

Research article

Open Access

A first insight into the genetic diversity of *Mycobacterium tuberculosis* in Dar es Salaam, Tanzania, assessed by spoligotyping

Vegard Eldholm^{1,2}, Mecky Matee³, Sayoki GM Mfinanga³, Manfred Heun² and Ulf R Dahle*¹

Address: ¹Division of Infectious Disease Control, Norwegian Institute of Public Health, Oslo, Norway, ²Institute of Nature Resource Management, Norwegian University of Life Sciences, Ås, Norway and ³Muhimbili Medical Research Centre, Dar es Salaam, Tanzania

Email: Vegard Eldholm - vegard.eldholm@fhi.no; Mecky Matee - mmatee@muchs.ac.tz; Sayoki GM Mfinanga - gsmfinanga@yahoo.com; Manfred Heun - manfred.heun@umb.no; Ulf R Dahle* - ulf.dahle@fhi.no

* Corresponding author

Published: 13 September 2006

Received: 23 June 2006

BMC Microbiology 2006, 6:76 doi:10.1186/1471-2180-6-76

Accepted: 13 September 2006

This article is available from: <http://www.biomedcentral.com/1471-2180/6/76>

© 2006 Eldholm et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Tanzania has a high tuberculosis incidence, and genotyping studies of *Mycobacterium tuberculosis* in the country are necessary in order to improve our understanding of the epidemic. Spoligotyping is a potentially powerful genotyping method due to fast generation of genotyping results, high reproducibility and low operation costs. The recently constructed SpolDB4 database and the model-based program 'Spotclust' can be used to assign isolates to families, subfamilies and variants. The results of a study can thus be analyzed in a global context.

Results: One hundred forty-seven pulmonary isolates from consecutive tuberculosis patients in Dar es Salaam were spoligotyped. SpolDB4 and 'Spotclust' were used to assign isolates to families, subfamilies and variants. The CAS (37%), LAM (22%) and EAI (17%) families were the most abundant. Despite the dominance of these three families, diversity was high due to variation within *M. tuberculosis* families. Of the obtained spoligopatterns, 64% were previously unrecorded.

Conclusion: Spoligotyping is useful to gain an overall understanding of the local TB epidemic. This study demonstrates that the extensive TB epidemic in Dar es Salaam, Tanzania is caused by a few successful *M. tuberculosis* families, dominated by the CAS family. Import of strains was a minor problem.

Background

In Tanzania, the tuberculosis (TB) incidence doubled between 1990 and 2004 [1]. The rate of all forms of the disease is estimated at 524/100,000 and the rate of new sputum smear positive disease is approximately 157/100,000 [1] with Dar es Salaam contributing about 26% of all TB cases [2]. The World Health Organization estimates that Tanzania has the 14th highest TB burden in the world [1]. Points of concern include the proportion of

patients lost to follow-up, currently at 9%, an average diagnostic delay of 6 months, decreasing case detection rate (from 55% in 1997 to 45% in 2004) and the continuing high prevalence of HIV [3]. The high case rate in many African countries has contributed to a rise of the global TB incidence, despite stable or declining rates in the rest of the world [1]. Tanzania with its 37 million inhabitants, has 701 district laboratories diagnosing TB, three laboratories culturing *M. tuberculosis* and one

National reference laboratory that perform drug susceptibility testing of *M. tuberculosis* isolates. Measures are undertaken to establish molecular genotyping methods such as spoligotyping [4], but currently no laboratory in Tanzania offers this service. Previous studies have described the molecular epidemiology of Tanzanian *M. tuberculosis* collections from the first half of the 1990s [5-7]. Spoligotyping is a PCR-based fingerprinting method that detects the presence or absence of 43 defined spacers situated between short direct repeat (DR) sequences in the genomes of members of the *M. tuberculosis* complex. Important advantages of spoligotyping are that it is cheap, easy to perform and fast. In addition, it has been demonstrated that the results are highly reproducible [8]. Unique to spoligotyping results are tools like the SpolDB4 database [9] and the web-based computer algorithm 'Spotclust' [10] that can be used to assign new isolates to families, subfamilies and variants (SpolDB4 only). SpolDB4 is the largest and most up to date available global database for spoligotypes. For previously not reported spoligopatterns, the 'Spotclust' database is a good additional tool in that it can assign these patterns to families by using a computer algorithm based on studies of SpolDB3 [10]. The results from local studies can thus be analyzed and compared to the global *M. tuberculosis* population. This may help us better understand the worldwide spread of common *M. tuberculosis* families and subfamilies. In order to improve our understanding of the TB epidemic in this high-incidence country, the current ongoing study included *M. tuberculosis* strains collected in Dar es Salaam during October and November 2005. We describe the diversity of *M. tuberculosis* isolates from Dar es Salaam, Tanzania, based on spoligotyping, and identify the families and subfamilies responsible for the current persistence and spread of TB in this high-incidence community.

Results

Genetic diversity and family assignment

The 147 analyzed isolates gave 76 different spoligopatterns resulting in an overall diversity of 52%: 57 spoligopatterns occurred only once and 19 patterns comprised 90 of the isolates (61%) (table 1). Forty-nine (64%) patterns had not been described previously. The SpolDB4 database assigns isolates to families, subfamilies and often to variants, whereas 'Spotclust' assigns isolates to families and subfamilies, but is not designed to assign isolates to variants. Four spoligopatterns were assigned to different families and nine patterns were assigned to different subfamilies by the two methods. SpolDB4 assigned names were used whenever a spoligopattern was found in the database, as this database is much larger than the SpolDB3 database, on which the 'Spotclust' algorithm is built. Patterns not found in SpolDB4 were assigned to families and subfamilies by 'Spotclust'. The family assign-

ment showed that 37% of the isolates belonged to the Central Asian (CAS) family, 22% to the Latin American Mediterranean (LAM) family, and 17% to the East-African Indian (EAI) family. These three main families thus accounted for 76% of the incidences in Dar es Salaam. This family assignment also includes the spoligopatterns not described before. Eight isolates lacked spacers 4-7, 10 and 20-35, typical of the CAS1-kili variant, but in addition, they all also lacked spacer 2 (table 2). This spacer is typically present in CAS1-kili lineages and its absence has not previously been reported in these variants. We propose to name these variants CAS1-DAR, since they appear to be abundant in Dar es Salaam.

The rate of diversity (number of spoligotypes divided by the number of isolates) within each main family varied substantially and was 27, 54 and 72% for CAS, LAM and EAI, respectively. This may indicate that the CAS family is best adapted to spread within this community. The diversity of the *M. tuberculosis* population in Dar es Salaam (52%) was comparable to that described in previous studies from Tanzania [5-7]. In Delhi, India the genetic diversity of the *M. tuberculosis* population is 42% [11], but it is only 25% in Harare, Zimbabwe [12]. Thus, the diversity in high-incidence countries varies greatly and may be difficult to estimate without molecular epidemiological studies.

Phylogenetic studies

A Neighbor-joining (NJ) tree of all the isolates is shown in figure 1. The main families were well distinguished and a high diversity within and between families were observed. To confirm the reliability of the NJ tree, the program 'Structure' was applied on the underlying 43-digit binary spacer codes. The open boxes in figure 1 demonstrate the nine groups found to be the most likely number; the NJ branches were supported by the grouping via 'Structure'.

Discussion

The current study demonstrated that most isolates had at least one other closely related isolate in Dar es Salaam. Based on these preliminary findings, the TB epidemic appeared to result from a gradually evolving *M. tuberculosis* population rather than imported strains. A spoligotyping study conducted in the Ouest province of Cameroon found that 193 of 413 *M. tuberculosis* isolates belong to the Cameroon family (LAM10-CAM) [13]. In Harare, Zimbabwe, 68 of 214 isolates are LAM11-ZWE variants [12]. Of the 147 isolates in this study, three and eight isolates belonged to these variants respectively. The scarcity of these strains, abundant in other African countries, also indicated that the TB epidemic in Dar es Salaam is local and well established.

Table 1: Spoligopatterns and family assignment

Spoligotype	Shared type	SpoIDB4 ID	SpotClust ID	SpotClust probability	No. of isolates with identical pattern
	1	Beijing	1		7
	4	LAM3/S	1		1
	1468	LAM11-ZWE		0,99	1
	NEW		LAM8	0,99	1
	964	LAM9		0,97	1
	NEW		LAM9	0,99	1
	NEW		LAM9	0,99	1
	NEW		LAM3	0,99	1
	NEW		LAM8	0,99	1
	NEW		LAM10	0,99	1
	NEW		LAM8	0,99	1
	NEW		LAM8	0,99	2
	150	LAM9		0,99	1
	811	LAM4		0,99	2
	59	LAM11-ZWE		0,99	8
	NEW		LAM9	0,99	1
	1530	LAM9		0,99	1
	42	LAM9		0,99	5
	61	LAM10-CAM		0,99	3
	288	CAS2		1	1
	NEW		CAS	0,99	1
	NEW		CAS	1	1
	1675	CAS1-kili		1	6
	NEW		CAS	1	1
	21	CAS1-kili		1	27
	NEW		CAS	1	1
	22	CAS		1	3
	486	CAS		1	1
	NEW		CAS	1	1
	NEW	CAS1-DAR		1	2
	NEW	CAS1-DAR		1	2
	NEW	CAS1-DAR		1	3
	NEW	CAS1-DAR		1	1
	NEW	CAS1-DAR		1	1
	NEW		EAI4	0,96	1
	NEW		EAI5	0,99	1
	733	EAI5		0,99	1

gopatterns and that the evolution of the region is unidirectional (spacers can be lost but not gained). Also, contiguous blocks of spacers and DRs can be lost in single events [14]. These facts may obscure phylogenetic analyses using simple distance based methods. Despite these weaknesses, spoligotypes have been shown to correlate quite well with single nucleotide polymorphisms (SNP), with the T family, constituting only 10 isolates in this study, as a notable exception. For these reasons a NJ-tree was used to illustrate the current results.

The success of the CAS family in particular, but also the LAM and EAI families in this community is intriguing. The low diversity of the highly prevalent CAS family in this study may indicate that the family is spreading rapidly, but could also reflect a slower evolution of the DR region which could possibly be a result of the missing spacers in the central part of the spoligopatterns of these strains.

The success of these three families suggests a possible co-evolution between specific *M. tuberculosis* families and host population, the molecular basis of which remains to be elucidated. A study conducted in San Francisco supports the idea of co-evolution between this pathogen and host populations [15]. In order to document such possible co-evolution, large populations should be preferred. Internationally standardized methods such as spoligotyping and MIRU-typing, as well as SNP and deligotyping, enable comparison of *M. tuberculosis* genotypes between studies conducted at different times and locations. This facilitates inter-study comparison and helps generate large populations for such evolutionary scenarios. It should be noted that the current study represents a short time period and a small collection of strains. This complicates interpretation of recent transmission and hampers comparisons of genetic diversity with that found in studies conducted over a longer period of time. The use of different genotyping methods also makes direct comparison with previous studies in Tanzania [5-7] difficult.

Recent findings suggest that the tubercle bacillus emerged in Africa and may have spread globally in parallel with the human migrations out of Africa [15,16]. Another study have however identified India as the center for the evolutionary radiation of *M. tuberculosis* [17]. These theories are not mutually exclusive; as the spread to India might represent an early and evolutionary important step in the radiation of *M. tuberculosis* out of Africa. The CAS- and EAI-families which this study found to be abundant in Dar es Salaam, have previously been identified to have the most ancestral roots [17]. We demonstrate that the Beijing family, which is highly prevalent in many Asian locations, is not common in the current population. It therefore appears unlikely that import of strains from Asia have had a major impact on the *M. tuberculosis* population in Dar es

Salaam. The sensitivity of spoligotyping alone is insufficient for pinpointing evolutionary origins and direction of movement, but the current findings lend support to a view of an early African origin of *M. tuberculosis*.

Spoligotyping is inexpensive, fast, simple and reliable. By using this method one can identify outbreaks, support community-based contact tracing, describe the diversity of a *M. tuberculosis* population, and compare this population to that in other parts of the world. Implementation of spoligotyping as a routine method for molecular epidemiological studies of *M. tuberculosis* isolates, appear to represent a valuable investment in many high-incidence countries.

Conclusion

Spoligotyping is very useful to gain an overall understanding of the local TB epidemic. This study demonstrated that the extensive TB epidemic in Dar es Salaam, Tanzania was caused by a few successful *M. tuberculosis* families, dominated by the CAS family. Import of new strains was a minor problem.

Methods

DNA extraction and spoligotyping

Isolates of *M. tuberculosis* were collected from sputum smear positive TB cases in consecutive patients in Dar es Salaam during October and November 2005. Heat-killed samples were shipped to Norway, DNA was extracted [18] and a total of 147 *M. tuberculosis* isolates were spoligotyped according to Kamerbeek et al. [4].

Family assignment

The obtained spoligopatterns were first compared to the SpolDB4 database [9] and assigned to families and subfamilies. Second, in order to assign names to the isolates not found in the SpolDB4 database, the spoligopatterns were analyzed with 'Spotclust'[10], using a mixture model built on the SpolDB3 database. This model takes into account knowledge of the evolution of the DR region and assigns spoligopatterns to families and subfamilies.

Phylogenetic analyses

A NJ-tree [19] was constructed by converting the presence or absence of 43 defined spacers of the 147 isolates into a Jaccard [20] based pair-wise distance matrix with the computer program 'NTSYSpc' (Exeter Software Co., New York). Without conversion to distance, to verify the NJ tree, the spacer data were directly used by the program 'Structure' [21] to identify groups into which the individual isolates fit best and to calculate the best number of groups explaining the whole data set (run with a no-admixture-model, and a burn-in of 100000 repeats and 400000 Markov Chain Monte Carlo repeats, 65%

assigned membership to a group was used as a threshold value in figure 1).

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

VE carried out the DNA extraction, genotyping, data analyses and participated in the design of the study. MM conceived the study and collected, cultured and identified the bacterial isolates. SGMM participated in the design of the study and collected, cultured and identified the bacterial isolates. MH participated in the data analyses and in the design of the study. URD conceived the study, supervised the DNA extraction, genotyping and data analyses. All authors contributed in the writing of the article, read and approved the final manuscript.

Acknowledgements

We acknowledge Jørn Henrik Sønstebo for valuable help with the data analyses and the contributors to the SpoIDB4 database and 'Spotclust'. This study is in part financed by the project "TB in the 21st century – an emerging pandemic" which is headed by Gunnar Bjune and Carol Holm-Hansen and funded by the Research Council of Norway. All participants of this consortium are acknowledged for valuable discussions.

References

1. WHO: **Global tuberculosis control: surveillance, planning, financing. WHO report 2005. WHO/HTM/TB/2005.349.** Geneva, World Health Organization; 2005.
2. **United Republic of Tanzania Ministry of Health: National Tuberculosis and Leprosy Programme. Annual Report.** Dar es Salaam; 2003.
3. Mookherji S WDESWHBA: **Motivating and Enabling Improved Tuberculosis Case Detection in Tanzania: Summary Report.** 2004.
4. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J: **Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology.** *J Clin Microbiol* 1997, **35(4)**:907-914.
5. Yang ZH, Mtoni I, Chonde M, Mwasekaga M, Fuursted K, Askgard DS, Bennedsen J, de Haas PE, van Soolingen D, van Embden JD: **DNA fingerprinting and phenotyping of Mycobacterium tuberculosis isolates from human immunodeficiency virus (HIV)-seropositive and HIV- seronegative patients in Tanzania.** *J Clin Microbiol* 1995, **33(5)**:1064-1069.
6. McHugh TD, Batt SL, Shorten RJ, Gosling RD, Uiso L, Gillespie SH: **Mycobacterium tuberculosis lineage: A naming of the parts.** *Tuberculosis* 2005, **85(3)**:127-136.
7. Gillespie SH, Kennedy N, Ngowi FI, Fomukong NG, Al-Maamary S, Dale JW: **Restriction fragment length polymorphism analysis of Mycobacterium tuberculosis isolated from patients with pulmonary tuberculosis in northern Tanzania.** *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1995, **89(3)**:335-338.
8. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PVM, Martin C, Palittapongarnpim P, Plikaytis BB, Riley LW, Yakrus MA, Musser JM, van Embden JDA: **Comparison of Methods Based on Different Molecular Epidemiological Markers for Typing of Mycobacterium tuberculosis Complex Strains: Interlaboratory Study of Discriminatory Power and Reproducibility.** *J Clin Microbiol* 1999, **37(8)**:2607-2618.
9. Brudey K, Driscoll J, Rigouts L, Prodinger W, Gori A, Al-Hajj S, Allix C, Aristimuno L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans J, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes H, Gutierrez MC, Hawkey P, van Helden P, Kadival G, Kreiswirth B, Kremer K, Kubin M: **Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpoIDB4) for classification, population genetics and epidemiology.** *BMC Microbiology* 2006, **6(1)**:23.
10. Vitol I, Driscoll J, Kreiswirth B, Kurepina N, Bennett KP: **Identifying Mycobacterium tuberculosis complex strain families using spoligotypes.** *Infection, Genetics and Evolution In Press, Corrected Proof.*
11. Singh UB, Suresh N, Bhanu NV, Arora J, Pant H, Sinha S, Aggarwal RC, Singh S, Pande JN, Sola C, Rastogi N, Seth P. UB, Suresh N, Bhanu NV, Arora J, Pant H, Sinha S, Aggarwal RC, Singh S, Pande JN, Sola C, Rastogi N, Seth P: **Predominant tuberculosis spoligotypes, Delhi, India.** *Emerging Infectious Diseases* 2004, **10(6)**:1138-1142.
12. Easterbrook PJ, Gibson A, Murad S, Lamprecht D, Ives N, Ferguson A, Lowe O, Mason P, Nduzo A, Taziwa A, Makombe R, Mbengeranwa L, Sola C, Rostogi N, Drobniewski F: **High Rates of Clustering of Strains Causing Tuberculosis in Harare, Zimbabwe: a Molecular Epidemiological Study.** *J Clin Microbiol* 2004, **42(10)**:4536-4544.
13. Niobe-Eyangoh SN, Kuaban C, Sorlin P, Thonnon J, Vincent V, Gutierrez MC: **Molecular Characteristics of Strains of the Cameroon Family, the Major Group of Mycobacterium tuberculosis in a Country with a High Prevalence of Tuberculosis.** *J Clin Microbiol* 2004, **42(11)**:5029-5035.
14. Warren RM, Streicher EM, Sampson SL, van der Spuy GD, Richardson M, Nguyen D, Behr MA, Victor TC, van Helden PD: **Microevolution of the Direct Repeat Region of Mycobacterium tuberculosis: Implications for Interpretation of Spoligotyping Data.** *J Clin Microbiol* 2002, **40(12)**:4457-4465.
15. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM: **Variable host-pathogen compatibility in Mycobacterium tuberculosis.** *PNAS* 2006, **103(8)**:2869-2873.
16. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omas B, Marmiesse M, Supply P, Vincent V: **Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis.** *PLoS Pathogens* 2005, **1(1)**:e5.
17. Filioli I, Motiwala AS, Cavatore M, Qi W, Hazbon MH, Bobadilla del Valle M, Fyfe J, Garcia-Garcia L, Rastogi N, Sola C, Zozio T, Guerrero MI, Leon CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendon A, Sifuentes-Osornio J, Ponce de Leon A, Cave MD, Fleischmann R, Whittam TS, Alland D: **Global Phylogeny of Mycobacterium tuberculosis Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set.** *J Bacteriol* 2006, **188(2)**:759-772.
18. van Soolingen DHPEWKK: **Restriction fragment length polymorphism (RFLP) typing of mycobacteria.** Bilthoven, National Institute of Public Health and the Environment; 1999.
19. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4(4)**:406-425.
20. Jaccard P: **Nouvelles recherches sur la distribution florale.** *Bulletin de la Société de Vaud Sciences Naturelles* 1908, **44**:223-270.
21. Pritchard JK, Stephens M, Donnelly P: **Inference of Population Structure Using Multilocus Genotype Data.** *Genetics* 2000, **155(2)**:945-959.