Enferm Infecc Microbiol Clin. 2011;**29(Supl 1)**:8-13

Enfermedades Infecciosas y Microbiología Clínica

www.elsevier.es/eimc

Innovations in the molecular epidemiology of tuberculosis

Darío García de Viedma[a,b,*], Igor Mokrousov[c] and Nalin Rastogi[d]

[a]*Servicio de Microbiología, Hospital General Universitario Gregorio Marañón, Madrid, Spain*
[b]*CIBER de Enfermedades Respiratorias (CIBERES), Spain*
[c]*Laboratory of Molecular Microbiology, St. Petersburg Pasteur Institute, St. Petersburg, Russia*
[d]*Tuberculosis and Mycobacteria Unit, WHO Supranational TB Reference Laboratory, Institut Pasteur de Guadeloupe Morne Joliviere, Abymes, Cedex, Guadeloupe, France*

Keywords:
Tuberculosis
Molecular epidemiology
Innovations
Recent transmission
Evolution

ABSTRACT

The application of genotyping tools to the analysis of tuberculosis (TB) has allowed us to identify clinical isolates of *Mycobacterium tuberculosis* to strain level. *M. tuberculosis* fingerprinting has been applied at different levels: *a)* in the laboratory, to optimize identification of cross-contamination events which can lead to a false diagnosis; *b)* in the patient, to determine whether recurrences are due to reactivations or exogenous reinfections or to identify cases coinfected by more than one strain; *c)* at the micropopulation level, to identify clusters of cases infected by the same strains (recent transmission) and to differentiate them from orphan cases that are most probably due to reactivations; and *d)* at the macropopulation level, to define the global distribution of *M. tuberculosis* lineages, to monitor the international spread of high-risk strains, and to explore the evolutionary features of *M. tuberculosis*. In recent years, important methodological and strategic advances have been applied at these different levels of analysis. Rather than provide an exhaustive review, the present study focuses on specific advances in micropopulation and macropopulation analysis.

Innovaciones en la epidemiología molecular de la tuberculosis

Palabras clave:
Tuberculosis
Epidemiología molecular
Innovaciones
Transmisión reciente
Evolución

RESUMEN

La aplicación de estrategias de genotipado al análisis de la tuberculosis (TB) ha permitido discriminar los aislados de *Mycobacterium tuberculosis* a nivel de cepa en distintos contextos: *a)* en el laboratorio, para optimizar eventos de contaminación cruzada; *b)* en el paciente, para discriminar recurrencias debidas a reactivaciones o reinfecciones e identificar casos con infecciones mixtas; *c)* en el contexto "micropoblacional", para identificar casos infectados por una misma cepa (transmisión reciente), y *d)* en el contexto "macropoblacional", para definir la distribución internacional de linajes de *M. tuberculosis*, de cepas de alto riesgo o analizar aspectos evolutivos. En los últimos años hemos asistido a avances metodológicos y analíticos en cada uno de los contextos mencionados. Esta revisión no pretende ofrecer un análisis exhaustivo de éstos, sino destacar algunos avances de especial interés en el contexto del análisis micro y macropoblacional.

Genotyping of *Mycobacterium tuberculosis* at the micropopulation level: analysis of epidemiologically-related populations

An understanding of how *Mycobacterium tuberculosis* is transmitted requires the analysis of well-defined epidemiologically consistent populations based on universal genotyping of *M. tuberculosis* isolates and long-term surveillance.

The way in which *M. tuberculosis* genotyping data are used for epidemiological purposes has changed. Most molecular epidemiology studies of TB have been performed using IS6110-RFLP (restriction fragment length polymorphism), due to its reproducibility, discriminatory power, and low cost. Initially, *M. tuberculosis* genotyping patterns provided a picture of the features of recent transmission in a population. In the present review, this approach is termed descriptive molecular epidemiology, which defines cases in the population infected by the same strain in order to calculate the percentage of clustered tuberculosis (TB) cases (recent transmission events). The traditional application of molecular epidemiology techniques is moving toward real-time *M. tuberculosis* genotyping, which will make it possible to take advantage of cluster data for epidemiological research. We call this approach interventionist epidemiology, namely, an attempt to exercise control over TB transmission.

*Corresponding author.
E-mail: dgviedma2@gmail.com (D. García de Viedma).

0213-005X/$ - see front matter

Descriptive molecular epidemiology

Molecular epidemiology allows us to identify recent transmission events, analyze the risk factors associated with recent transmission, detect non-conventional transmission contexts, identify weaknesses in the assignment of microepidemics, and measure the impact of TB control programs.

In recent years, changes in the socioepidemiological scenarios generated by immigration have revealed new challenges in molecular epidemiology. The key questions include comparison of the role of recent transmission with that of reactivation/importation in immigrant TB cases, the impact of potential importation of previously unidentified *M. tuberculosis* strains, and cross-transmission between cases from different nationalities. Without the support of molecular epidemiology, these questions would prove difficult to answer.

Scenarios in recent transmission and immigration are country-specific. Several studies have found that recent transmission plays a minor role in TB in immigrants,[1,2] thus suggesting that it is mainly due to reactivation/importation. However, recent transmission involving immigrants has been observed, and some authors have identified transmission between cases from different origins and between immigrant and autochthonous populations.[3-5] One recent study applied questionnaires to calculate an "integration-index," and revealed higher values for cases in multinational clusters.[6]

One concern in molecular epidemiology studies is the low frequency with which clusters are confirmed by epidemiological data. Some authors[7] consider that both, standard and molecular epidemiology, analyze independent subsets of cases, and discrepancies between their findings are therefore expected. Alternatively, a more frequent-than-expected role has been proposed for casual contacts, which are not identified by contact tracing. In fact, the difficulties in performing contact tracing in immigrant populations are usually considered responsible for the poor correlation between epidemiological data and molecular data.

Some authors are using novel, innovative approaches to resolve this lack of correlation. One strategy has been to refine the quality and amount of epidemiological data obtained from the cases. Questionnaires compiling detailed information about cases and their social networks[8] increase the limited data obtained using standard contact tracing. The clear improvements brought about by the application of these questionnaires[9] have led to a high correlation between standard and molecular epidemiology. Another approach involves the use of photographs of cases as a tool to establish epidemiological links through visual recognition.[10] When this strategy has been applied, clustered patients recognized more photographs of individuals in their cluster than from outside their cluster. These findings validate the strategy, and its application has made it possible to identify sites of TB transmission that could be open to intervention. Some authors have integrated the use of photographs with the application of detailed questionnaires, which have made it possible to identify transmission contexts that had not been identified by standard epidemiological approaches, thus increasing the percentage of clusters with demonstrated links.[11]

The descriptive application of molecular epidemiology has shown us that a certain proportion of transmission is expected to occur outside households, and this has generated interest in identifying transmission hot spots. In this sense, geographic information systems (GIS) to identify areas where transmission is likely to be occurring seem particularly useful. Cluster data obtained by genotyping have been integrated with those obtained by GIS to determine the spatial distribution of clustered cases in order to identify discrete geographic areas where ongoing transmission is actively occurring. Once hot spots were identified, targeted screening was applied and the number of cases of undiagnosed TB and latent TB infection identified were shown to exceed the number of cases normally revealed by standard screening programs.[12]

Interventionist molecular epidemiology

The refinements developed in descriptive studies have been followed by attempts to improve the speed with which genotyping data are made available, so that researchers can switch from retrospective descriptive designs to prospective real-time interventions. This approach makes it possible to identify ongoing transmission events in real time, to help epidemiologists identify sites of transmission —thus influencing their choices in contact tracing— and, to design efficient intervention strategies to improve TB control.

IS6110-RFLP-based fingerprinting has consolidated the role of molecular epidemiology. However, this technique is labor-intensive and requires well-grown cultures, thus leading to delays in obtaining genotypes. Alternative polymerase chain reaction (PCR)-based strategies can facilitate the switch between descriptive studies and challenging interventional approaches, and several PCR-based genotyping alternatives have been developed in recent years. The most representative techniques were compared several years ago in a multicenter study,[13] which concluded that variable number tandem repeat (VNTR)-based typing was the most robust approach. In parallel, more traditional approaches have also been improved, and amplified fragment length polymorphism (AFLP) including fluorescence targeting has also been used to analyze IS6110 locations.[14] Spoligotyping has evolved towards microbead-based designs and covers a higher number of targets.[15] Application of the multilocus VNTR-based methods mentioned above has been studied, and mycobacterial interspersed repetitive unit (MIRU)-VNTR[16] has been shown to meet new requirements.

Initially, MIRU-VNTR involved a 12-loci set which was considered[17] efficient for epidemiological purposes.[18] However, some authors found it to have limitations,[19,20] although it did have good discriminatory power when applied together with a second-line genotyping method.[20,21] A new improved set of 24 targets with enhanced discriminatory power has been recommended, and a 15-loci subset has been shown to ensure reasonable discrimination.[22,23] In addition, the initial design, which was based on simplex PCR, has been converted to a high-throughput format. Some authors apply high-performance liquid chromatography (HPLC),[24] although most use multiplex fluorescence-labelled PCR and capillary electrophoresis.[17] MIRU-15 and MIRU-24 have been considered an alternative to RFLP, and several studies have found a good correlation with RFLP-based typing data using different population-based approaches and independent settings.[22,25-27] Nevertheless, contradictory data have emerged.[19,28] Analysis of homogeneous lineages, such as the Beijing family, offers poorer results, and some authors have proposed suitable VNTR markers,[29,30] although MIRU-24 recently provided acceptable results in settings where the Beijing family is prevalent.[31] Inclusion of other hypervariable loci in standardized panels has recently been recommended in order to increase the quality of genotyping data.[32] The speed of MIRU-VNTR typing makes the technique suitable for intervention schemes.[25,33]

In the context of intervention, epidemiological resources are generally not sufficient to monitor all recent transmission chains. After identifying the more actively spread strain in a setting, some authors have developed targeted PCR-based genotypic strategies for specific monitoring.[34] Epidemiological control of TB can also be optimized by identifying clusters that are expected to be more robust indicators of recent transmission, as some clusters are never supported by epidemiological data, even in innovative approaches that refine the epidemiological survey. The lack of epidemiological support behind some clusters may be due to the fact that the genotyping tool applied to define clusters has insufficient resolution to determine genetic heterogeneity between the isolates. Alternatively, some epidemiologically unrelated cases could be independently infected by strains that are highly prevalent or endemic, and therefore appear as clustered cases. In this context,

several authors favor second-line genotyping methods to target epidemiological resources at clusters that are confirmed to be genotypically homogeneous by two genotyping tools (and therefore expected to be epidemiologically robust). IS6110-fAFLP has been applied on a selection of RFLP-defined clusters with various degrees of epidemiological support; it split mainly clusters with weak support, thus confirming those that had already been identified.[35] Similarly, MIRU-VNTR splits RFLP-defined clusters lacking epidemiological links and confirms homogeneity in clusters that are epidemiologically robust.[22,25,36,37] In addition to second-line genotyping, other analytical innovations have been applied to evaluate the robustness of RFLP-defined clusters and thus direct epidemiological resources more efficiently towards more epidemiologically significant clusters. Some research lines have determined whether the genotypes defining clusters in a population can also be involved in clusters in unrelated populations, thus minimizing their usefulness as markers of recent transmission. Strains belonging to the Haarlem lineage, which is prevalent in many settings, are preferentially split by MIRU,[22,37] and some RFLP patterns belonging to this lineage are prevalent in unrelated populations.[37]

Future challenges

A novel interventional molecular approach is necessary to reduce genotyping response times and thus enable ongoing transmission events to be analyzed in real time. Research will be aimed at direct genotyping of bacilli in clinical specimens. Innovative approaches in the characterization of resistance mutations and species identification by direct analysis of bacilli in clinical specimens now offer good specificity and sensitivity.[54] Therefore, future approaches should attempt to obtain fingerprints from bacilli in clinical samples before culture.

Application of molecular markers at the macropopulation level

The standard typing methods applied at the micropopulation level —IS6110 RFLP (based on mobile DNA elements), spoligotyping, and MIRU-VNTR (based on repetitive DNA elements)— have high discriminatory power. However, they are not completely suitable at the macropopulation level, because identical fingerprints can emerge in unrelated lineages (homoplasy) as a result of convergent evolution.[38-40] The slower molecular clock of large-sequence polymorphisms (LSPs) and single-nucleotide polymorphisms (SNPs) makes these markers stable enough to fulfill the conditions necessary to define phylogenetic associations unambiguously in *M. tuberculosis*. LSPs in mycobacteria can be detected by comparative whole-genome hybridization using DNA microarrays,[41] whereas SNPs can be identified using currently available in silico comparisons of the multiple whole-genome sequences of *M. tuberculosis*.[42] In view of the high clonality of *M. tuberculosis* species, it is not surprising that the use of different markers produces highly congruent phylogenetic trees consisting of a limited number of large phylogenetic/phylogeographic lineages.[43]

Evolutionary framework of M. tuberculosis

The genetically homogeneous *M. tuberculosis* complex is ecologically very diverse and includes *M. tuberculosis*, *Mycobacterium africanum* and *Mycobacterium canetti* (exclusively human pathogens), *Mycobacterium microti* (a rodent pathogen), and *Mycobacterium bovis* (with a wide host range), as well as *Mycobacterium pinnipedii* (seals) and *Mycobacterium caprae* (goats).

Analysis based on large deletions and other markers has made it possible to propose a new evolutionary scenario for *M. tuberculosis* complex, while a series of deletions have sequentially differentiated *M. africanum*, *M. microti*, and the various subspecies of *M. bovis*.[44,45] The RD1 region that contains the highly immunogenic ESAT 6 family of antigens has been found in both *M. tuberculosis* complex and other mycobacteria, but not in *M. microti* or *M. bovis*.[45,46] In particular, *M. bovis* has undergone many deletions of sequences that are present in *M. tuberculosis*; therefore, the hypothesis that *M. tuberculosis* evolved from *M. bovis* following the domestication of cattle is incorrect. Analysis based on large deletions also helps to elucidate the position of *M. africanum* and subdivide it into *africanum* I (an independent lineage within *M. tuberculosis* complex) and *M. africanum* II (a part of *M. tuberculosis sensu stricto*).

Horizontal gene transfer is virtually absent in *M. tuberculosis*, thus implying that its clonal population structure is presented by genetic families, namely, monophyletic clusters of genetically related strains. These families, or genotypes, originated in well-delimited geographic areas and were usually named according to the geographic, historical, or cultural name of the region/country where they were first isolated.

The low levels of sequence variation in *M. tuberculosis* have long precluded the use of multilocus sequence typing. However, recent advances in mycobacterial genomics show more substantial genetic variation at the whole-genome level. A recent study analyzing polymorphisms in DNA repair, recombination, and replication (3R) genes of a worldwide collection of tubercle bacilli[42] revealed a surprisingly high level of polymorphism for the 3R genes as compared to housekeeping genes. The study also underlined the usefulness of 3R-based trees for future discrimination between *M. tuberculosis* complex phylogenetic groups when more microbial genomes are sequenced. Indeed, suboptimal activity of the 3R genes (reportedly caused by a general negative/purifying selection) is reflected by their relaxed fidelity, which may in turn lead to adaptive variants, some of which will be able to survive. Niemann et al[47] compared the complete genomes of Beijing representatives in a high-incidence region (Karakalpakstan, Uzbekistan). One was drug-susceptible (isolated in 2001) and the other multidrug-resistant (MDR) isolated in 2004. Both isolates shared the same IS6110-RFLP pattern and the same allele at 23 out of 24 MIRU-VNTR loci, although they differed by 130 SNPs and one large deletion. The susceptible isolate had 55 specific SNPs, while the MDR variant had 75 specific SNPs, including resistance-conferring mutations. This finding underlines that an identical genotypic pattern may not denote clonality *sensu stricto*, even when multiple (and independent) genetic markers are used. Differences in genetic diversity using additional markers reveal remote links during earlier transmission events. Additionally, some of the strain-specific SNPs in the MDR isolate might represent mutations compensating for putative fitness effects of resistance-conferring mutations.

Human demography-influenced macropopulation structure of M. tuberculosis

The genetic diversity of *M. tuberculosis* may be linked to human demographic and migratory events; differences in the occurrence of given lineages and sublineages, as well as their local gradients, could be strongly influenced by the historical events affecting the human host.[48-51] Recent estimates[49] based on the application of Bayesian statistics to VNTR allelic diversity suggested that *M. tuberculosis* complex might be 40,000 years old, a figure which coincides with the expansion of "modern" human populations out of Africa.[52] Additionally, the strong and recent demographic spread of nearly all *M. tuberculosis* complex lineages, which coincided with the increase in the world's population over the last two centuries, has been corroborated by a coalescence analysis.[49]

Principal components analysis and its variant, multidimensional scaling (MDS), are widely used in human population genetics to visualize interpopulation relationships based on complex genetic data. The first application of MDS to *M. tuberculosis* VNTR data

highlighted strong geographic specificities of the local clonal variants of *M. tuberculosis* Beijing genotype.[53] The strong affinity observed for Russian strains, even among geographically distant *M. tuberculosis* complex populations, suggests relatively recent propagation of the Beijing strains presumably exacerbated by massive human migrations in 20th century Russia. Nonetheless, some weak and less expected affinities observed for Beijing strains in distant *M. tuberculosis* complex populations (northern Vietnam, South Africa, Beijing, and Hong Kong) are fascinating and should be closely analyzed to elucidate concealed patterns of human migration or as yet unfamiliar epidemiological links between distant regions. It has been suggested that dissemination of the *M. tuberculosis* Beijing genotype to other regions of the world was driven by population movements to Russia during the Middle Ages, or, more recently, to South Africa (since the 17th century) and to Australia (since the 19th century). Their differential dissemination within these areas, on the other hand, has been influenced by climatic factors in addition to demographic factors.[48,53]

Local M. tuberculosis clones and human-microbial co-adaptation

Interplay between human host genetics and microbial burden have led to co-adaptation. In Vietnam, individuals with the T597C allele of the human *TLR-2* gene were more likely to have tuberculosis caused by the East-Asian/Beijing genotype than other individuals.[54] In the Russian Slavic population in Siberia, the -336G allele of *CD209*, the gene encoding DC-SIGN was more common among patients infected with TB caused by Beijing strains than in those infected with non-Beijing strains.[55]

The acquisition of differential pathogenic characteristics by different *M. tuberculosis* complex lineages may lead to locally prevalent clones of the tubercle bacillus, some of which are better adapted to local human populations, such as a specific Beijing subtype in South Africa;[56] others may have evolved in response to selection factors, such as long-term mass BCG vaccination in Vietnam[57] and Tunisia.[58] At the same time, some clones may develop a stable association with a given population leading to non-competitive local circulation. In 2008, Namouchi et al[58] showed that >60% of TB cases were caused by a single genotype in each of the prevalent clades, in contrast to the more clustered ST50/Haarlem, which is predominant in northern Tunisia. The more widespread ST42/LAM (Latin-American-Mediterranean), with a low transmission rate and weak clustering, suggests its stable association with the Tunisian population.[58]

Local specificity of clones can be explained by recent importation and fast dissemination (due to specific pathogenic properties), outbreak conditions, or long-term historical presence in an area. The Beijing genotype is the best known, although it is not exceptional. The heterogeneous genetic family of *M. tuberculosis* LAM has remarkable pathogenic features in settings as geographically distant as Brazil, Russia and Cameroon.[59-61] Examples of the locally predominant but drug-susceptible clonal groups come from geographically diverse areas, both island and continental settings.[30,62]

New technologies and algorithms

Spoligotyping

A new microsphere (bead)-based laser technology (Luminex, Austin, Texas, USA) permitting the identification and quantification of each PCR product was applied to spoligotyping as an alternative to reverse line blot hybridization.[63] This method makes it possible to analyze a sample in less than 15 seconds and has recently been re-evaluated in France by Zhang et al[15], who found perfect agreement with the results of the membrane-based technique.

A novel alternative has recently been developed:[64] automated MALDI-TOF mass spectrometry (MALDI-TOF MS) was adopted for spoligotype detection and replaced the hybridization step with a multiplexed primer extension assay. A homogeneous assay format of PCR and multiplexed primer extension assay followed by MALDI-TOF MS detection on the MassARRAY® system (Sequenom, Inc.) streamlines sample processing by avoiding extensive washing steps and microsphere conjugation.

New algorithms

Sequence evolution models are not appropriate for many non-sequenced-based markers. In particular, adequate treatment of binary spoligotyping data is especially challenging, since the mode of evolution of the DR locus in *M. tuberculosis* is not completely clear. Numerical taxonomy is a simple approach that makes it possible to infer spoligotype-based phylogenies, although it is not very reliable. An interesting method has recently been proposed to visualize relationships between spoligotypes as a "spoligoforest" graph (http://www.emi.unsw.edu.au/spolTools/). This method is based on a statistically tested model showing that changes in the DR locus more frequently involve the loss of a single or a low number of adjacent spacers.

Bionumerics and PAUP (Phylogenetic Analysis Using Parsimony) packages are the most frequently used approaches to analyze VNTR data. More recently, the BURST algorithm (http://eburst.mlst.net), initially implemented for MLST data, has also been applied to infer VNTR-based phylogenies. This algorithm identifies mutually exclusive groups of related genotypes in the population and attempts to identify the founding genotype of each group. It then predicts the pattern of descent from the predicted founding genotype to the other genotypes in the group and displays the output as a radial diagram centered on the predicted founding genotype.

There is still a certain degree of controversy surrounding the treatment of VNTR data. First, mathematical modeling suggested that VNTR loci in *M. tuberculosis* more likely evolve via a single locus change (loss rather than gain); therefore, VNTR alleles should be treated as quantitative variables. Second, currently used programs still consider VNTR alleles as categorical variables, that is, any change is assumed to be equally likely.

Databasing in molecular tuberculosis control

Monitoring and timely reporting of circulating *M. tuberculosis* complex clones are essential in anti-TB strategies in order to pinpoint the subpopulations that are susceptible to a targeted, high-priority response by TB control programs. A suitable molecular marker must be chosen for macropopulation studies; thus LSPs and SNPs are well suited for phylogeographical classification of strains, but not for purely epidemiological purposes, as opposed to IS6110-RFLP and 24-loci MIRUs, which are well adapted for molecular epidemiology. Combined use of spoligotyping and MIRU typing is probably the best available compromise that enables a relatively good insight into the major genotypic lineages of *M. tuberculosis* complex, and makes it possible to efficiently investigate clustered cases to monitor ongoing TB transmission in a given setting.[65] Publicly available databases to compare the ever-increasing amount of genotyping data are necessary.

One of the first databases used for inter-laboratory comparison of IS6110-RFLP patterns was developed, maintained, and housed at the National Institute of Public Health and the Environment (Dutch: Rijksinstituut voor Volksgezondheid en Milieu or simply RIVM), Bilthoven, The Netherlands. This database was widely used under the auspices of a project on new-generation genetic markers for the study of the epidemiology of TB (EU project Q2K2-CT-2000-630). However, for reasons of security and confidentiality, this database was not made publicly available.

Although public MIRU-VNTR databases focusing on a global collection of *M. tuberculosis* complex strains did not become available until 2008, MIRU-VNTR patterns could be compared with those

included in an MLVA database at http://minisatellites.u-psud.fr/MLVAnet. This option has recently been complemented by a freely accessible web-based server (http://www.MIRU-VNTRplus.org) that provides information on geographical origin, drug susceptibility profile, corresponding genetic lineage, IS6110-RFLP, 24-locus MIRU-VNTR, spoligotyping, and SNP and LSP profiles for 186 reference strains.[66] However, it cannot compare genotyping results at a global level, since the information it provides serves to predict lineages for classification purposes rather than to describe the worldwide diversity of *M. tuberculosis* complex genotypes.

Spoligotyping is the backbone of the largest publicly available database, SpolDB4, which, on its release in 2006, described a total of 1939 clustered patterns (shared-types) representing 39,295 strains from 122 countries.[67] It tentatively classified *M. tuberculosis* complex into 62 clades/lineages using a mixed expert-based and bioinformatics approach (http://www.pasteur-guadeloupe.fr:8081/SITVITDemo). Developed and housed at Institut Pasteur de Guadeloupe, SpolDB4 has recently evolved to a proprietary multimarker database (SITVIT2) that contains genotyping information on nearly 75,000 *M. tuberculosis* complex isolates from 160 countries of isolation (including MIRU-VNTRs on about 15,000 isolates). A smaller version of this proprietary database will be released online in 2010 and will contain information on about 62,500 clinical isolates (105 countries of isolation and 153 countries of patient origin) and 3 markers (Spoligotyping, 5-loci ETR, and 12-loci MIRU-VNTR; personal communication from N. Rastogi).

Other databases that are not publicly available include: the database of the University of Zaragoza, Spain (5,694 IS6110-RFLP entries, of which 4,637 are from Spanish isolates);[68] the database of the Public Health Research Institute Tuberculosis Center (http://www.phri.org/programs/program_tbcenter.asp) that gives information on over 17,000 clinical isolates, as well as the Houston and the Centers for Disease Control and Prevention databases.

Future challenges

Recent advances in mycobacterial genomics have shown a more substantial genetic variation at the whole-genome level. Accordingly, whole-genome sequencing may become a tool for routine molecular epidemiology studies if its cost becomes comparable to that of traditional typing techniques.

Whereas several global TB databases collectively contain a large amount of genotypic information, discrepancies between them make information sharing difficult, if not impossible. Thus, a synchronization mechanism must be set up, in order to check similar genotypic patterns across databases and establish a common nomenclature.

Accumulation of new data on the global diversity of different molecular markers, together with application of the refined statistical approaches, should better define a time scale for the evolution of *M. tuberculosis* and its families. A truly quantitative approach to the co-evolution of *M. tuberculosis* and humans that takes into account the host-pathogen relationship has yet to be developed.

Conflict of interest

The authors declare they have not any conflict of interest.

References

1. Dahle UR, Sandven P, Heldal E, Caugant DA. Continued low rates of transmission of *Mycobacterium tuberculosis* in Norway. J Clin Microbiol. 2003;41:2968-73.
2. El Sahly HM, Adams GJ, Soini H, Teeter L, Musser JM, Graviss EA. Epidemiologic differences between United States- and foreign-born tuberculosis patients in Houston, Texas. J Infect Dis. 2001;183:461-8.
3. Diel R, Rusch-Gerdes S, Niemann S. Molecular epidemiology of tuberculosis among immigrants in Hamburg, Germany. J Clin Microbiol. 2004;42:2952-60.
4. Alonso Rodríguez N, Chaves F, Iñigo J, Bouza E, García de Viedma D, Andrés S, et al. Transmission permeability of tuberculosis involving immigrants, revealed by a multicentre analysis of clusters. Clin Microbiol Infect. 2009;15:435-42.
5. Borrell S, Español M, Orcau A, Tudó G, March F, Caylà JA, et al. Tuberculosis transmission patterns among Spanish-born and foreign-born populations in the city of Barcelona. Clin Microbiol Infect.2010;16:568-74.
6. Barniol J, Niemann S, Louis VR, Brodhun B, Dreweck C, Richter E, et al. Transmission dynamics of pulmonary tuberculosis between autochthonous and immigrant subpopulations. BMC Infect Dis. 2009;9:197.
7. Borrell S, Español M, Orcau A, Tudó G, March F, Caylà JA, et al. Factors associated with differences between conventional contact tracing and molecular epidemiology in study of tuberculosis transmission and analysis in the city of Barcelona, Spain. J Clin Microbiol. 2009;47:198-204.
8. Fitzpatrick LK, Hardacker JA, Heirendt W, Agerton T, Streicher A, Melnyk H, et al. A preventable outbreak of tuberculosis investigated through an intricate social network. Clin Infect Dis. 2001;33:1801-6.
9. Van Deutekom H, Hoijng SP, De Haas PE, Langendam MW, Horsman A, Van Soolingen D, et al. Clustered tuberculosis cases: do they represent recent transmission and can they be detected earlier? Am J Respir Crit Care Med. 2004;169:806-10.
10. Weis SE, Pogoda JM, Yang Z, Cave MD, Wallace C, Kelley M, et al. Transmission dynamics of tuberculosis in Tarrant county, Texas. Am J Respir Crit Care Med. 2002;166:36-42.
11. Martínez-Lirola M, Alonso-Rodríguez N, Sánchez ML, Herranz M, Andrés S, Peñafiel T, et al. Advanced survey of tuberculosis transmission in a complex socioepidemiologic scenario with a high proportion of cases in immigrants. Clin Infect Dis. 2008;47:8-14.
12. Moonan PK, Oppong J, Sahbazian B, Singh KP, Sandhu R, Drewyer G, et al. What is the outcome of targeted tuberculosis screening based on universal genotyping and location? Am J Respir Crit Care Med. 2006;174:599-604.
13. Kremer K, Van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, et al. 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;37:2607-18.
14. Thorne N, Evans JT, Smith EG, Hawkey PM, Gharbia S, Arnold C. An IS6110-targeting fluorescent amplified fragment length polymorphism alternative to IS6110 restriction fragment length polymorphism analysis for *Mycobacterium tuberculosis* DNA fingerprinting. Clin Microbiol Infect. 2007;13:964-70.
15. Zhang J, Abadía E, Refregier G, Tafaj S, Boschiroli ML, Guillard B, et al. *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping' with new spacers and a microbead-based hybridization assay. J Med Microbiol. 2010;59:285-94.
16. Mazars E, Lesjean S, Banuls AL, Gilbert M, Vincent V, Gicquel B, et al. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. Proc Natl Acad Sci USA. 2001;98:1901-6.
17. Supply P, Lesjean S, Savine E, Kremer K, Van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol. 2001;39:3563-71.
18. Sun YJ, Bellamy R, Lee AS, Ng ST, Ravindran S, Wong SY, et al. Use of mycobacterial interspersed repetitive unit-variable-number tandem repeat typing to examine genetic diversity of *Mycobacterium tuberculosis* in Singapore. J Clin Microbiol. 2004;42:1986-93.
19. Scott AN, Menzies D, Tannenbaum TN, Thibert L, Kozak R, Joseph L, et al. Sensitivities and specificities of spoligotyping and mycobacterial interspersed repetitive unit-variable-number tandem repeat typing methods for studying molecular epidemiology of tuberculosis. J Clin Microbiol. 2005;43:89-94.
20. García de Viedma D, Alonso Rodríguez N, Andrés S, Martínez-Lirola M, Ruiz Serrano MJ, Bouza E. Evaluation of alternatives to RFLP for the analysis of clustered cases of tuberculosis. Int J Tuberc Lung Dis. 2006;10:454-9.
21. Cowan LS, Diem L, Monson T, Wand P, Temporado D, Oemig TV, et al. Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. J Clin Microbiol. 2005;43:688-95.
22. Oelemann MC, Diel R, Vatin V, Haas W, Rusch-Gerdes S, Locht C, et al. Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol. 2007;45:691-7.
23. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 2006;44:4498-510.
24. Evans JT, Hawkey PM, Smith EG, Boese KA, Warren RE, Hong G. Automated high-throughput mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* strains by a combination of PCR and nondenaturing high-performance liquid chromatography. J Clin Microbiol. 2004;42:4175-80.
25. Alonso-Rodríguez N, Martínez-Lirola M, Sánchez ML, Herranz M, Peñafiel T, Bonillo MC, et al. Prospective universal application of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping to characterize *Mycobacterium tuberculosis* isolates for fast identification of clustered and orphan cases. J Clin Microbiol. 2009;47:2026-32.
26. Allix-Beguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of Mycobacterium tuberculosis. J Clin Microbiol. 2008;46:1398-406.
27. Valcheva V, Mokrousov I, Narvskaya O, Rastogi N, Markova N. Utility of new 24-locus variable-number tandem-repeat typing for discriminating *Mycobacterium tuberculosis* clinical isolates collected in Bulgaria. J Clin Microbiol. 2008;46:3005-11.

28. Blackwood KS, Wolfe JN, Kabani AM. Application of mycobacterial interspersed repetitive unit typing to Manitoba tuberculosis cases: can restriction fragment length polymorphism be forgotten? J Clin Microbiol. 2004;42:5001-6.
29. Mokrousov I, Narvskaya O, Vyazovaya A, Millet J, Otten T, Vishnevsky B, et al. *Mycobacterium tuberculosis* Beijing genotype in Russia: in search of informative variable-number tandem-repeat loci. J Clin Microbiol. 2008;46:3576-84.
30. Millet J, Miyagi-Shiohira C, Yamane N, Sola C, Rastogi N. Assessment of mycobacterial interspersed repetitive unit-QUB markers to further discriminate the Beijing genotype in a population-based study of the genetic diversity of *Mycobacterium tuberculosis* clinical isolates from Okinawa, Ryukyu Islands, Japan. J Clin Microbiol. 2007;45:3606-15.
31. Shamputa IC, Lee J, Allix-Beguec C, Cho EJ, Lee JI, Rajan V, et al. Genetic diversity of *Mycobacterium tuberculosis* isolates from a tertiary care tuberculosis hospital in South Korea. J Clin Microbiol. 2010;48:387-94.
32. Velji P, Nikolayevskyy V, Brown T, Drobniewski F. Discriminatory ability of hypervariable variable number tandem repeat loci in population-based analysis of *Mycobacterium tuberculosis* strains, London, UK. Emerg Infect Dis. 2009;15:1609-16.
33. Hawkey PM, Smith EG, Evans JT, Monk P, Bryan G, Mohamed HH, et al. Mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* compared to IS6110-based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. J Clin Microbiol. 2003;41:3514-20.
34. Ashworth M, Horan KL, Freeman R, Oren E, Narita M, Cangelosi GA. Use of PCR-based *Mycobacterium tuberculosis* genotyping to prioritize tuberculosis outbreak control activities. J Clin Microbiol. 2008;46:856-62.
35. Borrell S, Thorne N, Español M, Mortimer C, Orcau A, Coll P, et al. Comparison of four-colour IS6110-fAFLP with the classic IS6110-RFLP on the ability to detect recent transmission in the city of Barcelona, Spain. Tuberculosis (Edinb). 2009; 89:233-7.
36. Van Deutekom H, Supply P, De Haas PE, Willery E, Hoijng SP, Locht C, et al. Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. J Clin Microbiol. 2005;43:4473-9.
37. Alonso Rodríguez N, Martínez-Lirola M, Chaves F, Iñigo J, Herranz M, Ritacco V, et al. Differences in the robustness of clusters involving the *Mycobacterium tuberculosis* strains most frequently isolated from immigrant cases in Madrid. Clin Microbiol Infect. 2010;16:1544-54.
38. Gutacker MM, Smoot JC, Migliaccio CA, Ricklefs SM, Hua S, Cousins DV, et al. Genome-wide analysis of synonymous single nucleotide polymorphisms in *M. tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics. 2002;162:1533-43.
39. Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. Proc Natl Acad Sci USA. 2004;101:4871-6.
40. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbón MH, Bobadilla del Valle M, et al. 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;188:759-72.
41. 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;103:2869-73.
42. Dos Vultos T, Mestre O, Rauzier J, Golec M, Rastogi N, Rasolofo V, et al. Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. PLoS One. 2008;3:e1538.
43. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect Dis. 2007;7:328-37.
44. 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;99:3684-9.
45. Marmiesse M, Brodin P, Buchrieser C, Gutiérrez C, Simoes N, Vincent V, et al. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. Microbiology. 2004;150:483-96.
46. Mahairas, GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J Bacteriol. 1996;178:1274-82.
47. Niemann S, Köser CU, Gagneux S, Plinke C, Homolka S, Bignell H, et al. Genomic diversity among drug sensitive and multidrug resistant isolates of *Mycobacterium tuberculosis* with identical DNA fingerprints. PLoS One. 2009;4:e7407.
48. Mokrousov I, Ly HM, Otten T, Lan NN, Vyshnevskyi B, Hoffner S, et al. Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. Genome Res. 2005;15:1357-64.
49. 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:e1000160.
50 Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. 2008;6:e311.
51. Mokrousov I, Valcheva V, Sovhozova N, Aldashev A, Rastogi N, Isakova J. Penitentiary population of Mycobacterium tuberculosis in Kyrgyzstan: Exceptionally high prevalence of the Beijing genotype and its Russia-specific subtype. Infect Genet Evol. 2009;9:1400-5.
52. Cavalli-Sforza LL, Menozzi P, Piazza A. The History and Geography of Human Genes. Princeton: Princeton University Press; 1996.
53. Mokrousov I. Genetic geography of *Mycobacterium tuberculosis* Beijing genotype: a multifacet mirror of human history. Infect Genet Evol. 2008;8:777-85.
54. Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT, et al. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. PLoS Pathog. 2008;4:e1000034.
55. Ogarkov OB, Medvedeva TV, Nekipelov OM, Antipina SL, Men'shikov ML. Study of DC-SIGN gene polymorphism in patients infected with *Mycobacterium Tuberculosis* strains of different genotypes in the Irkutsk Region [article in Russian]. Probl Tuberk Bolezn Legk. 2007;(11):37-42.
56. Hanekom M, Van der Spuy GD, Gey Van Pittius NC, McEvoy CR, Ndabambi SL, Victor TC, et al. Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent. J Clin Microbiol. 2007;45:2263-6.
57. Kremer K, Van-der-Werf MJ, Au BK, Anh DD, Kam KM, Van-Doorn HR, et al. Vaccine-induced immunity circumvented by typical *Mycobacterium tuberculosis* Beijing strains. Emerg Infect Dis. 2009;15:335-9.
58. Namouchi A, Karboul A, Mhenni B, Khabouchi N, Haltiti R, Ben Hassine R, et al. Genetic profiling of *Mycobacterium tuberculosis* in Tunisia: predominance and evidence for the establishment of a few genotypes. J Med Microbiol. 2008;57:864-72.
59. Lazzarini LC, Spindola SM, Bang H, Gibson AL, Weisenberg S, Da Silva Carvalho W, et al. RDRio *Mycobacterium tuberculosis* infection is associated with a higher frequency of cavitary pulmonary disease. J Clin Microbiol. 2008;46:2175-83.
60. Dubiley S, Ignatova A, Mukhina T, Nizova A, Blagodatskikh S, Stepanshina V, et al. Molecular epidemiology of tuberculosis in the Tula area, central Russia, before the introduction of the Directly Observed Therapy Strategy. Clin Microbiol Infect. 2010;16:1421-6.
61. Niobe-Eyangoh SN, Kuaban C, Sorlin P, Thonnon J, Vincent V, Gutiérrez 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;42:5029-35.
62. Valcheva V, Mokrousov I, Panaiotov S, Bachiiska E, Zozio T, Sola C, et al. Bulgarian specificity and controversial phylogeography of *Mycobacterium tuberculosis* spoligotype ST125_BGR. FEMS Immunol Med Microbiol. 2010;59:90-9.
63. Cowan LS, Diem L, Brake MC, Crawford JT. Transfer of a *Mycobacterium tuberculosis* genotyping method, Spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. J Clin Microbiol. 2004;42:474-7.
64. Honisch C, Mosko M, Arnold C, Gharbia SE, Diel R, Niemann S. Replacing reverse line-blot hybridization spoligotyping of the *Mycobacterium tuberculosis* complex. J Clin Microbiol. 2010;48:1520-6.
65. Rastogi N, Sola C. Molecular Evolution of the *Mycobacterium tuberculosis* Complex In: Palomino JC, Cardoso Leão S, Ritacco V, editors. Tuberculosis 2007. Available at: http://www.tuberculosistextbook.com/index.htm
66. Allix-Béguec C, Harmsen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol. 2008;46:2692-9.
67. Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, et al. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. 2006;6:23.
68. López-Calleja AI, Gavín P, Lezcano MA, Vitoria MA, Iglesias MJ, Guimbao J, et al. Unsuspected and extensive transmission of a drug-susceptible *Mycobacterium tuberculosis* strain. BMC Pulm Med. 2009;9:3