Introduction

When repetitive sequences of the genome of *Mycobacterium tuberculosis* were first successfully obtained in 1991, it became possible to compare the DNA fingerprints of different isolates and, therefore, distinguish different strains from one another. Differentiation of isolated strains has allowed new information on the epidemiology of the disease to be collected. With this new information, our understanding of how tuberculosis is transmitted has advanced, overturning long-standing preconceptions. Today, we have information on the relative importance of reactivation and recent transmission in areas with differing prevalence and on risk factors, such as the frequency of exogenous reinfection and the frequency of such occurrences in different zones. We also have new data on the infectivity of tuberculosis in situations in which sputum staining gives negative results, and on differences in the transmissibility and other phenotypic characteristics between *M. tuberculosis* strains.

This review, which focuses on aspects of clinical interest, first describes the most common methods for molecular characterization of *M. tuberculosis* and then summarizes the most novel information gleaned from molecular epidemiology studies, with particular reference to studies done in Spain.

Methods for Molecular Characterization of *M. tuberculosis*

For a long time, strains have been characterized by markers based on characteristics expressed by the microorganisms (phage typing, resistance typing, serotyping, and plasmid analysis). These phenotyping techniques are currently being superseded by molecular techniques that analyze the DNA fingerprint. The new approaches are varied—some characterize the degree of similarity and the distribution of these variable elements among the isolates, whereas others study the presence or absence of certain DNA fragments or compare the complete genome of the microorganism.

Restriction Fragment Length Polymorphism

The reference method for strain typing is considered to be restriction fragment length polymorphism (RFLP). The technique is based on a standard protocol which analyzes sequence IS6110. It is very reproducible and useful for differentiating between epidemiologically related and unrelated isolates, and so has been used as an indicator of recent transmission. The technique is based on counting the number times the IS6110 restriction fragment is repeated in the genome of the mycobacterium (usually between 0 and 25 times). This technique presents certain limitations; for example, if the mycobacterium has less than 6 copies of this restriction fragment other techniques must be used. Furthermore, large amounts of DNA are required (with the corresponding need for subculturing which may take several weeks), and the procedure is complex and expensive.

Spoligotyping

A method widely used, usually as a complement to RFLP, is spoligotyping because of its relative simplicity, speed, and low cost. Spoligotyping studies the presence or absence of 43 DNA fragments known as spacers. Less DNA is required for this technique than for RFLP and, because expression (of each spacer) is either positive or negative, digital analysis is possible. An international database is available with more than 11,000 standard isolates (“spoligotypes”) taken from more than 90 countries. However, spoligotyping will not completely replace RFLP because it provides lower discrimination.
Variable Number Tandem Repeat

The technique known as variable number tandem repeat (VNTR) detects the number of times that several sequences occur adjacent within the genome of the mycobacterium. The most common approach is mycobacterial interspersed repetitive unit VNTR (MIRU-VNTR), which determines repetitions at 12 loci by polymerase chain reaction. From 2 to 8 alleles are located at each of the 12 loci, generating approximately 20 million possible allele combinations. The discrimination afforded by MIRU-VNTR is better than that of spoligotyping and similar to that of techniques based on RFLP. Given that MIRU-VNTR can be readily automated and the technology is simpler, it may well become the reference method. Moreover, findings can be compared easily with an international database accessible via the Internet.

Amplified Fragment Length Polymorphism

Amplified fragment length polymorphism (AFLP) differs from other methods in that the whole mycobacterial genome is studied, making it potentially more specific than RFLP for confirming recent transmission of a strain. However, this technique, which is based on polymerase chain reaction, is expensive, time-consuming, and requires highly specialized personnel. Furthermore, application of AFLP in molecular epidemiology studies and interpretation of the results are still pending more detailed investigation.

Proportion of Recent Tuberculosis Transmission

Recent transmission is assumed when several isolates from M. tuberculosis have identical or similar DNA fingerprints—the so-called cluster cases. On the other hand, isolates with unique DNA fingerprints are considered to indicate reactivation of a previously acquired infection. The DNA of M. tuberculosis undergoes progressive changes over time. The time needed for a DNA fingerprint to change (half-life), according to studies with IS6110 restriction fragment RFLP, ranges from 2 years to over 30 years. Several factors can be assumed to influence the time needed, such as the efficacy of the treatment, the interval between disease onset and treatment, and, probably, the relative predominance of more or less stable strains of the pathogen in each region, as well as whether the disease is latent or active.

Several studies have been published that investigate cases attributable to recent transmission according to the principles outlined above. Interestingly, the proportion of cluster cases can be as high as 40% in the United States and Europe, even in areas where the incidence of tuberculosis is low, suggesting that recent transmission of M. tuberculosis could have a greater relative importance than was previously thought. Countries with a high incidence of tuberculosis and an inadequate health care system would be expected to have a high percentage of tuberculosis cases associated with recent transmission. However, the results of studies done to test this have been very variable—the percentage of recent transmission ranges from 30% to 80% in Africa and from 50% to 80% in Asia, but, surprisingly, is only 20% according to other studies. This paradoxical finding can be explained by the methodology used, as small studies in areas with high incidences greatly underestimate recent transmission.

In Spain, 9 molecular epidemiology studies of tuberculosis transmission have been published. These studies have used RFLP alone or in combination with spoligotyping (Table 1). The percentage corresponding to cluster cases (recent transmission) ranged from 28% in Segovia to 58% on Gran Canaria Island. These percentages might reflect true differences in the disease situation in the different areas, but the characteristics of the studies themselves (such as the study duration and number of cases not included) could also have influenced the results. With the new molecular methods, investigators have also been able to demonstrate the epidemiological connection between outbreaks of multidrug-resistant tuberculosis in Spain.

<table>
<thead>
<tr>
<th>Author and Year</th>
<th>Study Period, Years</th>
<th>Place</th>
<th>Method</th>
<th>No. of Patients</th>
<th>Percentage of Cluster Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safi et al.1997</td>
<td>3</td>
<td>Seville</td>
<td>RFLP</td>
<td>175</td>
<td>38</td>
</tr>
<tr>
<td>Samper et al.1998</td>
<td>1</td>
<td>Zaragoza</td>
<td>RFLP</td>
<td>226</td>
<td>39</td>
</tr>
<tr>
<td>Íñigo Martínez et al.2000</td>
<td>2</td>
<td>Madrid</td>
<td>RFLP and spoligotyping</td>
<td>148</td>
<td>42</td>
</tr>
<tr>
<td>Solsona et al.2001</td>
<td>2</td>
<td>Ciutat Vella, Barcelona</td>
<td>RFLP</td>
<td>171</td>
<td>46</td>
</tr>
<tr>
<td>Fernández de la Hoz et al.2001</td>
<td>2</td>
<td>Urban population in south Madrid and prison population</td>
<td>RFLP</td>
<td>231</td>
<td>48</td>
</tr>
<tr>
<td>Elizaga et al.2002</td>
<td>5</td>
<td>Segovia</td>
<td>RFLP and spoligotyping</td>
<td>87</td>
<td>28</td>
</tr>
<tr>
<td>Ruiz et al.2002</td>
<td>7</td>
<td>Elche</td>
<td>RFLP</td>
<td>147</td>
<td>52</td>
</tr>
<tr>
<td>Pena et al.2003</td>
<td>4</td>
<td>Gran Canaria Island</td>
<td>RFLP</td>
<td>409</td>
<td>58</td>
</tr>
<tr>
<td>Íñigo Martínez et al.2003</td>
<td>3</td>
<td>Madrid</td>
<td>RFLP and spoligotyping</td>
<td>233</td>
<td>42</td>
</tr>
</tbody>
</table>

*RFLP indicates restriction fragment length polymorphism.
Many authors prefer to express their results as an incidence of cluster cases per 100,000 inhabitants per year rather than as percentages. Thus, for example, the decrease in the incidence of cluster cases in San Francisco between 1991, when it was 10.4 cases per 100,000 inhabitants, and 1997, when it was 3.8 cases per 100,000 inhabitants, was a clear reflection of the decrease in recent transmission over this period.

These findings, in combination with conventional epidemiological methods, can identify, for example, transmission between different ethnic and social groups. According to estimates for the Netherlands, 17% of the cases in the country are due to recent transmission from foreigners, and by 2030, investigators forecast that at least 60% of tuberculosis cases in autochthonous Dutch patients will be due to transmission by immigrants.

Murray and Nardell, however, insist that the cluster cases should be interpreted with caution when estimating the true transmission rate of tuberculosis because case clustering could vary according to the characteristics of the host and the population. For example, in rural areas, similar genotypes may not necessarily be an indication of a recent infection. Several factors in addition to recent transmission of the disease can give rise to similar DNA fingerprints; examples of such factors are the simultaneous reactivation of an infection with the same organism that was acquired long before (endemic strains), a change in the location of the insertion sequence at the same site, and, of course, laboratory errors. The opposite situation is also true: an isolate may be mistaken for an endogenous reactivation when it is actually associated with other undetected cases. A single DNA fingerprint is more likely to be erroneously interpreted as endogenous reactivation in shorter studies (the duration of a study should not be less than 2 years) and in studies that analyze small numbers of cases.

**Risk Factors for Recent Transmission**

Some authors have tried to use case cluster data to determine risk factors for recent transmission. According to a study by Alland et al., younger subjects, those born in the United States of America, Hispanics, and those infected with the human immunodeficiency virus (HIV) were at greater risk. However, there are often discrepancies among different studies, reflecting both the diversity of disease transmission in different communities and methodological shortcomings.

The duration of symptoms, which could theoretically be a reflection of risk of transmission, was not associated with cluster size defined by molecular techniques. It is important to point out that although recent transmission is more frequent when patients have positive results from sputum smears, subjects with negative results can be responsible for a relatively large number of transmissions.

In a study which analyzed clinical and socioeconomic factors of tuberculosis in Elche, Spain, age less than 25 years, a high percentage of infections from inner circle contacts, urban residence, use of bronchoscopy to obtain samples, and working in contact with many people were all independent variables predictive of belonging to a case cluster. In an extensive study done on Gran Canaria Island, Spain, the only risk factor associated with belonging to a case cluster was young age.

**Risk of Recent Transmission in Patients With HIV Infection**

Patients with HIV infection are known to be at high risk of contracting tuberculosis, although current antiretroviral treatments may help to lower the incidence of the disease in this population group.

Different molecular epidemiological studies have found HIV infection to be the biggest risk factor for recent transmission. Although most studies indicate that HIV infection is a strong risk factor, not all studies have confirmed this finding. Variable social factors among these patients (such as use of intravenous drugs or, in particular, imprisonment) in a given country may influence the risk of transmission.

**Exogenous Reinfection**

One of the long-standing debates about tuberculosis transmission was the true importance of exogenous reinfection. For a long time, it was thought that around 90% of the cases of tuberculosis in developed countries were due to reactivation of a previously acquired infection. Nevertheless, the occurrence of exogenous reinfection had already been demonstrated with conventional epidemiological methods and phenotyping (for resistance) of the mycobacteria. It clearly used to be difficult to establish whether a case of tuberculosis was due to activation of latent disease or to a new exogenous infection, but our knowledge of this aspect has been improved by molecular epidemiological data.

In a region with a high incidence of tuberculosis, investigators analyzed 16 cases of patients with disease recurrence after correctly completing treatment and identified strains other than those responsible for the initial disease in 12 of them. Reinfection is also important in countries with a lower incidence of tuberculosis. In Spain, Caminero et al. performed genotyping of *M. tuberculosis* isolated from the 18 patients with positive tuberculosis cultures taken at least 12 months apart and found that 44% of these cultures had mycobacteria with different genotypes, indicating exogenous reinfection. An extensive study of recurrence in cases of tuberculosis in Madrid found that recurrence in 14 of the 43 patients analyzed (33%) in a period of 12 years was due to exogenous reinfection. In contrast, in countries with a low incidence of tuberculosis, such as the United States of America or Canada, the proportion of reinfections appears lower.
**Tuberculosis Control Programs**

With molecular epidemiological studies, investigators have been able to detect outbreaks of related tuberculosis cases which would have been missed by conventional epidemiology. This has improved contact investigations for tuberculosis and one study was even able to identify 3 times as many contacts for each tuberculosis case. The application of a capture-recapture method in Spain that combined conventional and molecular epidemiology greatly improved the detection of related cases in recent transmission.

In addition to contact investigations, information from molecular epidemiology studies is useful for assessing the results of disease control programs. As mentioned earlier, between 1991 and 1997, the number of cluster cases in San Francisco decreased, suggesting that the approaches for controlling the disease reduced the spread of tuberculosis.

The epidemiological analyses done in some countries to guide the implementation of specific policies to combat tuberculosis is an obvious example of the need for this type of study. It is important to determine whether the active disease among the population in a given area is due to recent infection or reactivation of latent disease. If the number of cases of recent infection turns out to be significant, case finding and treatment should be stepped up, but if latent infection is the main cause, then measures should be focused more on preventing reactivation.

Doubts remain about the methods for genotyping (selection of ideal method with reproducible information at a reasonable cost) and interpretation of the findings, but application of these techniques in normal practices of prevention and control of tuberculosis is currently considered justified. Reference centers seem particularly useful, as they allow studies that extend beyond the limits of a given health area.

**Risk of Transmission of Tuberculosis to Health Care Workers**

In a recent review, Seidler et al found only 2 molecular epidemiological studies that separately calculated the percentage of tuberculosis cases among health care workers and nonhealth care workers corresponding to cluster cases. A study in New York that included 142 tuberculosis cases found that 65% of infected health care workers and 41% of infected patients had a clustering pattern. The authors concluded that many of the cases among health care workers could be considered as occupational infection. In contrast, another study in Amsterdam, the Netherlands, concluded that health care workers had a lower chance of belonging to a cluster, although the authors could not explain this finding.

In Spain, Tudó et al used RFLP to analyze patients admitted to hospital with tuberculosis in search of a possible undetected nosocomial transmission of the disease. They found no cluster cases resulting from transmission by these patients in the hospital. Likewise, the authors reported no cases of transmission to patients who shared a room with tuberculosis patients. These findings suggest that the risk of infection is very low. However, other Spanish studies have also reported an increased accumulated incidence of tuberculosis among health care workers, suggesting that there is a risk due to occupational exposure.

All these findings underline the need to perform prospective studies that combine methods of conventional and molecular epidemiology to establish the true risk of tuberculosis transmission in health care workers.

**Transmission of Different Strains of M tuberculosis**

It used to be thought that the different strains of M. tuberculosis had a similar virulence, but molecular studies have been able to confirm the existence of strains with specific and characteristic phenotypes. Several of the strains identified have been associated with epidemic outbreaks in different regions and at different times and it might be assumed that these strains are more virulent or easily transmitted (superspreaders).

One example is the M. tuberculosis family denominated Beijing/W. The Beijing genotype strain has been implicated in large-scale transmissions in the United States of America, Asia, Eastern Europe, and the Russian Federation. This strain is widely distributed, especially in Asia, where it is predominant, but also in Europe, the United States, South America, Caribbean countries, and Africa. Spoligotyping has shown that strain W, detected in the 1990s in the United States and responsible for outbreaks of multiresistant disease, belongs to an evolutionary branch of strains of the Beijing genotype, and so the Beijing/W family was established. The predominance of Beijing strains in many regions might indicate that they have a selective advantage over other strains. On the island of Gran Canaria, Spain, this strain was shown to have spread rapidly from a single patient to account for 21.7% of isolates of M. tuberculosis within 4 years. The Beijing strain has often, though not always, been associated with drug-resistance in a number of studies. These strains of M. tuberculosis have a greater ability to replicate in human macrophages and this could be the main propagation mechanism. However, apparent differences in virulence of the mycobacterium could also be due to variations in immunogenicity, transmissibility, growth rates, or characteristics of the exposed population.

**Transmission of Drug-Resistant Strains of M tuberculosis**

The transmission rate of drug-resistant strains is often assumed to be lower than that of other strains. Several studies have observed a negative correlation between case clusters and resistance to tuberculosis.
Molecular Epidemiology of Tuberculosis: Main Findings and Their Application in Spain

GARCÍA-PACHÓN E. ET AL. MOLECULAR EPIDEMIOLOGY OF TUBERCULOSIS: MAIN FINDINGS AND THEIR APPLICATION IN SPAIN


However, not all findings are consistent—some studies describe outbreaks that include many cases of multiresistant tuberculosis. Indeed, in a study in Norway, infection with a resistant strain was independently associated with forming part of a cluster (recent transmission). A high risk of transmission of resistant stains of the Beijing/W family has also been reported.

Murray and Nardell propose several possibilities to explain why clusters of tuberculosis resistant to one or more drugs could be smaller than the clusters of susceptible patients. For example, patients with multiresistant tuberculosis might have worse access to health care systems in many countries, they might have fewer social contacts, or they might have been exposed to the pathogen more often.

Other Uses of Genotyping

A large study in London found that only 0.93% of false positives in tuberculosis cultures were due to cross-contamination in the laboratory, whereas some studies indicate that up to 3% of the cultures of tuberculosis are false positives, particularly with negative sample staining and single culture growth. In such cases, and when clinical suspicion does not support a diagnosis of tuberculosis, typing of isolates could provide information that avoids unnecessary treatments. If the genotype of the strain of doubtful clinical significance corresponds with that of another strain handled at the same time in the laboratory, it is likely that contamination has occurred.

Another possible use is in patients who are receiving or who have received tuberculosis treatment and for whom resistances appear in new cultures. Genotyping could distinguish between exogenous reinfection and a strain that has developed resistance. In the latter instance, the treatment compliance of the patient, the possibility of malabsorption of the drugs, or pharmacological interactions would have to be investigated. Poor treatment compliance has been shown to be a cause of spread of the disease.

In clinical studies of tuberculosis drugs, it is important (and now possible) to determine whether therapeutic failures are due, in fact, to new infections or not. Genotyping also allows comparison of families of M tuberculosis in international databases already available or in development, and so it will be possible to detect and control large outbreaks that might not be suitably identified in local studies.

Conclusions and Outlook

The molecular epidemiology of tuberculosis has provided novel and often surprising information on the transmission of the disease. Of course, this method should be used as a complement to conventional epidemiology and not as a replacement. It is important to know which strains predominate in the community and how the cases attributable to recent transmission change over time as well as to determine why transmission occurs and which social groups require the most care. Growing immigration, which brings new strains of M tuberculosis, will challenge our epidemiological surveillance system and molecular studies are going to be essential. To meet this challenge, long-term studies are necessary in Spain and reference centers should probably be set up to coordinate this information.

Automated techniques are now available to analyze the results with reference to international databases at a reasonable cost. In a few years, our approach to tuberculosis will be influenced by this new information. The application of molecular epidemiology in combination with conventional epidemiological methods is a reasonable aim for the coming years in developed countries. To achieve this, close collaboration of pulmonologists, epidemiologists, and microbiologists will be necessary.

REFERENCES


GARCÍA-PACHÓN E, ET AL. MOLECULAR EPIDEMIOLOGY OF TUBERCULOSIS: MAIN FINDINGS AND THEIR APPLICATION IN SPAIN


