Conference; 2011.

30. Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, et al. TB-Lineage: an online tool for classification and analysis of strains of Mycobacterium tuberculosis complex. Infect Genet Evol. 2012; 12(4):789–97. doi: 10.1016/j.meegid.2012.02.010 PMID: 22406225

31. Mallard K, McNerney R, Crampin AC, Houben R, Ndlovu R, Munthali L, et al. Molecular Detection of Mixed Infections of Mycobacterium tuberculosis Strains in Sputum Samples from Patients in Karonga District, Malawi. J Clin Microbiol. 2010; 48(12):4512–8. doi: 10.1128/JCM.01683-10 PMID: 20962138

32. Ignatova A, Dubiley S, Stepanshina V, Shemyakin I. Predominance of multi-drug-resistant LAM and Beijing family strains among Mycobacterium tuberculosis isolates recovered from prison inmates in Tula Region, Russia. J Med Microbiol. 2006; 55(10):1413–8.

33. Belay M, Ameni G, Bjune G, Couvin D, Rastogi N, Abebe F. Strain Diversity of Mycobacterium tuberculosis Isolates from Pulmonary Tuberculosis Patients in Afar Pastoral Region of Ethiopia doi: 10.1155/2014/238532 BioMed Res Intern. 2014; 2014(ID 238532):12 pages.

34. Cadmus S, Hill Vr, van Soolingen D, Rastogi N. Spoligotype Profile of *Mycobacterium tuberculosis* Complex Strains from HIV-Positive and -Negative Patients in Nigeria: a Comparative Analysis. J Clin Microbiol. 2011 January 1, 2011; 49(1):220–6. doi: 10.1128/JCM.01241-10 PMID: 21048016

35. Guerra-Assunção JA, Crampin AC, Houben RM, Mzembe T, Mallard K, Coll F, et al. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. eLife 2015 2015:Mar 3; 4 doi: 10.7554/eLife.05166

36. Lukoye D, Katabazi FA, Musisi K, Kateete DP, Asiimwe BB, Okee M, et al. The T2 *Mycobacterium tuberculosis* Genotype, Predominant in Kampala, Uganda, Shows Negative Correlation with Antituberculosis Drug Resistance. Antimicrob Agents Chemother. 2014 July 1, 2014; 58(7):3853–9. doi: 10.1128/AAC.02338-13 PMID: 24777100

37. Viegas SO, Machado A, Groenheit R, Ghebremichael S, Pennhag A, Gudo PS, et al. *Mycobacterium tuberculosis* Beijing Genotype Is Associated with HIV Infection in Mozambique. PLoS ONE. 2013; 8(8): e71999. doi: 10.1371/journal.pone.0071999 PMID: 23940801

38. Suchindran S, Brouwer ES, Van Rie A. Is HIV Infection a Risk Factor for Multi-Drug Resistant Tuberculosis? A Systematic Review. PLoS ONE. 2009; 4(5):e5561. doi: 10.1371/journal.pone.0005561 PMID: 19440304

39. Kaufmann SH, McMichael AJ. Annulling a dangerous liaison: vaccination strategies against AIDS and tuberculosis. Nat Med. 2005; 11(4 Suppl):S33–44. PMID: 15812488

40. Pawlowski A, Jansson M, Sköld M, Rottenberg ME, Källenius G. Tuberculosis and HIV Co-Infection. PLoS Pathogens. 2012; 8(2):e1002464. doi: 10.1371/journal.ppat.1002464 PMID: 22363214

41. Gagneux S, Small PM. Global phylogeography of mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis 2007; 7:328–73. PMID: 17448936

42. Albanna AS, Reed MB, Kotar KV, Fallow A, McIntosh FA, Behr MA, et al. Reduced Transmissibility of East African Indian Strains of *Mycobacterium tuberculosis*. PLoS ONE. 2012; 6(9):e25075.

43. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, et al. Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex. PLoS Pathog. 2008; 4(9):e1000160. doi: 10.1371/journal.ppat.1000160 PMID: 18802459

44. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host–pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci USA. 2006 February 21, 2006; 103(8):2869–73. PMID: 16477032

45. Brudey K, Driscoll J, Rigouts L, Prodinger W, Gori A, Al-Hajoj S, et al. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiology C7–23. 2006; 6(1):1–17.

46. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci USA. 2002; 99(6):3684–9. PMID: 11891304

47. Sola C, Filliol I, Legrand E, Lesjean S, Locht C, Supply P, et al. Genotyping of the Mycobacterium tuberculosis complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. Infection, Genetics and Evolution. 2003; 3(2):125–33. PMID: 12809807

48. Vellarikkal SK, Singh AV, Singh PK, Garg P, Katoch VM, Katoch K, et al. Draft Genome Sequence of Multidrug-Resistant Mycobacterium tuberculosis Clinical Isolate OSDD515, Belonging to the Uganda I Genotype. Genome Announcements. 2013 December 26, 2013; 1(6).

49. Thomas SK, Iravatham CC, Moni BH, Kumar A, Archana BV, Majid M, et al. Modern and Ancestral Genotypes of *Mycobacterium tuberculosis* from Andhra Pradesh, India. PLoS ONE. 2011; 6(11): e27584. doi: 10.1371/journal.pone.0027584 PMID: 22114678

50. Helal ZH, El-Din Ashour MS, Eissa SA, Abd-Elatef G, Zozio T, Babapoor S, et al. Unexpectedly High Proportion of Ancestral Manu Genotype Mycobacterium tuberculosis Strains Cultured from Tuberculosis Patients in Egypt. J Clin Microbiol. 2009 September 1, 2009; 47(9):2794–801. doi: 10.1128/JCM.00360-09 PMID: 19553569

51. Jamison DT, Feachem RG, Makgoba MW, Bos ER, Baingana FK, Hofman KJ, et al. Disease and Mortality in Sub-Saharan Africa. 2nd ed. Washington (DC): World Bank; 2006.

52. Ahmed MM, Mohammed SH, Nasurallah HAA, Ali MM, Couvin D, Rastogi N. Snapshot of the genetic diversity of Mycobacterium tuberculosis isolates in Iraq. International Journal of Mycobacteriology. 2014 2015/04/21; 3(3):184–96. doi: 10.1016/j.ijmyco.2014.07.006 PMID: 26786487

53. Ismail F, Couvin D, Farakhin I, Abdul Rahman Z, Rastogi N, Suraiya S. Study of Mycobacterium tuberculosis Complex Genotypic Diversity in Malaysia Reveals a Predominance of Ancestral East-African-Indian Lineage with a Malaysia-Specific Signature. PLoS ONE. 2014; 9(12):e114832. doi: 10.1371/journal.pone.0114832 PMID: 25502956

54. Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, et al. Predominance of Ancestral Lineages of Mycobacterium tuberculosis in India. Emerg Infect Dis. 2006; 12(9):1367–74. PMID: 17073085

55. Levy S. The Evolution of Tuberculosis: Genetic analysis offers new insight on the spread of an ancient disease. BioScience. 2012 July 1, 2012; 62(7):625–9.

56. Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, et al. Clade-Specific Virulence Patterns of Mycobacterium tuberculosis Complex Strains in Human Primary Macrophages and Aerogenically Infected Mice mBio. 2013 August 30, 2013; 4(4).

57. Brown T, Nikolayevskyy V, Velji P, Drobniewski F. Associations between Mycobacterium tuberculosis Strains and Phenotypes. Emerging Infectious Diseases. 2010; 16(2):272–80. doi: 10.3201/eid1602.091032 PMID: 20113558

58. Filliol I, Driscoll JR, van Soolingen D, Kreiswirth BN, Kremer K, Valétudie G, et al. Global Distribution of Mycobacterium tuberculosis Spoligotypes. Emerging Infectious Diseases. 2002; 8(11):1347–9. PMID: 12453368

59. Chang J-R, Chen Y-Y, Huang T-S, Huang W-F, Kuo S-C, Tseng F-C, et al. Clonal Expansion of Both Modern and Ancient Genotypes of Mycobacterium tuberculosis in Southern Taiwan. PLoS ONE. 2012; 7(8):e43018. doi: 10.1371/journal.pone.0043018 PMID: 22937008

60. Arora J, Singh UB, Suresh N, Rana T, Porwal C, Kaushik A, et al. Characterization of predominant Mycobacterium tuberculosis strains from different subpopulations of India. Infection, Genetics and Evolution. 2009; 9(5):832–9. doi: 10.1016/j.meegid.2009.05.008 PMID: 19463970

61. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat Genet. 2012; 47(3):242–9.

62. Flores-Treviño S, Morfín-Otero R, Rodríguez-Noriega E, González-Díaz E, Pérez-Gómez HR, Bocanegra-García V, et al. Genetic Diversity of *Mycobacterium tuberculosis* from Guadalajara, Mexico and Identification of a Rare Multidrug Resistant Beijing Genotype. PLoS ONE. 2015; 10(2):e0118095. doi: 10.1371/journal.pone.0118095 PMID: 25695431

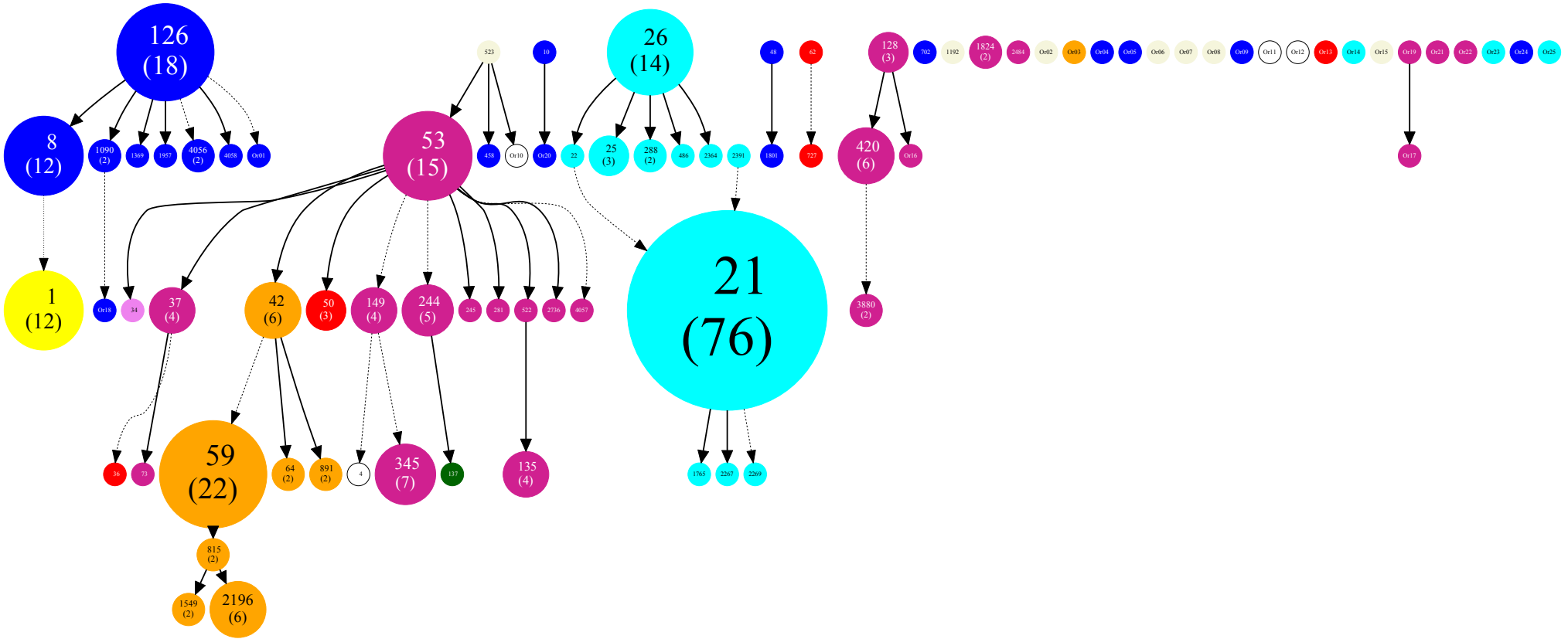
63. Phyu S, Stavrum R, Lwin T, Svendsen ÃyS, Ti T, Grewal HMS. Predominance of Mycobacterium tuberculosis EAI and Beijing Lineages in Yangon, Myanmar. J Clin Microbiol. 2009; 47(2):335–44. doi: 10.1128/JCM.01812-08 PMID: 19036933

64. Maeda S, Hang NT, Lien LT, Thuong PH, Hung NV, Hoang NP, et al. *Mycobacterium tuberculosis* strains spreading in Hanoi, Vietnam: Beijing sublineages, genotypes, drug susceptibility patterns, and host factors. Tuberculosis 2014; 94(6):649–56. PMID: 25459163

65. Douglas JT, Qian L, Montoya JC, Musser JM, Van Embden JDA, Van Soolingen D, et al. Characterization of the Manila Family of Mycobacterium tuberculosis. J Clin Microbiol. 2003; 41(6):2723–6. PMID: 12791915

66. Ferdinand S, Sola C, Chanteau S, Ramarokoto H, Rasolonavalona T, Rasolofo-Razanamparany V, et al. A study of spoligotyping-defined Mycobacterium tuberculosis clades in relation to the origin of peopling and the demographic history in Madagascar. Infection, Genetics and Evolution. 2005; 5(4):340–8. PMID: 16168940

67. Rahim Z, Zaman K, van der Zanden AGM, Möllers MJ, van Soolingen D, Raqib R, et al. Assessment of Population Structure and Major Circulating Phylogeographical Clades of Mycobacterium tuberculosis Complex in Bangladesh Suggests a High Prevalence of a Specific Subclade of Ancient M. tuberculosis Genotypes. J Clin Microbiol. 2007; 45(11):3791–4. PMID: 17804653

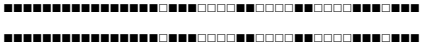
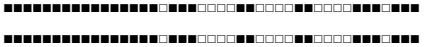
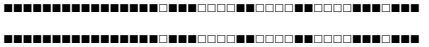

126
(18)
8
(12)
1
(12)
1090
(2)
1369
1957
4056
(2)
4058
Or01
Or18
34
37
(4)
36
73
59
(22)
815
(2)
1549
(2)
2196
(6)
53
(15)
42
(6)
64
(2)
891
(2)
4
50
(3)
149
(4)
345
(7)
137
244
(5)
135
(4)
245
281
522
2736
4057
523
458
Or10
Or20
10
26
(14)
22
25
(3)
288
(2)
486
2364
2391
21
(76)
1765
2267
2269
48
1801
62
727
128
(3)
420
(6)
Or16
3880
(2)
Or17
702
1192
1824
(2)
2484
Or02
Or03
Or04
Or05
Or06
Or07
Or08
Or09
Or11
Or12
Or13
Or14
Or15
Or19
Or21
Or22
Or23
Or24
Or25

IsoNumber	Year	Strain	Spoligotype Description	Octal code	SIT	Lineage	Isolation city	Sex/Age	HIV Status
TZA0220121B0725	2012	B0725	□□■■■■■■■■■	000000000003771	1	Beijing	Bunda	F/41	positive
TZA0220101B0656	2010	B0656	□□■■■■■■■■■	000000000003771	1	Beijing	Bunda	M/31	negative
TZA0220111S0697	2011	S0697	□□■■■■■■■■■	000000000003771	1	Beijing	Serengeti	M/33	negative
TZA0220111S0705	2011	S0705	□□■■■■■■■■■	000000000003771	1	Beijing	Serengeti	M/22	negative
TZA0220111S0710	2011	S0710	□□■■■■■■■■■	000000000003771	1	Beijing	Serengeti	M/46	negative
TZA0220121S0700	2012	S0700	□□■■■■■■■■■	000000000003771	1	Beijing	Serengeti	M/51	negative
TZA0220121B0724	2012	B0724	□□■■■■■■■■■	000000000003771	1	Beijing	Bunda	M/37	negative
TZA0220121N1254	2012	N1254	□□■■■■■■■■■	000000000003771	1	Beijing	Ngorongoro	F/29	negative
TZA0220121D0167	2012	D0167	□□■■■■■■■■■	000000000003771	1	Beijing	Dar es Salaam	M/19	Not recorded
TZA0220111D0122	2011	D0122	□□■■■■■■■■■	000000000003771	1	Beijing	Dar es Salaam	M/27	Not recorded
TZA0220121D0150	2012	D0150	□□■■■■■■■■■	000000000003771	1	Beijing	Dar es Salaam	M/34	Not recorded
TZA0220111D0106	2011	D0106	□□■■■■■■■■■	000000000003771	1	Beijing	Dar es Salaam	F/43	Not recorded
TZA0220121N1012	2012	N1012	□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■□□□□■■■■■■■	000000007760771	4	Unknown	Ngorongoro	F/25	negative
TZA0220111B0721	2011	B0721	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Bunda	M/36	negative
TZA0220111N1273	2011	N1273	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Ngorongoro	F/45	negative
TZA0220101B0587	2010	B0587	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Bunda	F/28	positive
TZA0220101S0012	2010	S0012	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Serengeti	F/25	negative
TZA0220121D0155	2012	D0155	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Dar es Salaam	F/41	Not recorded
TZA0220121D0173	2012	D0173	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Dar es Salaam	M/33	Not recorded
TZA0220121D0179	2012	D0179	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Dar es Salaam	F/39	Not recorded
TZA0220121D0181	2012	D0181	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Dar es Salaam	F/20	Not recorded
TZA0220121D0142	2012	D0142	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Dar es Salaam	M/35	Not recorded
TZA0220121D0168	2012	D0168	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Dar es Salaam	M/32	Not recorded
TZA0220111D0127	2011	D0127	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Dar es Salaam	F/39	Not recorded
TZA0220111D0116	2011	D0116	■□□□□□□□□□□□□■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	8	EAI5	Dar es Salaam	M/27	Not recorded
TZA0220111D0115	2011	D0115	■□□■■■■■■■■■■■■■■□■□■■■■■■■■□□□□■□■■■■■■■■■	477777277413771	10	EAI8-MDG	Dar es Salaam	M/26	Not recorded
TZA0220111N1172	2011	N1172	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/42	negative
TZA0220101B0586	2010	B0586	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Bunda	F/38	negative
TZA0220101N0481	2010	N0481	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/39	negative
TZA0220101N1265	2010	N1265	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/32	negative
TZA0220101N1161	2010	N1161	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/35	negative
TZA0220111B0588	2011	B0588	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Bunda	M/33	negative
TZA0220121N1175	2012	N1175	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/38	negative
TZA0220101B0368	2010	B0368	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Bunda	M/31	negative
TZA0220121B0727	2012	B0727	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Bunda	F/43	negative
TZA0220101N0016	2010	N0016	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/60	negative
TZA0220121N1010	2012	N1010	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/44	positive
TZA0220121N1116	2012	N1116	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/29	negative
TZA0220101N010-2	2010	N010-2	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/51	negative
TZA0220101N010-5	2010	N010-5	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/39	negative
TZA0220121N1014	2012	N1014	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/37	negative
TZA0220111B0572	2011	B0572	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Bunda	F/47	negative
TZA0220121N1115	2012	N1115	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/55	negative
TZA0220121N1131	2012	N1131	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/52	negative
TZA0220101N1128	2010	N1128	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/28	negative
TZA0220121N1130	2012	N1130	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/47	negative
TZA0220121N1126	2012	N1126	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/52	negative
TZA0220121N1127	2012	N1127	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/45	negative
TZA0220111N1117	2011	N1117	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/31	negative
TZA0220121N1124	2012	N1124	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/33	positive
TZA0220121N1137	2012	N1137	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/24	negative
TZA0220111N1118	2011	N1118	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	M/28	negative
TZA0220111N1119	2011	N1119	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/37	negative
TZA0220101N0487	2010	N0487	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/25	negative
TZA0220121S0703	2012	S0703	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Serengeti	M/46	negative
TZA0220111S0709	2011	S0709	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Serengeti	F/37	negative
TZA0220111S0698	2011	S0698	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Serengeti	M/30	negative
TZA0220101N0485_b	2010	N0485_b	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Ngorongoro	F/30	negative
TZA0220111B0594	2011	B0594	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■■	703377400001771	21	CAS1-Kili	Bunda	F/39	negative

TZA0220101B0370_a	2010	B0370_a		703377400001771	21	CAS1-Kili	Bunda	F/24	negative
TZA0220111S0655	2011	S0655		703377400001771	21	CAS1-Kili	Serengeti	M/38	negative
TZA0220111B0513	2011	B0513		703377400001771	21	CAS1-Kili	Bunda	M/28	negative
TZA0220101N0482	2010	N0482		703377400001771	21	CAS1-Kili	Ngorongoro	F/39	negative
TZA0220101B0370_b	2010	B0370_b		703377400001771	21	CAS1-Kili	Bunda	M/35	negative
TZA0220121N1252	2012	N1252		703377400001771	21	CAS1-Kili	Ngorongoro	M/22	negative
TZA0220121D0175	2012	D0175		703377400001771	21	CAS1-Kili	Dar es Salaam	M/43	Not recorded
TZA0220121D0174	2012	D0174		703377400001771	21	CAS1-Kili	Dar es Salaam	F/37	Not recorded
TZA0220121D0161	2012	D0161		703377400001771	21	CAS1-Kili	Dar es Salaam	F/48	Not recorded
TZA0220121D0153	2012	D0153		703377400001771	21	CAS1-Kili	Dar es Salaam	F/45	Not recorded
TZA0220121D0162	2012	D0162		703377400001771	21	CAS1-Kili	Dar es Salaam	F/27	Not recorded
TZA0220121D0172	2012	D0172		703377400001771	21	CAS1-Kili	Dar es Salaam	M/36	Not recorded
TZA0220121D0145	2012	D0145		703377400001771	21	CAS1-Kili	Dar es Salaam	M/42	Not recorded
TZA0220121D0170	2012	D0170		703377400001771	21	CAS1-Kili	Dar es Salaam	M/65	Not recorded
TZA0220111D0137	2011	D0137		703377400001771	21	CAS1-Kili	Dar es Salaam	M/22	Not recorded
TZA0220121D0166	2012	D0166		703377400001771	21	CAS1-Kili	Dar es Salaam	F/31	Not recorded
TZA0220121D0146	2012	D0146		703377400001771	21	CAS1-Kili	Dar es Salaam	M/24	Not recorded
TZA0220121D0148	2012	D0148		703377400001771	21	CAS1-Kili	Dar es Salaam	M/33	Not recorded
TZA0220121D0164	2012	D0164		703377400001771	21	CAS1-Kili	Dar es Salaam	F/25	Not recorded
TZA0220111D0113	2011	D0113		703377400001771	21	CAS1-Kili	Dar es Salaam	M/23	Not recorded
TZA0220121D0171	2012	D0171		703377400001771	21	CAS1-Kili	Dar es Salaam	F/21	Not recorded
TZA0220121D1100	2012	D1100		703377400001771	21	CAS1-Kili	Dar es Salaam	M/27	Not recorded
TZA0220111D0126	2011	D0126		703377400001771	21	CAS1-Kili	Dar es Salaam	F/31	Not recorded
TZA0220111D0104	2011	D0104		703377400001771	21	CAS1-Kili	Dar es Salaam	F/30	Not recorded
TZA0220121D0160	2012	D0160		703377400001771	21	CAS1-Kili	Dar es Salaam	F/27	Not recorded
TZA0220121D0141	2012	D0141		703377400001771	21	CAS1-Kili	Dar es Salaam	M/39	Not recorded
TZA0220111D0136	2011	D0136		703377400001771	21	CAS1-Kili	Dar es Salaam	F/46	Not recorded
TZA0220111D0120	2011	D0120		703377400001771	21	CAS1-Kili	Dar es Salaam	M/23	Not recorded
TZA0220111D0131	2011	D0131		703377400001771	21	CAS1-Kili	Dar es Salaam	F/49	Not recorded
TZA0220111D0128	2011	D0128		703377400001771	21	CAS1-Kili	Dar es Salaam	F/37	Not recorded
TZA0220121D0156	2012	D0156		703377400001771	21	CAS1-Kili	Dar es Salaam	F/23	Not recorded
TZA0220121D0151	2012	D0151		703377400001771	21	CAS1-Kili	Dar es Salaam	F/28	Not recorded
TZA0220121D0140	2012	D0140		703377400001771	21	CAS1-Kili	Dar es Salaam	F/39	Not recorded
TZA0220111D0125	2011	D0125		703377400001771	21	CAS1-Kili	Dar es Salaam	M/21	Not recorded
TZA0220111D0108	2011	D0108		703377400001771	21	CAS1-Kili	Dar es Salaam	M/42	Not recorded
TZA0220111D0105	2011	D0105		703377400001771	21	CAS1-Kili	Dar es Salaam	M/41	Not recorded
TZA0220111D0102	2011	D0102		703377400001771	21	CAS1-Kili	Dar es Salaam	F/62	Not recorded
TZA0220111D0112	2011	D0112		703377400001771	21	CAS1-Kili	Dar es Salaam	M/33	Not recorded
TZA0220111D0123	2011	D0123		703377400001771	21	CAS1-Kili	Dar es Salaam	M/49	Not recorded
TZA0220121D1110	2012	D1110		703377400001771	21	CAS1-Kili	Dar es Salaam	M/51	Not recorded
TZA0220111D0135	2011	D0135		703377400001771	21	CAS1-Kili	Dar es Salaam	M/23	Not recorded
TZA0220111D0124	2011	D0124		703377400001771	21	CAS1-Kili	Dar es Salaam	F/35	Not recorded
TZA0220121D0165	2012	D0165		703377400001771	21	CAS1-Kili	Dar es Salaam	M/39	Not recorded
TZA0220121D0177	2012	D0177		703777400001771	22	CAS1-Delhi	Dar es Salaam	M/22	Not recorded
TZA0220111N1144	2011	N1144		703777740003171	25	CAS1-Delhi	Ngorongoro	M/52	positive
TZA0220121N1367	2012	N1367		703777740003171	25	CAS1-Delhi	Ngorongoro	M/32	negative
TZA0220111D0110	2011	D0110		703777740003171	25	CAS1-Delhi	Dar es Salaam	F/31	Not recorded
TZA0220101S0225	2010	S0225		703777740003771	26	CAS1-Delhi	Serengeti	F/22	negative
TZA0220111S0548	2011	S0548		703777740003771	26	CAS1-Delhi	Serengeti	F/27	negative
TZA0220101S0026_a	2010	S0026_a		703777740003771	26	CAS1-Delhi	Serengeti	M/54	negative
TZA0220111N1013	2011	N1013		703777740003771	26	CAS1-Delhi	Ngorongoro	F/27	negative
TZA0220121N1123	2012	N1123		703777740003771	26	CAS1-Delhi	Ngorongoro	M/42	negative
TZA0220101S0256	2010	S0256		703777740003771	26	CAS1-Delhi	Serengeti	M/38	negative
TZA0220111S0696	2011	S0696		703777740003771	26	CAS1-Delhi	Serengeti	F/37	negative
TZA0220111S0699	2011	S0699		703777740003771	26	CAS1-Delhi	Serengeti	F/35	negative
TZA0220121B0714_a	2012	B0714_a		703777740003771	26	CAS1-Delhi	Bunda	M/39	negative
TZA0220111S0707	2011	S0707		703777740003771	26	CAS1-Delhi	Serengeti	F/21	negative
TZA0220101S0026_b	2010	S0026_b		703777740003771	26	CAS1-Delhi	Serengeti	M/21	negative
TZA0220121D0176	2012	D0176		703777740003771	26	CAS1-Delhi	Dar es Salaam	M/21	Not recorded
TZA0220111D0133	2011	D0133		703777740003771	26	CAS1-Delhi	Dar es Salaam	F/38	Not recorded
TZA0220121D0154	2012	D0154		703777740003771	26	CAS1-Delhi	Dar es Salaam	M/75	Not recorded
TZA0220111B0510	2011	B0510		776377777760771	34	S	Bunda	F/41	positive
TZA0220101N0055	2010	N0055		777737777720771	36	H3	Ngorongoro	M/39	negative

TZA0220111B0647	2011	B0647		777737777760771	37	T3	Bunda	F/34	negative
TZA0220121B0646	2012	B0646		777737777760771	37	T3	Bunda	F/33	negative
TZA0220111D0117	2011	D0117		777737777760771	37	T3	Dar es Salaam	F/40	Not recorded
TZA0220111D0109	2011	D0109		777737777760771	37	T3	Dar es Salaam	F/25	Not recorded
TZA0220111N1372	2011	N1372		777777607760771	42	LAM9	Ngorongoro	M/30	negative
TZA0220101N1256	2010	N1256		777777607760771	42	LAM9	Ngorongoro	M/22	positive
TZA0220111N1259	2011	N1259		777777607760771	42	LAM9	Ngorongoro	M/37	negative
TZA0220111N0584	2011	N0584		777777607760771	42	LAM9	Ngorongoro	M/34	negative
TZA0220121N1276	2012	N1276		777777607760771	42	LAM9	Ngorongoro	M/33	positive
TZA0220111D0130	2011	D0130		777777607760771	42	LAM9	Dar es Salaam	F/48	Not recorded
TZA0220121D0139	2012	D0139		777777777413731	48	EAI1-SOM	Dar es Salaam	M/30	Not recorded
TZA0220121S0663	2012	S0663		777777777720771	50	H3	Serengeti	M/35	negative
TZA0220111B0666	2011	B0666		777777777720771	50	H3	Bunda	F/45	negative
TZA0220111S0664	2011	S0664		777777777720771	50	H3	Serengeti	M/40	negative
TZA0220101B0385	2010	B0385		777777777760771	53	T1	Bunda	M/21	negative
TZA0220101B0133	2010	B0133		777777777760771	53	T1	Bunda	F/22	negative
TZA0220111B0581	2011	B0581		777777777760771	53	T1	Bunda	F/50	negative
TZA0220111B0577	2011	B0577		777777777760771	53	T1	Bunda	F/21	negative
TZA0220121N1136	2012	N1136		777777777760771	53	T1	Ngorongoro	F/32	negative
TZA0220101B0392	2010	B0392		777777777760771	53	T1	Bunda	M/41	negative
TZA0220111S0704	2011	S0704		777777777760771	53	T1	Serengeti	M/38	negative
TZA0220111B0693	2011	B0693		777777777760771	53	T1	Bunda	M/38	negative
TZA0220121B0714_b	2012	B0714_b		777777777760771	53	T1	Bunda	M/45	negative
TZA0220111B0582	2011	B0582		777777777760771	53	T1	Bunda	M/59	negative
TZA0220101B0391	2010	B0391		777777777760771	53	T1	Bunda	M/33	negative
TZA0220101B0387	2010	B0387		777777777760771	53	T1	Bunda	M/51	negative
TZA0220121D0169	2012	D0169		777777777760771	53	T1	Dar es Salaam	F/41	Not recorded
TZA0220111D0118	2011	D0118		777777777760771	53	T1	Dar es Salaam	M/34	Not recorded
TZA0220121D0157	2012	D0157		777777777760771	53	T1	Dar es Salaam	M/63	Not recorded
TZA0220111N1164	2011	N1164		777777606060771	59	LAM11-ZWE	Ngorongoro	M/21	negative
TZA0220111N1163	2011	N1163		777777606060771	59	LAM11-ZWE	Ngorongoro	F/40	positive
TZA0220111B0636	2011	B0636		777777606060771	59	LAM11-ZWE	Bunda	F/41	negative
TZA0220111B0695	2011	B0695		777777606060771	59	LAM11-ZWE	Bunda	M/51	negative
TZA0220121N1243	2012	N1243		777777606060771	59	LAM11-ZWE	Ngorongoro	F/20	positive
TZA0220101S0264	2010	S0264		777777606060771	59	LAM11-ZWE	Serengeti	M/40	negative
TZA0220101N0068	2010	N0068		777777606060771	59	LAM11-ZWE	Ngorongoro	F/25	negative
TZA0220101S0263	2010	S0263		777777606060771	59	LAM11-ZWE	Serengeti	F/33	negative
TZA0220101N010-1	2010	N010-1		777777606060771	59	LAM11-ZWE	Ngorongoro	F/45	negative
TZA0220101N010-3	2010	N010-3		777777606060771	59	LAM11-ZWE	Ngorongoro	M/53	negative
TZA0220121N1121	2012	N1121		777777606060771	59	LAM11-ZWE	Ngorongoro	F/34	negative
TZA0220121B0632	2012	B0632		777777606060771	59	LAM11-ZWE	Bunda	M/23	negative
TZA0220121B0631	2012	B0631		777777606060771	59	LAM11-ZWE	Bunda	M/32	negative
TZA0220121N1133	2012	N1133		777777606060771	59	LAM11-ZWE	Ngorongoro	F/40	negative
TZA0220101B0374_b	2010	B0374_b		777777606060771	59	LAM11-ZWE	Bunda	M/22	negative
TZA0220101B0375	2010	B0375		777777606060771	59	LAM11-ZWE	Bunda	F/24	negative
TZA0220101B0373	2010	B0373		777777606060771	59	LAM11-ZWE	Bunda	M/18	negative
TZA0220111B0686	2011	B0686		777777606060771	59	LAM11-ZWE	Bunda	F/35	negative
TZA0220121N1242	2012	N1242		777777606060771	59	LAM11-ZWE	Ngorongoro	F/27	negative
TZA0220121D0147	2012	D0147		777777606060771	59	LAM11-ZWE	Dar es Salaam	M/38	Not recorded
TZA0220111D0138	2011	D0138		777777606060771	59	LAM11-ZWE	Dar es Salaam	F/25	Not recorded
TZA0220121D0178	2012	D0178		777777606060771	59	LAM11-ZWE	Dar es Salaam	F/29	Not recorded
TZA0220121N1365	2012	N1365		777777774020731	62	H1	Ngorongoro	M/30	negative
TZA0220121N1249	2012	N1249		777777607560771	64	LAM6	Ngorongoro	M/39	negative
TZA0220111D0114	2011	D0114		777777607560771	64	LAM6	Dar es Salaam	M/36	Not recorded
TZA0220121N1120	2012	N1120		777737777760731	73	T	Ngorongoro	M/29	negative
TZA0220101B0374_a	2010	B0374_a		477777777413771	126	EAI5	Bunda	M/41	negative
TZA0220121B0732	2012	B0732		477777777413771	126	EAI5	Bunda	F/21	negative
TZA0220111S0677	2011	S0677		477777777413771	126	EAI5	Serengeti	F/24	negative
TZA0220111B0613	2011	B0613		477777777413771	126	EAI5	Bunda	M/19	negative
TZA0220121N1122	2012	N1122		477777777413771	126	EAI5	Ngorongoro	F/40	negative
TZA0220101B0494	2010	B0494		477777777413771	126	EAI5	Bunda	M/38	negative
TZA0220111N0564	2011	N0564		477777777413771	126	EAI5	Ngorongoro	M/20	negative
TZA0220111B0379	2011	B0379		477777777413771	126	EAI5	Bunda	F/29	negative

TZA0220101B0384	2010	B0384		477777777413771	126	EAI5	Bunda	M/54	negative
TZA0220121N0589	2012	N0589		477777777413771	126	EAI5	Ngorongoro	F/25	negative
TZA0220111S0678	2011	S0678		477777777413771	126	EAI5	Serengeti	F/20	negative
TZA0220111B0675	2011	B0675		477777777413771	126	EAI5	Bunda	M/40	negative
TZA0220111S0676	2011	S0676		477777777413771	126	EAI5	Serengeti	F/25	negative
TZA0220101B0383	2010	B0383		477777777413771	126	EAI5	Bunda	F/30	negative
TZA0220101B0378	2010	B0378		477777777413771	126	EAI5	Bunda	F/27	negative
TZA0220101B0307	2010	B0307		477777777413771	126	EAI5	Bunda	F/35	negative
TZA0220121D0159	2012	D0159		477777777413771	126	EAI5	Dar es Salaam	F/50	Not recorded
TZA0220111D0144	2011	D0144		477777777413771	126	EAI5	Dar es Salaam	F/19	Not recorded
TZA0220101S0011_a	2010	S0011_a		637775777760730	128	T2-uganda	Serengeti	M/46	negative
TZA0220101S0011_b	2010	S0011_b		637775777760730	128	T2-uganda	Serengeti	F/39	negative
TZA0220111D0103	2011	D0103		637775777760730	128	T2-uganda	Dar es Salaam	M/38	Not recorded
TZA0220111S0684	2011	S0684		777777777760730	135	T2-uganda	Serengeti	M/39	positive
TZA0220111S0681	2011	S0681		777777777760730	135	T2-uganda	Serengeti	M/22	negative
TZA0220111B0679	2011	B0679		777777777760730	135	T2-uganda	Bunda	M/31	negative
TZA0220111S0683	2011	S0683		777777777760730	135	T2-uganda	Serengeti	F/28	negative
TZA0220101N0052	2010	N0052		777776777760601	137	X2	Ngorongoro	F/37	negative
TZA0220101B0652	2010	B0652		777000377760771	149	T3-ETH	Bunda	M/36	negative
TZA0220121B0715	2012	B0715		777000377760771	149	T3-ETH	Bunda	M/20	positive
TZA0220111B0712	2011	B0712		777000377760771	149	T3-ETH	Bunda	F/41	negative
TZA0220121B0716	2012	B0716		777000377760771	149	T3-ETH	Bunda	F/47	negative
TZA0220101N010-6	2010	N010-6		777777777760601	244	T1	Ngorongoro	M/36	negative
TZA0220111B0630	2011	B0630		777777777760601	244	T1	Bunda	M/49	negative
TZA0220101N0196	2010	N0196		777777777760601	244	T1	Ngorongoro	M/33	negative
TZA0220111N0419	2011	N0419		777777777760601	244	T1	Ngorongoro	F/29	negative
TZA0220101N0051	2010	N0051		777777777760601	244	T1	Ngorongoro	F/40	negative
TZA0220101N1239	2010	N1239		777777777760671	245	T1	Ngorongoro	F/29	negative
TZA0220111D0134	2011	D0134		777775777760771	281	T1	Dar es Salaam	M/39	Not recorded
TZA0220111B0603	2011	B0603		700377740003771	288	CAS2	Bunda	F/30	negative
TZA0220121D0149	2012	D0149		700377740003771	288	CAS2	Dar es Salaam	M/35	Not recorded
TZA0220111N0483	2011	N0483		777000377760731	345	T3-ETH	Ngorongoro	F/29	negative
TZA0220101N0484	2010	N0484		777000377760731	345	T3-ETH	Ngorongoro	F/41	negative
TZA0220111N0485_a	2011	N0485_a		777000377760731	345	T3-ETH	Ngorongoro	F/22	negative
TZA0220101N0486	2010	N0486		777000377760731	345	T3-ETH	Ngorongoro	F/20	negative
TZA0220121S0701	2012	S0701		777000377760731	345	T3-ETH	Serengeti	F/33	negative
TZA0220101N0479	2010	N0479		777000377760731	345	T3-ETH	Ngorongoro	M/21	negative
TZA0220111S0558	2011	S0558		777000377760731	345	T3-ETH	Serengeti	M/51	negative
TZA0220101S0005	2010	S0005		637774777760730	420	T2-uganda	Serengeti	M/30	negative
TZA0220101S0006	2010	S0006		637774777760730	420	T2-uganda	Serengeti	M/22	negative
TZA0220101S0010	2010	S0010		637774777760730	420	T2-uganda	Serengeti	M/35	negative
TZA0220101S0002_a	2010	S0002_a		637774777760730	420	T2-uganda	Serengeti	M/39	positive
TZA0220101S0009	2010	S0009		637774777760730	420	T2-uganda	Serengeti	F/34	negative
TZA0220101S0002_b	2010	S0002_b		637774777760730	420	T2-uganda	Serengeti	M/44	positive
TZA0220101B0057	2010	B0057		777777777403771	458	EAI5	Bunda	M/37	negative
TZA0220121D0143	2012	D0143		703777740000371	486	CAS	Dar es Salaam	M/51	Not recorded
TZA0220111B0576	2011	B0576		777777777760770	522	T1	Bunda	F/45	negative
TZA0220101N010-4	2010	N010-4		777777777777771	523	Manu_ancestor	Ngorongoro	M/22	negative
TZA0220121D0158	2012	D0158		700775747413771	702	EAI6-BGD1	Dar es Salaam	F/44	Not recorded
TZA0220111N2080	2011	N2080		777737774020731	727	H1	Ngorongoro	F/41	negative
TZA0220121B0719	2012	B0719		777777606060731	815	LAM11-ZWE	Bunda	M/40	negative
TZA0220121N1238	2012	N1238		777777606060731	815	LAM11-ZWE	Ngorongoro	F/45	negative
TZA0220121B0729	2012	B0729		777777607660771	891	LAM9	Bunda	F/31	negative
TZA0220121B0735	2012	B0735		777777607660771	891	LAM9	Bunda	M/25	negative
TZA0220101N010-7	2010	N010-7		077777777413771	1090	EAI5	Ngorongoro	M/20	negative
TZA0220111B0604	2011	B0604		077777777413771	1090	EAI5	Bunda	M/36	negative
TZA0220101N010-9	2010	N010-9		777777677763771	1192	Manu2	Ngorongoro	F/35	negative
TZA0220111B0304	2011	B0304		477777777413671	1369	EAI5	Bunda	F/37	negative
TZA0220101N1062	2010	N1062		775777606060731	1549	LAM11-ZWE	Ngorongoro	F/39	negative
TZA0220111B0573	2011	B0573		775777606060731	1549	LAM11-ZWE	Bunda	F/32	positive
TZA0220121D0152	2012	D0152		703337400001771	1765	CAS1-Kili	Dar es Salaam	M/41	Not recorded
TZA0220121N1125	2012	N1125		777777777413730	1801	EAI1-SOM	Ngorongoro	M/22	positive
TZA0220101S002M	2010	S002M		760377777760701	1824	T1	Serengeti	M/54	negative

TZA0220101S0025	2010	S0025		760377777760701	1824	T1	Serengeti	M/41	negative
TZA0220111D0129	2011	D0129		477777777013771	1957	EAI5	Dar es Salaam	F/33	Not recorded
TZA0220111B0657	2011	B0657		777775606060731	2196	LAM11-ZWE	Bunda	F/42	negative
TZA0220121B0659	2012	B0659		777775606060731	2196	LAM11-ZWE	Bunda	F/30	negative
TZA0220121S0711	2012	S0711		777775606060731	2196	LAM11-ZWE	Serengeti	F/43	negative
TZA0220111S0702	2011	S0702		777775606060731	2196	LAM11-ZWE	Serengeti	M/30	negative
TZA0220111S0706	2011	S0706		777775606060731	2196	LAM11-ZWE	Serengeti	M/42	negative
TZA0220111S0708	2011	S0708		777775606060731	2196	LAM11-ZWE	Serengeti	M/31	positive
TZA0220121N1247	2012	N1247		701377400001771	2267	CAS	Ngorongoro	M/23	negative
TZA0220101N1171	2010	N1171		703377400000771	2269	CAS	Ngorongoro	F/45	negative
TZA0220111N1071	2011	N1071		703601740003771	2364	CAS1-Delhi	Ngorongoro	F/22	negative
TZA0220111D0121	2011	D0121		703377400003771	2391	CAS1-Kili	Dar es Salaam	M/29	Not recorded
TZA0220121N1011	2012	N1011		737377677760771	2484	T1	Ngorongoro	F/22	positive
TZA0220121N1132	2012	N1132		777777777760741	2736	T1	Ngorongoro	M/39	negative
TZA0220101B0306	2010	B0306		437774777760730	3880	T2-uganda	Bunda	M/38	positive
TZA0220111B0302	2011	B0302		437774777760730	3880	T2-uganda	Bunda	F/37	negative
TZA0220121B0720	2012	B0720		477777777413411	4056*	EAI5	Bunda	F/22	negative
TZA0220121B0722	2012	B0722		477777777413411	4056*	EAI5	Bunda	M/38	negative
TZA0220121N1253	2012	N1253		777017777760771	4057*	T1	Ngorongoro	M/55	negative
TZA0220121B0398	2012	B0398		477777777413371	4058*	EAI5	Bunda	M/41	negative
TZA0220101N1240	2010	N1240		477777776003771	Or01	EAI	Ngorongoro	M/37	negative
TZA0220101B0046	2010	B0046		733357777773771	Or02	Manu1	Bunda	M/32	negative
TZA0220111N1251	2011	N1251		777356607760671	Or03	LAM9	Ngorongoro	F/40	negative
TZA0220121N1278	2012	N1278		733647672403771	Or04	EAI	Ngorongoro	M/30	negative
TZA0220111N1246	2011	N1246		516357777413731	Or05	EAI1-SOM	Ngorongoro	F/21	negative
TZA0220101N1244	2010	N1244		773257677743771	Or06	Manu2	Ngorongoro	M/39	negative
TZA0220121N1370	2012	N1370		773277606763671	Or07	Manu2	Ngorongoro	F/44	negative
TZA0220121N1371	2012	N1371		777357676263771	Or08	Manu2	Ngorongoro	M/29	negative
TZA0220111N1160	2011	N1160		773357776003771	Or09	EAI	Ngorongoro	M/22	negative
TZA0220121N1266	2012	N1266		777777777601771	Or10	Manu2	Ngorongoro	F/35	negative
TZA0220101N1057	2010	N1057		777777637563771	Or11	Manu2	Ngorongoro	M/24	negative
TZA0220111N1155	2011	N1155		770217777513760	Or12	Manu1	Ngorongoro	M/27	negative
TZA0220111N1135	2011	N1135		700002210320300	Or13	H3	Ngorongoro	F/37	positive
TZA0220101N1129	2010	N1129		703245020000640	Or14	CAS	Ngorongoro	F/49	negative
TZA0220111N1134	2011	N1134		777747777763771	Or15	Manu2	Ngorongoro	F/53	negative
TZA0220101B0660	2010	B0660		437775777760730	Or16	T2-uganda	Bunda	M/51	negative
TZA0220111B0713	2011	B0713		637734477760730	Or17	T2-uganda	Bunda	F/42	negative
TZA0220101B0393	2010	B0393		077777777413371	Or18	EAI5	Bunda	M/35	negative
TZA0220121B0723	2012	B0723		637735477760730	Or19	T2-uganda	Bunda	F/29	positive
TZA0220101B0542	2010	B0542		477777277413731	Or20	EAI8-MDG	Bunda	M/23	negative
TZA0220111B0590	2011	B0590		577777774060771	Or21	T1	Bunda	M/31	negative
TZA0220101B0389	2010	B0389		637777477760730	Or22	T2-uganda	Bunda	F/25	negative
TZA0220111D0132	2011	D0132		703775000003771	Or23	CAS	Dar es Salaam	M/45	Not recorded
TZA0220111D0119	2011	D0119		777677757413371	Or24	EAI6-BGD1	Dar es Salaam	F/21	Not recorded
TZA0220121D0163	2012	D0163		703301000001771	Or25	CAS	Dar es Salaam	F/61	Not recorded