Genotyping of *Mycobacterium tuberculosis* in China and Missing Links in the Chain of Ongoing Transmission of Tuberculosis

Sandra M. Arend¹ and Dick van Soolingen²,³

¹Department of Infectious Diseases, Leiden University Medical Center, ²Tuberculosis Reference Laboratory, National Institute for Public Health and Environment (RIVM), Bilthoven, and ³Departments of Pulmonary Diseases and Medical Microbiology, Radboud University Nijmegen Medical Center, The Netherlands

(See the Major Article by Yang et al on pages 219–27.)

*Keywords.* tuberculosis; genotyping; variable number of tandem repeats; China; transmission.

Worldwide, the epidemiology of tuberculosis varies widely from an uncommon disease refueled by immigration to highly endemic, with ongoing transmission as the driving force. Countries with the highest incidence rates are found in sub-Saharan Africa, whereas the highest absolute numbers of cases are recorded in India and China. The latter countries experience intermediate incidence rates, but have population sizes of >1 billion and suboptimal healthcare systems.

In China, the yearly number of new tuberculosis cases is estimated at approximately 1 million, the incidence rate being around 70 per 100 000 population per year in 2013 [1]. However, the true magnitude of this problem may be much larger, as the diagnosis of tuberculosis is still mainly based on the insensitive Ziehl–Neelsen microscopy. Notwithstanding the recent figures, China has made important progress in tuberculosis control and prevention during the past 20 years, with a reduction of the prevalence and mortality rates by half [2]. Global tuberculosis control targets set by the World Health Organization and the United Nations’s Millennium Development Goals were successively met through establishment of a nationwide tuberculosis control network, implementation of the directly observed treatment, short-course strategy, and standardization of the diagnosis and treatment. The aims for 2009–2014 were to use innovative technologies to improve diagnosis and treatment, but the implementation of this strategy has proven more complicated than anticipated. Although there is progress in the campaign against tuberculosis in China, the rates of multidrug-resistant (MDR) tuberculosis in particular are alarmingly high.

In the fight against tuberculosis, the natural history of the infection is key (Figure 1). New cases with active tuberculosis either result from early progression (ie, within 2 years after infection, during which 80% of cases occur), or following endogenous reactivation of an infection acquired in the remote past (Figure 1, phases 2 → 3). Early progression relates to recent transmission, whereas reactivation tuberculosis is a late effect of the epidemiology in the past. Knowing the proportion of both determines which interventions can be expected to contribute to control of tuberculosis. This knowledge can be gained by applying DNA fingerprinting, as shown by the important study by Yang and colleagues in this issue of *Clinical Infectious Diseases* [3].

**GENOTYPING OF MYCOBACTERIUM TUBERCULOSIS**

Genotyping of cultured isolates is invaluable to differentiate between late reactivation and recent transmission. Until recently, this was mainly based on IS6110 restriction fragment length polymorphism (RFLP) typing and spoligotyping. Although suited for many settings, these methods lack discriminatory power, especially for Beijing genotype isolates, which constitute the majority of isolates in China and are increasing elsewhere. This limitation was recently overcome by the exploration of the variable number of tandem repeats (VNTR) typing technique, including an additional set of primers of 4 hypervariable loci in Beijing genotype strains, as based on recent research [4, 5]. This paved the way for population-based
genotyping studies in Chinese endemic settings, of which the study by Yang et al is the first of its kind [3]. The authors genotyped 2274 isolates that had been collected at 5 noncontiguous study sites in China; genotypes were analyzed per site and classified as clustered or unique under the assumption that the first isolate in a cluster represented the source (index) patient. Thirty-one percent of isolates were clustered, leading to the conclusion that recent transmission contributes significantly to the tuberculosis burden in China. Other relevant findings were that cluster size was mostly very small; secondary cases were rarely close contacts; Beijing genotype and MDR tuberculosis were associated with clustering; and, unexpectedly, transmission by smear-positive and smear-negative index patients was equally likely. Each of these observations deserves further consideration.

**CLUSTERED CASES**

Regarding the accuracy of the observed proportion of clustered cases, the authors mention that this figure likely underestimates the true value. Apart from the self-evident exclusion of culture-negative tuberculosis cases in a genotyping study, there are other arguments to assume a higher real-life clustering rate than was observed. First, the study relied on passive case finding. As a consequence, patients with symptomatic pulmonary tuberculosis who did not seek medical care were missed. Next, children were excluded, while these cases always represent recent infections and hence clustered cases. Furthermore, the time frame of the study involved only 3 years. Because sources of infection from before the study period are missed, as are secondary cases infected by sources in the study but first diagnosed after the study period, the clustering of cases will have been underestimated. Last, but not least, migration may have played an important role by which cases may have escaped the study, whereas cases resulting from transmission outside the study sites may have been included. In this regard, a specific sociodemographic group in China that needs consideration is that of migrant workers, the number of which is reported to exceed 150 million. These vulnerable individuals, often subject to crowded living conditions, could have been infected and develop active tuberculosis while at work and return to the region where they are registered to seek medical care. In this scenario, the classification as unique isolate would be technically unjust. Therefore, the question could be: Is the observed 31% of clustered cases accurate or just the tip of an iceberg? To know the true value would require further study.

---

**Figure 1.** Overview of the possible phases in the course of pulmonary tuberculosis (TB) and corresponding potential prevention and control measures. Individual patients do not necessarily go through all stages, and spontaneous reversion to an earlier stage (except stage 0) is possible. *The arrows at the bottom are intended as a challenge: Possibly not every mentioned control measure can be directly linked to genotyping. Abbreviations: →→, also applies to next phase; AFB, acid-fast bacilli; DOTS, directly observed therapy, short-course; DST, drug susceptibility testing; HCW, healthcare worker; LTBI, latent tuberculosis; MTB, *Mycobacterium tuberculosis*; VNTR, variable number of tandem repeats.*
with extension of both the time frame and spatial coverage. In addition, more demographic data, including information on migration of the patients, are needed.

CLUSTER ANALYSIS

In the study by Yang and colleagues, 80% of all clusters consisted of just 2 patients; the largest cluster comprised 13 patients. This suggests that transmission relied on many infectious cases, each resulting in a limited number of secondary cases. However, in view of the above-mentioned reasons why clustered cases may have been missed, cluster sizes could have been underestimated as well.

Epidemiological links between clustered cases were rarely found, which could relate to lack of sensitivity of the questionnaire or to missed links in the chain of transmission. Only 6.1% of clustered cases were family members, which is lower than observed in cluster investigations in low-incidence settings, but in agreement with findings in South Africa [6]. This phenomenon may in part be explained by the smaller family size in China compared to other countries. The consequence would be that, if contact investigations are considered, the traditional “stone-in-the-pond” principle of Europe might not apply in China [7].

MULTIDRUG-RESISTANT TUBERCULOSIS

It is estimated that one-third of the world’s MDR tuberculosis cases occur in China, amounting to >60,000 cases yearly. In the study by Yang et al, 6% of isolates were MDR tuberculosis, which is in accordance with the reported rate among new cases in China [1]. The relative rate of transmission of MDR tuberculosis was higher than that of drug-susceptible isolates. This indicates that, in contrast to the past in Europe where resistant bacteria were assumed to have a reduced fitness, the MDR bacteria circulating in China seem to be more transmissible than do susceptible bacteria. Although there are new highly promising drugs for the treatment of MDR tuberculosis, such as bedaquiline and delamanid, transmission of MDR tuberculosis likely occurs mostly before the initiation of treatment, underscoring the need for additional measures other than effective treatment.

BEIJING AS PREDOMINANT GENOTYPE

As expected, most isolates in the study by Yang and colleagues were of the Beijing genotype, with a moderate association with clustering. It is not precisely known why the Beijing genotype is predominant in China. It has been suggested that this genotype represents a vaccine escape mutant [8] and that it induced a differential cytokine response signature in animal models but also in humans [9,10], linked to mutations in regulatory genes [11]. Preferential transmission of particular strains such as the Beijing genotype and/or MDR tuberculosis likely depends on selection of strains with compensatory mutations under pressure of vaccination or antibiotic use. A more detailed study into the genomic aspects of Beijing and MDR tuberculosis strains may shed light on this important issue.

TRANSMISSION BY SMEAR-NEGATIVE CASES

Previous studies from areas of low endemicity consistently showed that transmission by smear-negative patients occurs, but at a several-fold lower rate than for smear-positive cases [12–14]. A gradient of risk of transmission exists, increasing with the number of bacilli in sputum [15]. Therefore, it is rather unexpected that the relative rate of transmission by sputum smear–negative patients in the study by Yang and colleagues did not significantly differ from smear-positive index cases. An explanation of this discrepancy with previous studies could be that the assumption of the first case in a cluster being the index case was not always correct. Even in the simplest scenario with a cluster of 2 patients, either could be the index case depending on who first became infectious, or both could be secondary cases infected by an index case that did not end up in the study population. In real life, transmission of tuberculosis is often complex, more resembling a forked chain and progressing in various directions rather than from a single point source or via a linear chain. To ascertain the actual order and route of transmission, more comprehensive studies would be required, including detailed information on, for example, duration of cough before diagnosis. An alternative and disturbing hypothesis to explain preferential transmission of the Beijing genotype is that many fewer bacteria could be required to establish successful transmission, which may be tested in an animal model.

POSSIBLE INFLUENCE OF THE GENOTYPING METHOD

Although VNTR typing is more discriminatory than RFLP and spoligotyping, the overall variation between Beijing genotype isolates is still lower than among strains of other genotype families, conceivably due to the more clonal structure of the Beijing genotype. In a setting with intensive transmission, the genetic diversity of circulating strains may be too low to serve the molecular epidemiology. This concept deserves thorough study of the population structure and genetic turnover in the Beijing genotype family.

Ultimately, whole-genome sequencing could detect minor genetic variations and accommodate clustered isolates in an exact phylogenetic tree. With ever-lower costs of sequencing, this may be within reach sooner than we think.

HOW CAN GENOTYPING CONTRIBUTE TO TUBERCULOSIS CONTROL?

Genotyping in itself has no direct effect on transmission, but rather serves as a
smoke alarm. The potential of genotyping lies in the disclosure of the natural history of tuberculosis infection. Active case finding among contacts is a relatively straightforward approach in combating pulmonary tuberculosis and can result in timely diagnosis before an infectious stage is reached. Targeted testing for latent tuberculosis is another approach, and what is effective in low-endemic regions is not necessarily transferable to other regions [16]. As the tuberculin skin test is unreliable after BCG vaccination, interferon-γ release assays (IGRAs) are highly specific, but the background rate of positive IGRA results is high in endemic areas (approximately 20%) due to previous exposure [17–19]. IGRAs cannot differentiate between recent and remote infection [20]. Finally, the predictive value of IGRA results for the risk of progression to active tuberculosis is modest due to ineffective drugs and introduction of the BCG vaccine followed improvements in socioeconomic development. Smoking, which was not a risk factor for clustering in the study by Yang and colleagues, does increase the risk of active tuberculosis [25], and smoking is highly prevalent among Chinese men.

In view of the complexity of tuberculosis prevention and control in regions with a high prevalence of latent tuberculosis, the international efforts that are presently directed toward the development of improved vaccines aimed at different stages in the course of tuberculosis (Figure 1, stages 0, 2, and 6) show promise [26, 27].

Conclusions

Improving tuberculosis prevention and control is an international challenge because it is multifaceted and requires integrated actions at several levels. Regarding tuberculosis in China, there is still a battle to win, and the study by Yang and colleagues will hopefully lead to funding of additional studies in China using molecular epidemiology and result in a centralized database allowing investigation of the transmission and population structure of Mycobacterium tuberculosis on a nationwide scale. Ideally, international collaborations will ultimately result in a worldwide database of genomic information on M. tuberculosis, so that effective measures can be taken to counteract the march of this plague of humankind.

Note

Potential conflict of interest. Both authors: No potential conflicts of interest.

Both authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References