Highly structured genetic diversity of the *Mycobacterium tuberculosis* population in Djibouti

S. Godreuil[1,2], F. Renaud[2], M. Choisy[2], J. J. Depina[3,4], E. Garnotel[3,4], M. Morillon[3], P. Van de Perre[1] and A. L. Bañuls[2]

1) *Université Montpellier 1, EA 4205 'Transmission, Pathogenèse et Prévention de l'Infection par le VIH', and CHU Montpellier, Laboratoire de Bactériologie-Virologie Arnaud de Villeneuve, Montpellier,* 2) *GEMI, UMR CNRS-IRD 2724, Centre IRD de Montpellier,* 3) *Laboratoire de Biologie Médicale, Service de Biologie, HIA Laveran, Marseille, France and* 4) *Laboratoire de Biologie, Hôpital Paul Faure, Djibouti Ville, Djibouti*

Abstract

Djibouti is an East African country with a high tuberculosis incidence. This study was conducted over a 2-month period in Djibouti, during which 62 consecutive patients with pulmonary tuberculosis (TB) were included. Genetic characterization of *Mycobacterium tuberculosis*, using mycobacterial interspersed repetitive-unit variable-number tandem-repeat typing and spoligotyping, was performed. The genetic and phylogenetic analysis revealed only three major families (Central Asian, East African Indian and T). The high diversity and linkage disequilibrium within each family suggest a long period of clonal evolution. A Bayesian approach shows that the phylogenetic structure observed in our sample of 62 isolates is very likely to be representative of the phylogenetic structure of the *M. tuberculosis* population in the total number of TB cases.

Keywords: Djibouti, genetic diversity, *Mycobacterium tuberculosis*, population structure, spoligotyping/MIRU-VNTR

Original Submission: 11 April 2009; **Revised Submission:** 26 June 2009; **Accepted:** 3 August 2009
Editor: M. Drancourt

Clin Microbiol Infect

Corresponding author and reprint requests: S. Godreuil, CHU Montpellier, Laboratoire de Bactériologie-Virologie Arnaud de Villeneuve, 371 Avenue du Doyen Gaston Giraud, Montpellier, France
E-mail: godreuil@yahoo.fr

Djibouti is an East African country with a total population of over 819 000. In 2004, the estimated tuberculosis (TB) incidence was 951 cases per 100 000 inhabitants, which is one of the highest incidences in the world [1]. The objectives of this study were to identify the *Mycobacterium tuberculosis* families responsible for the TB cases, and to analyse their genetic diversity and the structure of the *M. tuberculosis* population in an area with this high TB incidence.

The study was conducted over a 2-month period at Paul Faure Hospital in Djibouti City. During this period, 62 consecutive patients with symptomatic disease and sputum culture positive for *M. tuberculosis* complex were included. Spoligotyping [2] and mycobacterial interspersed repetitive-unit variable-number tandem-repeat (MIRU-VNTR) typing [3] was performed with DNA from each isolate. To study the genetic variability, a set of diversity indices, including genotypic diversity and mean genetic diversity (*H*), was evaluated using F-STAT version 2.9.3 [4]. The population structure was explored by analysis of linkage disequilibrium (*LD*) and calculation of *Fst* (index of genetic differentiation between samples) using F-STAT version 2.9.3 [4]. Phylogenetic relationships among the isolates were inferred from spoligotyping and MIRU-VNTR data using UPGMA method and bootstrapping procedures. Tree was built using PAUP 4.0 [5], and TreeDyn software [6] was used for tree visualization and annotation.

The molecular *M. tuberculosis* complex identification methods assigned all 62 isolates to the *M. tuberculosis* complex and to *M. tuberculosis sensu stricto*. Twenty spoligotypes were detected, of which 14 were already known in SpolDB4 [7], and six were undescribed and unique. Three major types were represented: the T family, the Delhi or Central Asian (CAS) family and the East African Indian (EAI) family (Fig. 1). The combined data allowed the generation of 57 distinct patterns, with nine isolates grouped into four clusters (identical genotypes) and 53 isolates with unique patterns (Fig. 1).

All trees built from the different datasets and using different phylogenetic methods clearly distinguished three groups, i.e. the EAI, T and CAS families, sustained by high bootstrap values (>80). The genetic differentiation among these three families was high and significant (EAI vs. T, *Fst* = 0.65; EAI vs. CAS, *Fst* = 0.73; T vs. CAS, *Fst* = 0.72; $p < 0.05$).

An important polymorphism was revealed in this population. The *H* index was not significantly different in each group and in the whole sample ($p \leq 0.05$; 0.34). The genotypic diversity index varied with the group (EAI and CAS = 100%; T = 83%; and 92% for the total sample). Moreover, the *LD* calculated on the basis of MIRU-VNTR data was highly significant in the entire population and in each group ($p = 7.2 \times 10^{-4}$), suggesting, as already proposed for *M. tuberculosis*, clonal and independent propagation within these phylogenetic lineages.

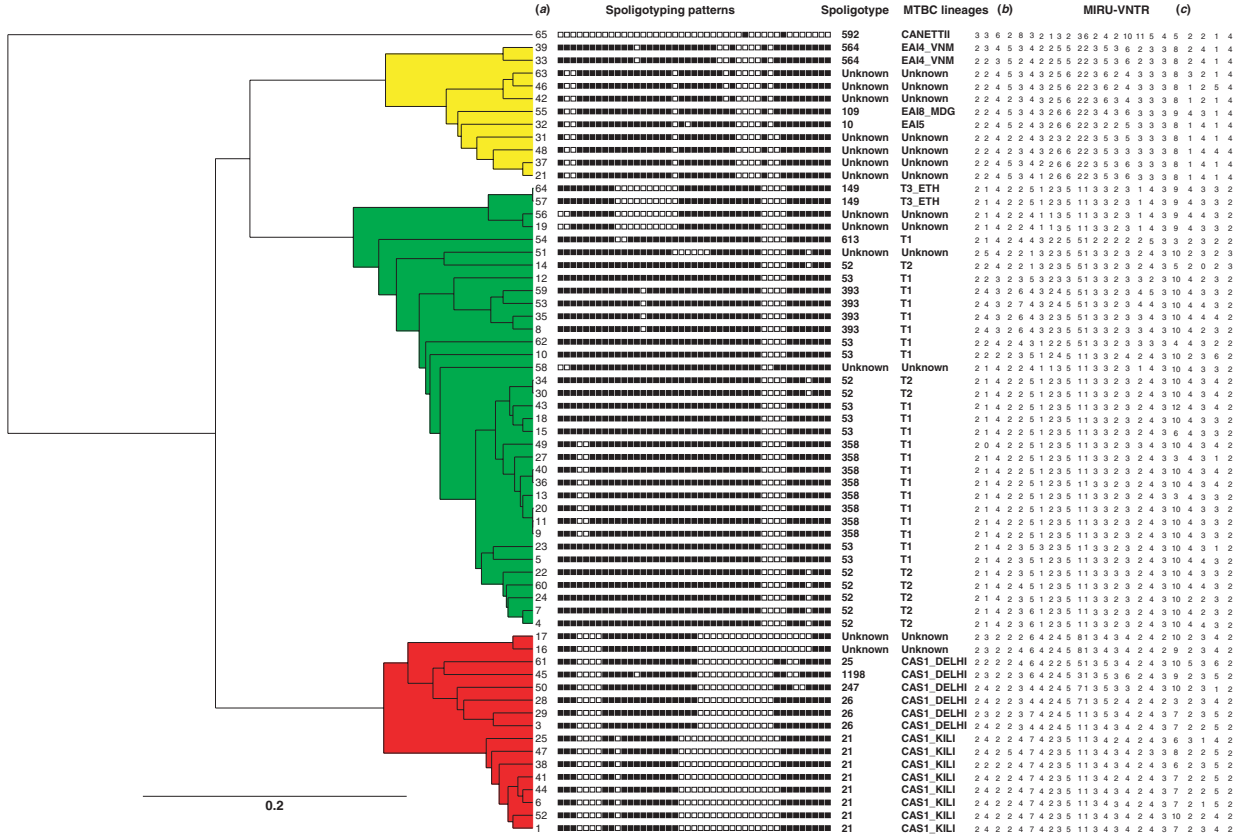

FIG. 1. UPGMA tree based on combined data, spoligotypes and mycobacterial interspersed repetitive-unit variable-number tandem-repeats (MIRU-VNTR) (24 loci) of the 62 samples under study (a *Mycobacterium canettii* stock was used as outgroup (a)). The relationships among patterns were assessed using the UPGMA dendrogram. The spoligotypes listed correspond to the designation in the SpolDB4 database ((b)*Mycobacterium tuberculosis* complex (MTBC) lineages) and lineages were determined using SpolDB4 rules [7]. Unknown patterns match none of the spoligotypes described in the SpolDB4 database. (c) MIRU-VNTR patterns for the 24 loci and for each isolate are displayed in each line. The columns correspond to the MIRU-VNTR loci in the following order: MIRU02, VNTR0424, ETR C, MIRU04, MIRU40, MIRU10, MIRU16, MIRU20, VNTR1955, MIRU23, MIRU24; MIRU26, MIRU27, MIRU31, MIRU39, ETR A, QUB-11b, VNTR2347, VNTR3171, QUB-26, VNTR2401, VNTR3690, VNTR4156, and ETR B.

A major question when dealing with small sample sizes is whether the conclusions reached from the analysis of the sample can be generalized to the total population. In the context of our study, whether the observed phylogenetic structure is an artefact of a small sample size or representative of the total population is a major point to address. Using a Bayesian framework [8,9], we estimated the probability distributions of the proportions of total TB cases caused by each of the three phylogenetic groups observed in Fig. 1 (marked 1, 2 and 3 in Fig. 2) and of total TB cases that were not caused by any of the observed three groups in Fig. 1 (marked 0 in Fig. 2). As can be seen in Fig. 2, the posterior proportions of sites not conforming to our observed phylogenetic structure are very low (mean 0.01474; 95% CI 0–0.19203). The phylogenetic structure observed in our sample of 62 isolates is thus very likely to be representative of the phylogenetic structure in the total number of TB cases.

This study provides the first analysis of *M. tuberculosis* families in Djibouti. As compared with SpolDB4, only three major lineages, CAS, T and EAI, were identified, with no genotype external to these three lineages (Fig. 1). The *M. tuberculosis* molecular studies, in both developed and developing countries, usually reveal few clusters associated with a high variety of genotypes and families without strong structuring [10–13]. This is all the more remarkable in that Djibouti has long been considered to be a cosmopolitan country with important immigration from Asian and African countries. Moreover, other major families (LAM, Haarlem and Beijing) circulate in neighbouring African countries (Ethiopia, Kenya and Sudan) and Saudi Arabia [7,14,15].

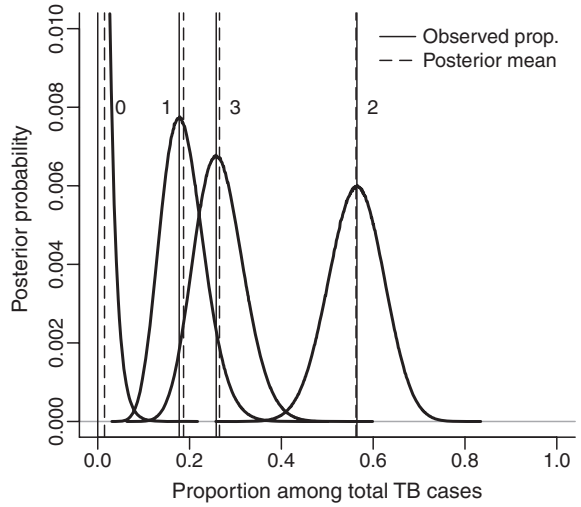

FIG. 2. Distribution of the posterior probabilities of a strain in the total number of tuberculosis (TB) cases belonging to one of the three observed lineages (marked 1, 2 and 3 for the EAI, T and CAS groups of Fig. 1) or none of them (marked 0). The vertical full lines are the observed proportions, and the vertical dashed lines are the posterior means obtained using a uniform prior distribution.

The CAS and EAI families are prevalent in Central Asia and East Asia, respectively [7]. Two hypotheses could explain the presence of these families in Djibouti: (i) the large Pakistani and South Asian communities in Djibouti may have participated in the introduction of these families; or (ii) these families could have emerged from Djibouti and migrated through Asia, a hypothesis that is in agreement with the suggestion that East Africa is the cradle of *M. tuberculosis* complex species [16].

The *LD* observed in the entire sample but also in each group and in each area is in agreement with the clonal structure proposed for *M. tuberculosis*. Thus, each lineage can be considered to be a clone that evolves independently. In countries where the TB incidence is low and therapeutic management is effective, the spread of new genotypes is normally rapidly stopped, and these new strains find no opportunity to propagate and evolve [10,17,18]. This would explain the large variety of genotypes with only a few clusters or lineages observed in these low-TB-incidence countries. In Djibouti, we noted only three lineages, and high genetic diversity within each of them. A high incidence of TB and the difficulties in treating the disease effectively result in a relatively free circulation of genotypes, generating important genetic diversity in each lineage.

The population structure observed in this study, three individualized lineages with high genetic diversity, could reflect evolution over a long period of time and a high transmission level of circulating clones in Djibouti. In conclusion, only three major *M. tuberculosis* families were identified in our patients in Djibouti. The high diversity and the strong *LD* within each family suggest a long period of clonal evolution of the three lineages T, CAS and EAI in Djibouti.

Acknowledgements

The authors acknowledge the assistance of L. Northrup, who edited the manuscript.

Transparency Declaration

We are grateful to the IRD (Institut de Recherche pour le Développement), the CNRS (Centre National de la Recherche Scientifique), the Laboratoire de bactériologie, Hôpital Arnaud-de-Villeneuve, Montpellier France and the Laboratoire de Biologie, Hôpital Paul Faure, Djibouti for financial and technical support. None of the authors has a conflict of interest or any commercial association that may pose a conflict of interest.

References

1. World Health Organization. *Global tuberculosis control—surveillance, planning, financing*. WHO Report, WHO/HTM/TB/2004.331, 2004. Available at: http://www.who.int/tb/publications/global_report/en/.
2. Kamerbeek J, Schouls L, Kolk A *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997; 35: 907–914.
3. Supply P, Allix C, Lesjean S *et al.* Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2006; 44: 4498–4510.
4. Goudet J. Fstat (vers.1.2): a computer program to calculate F-statistics. *J Hered* 1995; 86: 485–486.
5. Swofford DL. PAUP: Phylogenetic analysis using parsimony (and other methods). Sunderland, MA: Sinauer Associates.1998.
6. Chevenet F, Brun C, Bañuls AL, Jacq B, Christen R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 2006; 7: 1–9.
7. Brudey K, Driscoll JR, Rigouts L *et al. Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 2006; 6: 1–17.
8. Hilborn R, Mangel M. eds. *The ecological defective: confronting models with data*. Princeton University Press, New Jersey. 1997.
9. McCarthy MA. Bayesian methods for ecology. Cambridge University Press, Cambridge. 1997.
10. Godreuil S, Torrea G, Terru D *et al.* First molecular epidemiology study of *Mycobacterium tuberculosis* in Burkina Faso. *J Clin Microbiol* 2007; 45: 921–927.
11. Homolka S, Post E, Oberhauser B *et al.* High genetic diversity among *Mycobacterium tuberculosis* complex strains from Sierra Leone. *BMC Microbiol* 2008; 8: 1–8.

12. Tazi L, El Baghdadi J, Lesjean S *et al.* Genetic diversity and population structure of *Mycobacterium tuberculosis* in Casablanca, a Moroccan city with high incidence of tuberculosis. *J Clin Microbiol* 2004; 42: 461–466.
13. Valcheva V, Mokrousov I, Rastogi N, Narvskaya O, Markova N. Molecular characterization of *Mycobacterium tuberculosis* isolates from different regions of Bulgaria. *J Clin Microbiol* 2008; 46: 1014–1018.
14. Al-Hajoj SA, Zozio T, Al-Rabiah F *et al.* First insight into the population structure of *Mycobacterium tuberculosis* in Saudi Arabia. *J Clin Microbiol* 2007; 45: 2467–2473.
15. Githui WA, Jordaan AM, Juma ES *et al.* Identification of MDR-TB Beijing/W and other *Mycobacterium tuberculosis* genotypes in Nairobi, Kenya. *Int J Tuberc Lung Dis* 2004; 8: 352–360.
16. Gutierrez MC, Brisse S, Brosch R *et al.* Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog* 2005; 1: 1–5.
17. Tazi L, Kreiswirth B, Carriere C, Tibayrenc M. Molecular epidemiology of *Mycobacterium tuberculosis* and its relevance to the surveillance and control of TB: an e-debate. *Infect Genet Evol* 2002; 2: 153–158.
18. Tazi L, Reintjes R, Bañuls AL. Tuberculosis transmission in a high incidence area: a retrospective molecular epidemiological study of *Mycobacterium tuberculosis* in Casablanca, Morocco. *Infect Genet Evol* 2007; 7: 636–644.