Journal of Infectious Diseases and Therapy

Oudghiri et al., J Infect Dis Ther 2018, 6:6
DOI: 10.4172/2332-0877.1000386

Review article Open Access

Molecular Epidemiology of Tuberculosis: A Review of Tools and Applications

Amal Oudghiri[1,2], **Imane Chaoui**[1*] **and Mohammed Elmzibri**[1]

[1]*Unité de Recherches Médicales et Biologiques, Département des Sciences du Vivant,Centre National de l'Energie, des Sciences et Techniques Nucléaires, Rabat, Morocco*

[2]*Equipe de recherche biotechnologie et génie des biomolécules, FST de Tanger, Ancienne Route de l'Aéroport, Tanger, Morocco*

[*]**Corresponding author**: Imane Chaoui, Unité de Recherches Médicales et Biologiques, Département des Sciences du Vivant,Centre National de l'Energie, des Sciences et Techniques Nucléaires, Rabat, Morocco, Tel: +212 537 712 031; E-mail: im_chaoui@yahoo.fr

Received date: September 19, 2018; **Accepted date:** October 7, 2018; **Published date:** October 21, 2018

Abstract

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (MTB), is a complex infectious disease driven by multiple biological, socioeconomic and environmental factors. Molecular epidemiology (ME) has contributed extensively to our current knowledge of TB through several studies conducted since the early 1990s. Interestingly, MTB strain typing has been used for various purposes: i) to identify misdiagnosis results from laboratory cross contamination; ii) to differentiate cases of TB relapse from re-infection; iii) to trace ongoing chains of TB transmission; (iv) to detect the relation between drug resistance and a specific genotype; v) to define the global distribution of MTB lineages, to monitor the international spread of MTB strains especially virulent ones, and to explore the evolutionary features of MTB. This review outlines the main methods, concepts and applications of molecular approaches used to gain insight into the epidemiology of TB over the world.

Keywords: Tuberculosis; *Mycobacterium tuberculosis*; Molecular epidemiology; IS6110 RFLP; Spoligotyping; MIRU-VNTR; Whole Genome Sequencing (WGS)

Introduction

Countless millions of people have died from tuberculosis (TB), a chronic infectious disease caused by the *Mycobacterium tuberculosis* (MTB) bacillus. Despite the availability of Diagnostic tools, chemotherapy, vaccination and prophylaxis strategies, the disease is still yet a public health issue [1]. During the last decades, two relevant events have changed the epidemiology of the disease: the spread of the TB/HIV (Human Immunodeficiency Virus) co-infection and the emergence and spread of the multi-drug resistance tuberculosis (MDR-TB) (resistance to Rifampicin (RIF) and Isoniazid (INH)). The latter phenomenon has been generated mainly by the inappropriate management of the anti-tuberculosis drugs. Currently, the World Health Organization (WHO) estimates at least 480 000 MDR-TB cases worldwide, particularly in China, India, South Africa, and in former Soviet Union countries [2]. In 2006, a bleaker picture has emerged for TB control programs with the discovery of extremely drug resistant TB (XDR-TB) strains. These strains are not only MDR but are also resistant to the second line drugs (Fluoroquinolones and at least one of the three injectible drugs: Kanamycin, Amikacin or Capreomycin) [3,4].

Unfortunately, more severe forms of drug resistant TB emerged, named totally drug-resistant tuberculosis (TDR-TB) which refers to MTB clinical strains that show in vitro resistance to all first- and second-line drugs tested. Indeed, MTB acquires drug resistance mutations in a sequential fashion under suboptimal drug pressure due to monotherapy, inadequate or incomplete treatments and drug interactions [5-7].

From an evolutionary window, the global TB epidemic presents a dynamic picture fueled by the most successful genotypes. Selection of the MTB genotypes is monitored by Pathogen-related factors mainly the degree of transmissibility from person to person, the pathogenicity, the level of protection afforded by BCG vaccination and the ability to acquire drug resistance phenotype [8].

Recent advances in molecular biology have vastly increased the possibilities for studying the epidemiology of TB. Within this context, genotyping is widely used to track specific MTB isolates in a community. It has been successfully used in epidemiologic research "molecular epidemiology" to study the transmission dynamics of TB [9]. Molecular epidemiology (ME) is a combination of both molecular biology and epidemiology, which involves the study of distribution of the diseases in human populations, identified at the molecular level [10]. It is a powerful tool for monitoring infectious diseases such as TB, where patients infected with a given strain may undergo relapse due to reactivation of the same strain or a different strain after cure [11-12]. ME has revolutionized our understanding of the pathogenesis of TB and has also provided unique insights into the international dissemination of TB through geographic comparison and evolutionary analysis of widespread MTB populations [13].

In this review, we summarize the main concepts and methods for genotyping MTBC strains and discuss the relevance of genotyping to the control and understanding the transmission dynamics of TB.

Current Methods and Genetic Markers used for the Molecular Epidemiological Studies

The availability of multiple and polymorphic molecular markers within the Mycobacterium The availability of multiple and polymorphic molecular markers within the Mycobacterium tuberculosis complex (MTBC) genomes has permitted to "zoom in" to detect transmission chains of TB (epidemiological studies) and to

"zoom out" to track regional and global spread of MTB strains (phylogenetic studies) [14]. Ideal molecular marker must meet the following criteria (i) sufficient variability to differentiate unrelated cases (ii) satisfactory clonal stability to trace accurately transmission chains, (iii) robustness to be applicable to a wide range of strains [15-16]. Genetic diversity of MTBC is by far much higher than previously anticipated. Insertions/deletions/duplications and single nucleotide polymorphisms (SNPs) are the driving force of MTB genomic variance and many of them are likely to impact the pathobiological phenotype [17-18].

Actually, a plethora of molecular methods are available to measure the genetic relationship between MTB strains [19]. Each method yields a genetic profile termed "fingerprint" or "genotype" that is strain-specific. These classical genotyping tools include IS6110-RFLP [20], CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats)-based spoligotyping [21], MIRU-VNTRs (Mycobacterial Interspersed Repetitive Unit-Variable Number of Tandem Repeats) [15,22] and RD-LSPs (Regions of Differences-Large Sequence Polymorphisms)-based deligotyping [23].

IS6110-Restriction Fragment Length Polymorphism

Earlier TB molecular epidemiological investigations were based on the IS6110 Restriction Fragment Length Polymorphism (RFLP) analysis. This technique was long recognized as the gold standard for MTBC strain differentiation because of its highest discriminatory power compared to all available MTBC genotyping methods [20,24].

The IS6110 sequence belongs to IS3 family of mobile elements and is a 1355 bp long, found only in MTBC [19]. The genotyping method is based on the variability of the number of copies of IS6110 within MTBC genome. Briefly, the method includes a restriction enzyme (PvuII) that cleaves the IS6110 element once yielding DNA fragments between 0.9 and 10 kilobases which are separated by gel electrophoresis, then transferred onto a membrane and and hybridized with a peroxydase-labelled IS6110 specific probe. As a result, the visualized fragments represent each a single copy of IS6110 surrounded in length flanking DNA. The analysis of the resulting IS6110-RFLP patterns is performed by software applications which allow the intra- and interlaboratory comparisons of patterns and the establishment of huge national and international databases [25-26]. However, the IS6110-RFLP typing is poorly discriminative for strains with less than five copy number of IS6110 named Low Copy Clades "LCC" and for Non-Tuberculosis Mycobacteria (NTM) harbouring multiple copies of homologous sequences to IS6110 [27].

Although it still constitutes one of the most reliable genotyping methods, the complexity of this cumbersome method along with the absence of IS6110 element in some geographic variants of MTBC led to its replacement by PCR-based methods, which are easier to perform and requiring a few bacteria as targets for typing and have the advantage that, in principle, few bacteria are sufficient as targets for typing. These methods include mainly spoligotyping and MIRU typing which have together been recently defined as the new gold standard for molecular epidemiological investigations of TB [28,29].

Spoligotyping

Direct Repeats (DR) loci are members of a universal family of sequences, designated as clustered regularly Interspaced Short Palindromic Repeats CRISPR, whose physiological role is poorly known [35]. It is postulated that these elements are reminiscent of centromer like structure with a possible role in replication partitioning [31]. The DR locus, initially identified by Hermans et al (1991) in the M. bovis BCGP3 strain [32], contains 10 to 50 copies of a 36-bp identical Direct Repeats (DRs), which are separated by spacers having different sequences. Of particular interest, the spacer sequences between any two specific DRs are conserved among strains. Hence, the difference between strains is generated by the presence or absence of specific spacers yielding a Spoligotype through the application of spacer oligonucleotide typing (spoligotyping) method [21].

The Original form of spoligotyping developed by Kamerbeek et al (1997) [21], comprises a PCR-based reverse line blot hybridization method that targets the 43 spacers of DR locus.

As an alternative, the standard spoligotyping assay was transferred to a luminex multianalyte profiling platform. The Microbead based-spoligotyping technique consists on the immobilization of the synthetic spacers' probes on microspheres by covalent coupling and the detection is achieved via fluorochromes attached to the beads and hybridized PCR product [33]. Such high throughput technology provides greater flexibility, robustness and cost effectiveness than the classical spoligotyping. Later on, concerns have been raised regarding suboptimal spoligotyping. In this regard, the 68- spacer format, with 25 out of 51 new spacers, improved the discrimination for Principal Genetic Group1 (PGG1) MTB isolates, namely the East African-Indian (EAI) clade of MTB [34].

In general, Spoligotyping has many advantages compared to IS6110-based genotyping: (i) small amounts of DNA are required (10 fg of chromosomal DNA), (ii) can be performed directly on clinical samples or on MTB strains shortly after their inoculation into liquid culture, which means a "real-time" MTB genotyping, (iii) typing even non-viable cultures, Zhiel-Neelsen smear slides, or paraffin-embedded tissue sections [14,35] and (iv) the results of spoligotyping are presented in a binary format (positive or negative of each spacer).

However, Spoligotyping presents several drawbacks: (i) limited discriminatory power since the method targets only a single genetic locus, covering less than 0.1% of the MTBC genome [16], (ii) a lower level of discrimination than the IS6110 RFLP for strains harboring more than five copies IS6110, (iii) a tendency to overestimate the proportion of clustered strains in case of highly homogeneous MTB population structure [14,36] (iv) inability to detect contaminations and multi-strain infections when the technique is performed directly on clinical samples.

The MTBC is made up of a group of closely related species: M. africanum, M. bovis, M. caprae, M. tuberculosis, M. microti, M. canettii and M. pinnipedi. Spoligotyping allows identification and strain differentiation of MTBC isolates at the (sub) species level given that the presence or the absence of certain spacer sequences acts as a signature for presumptive species identification. Over the years, the signatures given by the 43 spacer-spoligotyping patterns contributed to nominate circulating MTBC genotypes, to classify them into well defined strain families collections and to release huge databases and web services of spoligotyping patterns: SpolDB4 [30], SpolTools [37], MIRU-VNTRplus [38], and SITVIT [39].

Mycobcterial Interspersed Repetitive Units-Variable Number of Tandem Repeat Typing (MIRU-VNTR)

Variable numbers of tandem repeats (VNTRs) are markers that provide data in a simple and format based structure on the number of

repetitive sequences in polymorphic micro- and mini-satellite regions [40]. Mycobacterial interspersed repetitive units (MIRUs) are a VNTR introduced for MTBC by Supply et al [41]. Of the 41 different MIRU loci, 12 loci were initially identified as hypervariable repetitive units [42].

The typing method relies on PCR amplification of 12 loci using primers specific for the flanking regions of each repeat locus, the sizes of the amplicons reflect the numbers of copies of the targeted MIRUs-VNTR repeat units. The results yielded 12-digit number corresponding to the number of repeats at each MIRU locus which facilitates interlaboratory comparaison and exchange. There is a technical difficulty of sizing the multiple small PCR fragments, which is now overcome by combining multiplex PCR with a fluorescence-based DNA analyzer. MIRUs-VNTR is considerably faster than IS6110-RFLP typing, is applicable to crude DNA extracts from early mycobacterial cultures and has been adapted to high throughput conditions [22-43]. Interestingly, a global epidemiological database is available online (SITVIT database and MIRU-VNTRplus) to assess into the distribution (epidemiological analysis) and evolution of MTBC (phylogenetic analysis) [38-39].

In order to exchange MIRUs-VNTR data, MIRU-VNTRplus web service provides an expanding nomenclature that assigns a numerical code to MIRUs-VNTR patterns named "MLVA MTB15-9" which is a juxtaposition of two subtypes MTB15 (set of the most discriminatory loci) and MTBC9 (set of ancillary loci) [38].

The discriminatory power of MIRUs-VNTR analysis is typically proportional to the number of loci evaluated. Interestingly, when more than 12 loci are used, or MIRU analysis is combined with spoligotyping, the discriminatory power approximates that of IS6110 RFLP. An optimized set of 24 MIRUs-VNTR loci, including a subset of 15 discriminatory loci using the Genoscreen MIRU typing kit (Genoscreen, Lille, France) has been proposed to be used as a first-line typing method and has substantially improved the discrimination of MTBC isolates compared to the original 12-locus set [15].

Nowadays, MIRUs-VNTR typing is considered to be the new reference standard for molecular epidemiological studies given its satisfactory discriminatory power and the digitalization of data generated [16].

However, standard 24-locus MIRUs VNTR typing lacks resolution power for discriminating closely related clonal MTB groups, particularly MTB Beijing lineage. For instance, a consensus set of for novel MIRUs-VNTR loci termed hyper variable loci i.e. 1982 (alias QUB-18); 3232 (alias QUB-3232); 3820; 4120 has been suggested for subtyping clustered samples belonging to highly homogeneous Beijing lineage in order to improve the true cluster definition and to achieve a high typeability [44].

Nevertheless, scarce informations are available on the possible contribution of these markers to the improvement of discrimination among non-Beijing MTB lineages. Trovato et al (2016) assessed the epidemiological value of the inclusion of these hypervariable MIRUs-VNTR loci for subtyping MTB strains already clustered by a conventional genotyping strategy (spoligo-24-locus MIRUs-VNTR typing strategy) regardless of their lineage [45]. The results are promising.

Next Generation Sequencing

Up to date, various strain typing methods have been established and successfully applied during the two last decades in molecular epidemiological studies to answer a wide variety of research questions discussed below. These typing methods have an intrinsically restricted discriminatory power as they rely on polymorphic genetic markers, but they interrogate less than 1% of the genome. Hence, they are not able to fully detect and resolve recent transmission chains, they often fail to distinguish between genetically closely related strains and the turnover of these markers is variable. These limitations could be overcome by the application of high-resolution molecular typing tools such as Next Generation Sequencing (NGS) or Second Generation Sequencing (SGS) for whole genome-based epidemiology.

Whole Genome Sequencing using NGS has yielded comprehensive genetic information including genomic markers, drug resistance profile, virulence determinants, and genome evolution. It has been explored using several platforms [46-47]. The process includes the fragmentation and the amplification of the genome on a solid support to increase the signal emission. During sequencing, the signal generated by the incorporation of a nucleotide/ oligonucleotide is read in real time. SGS sequencers read out short sequences ranging from 150 bp to 800 bp. Finally, these generated sequences are further either mapped to an existing template or assembled de novo based on the overlapping regions [48].

WGS technology is increasingly affordable; it offers (i) the highest possible resolution for large TB outbreaks [49] (ii) an efficient tool to track the dynamics of TB transmission in human populations without prior recourse to epidemiological data [50]. Nevertheless, the cost of WGS is still high and the standardization as well as quality assurance programs are yet to be established.

Recently, an all-in-one web-based tool for genotyping MTBC named "Total Genotyping Solution for TB (TGS-TB)" was developed. It incorporates multiple genotyping platforms namely spoligotyping, detection of phylogenies with single nucleotide variations, IS6110 insertion sites, and 43 customized loci for VNTR [51]. TGS-TB is more accurate and has the greatest discriminatory power than all genotyping method taken separately.

All molecular methods discussed above and used to study the molecular epidemiology of TB are summarized in table 1.

Genotyping method	Principle	Advantages	Disadvantages
IS6110 RFLP typing	Based on the copy number of IS6110 (0 to 25) that depend on the frequency of transposition conditioned by the nature of targeted genomic region in which transposition occurs.	- Considered as gold standard method. - Detection of cross-contamination and mixed infection. - The stability of its profiles over time, allowing distinguishing epidemiologically related from unrelated isolates.	- Requires subculturing to get large amounts of high quality DNA. - Laborious (restriction, gel electrophoresis, blotting, hybridization, and exposure of X-ray film). - Time consuming.

		- Very appropriate element to study TB transmission owing to its Biological clock referring to biomarker stability. - Method used for molecular epidemiology, evolutionary and phylogeny studies.	- May provide poor resolution in electrophoretic separation of fragments. - Sophisticated and expensive computer software are needed. - Interlaboratory comparative analysis RFLP patterns is difficult. - A high discriminatory power is obtained only for strains with more than 6 copy number of IS6110 element.
Spoligotyping	Based on the variability of 43 spacers within the direct repeat locus in the genome.	- Performed directly on cell lysate. - High discriminatory power for strains with low (<6) copy number of IS6110 elements. - Genotyping data are presented in binary format making them easy and simple to interpret. -Comparison Data inter and intra laboratory for analysis using available databases.	- Poor discriminatory power - Need to be combined with MIRU-VNTRs method especially in studies including predominant or endemic strains. - Cannot detect mixed infections.
MIRUs-VNTR	Based on 10- to 100-bp of hypervariable sequences often found in tandem, dispersed in intergenic regions in the genome.	- Requires by far less DNA because it is based on DNA amplification. - Performed on cell lysate. - Less time-consuming. - Automated. - Data exchange for an inter or intra laboratory comparison is possible. - High-throughput method compared to IS6110-RFLP typing and spoligotyping when using 24 loci.	-The 12 loci based method is less discriminatory than IS6110-RFLP typing. -Sequencing based method is expensive
NGS	Based on capillary electrophoresis with highly parallelized through-put sequencing (enabling to sequence thousands to millions of short reads of the entire genome)	- Ultimate level of discrimination. - Less time-consuming compared to all techniques. - Cost effective.	- Requires sophisticates bioinformatics systems, fast data software processing and large storage capabilities to deliver data to analytics tools. - The use of short sequencing reads may alter the results which requires integrative approach to all relevant information of the studied genome locus (presence of CRISPR spacers, presence of interfering IS6110 insertions)

Table 1: Comparison of methods widely used to study the molecular epidemiology of TB.

Genetic markers for phylogenetic analysis

Phylogeny and population genetic analyses rely on the robustness of the molecular markers, exhibiting low homoplasy and minimal rate of convergent evolution. Interestingly, genomic deletions namely RDs or LSPs have been used as markers to classify groups of MTBC strains into phylogenetic lineages and sublineages [23,52-54]

MTBC comprises seven human adapted lineages (Lineages 1 to 7) which show a strong phylogeographical population structure, with the difference lineage associated with distint geographic regions : Indo-Oceanic (lineage 1), East-Asian including Beijing (lineage 2), East-African-Indian (lineage 3), Euro-American (lineage 4), West Africa 1 or M. africanum I (lineage 5), and West Africa 2 or *M. africanum II* (lineage 6). Furthermore, a new lineage referred to as lineage 7 was confined to Ethiopia and recent immigrants from this country [55].

Clades of MTBC are assigned as “ancestral”, “modern” or “evolutionary-recent” lineages based on the presence or absence of specific deletion TbD1 [52], whereas the determination of the Principal Genetic Groups (PGGs) is based on KatG-gyrA polymorphism [56]. Briefly, MTBC can be divided into “ancestral” TbD1-positive and “modern” TbD1-negative strains. The TbD1-positive strains are invariably classified as PGG1 upon KatG-gyrA polymorphism, as opposed to “modern” TbD1 negative strains that may belong to the one of the three PGG subgroups. The PGG2 and PGG3 subgroups are also termed as evolutionary recent (Latino American and Mediterranean [LAM], Haarlem, [X] for “Anglo Saxon and [T] for “ill defined”) as opposed to the PGG1 subgroup (Beijing, East-African Indian [EAI], Central Asian [CAS], and *Africanum*); however, only EAI and *M. africanum* are classified as “ancestral” sensu strict [52,57].

Population Structure and Clinical Relevance

MTBC has clonal population structure, and some genetic families that were initially endemic within specific geographical areas have become ubiquitous e.g. Beijing family/genotype, the latter was initially identified in the Beijing area of China and now found on all continents with predominance in East Asia, Northern Eurasia, and Southern Africa, and thus may be defined as endemic, epidemic, or sporadic [30,58]. In addition to the lineage/family associated strain properties, it has been recognized that some of the clonal clusters termed “sub

lineages" within the same family may have a greater capacity to transmit than others. In particular, while the Beijing genotype is frequently associated with drug resistance and hypervirulence, some of its variants demonstrate even more remarkable pathogenic properties, especially in settings with high burden of drug resistant TB [59].

Molecular Epidemiology and Mixed Infections

MTB infection has been recognized for long as clonally simple. However, the introduction of molecular biology tools into the clinical microbiology laboratory revealed that this infection is sometimes clonally complex. Although clinicians in the past assumed that clonal complexity did not interfere with the management of TB, differences in drug-susceptibility patterns of bacterial populations within a same patient may alter the treatment outcome [60]. Within this context, fingerprinting of clinical cultures has revealed (i) recurrence due to exogenous reinfection more than previously thought, (ii) simultaneous co infection with two or more MTB strains, (iii) microevolution phenomena and iv) compartmentalization of the infection, with different strains infecting either different tissues, or independent lung areas.

Another issue is the occurrence of mixed infections with both a drug sensitive and drug resistant strain which might influence the diagnosis of drug resistance and the treatment outcome as well as the selection of drug resistant strains leading to drug resistant TB. Fortunately, the use of MIRU-VNTR minisatellites has provided an efficient molecular tool to illustrate that infection with MTB may involve more than one strain [61]. Nonetheless, there is a difference between polyclonal infections which involves clonal complexes of highly related parental strain(s) due to micro-evolutionary events [60] versus mixed-infections that involve two or more genetically unrelated strains [59]. Generally, infections are categorized as mixed-strain infections when more than one allele at more than one MIRU-VNTR locus is identified, and as polyclonal when there is more than one allele at a single locus [36,60].In a mixed infection, the spoligotype pattern may either reflect the cumulative presence of spacers of two or more strains, or may report the spoligotype pattern of the dominant strain [14,62,63].

Application of Molecular Strain Typing

Genotyping has become an indispensable tool in medical microbiology and epidemiology and one of the first targets has been MTB. Over the past 20 years, more than 1000 pertinent publications have substantiated the value of the genotyping approach for short-term (local epidemiological) and long term (global epidemiological) investigations for tuberculosis control programs.

Molecular strain typing for public health

From molecular window, studies of MTB transmission rely on the postulate that strains exhibiting the same DNA fingerprint belong to the same cluster [60-65]. However, the proportion of clustered strains in a given study could be under or overestimated depending on the completeness of the sample and thus may affect recent transmission of TB in a given setting. It is noteworthy to correlate between the molecular method used, the study question and the population in mind to gain insight into transmission dynamics of TB; e.g. in regions where the genetic diversity of MTB population is limited, e.g. W-Beijing East Asia, spoligotyping will overestimate the proportion of clustered isolates [26]. Hence, a single/particular approach to the analysis of genotyping data is obsolete and may obscures the picture of recent transmission versus distant transmission, thereby necessitating the use of a combination of genotyping methods/specific method in a specific area to glean insight into transmission dynamics of TB.

Molecular strain typing for patient management

The applications of typing methods have permitted to recognize nosocomial infections and Laboratory cross contamination to differentiate a true outbreak of TB from a pseudo outbreak based on false positive MTB laboratory cultures [61], as well as to distinguish exogenous infection from endogenous reactivation.

Selective vs. Universal genotyping

Universal genotyping presents several advantages compared to selective one as it allows (i) earlier identification of false positive MTB cultures, (ii) detection of unsuspected cases/chains of TB transmission, and (iii) monitoring the success of TB control programs in a given setting through the establishment of huge strain diversity databases. Of note, universal genotyping has been implemented in some countries. The largest genotyping program currently in operation was developed by the Centers for Disease Control and Prevention USA [14]. By contrast, selective genotyping is performed, under request, weeks or months after the specimen reception. In these circumstances, archival storage of MTB specimens is a critical criterion in genotyping accomplishment [14].

Interpretation of Molecular Epidemiological Data

Since the early 1990s, global strain genotyping have been widely used to study the molecular epidemiology of TB, However, it has several limitations especially in settings with endemic or LCC strains or in settings where the genetic diversity of MTB is known to be low [66-68]. Indeed, molecular epidemiologic studies (i) often have a hierarchical data structure (e.g., Individual clustered cases occurring within households); (ii) may be affected by sampling bias [9]. (ii) may suffer from strain misclassification (clustered versus unique cases) and (iv) are strongly affected by the lengthy of study duration which lead to increase of cluster proportion, and thus overestimation of recent transmission rate in a given setting. To overcome this limitation, the "Cluster window" (usually 1-2 years) was established; it refers to the maximum allowable time between two matching strains in molecular epidemiologic studies, and is a critical parameter to establish the true chains of TB transmission [9]. In addition, multiples studies reported strains with small variations detected by one or multiple genotyping methods and were concluded to be clonal taking into account the clinical and epidemiological data [9]. These small variations result from microevolutionary events within MTB genome occurring during transmission or during dissemination within an individual patient. Thus, clinical and epidemiological data are critical to establish molecular fingerprints.

Conclusions

The classical epidemiology of tuberculosis was revolutionized by molecular tools; the latter have permitted to understand the complex transmission dynamics within populations and between hosts. However, molecular studies were mostly conducted in resource rich areas with low incidences of TB. Therefore, their applications are limited. Further steps are (i) to conduct molecular epidemiology studies in countries with highest burden of disease to better

understand the transmission dynamics of TB, (ii) to combine species determination, drug resistance testing, detection of pathogenicity factors and fingerprinting in a single assay. Fortunately, new opportunities have emerged with rapid technical developments especially next generation sequencing making possible to view the complete genetic information of the bacteria, which should improve the accuracy of efforts to monitor strains of MTB as they move through space and time.

References

1. Kaufmann SHE, Schaible UE (2005) 100th anniversary of Robert Koch's Nobel Prize for the discovery of the tubercle bacillus. Trends Microbiol 13: 469-475.
2. World Health Organization, 20th edition Global tuberculosis report. 2015 WHO, Geneva, Switzerland.
3. Gandhi NR, Moll A, Sturm AW (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet 368: 1575-1580.
4. Wright ABG, Barrera L, Boulahbal F (2006) Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs Worldwide, 2000–2004. M M W R 55: 301-3055.
5. Ramaswamy S, Musser JM (1998) Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. Tuber Lung Dis 79: 3-29.
6. Parida SK, Axelsson-Robertson R, Rao MV, Singh N, Master I, et al. (2015) Totally drug-resistant tuberculosis and adjunct therapies. J Intern Med 277: 388-405.
7. Nguyen L (2016) Antibiotic resistance mechanisms in M. tuberculosis: an update. Arch. Toxicol 90:1 585-1604.
8. Marais BJ, Hesseling AC, Schaaf HS, Gie RP, van Helden PD et al. (2009) Mycobacterium tuberculosis transmission is not related to household genotype in a setting of high endemicity. J Clin Microbiol 47: 1338-1343.
9. Kato-Maeda M, Metcalfe JZ, Flores L (2011) Genotyping of Mycobacterium tuberculosis: application in epidemiologic studies. Future Microbiol 6: 203-216.
10. Narayanan S (2004) Molecular epidemiology of Mycobacterium tuberculosis. Indian J Med Res 20: 233-47.
11. van Embden JD, van Gorkom T, Kremer K, Jansen R, van der Zeijst BAM et al. (2000) Genetic variation and evolutionary origin of the direct repeat locus of Mycobacterium tuberculosis complex bacteria. J Bacteriol 182: 2393-2401.
12. Foxman B, Riley L (2001) Molecular epidemiology: focus on infection. Am J Epidemiol 153:1135-1141.
13. Desikan S, Narayanan S (2015) Genetic markers, genotyping methods & next generation sequencing in Mycobacterium tuberculosis. Indian J Med Res 141: 761-874.
14. Driscoll JR (2009) Spoligotyping for molecular epidemiology of the Mycobacterium tuberculosis complex. Methods Mol Biol 551: 117-28.
15. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S et al. (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 44: 4498-4510.
16. Jagielski T, van Ingen J, Rastogi N, Dziadek J, Mazur PK et al. (2014) Current Methods in the molecular typing of Mycobacterium tuberculosis and Other Mycobacteria. Biomed Res Int 2014:645802.
17. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, et al. (2008) High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol 6: e311.
18. Nicol MP, Wilkinson RJ (2008) The clinical consequences of strain diversity in Mycobacterium tuberculosis. Trans R Soc Trop Med Hyg 102: 955-965.
19. Moström P, Gordon M, Sola C, Ridell M, Rastogi N (2002) Methods used in the molecular epidemiology of tuberculosis. Clin Microbiol Infect 8: 694-704.
20. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD et al. (1993) Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31: 406-409. 21.
21. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D et al. (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35: 907-914.
22. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D et al. (2001) Automated high throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive units. J Clin Microbiol 39: 3563-3571.
23. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC et al. (2006) Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci USA 103: 2869-2873.
24. Christianson S, Wolfe J, Orr P, Karlowsky J, Levett PN et al. (2010) Evaluation of 24 locus MIRU-VNTR genotyping of Mycobacterium tuberculosis isolates in Canada. Tuberculosis 90: 31-38.
25. Kanduma E, McHugh TD, Gillespie SH (2003) Molecular methods for Mycobacterium tuberculosis strain typing: a users guide. J Appl Microbiol 94: 781-791.
26. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN (2006) Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev 19: 658-685.
27. McHugh TD, Newport LE, Gillespie SH (1997) IS6110 homologs are present in multiple copies in Mycobacteria other than tuberculosis: causing mycobacteria. J Clin Microbiol 35: 1769-1771.
28. Schürch AC, van Soolingen D. DNA fingerprinting of Mycobacterium tuberculosis: from phage typing to whole-genome sequencing. Infect Genet Evol 2012 12:602-609.
29. Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of Genetically Monomorphic Bacteria: DNA Sequencing in Mycobacterium tuberculosis Highlights the Limitations of Current Methodologies. PLoS One 4: e7815.
30. Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, et al. (2006) Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol 6: 23.
31. Skuce RA, Neill SD (2001) Molecular epidemiology of Mycobacterium bovis: exploiting molecular data. Tuberculosis 81: 169-175.
32. Hermans PW, van Soolingen D, Bik EM, de Haas PE, Dale JW et al. (1991) Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. Infect Immun 59: 2695-2705.
33. Cowan LS, Hooks DP, Christianson S, Sharma MK, Alexander DC et al. (2012) Evaluation of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping as performed in laboratories in Canada, France, and the United States. J Clin Microbiol 50:1830-1831.
34. Zhang J, Abadia E, Refregier G, Tafaj S, Boschiroli ML et al. (2010) 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 59: 285-294.
35. van der Zanden AG, Hoentjen AH, Heilmann FG, Weltevreden EF, Schouls LM et al. (1998) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis complex in paraffin wax embedded tissues and in stained microscopic preparations. Mol Pathol 51: 209-214.
36. Chaoui I, Zozio T, Lahlou O, Sabouni R, Abid M et al. (2014) Contribution of spoligotyping and MIRU-VNTRs to characterize prevalent Mycobacterium tuberculosis genotypes infecting tuberculosis patients in Morocco. Infect. Genet Evol 21: 463-471.

37. Tang C, Reyes JF, Luciani F, Francis AR, Tanaka MM (2008) spolTools: online utilities for analyzing spoligotypes of the Mycobacterium tuberculosis complex. Bioinformatics 24: 2414-2415.
38. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. Nucleic Acids Res 38:326-331.
39. Demay C, Liens B, Burguière T, Hill V, Couvin D et al. (2012) SITVITWEB--a publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. Infect Genet Evol 12:755-766.
40. Evans JT, Hawkey PM, Smith EG, Boese KA, Warren RE et al. (2004) 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 42: 4175-4180.
41. Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, et al. (2000) Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. Mol Microbiol 36: 762-771.
42. Mazars E, Lesjean S, Banuls AL, Gilbert M, Vincent V, et al. (2001) High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. Proc Natl Acad Sci USA 98: 1901-1906.
43. Allix C, Supply P, Fauville-Dufaux M (2004) Utility of fast mycobacterial interspersed repetitive unit-variable number tandem repeat genotyping in clinical mycobacteriological analysis. Clin Infect Dis 39: 783-789.
44. Allix-Béguec C, Wahl C, Hanekom M, Nikolayevskyy V, Drobniewski F, et al. (2014) Proposal of a consensus set of hypervariable mycobacterial interspersed repetitive-unit-variable-number 21 tandem-repeat loci for subtyping of Mycobacterium tuberculosis Beijing isolates. J Clin Microbiol 52: 164-172.
45. Trovato A, Tafaj S, Battaglia S, Alagna R, Bardhi D, et al. (2016) Implementation of a Consensus Set of Hypervariable Mycobacterial Interspersed Repetitive-Unit-Variable-Number Tandem-Repeat Loci in Mycobacterium tuberculosis Molecular Epidemiology. J Clin Microbiol 54: 478-482.
46. Liu L, Li Y, Li S, Hu N, He Y, et al. (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 251364.
47. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, et al. (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol 10: 599-606.
48. Jagielski T, Minias A, van Ingen J, Rastogi N, Brzostek A, et al. (2016) Methodological and clinical aspects of the molecular epidemiology of Mycobacterium tuberculosis and other mycobacteria. Clin Microbiol Rev 29: 239-290.
49. Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, et al. (2013) Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS Med 10: e1001387.
50. Walker TM, Monk P, Smith EG, Peto TE (2013) Contact investigations for outbreaks of Mycobacterium tuberculosis: advances through whole genome sequencing. Clin Microbiol Infect 19: 796-802.
51. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, et al. (2015) TGS-TB: Total Genotyping Solution for Mycobacterium tuberculosis Using Short-Read Whole Genome Sequencing. PLoS One 10: e0142951
52. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, et al. (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci USA 99: 3684-3689.
53. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, et al. (2004) Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. Proc. Natl Acad Sci USA 101: 4865-4870.
54. Coscolla M, Gagneux S (2014) Consequences of genomic diversity in Mycobacterium tuberculosis. Semin Immunol 26: 431-444.
55. Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, et al. (2012) Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching Mycobacterium tuberculosis Clade. PLoS ONE 7: e52841.
56. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, et al. (1997) Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A 94: 9869-9874.
57. Rastogi N, Sola C. Molecular evolution of the Mycobacterium tuberculosis complex 53-91. In: Palomino JC, Leão SC, and Ritacco V, (ed). Tuberculosis 2007: from basic science to patient care. Amedeo Online Textbooks.
58. Mokrousov I (2012) Human migratory history: through the looking glass of genetic geography of Mycobacterium tuberculosis. 317-341.
59. Lasunskaia E, Ribeiro SC, Manicheva O, Gomes LL, Suffys PN, et al. (2010) Emerging multi-drug resistant Mycobacterium tuberculosis strains of the Beijing genotype circulating in Russia express a pattern of biological properties associated with enhanced virulence. Microbes Infect 12: 467-475.
60. Martín A, Herránz M, Serrano MJ, Bouza E, García de Viedma D (2007) Rapid clonal analysis of recurrent tuberculosis by direct MIRU-VNTR typing on stored isolates. BMC Microbiol 7: 73.
61. García de Viedma D, Mokrousov I, Rastogi N (2011) Innovations in the molecular epidemiology of tuberculosis. Enferm Infecc Microbiol Clin 29: 8-13.
62. Lazzarini LC, Rosenfeld J, Huard RC, Hill V, Lapae Silva JR, et al. (2012) Mycobacterium tuberculosis spoligotypes that may derive from mixed strain infections are revealed by a novel computational approach. Infect Genet Evol 12: 798-806.
63. Cohen T, Murray M, Abubakar I, Zhang Z, Sloutsky A, et al. (2011) Multiple introductions of multidrug-resistant tuberculosis into households, Lima, Peru. Emerg Infect Dis 17: 969-975.
64. Van Soolingen D (2001) Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J Intern Med 249: 1-26.
65. Bhanu NV, Banavalikar JN, Kapoor SK, Seth P (2004) Suspected small-scale interpersonal transmission of Mycobacterium tuberculosis in wards of an urban hospital in Delhi, India. Am J Trop Med Hyg 70: 527-531.
66. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, et al. (1993) Molecular strain typing of Mycobacterium tuberculosis to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false- positive cultures. J Clin Microbiol 31: 1677-1682.
67. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, et al. (1997) Interpretation of restriction fragment length polymorphism analysis of Mycobacterium tuberculosis isolates from a state with a large rural population. J Infect Dis 175: 1446-1452.
68. Chin DP, Crane CM, Diul MY, Sun SJ, Agraz R, et al. (2000) Spread of Mycobacterium tuberculosis in a community implementing recommended elements of tuberculosis control. JAMA 283: 2968-2974.