Re-analysis of epidemiologically linked tuberculosis cases not supported by IS6110-RFLP-based genotyping

A. Martín1, J. Iñigo3, F. Chaves4, M. Herranz1,2, M. J. Ruiz-Serrano1,2, E. Palenque4, E. Bouza1,2 and D. García de Viedma1,2
1) Servicio de Microbiología y Enfermedades Infecciosas, Hospital Gregorio Marañón, Madrid, 2) CIBER Enfermedades Respiratorias-CIBERES, 3) Consejería de Sanidad Comunidad de Madrid and 4) Servicio de Microbiología, Hospital Doce de Octubre, Madrid, Spain

Abstract

Tuberculosis microepidemics are considered as such when a proven epidemiological link is identified between the cases. However, some studies have found microepidemics that were not supported by genotyping data. In a cross-sectional study, 44 linked pairs from 33 microepidemics identified during a 5-year period in Madrid, Spain were analysed to evaluate whether the epidemiological findings were consistent with the molecular analysis by IS6110-RFLP. Twelve pairs (27.3%) were not initially confirmed by molecular typing, and a refined re-analysis was performed to identify the reasons for the discrepancies. The possible causes were as follows: (i) laboratory errors or cross-contamination events, (ii) undetected clonally complex infections, and (iii) lack of refinement in the genotyping analysis that could be clarified by applying second-line fingerprinting tools. One discrepant pair was caused by laboratory error. No discrepant pairs were the result of incorrect assignment of genotypes due to clonally complex infections. The application of spoligotyping, MIRU-15 and RFLP enabled the establishment of matching shared genotypes in four linked pairs initially considered as discrepant; therefore, the percentage of discrepant pairs was reduced from 27.3% to 15.9% (7/44). However, in the remaining seven pairs, second-line fingerprinting identified differences with at least two of the three genotyping tools applied. This finding alerts us to the need to (i) refine as much as possible the molecular analysis to establish more accurate identification of truly discrepant cases, and (ii) broaden the search for epidemiological links to include non-conventional contexts outside the household or work/school settings.

Keywords: Discrepancies, Microepidemics, molecular epidemiology, tuberculosis

Introduction

Genotyping of Mycobacterium tuberculosis (MTB) isolates using restriction fragment length polymorphism (RFLP) with the insertion sequence IS6110 is the standard molecular approach to analysing tuberculosis (TB) outbreaks, [1–3] recent transmission patterns [4], and prevalent circulating MTB strains [5]. In general, MTB isolates from epidemiologically linked cases share identical RFLP patterns, and are considered to be the result of recently transmitted infections.

However, the application of molecular techniques in population studies has produced paradoxical observations. First, some authors have observed that clustering of TB cases may not always be the result of recently transmitted infection and identical DNA fingerprints can be found among patients who have never been in contact [6,7]. Others, however, suggest that in the vast majority (86%) of patients, clustering indeed represents recent transmission [8]. Second, in studies that have tested the assumption that TB in a secondary case is the result of transmission from the corresponding source case, DNA fingerprinting has shown that in up to 30% of pairs the genotypic patterns for the MTB isolates from the source and secondary cases are different [9,10].

The present study analyses the microepidemics (defined as the existence of one or more secondary cases arising from the defined index case) detected in the south of Madrid over a 5-year period. We evaluated the correlation between epidemiological and molecular data to identify discrepant pairs, which were then studied in depth to clarify the reasons for the discrepancies by refining the fingerprinting methods and strategies.
Materials and Methods

Study design and population
This cross-sectional population-based study of culture-positive TB patients was conducted in nine urban districts in the south of Madrid (Spain). The total population of almost 1.5 million inhabitants receives specialized care at Hospital 12 de Octubre and Hospital Gregorio Marañón.

The target population consisted of all residents diagnosed with TB from January 2002 to December 2006. Collaborative surveillance was carried out by two mycobacterial laboratories that were integrated into the TB control programme, together with other hospital medical departments, public health services, and primary health care.

Sociodemographic and clinical data were obtained from the Regional Tuberculosis Case Register using a standardized protocol. Data were recorded on patient characteristics (age, sex, country of origin, number of years of residence in Madrid or elsewhere in Spain, and district of residence), clinical characteristics (date of onset of symptoms, date of diagnosis, site of disease, and previous TB), risk factors for TB (homelessness, injection drug use, presence of HIV infection, previous imprisonment, diabetes, silicosis, gastrectomy, and malignant neoplasm), and information from follow-up and contact investigations.

The study was approved by the hospital ethics committee.

Definition of microepidemic
We considered an outbreak or microepidemic as the existence of one or more secondary cases arising from the same source over a 1-year period after the index case was diagnosed [11]. We preferred the term microepidemic because most of the clusters we studied included two to three cases and transmission occurred mainly between household members.

In all cases, the likelihood of transmission was classified as ‘definite’ because patients were found to be in the same place at the same time, when one was infectious or a patient identified another patient as being the source of TB. This classification was made by epidemiologists before the molecular study was carried out.

We used the definition of epidemiologically linked case pairs reported by Bennett et al. [10].

DNA fingerprinting
DNA extraction, IS6110-RFLP typing, and spoligotyping were performed according to standard methods [12,13]. Mycobacterial interspersed repetitive units-variable number of tandem repeats (MIRU-VNTR) was performed by amplifying the 15 MIRU-VNTR loci as described elsewhere [14].

Pairs were classified as infected by the same strain if (i) their IS6110 patterns had more than five bands and were identical or (ii) their IS6110 patterns had five or fewer bands of identical sizes and also shared identical spoligotypes. RFLP patterns were compared by computer-assisted analysis (BioNumerics 5.1 software; Applied Maths, Sint-Martens-Latem, Belgium). For the pairs that did not fit the previous criteria, we applied a second-line typing approach based on spoligotyping and VNTR-MIRU with the set of 15 loci [15]. For these discrepant case pairs, we also analysed, when available, more than one sample from each patient.

The second-line genotyping analysis of the pairs showing differences in their RFLP types was used to classify them as follows: (i) matched pairs, if they shared both spoligotypes and MIRU types; (ii) unmatched pairs, if they showed differences both in spoligotypes and MIRU types; and (iii) partially unmatched pairs, if they showed differences either in spoligotypes or MIRU types.

Spoligotypes were considered as shared if they were identical and no variations were tolerated because of the low discriminatory power of this technique. MIRU types were considered shared if they were identical, but an allelic variation of one repetition in a single locus was tolerated.

Molecular strategy to analyse discrepant microepidemics
For the microepidemics with molecular discrepancies, we evaluated different possibilities: (i) false unclustering of the cases by involvement of some laboratory cross-contamination event or laboratory error; (ii) unclustering in the cases involved in the microepidemic caused by incorrect assignment of genotypes, due to clonally complex infections (mixed or compartmentalized infections); and (iii) lack of accuracy in the genotyping analysis that could be clarified by applying second-line fingerprinting tools, and which may have considered pairs with a similar but not identical genotype as unclustered cases.

Results

During the study period, 1768 TB cases were diagnosed, of which 1333 had a positive culture for MTB. Among the culture-positive cases, 956 isolates were available for genotyping. The study population with molecular data available did not differ with regard to clinical or demographic characteristics from the total number of culture-positive cases, with the exception that a greater proportion of isolates from immigrants was available for genotyping (36.6% vs. 30.8%; \( p = 0.04 \)). Three hundred and thirty-one cases (34.6%) were grouped in 91 clusters involving between two and 20...
During this period, 33 epidemiologically proven microepidemics were detected. These involved 44 linked case pairs ranging from one to three linked pairs per microepidemic and including a total of 77 cases. In 29 pairs (65.9%), the contact occurred in the same household; the remaining linked pairs corresponded to family links (nine pairs), friends (four pairs), occupational contact (one pair), homeless sharing a shelter (one pair).

When we compared the epidemiological and molecular findings, we found that 32 linked pairs (72.7%) were initially confirmed by the molecular analysis (Fig. 1). The remaining 12 (27.3%) case pairs were not initially confirmed by RFLP-based molecular typing and eight of these (75%) were household contacts (Fig. 2).

We decided to analyse in more detail the subset of microepidemics that did not correlate with the fingerprinting data. The analysis of additional samples allowed us to identify genotyping matches in one of the pairs initially considered as discrepant. The analysis in pair 500-a of an additional sample from case 500-3 allowed us to detect the same isolate that was involved in the infection of the related case (Fig. 2a). The genotype that had been assigned to patient 500-3 was shared with another case whose sample had been processed in the laboratory on the same day, suggesting an incorrect assignment of a specimen from the other patient.

In order to examine clonal complexity as a cause of discrepancies in some microepidemics, we searched for potential errors in the assignment of RFLP types due to mixed infections by more than one MTB strain in the same sputum sample or with different strains isolated in independent samples from a single patient. The first possibility could be explored by the application of MIRU-VNTR, which enables us to reveal clonal complexity in a sample. A sample with two clonal variants was detected (5112; Fig. 2d) in only one microepidemic, and the differences between the variants were so subtle that they were not expected to be responsible for the discrepancies in this microepidemic. As for the second possibility—different strains/clonal variants in

<table>
<thead>
<tr>
<th>Microepidemic</th>
<th>N</th>
<th>RFLP band-pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>155</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>158</td>
<td>3</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>161</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>148</td>
<td>3</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>147</td>
<td>4</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>128</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>151</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>160</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>1167</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>139</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>157</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>149</td>
<td>3</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>5102</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>1182</td>
<td>2</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>1172</td>
<td>3</td>
<td>[1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>1171</td>
<td>4</td>
<td>[1 1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>1163</td>
<td>2</td>
<td>[1 1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>156</td>
<td>2</td>
<td>[1 1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>5122-b</td>
<td>2</td>
<td>[1 1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>152-b</td>
<td>2</td>
<td>[1 1 1 1 1 1 1 1]</td>
</tr>
<tr>
<td>500-b</td>
<td>2</td>
<td>[1 1 1 1 1 1 1 1]</td>
</tr>
</tbody>
</table>

FIG. 1. Microepidemics showing cases matched using molecular techniques. N, number of cases involved in each microepidemic; RFLP, restriction fragment length polymorphism.
FIG. 2. Linked pairs initially considered not supported by the molecular analysis. The figure shows the number identifying each linked pair, the patients involved, the number of isolates available (I.A.) for analysis from independent specimens, the restriction fragment length polymorphism (RFLP), spoligotype, and the Mycobacterial interspersed repetitive unit (MIRU) patterns. MIRU alleles with differences between the cases involved are highlighted in bold. For case 5112-2, who was infected with two clonal variants, both alleles are shown. Pairs are grouped as follows according to the redistribution in different interpretation categories after the refined molecular re-analysis: (a) proved linked pairs after identifying a laboratory error, (b) proved linked pairs for which a molecular match had been established after second-line genotyping, (c) pairs with partial molecular differences in second-line genotyping, and (d) pairs with marked molecular differences in second-line genotyping.
different samples—we found more than one sample for analysis in only four microepidemics. We detected two clonal variants (identified by MIRU) in only one case (152-3; Fig. 2d), each in independent samples. Again, the differences between the two variants were so subtle that they were not expected to be responsible for the discrepancies in this microepidemic.

For the remaining 11 pairs that had not been clarified after consideration of laboratory errors or clonal complexity, we refined the genotyping analysis by applying a second-line approach. This allowed us to classify the pairs in three different interpretation categories:

1) Matched pairs: four pairs that shared spoligotypes and MIRU types (1169, 5104, 5120, 5123; Fig. 2b) and that had shown RFLP types differing in only one or two bands (except for 5120).

2) Partially unmatched pairs: three linked pairs sharing MIRU types but no spoligotypes (1164, 5107, 5122-a; Fig. 2c) and that had shown RFLP types differing in more than two bands.

3) Unmatched pairs: four pairs with differences both by spoligotyping and MIRU typing (144, 152-a, 164, 5112; Fig. 2d) and that had shown RFLP types differing in more than two bands.

A review of the epidemiological links supporting the definition of these microepidemics showed that they were definite. Only in one pair (164, which had shown marked molecular differences) were we able to identify a possible weakness in the assignment, because the patients were homeless.

The results of our refined re-analysis are summarized in Fig. 3. These results enabled us to reinterpret five of the 12 linked pairs analysed, which increased the percentage of pairs that were confirmed by molecular techniques from 72.7% to 84.1%. One pair was confirmed after identification of a laboratory error. Four pairs could have been considered concordant if spoligotyping and MIRU-15 had been applied or if minor differences between the RFLP patterns had been tolerated. Four pairs were confirmed as unrelated. For the remaining three discrepant linked pairs, one of the second-line genotyping techniques identified shared genotypes, whereas the other found differences between them.

**Discussion**

Microepidemics are generally defined using standard epidemiological surveys and involve cases with definite links, mostly in a household context. The introduction of molecular analysis into the study of TB transmission dynamics has revealed that a proportion of these theoretically linked cases are not supported by fingerprinting [9,10].

**FIG. 3.** Summary of the re-analysis of the 44 pairs initially considered discrepant using molecular data. From the total number of pairs in each panel, the number of pairs that involved household contacts is also indicated.
Rapid genotyping methods are available in certain centres as first-line approaches for molecular epidemiology purposes; however, studies using these techniques are scarce and RFLP is still the method most commonly used worldwide for the analysis of TB microepidemics. Considering this, we also followed the standard scheme of applying RFLP as the reference genotyping method, and MIRU-15 and spoligotype were used as second-line methods to clarify discrepancies.

We studied the proportion of microepidemics without fingerprint matches during a 5-year period in Madrid and refined the molecular analysis to evaluate the robustness of these discrepancies. Initially, we identified 12 of 44 epidemiologically linked pairs (27.3%) with discrepancies at the molecular level. After re-analysis, five of the 12 pairs (41.7%) initially considered not to be supported by genotyping were finally considered matched, thus reducing the percentage of discrepant pairs to 15.9% (7/44). However, for these remaining seven pairs, we still found genotyping differences when one or even two of the second-line genotyping tools were applied.

At first, we suspected laboratory cross-contamination [16] or errors in the assignment of some samples as causes of the discrepancies. After this analysis, one of the 12 pairs (8.3%) initially considered to be discrepant was revealed to be concordant after an unnoticed incorrect assignment of samples from two patients was identified in the laboratory.

Mixed infections by more than one strain, and compartmentalization of the infection with different strains infecting different tissues have been reported [17]. We took advantage of MIRU’s ability to identify clonally complex infections directly [18], which is not easy with RFLP or spoligotyping. This analysis revealed no microepidemics in which the discrepancies could be explained by clonally complex infections.

To determine the robustness of the discrepancies in the 12 pairs that had not yet been clarified, we applied a second-line genotyping strategy based on spoligotyping and MIRU-VNTR. MIRU-15 has proven to be as discriminative as RFLP [14] and to have a good fit with the epidemiological data [19].

After our refined analysis of the discrepancies, we found that a proportion (five of the 12 linked pairs) of the microepidemics that had initially been considered not to match the molecular data could be solved and were finally shown to be matched pairs. Nevertheless, in one of these pairs (S120), matching by spoligotyping and MIRU showed differences in more than two RFLP bands, thus suggesting some potentially unexpected rearrangements within the microepidemic only revealed by this technique. Similar results from other authors might suggest the application of an excessively strict criterion that assumed as clustered only those cases infected by MTB isolates sharing identical RFLP patterns. Subtle genetic changes can arise from the adaptation of a strain to the different hosts in a transmission chain [20], with the result that strains showing subtle variations might be considered clustered [21].

Although the percentage of linked pairs with no molecular evidence fell from 27.3% to 15.9%, it is interesting to examine the remaining seven pairs with molecular differences in the second-line genotyping scheme. Four of these pairs had molecular mismatches by the two second-line genotyping tools and three pairs had differences by only one of them. We could only posit links other than those expected in one pair, as the cases involved were homeless people; for the remaining cases, the epidemiological survey established a definite link, and in five of these seven pairs not supported by genotyping the theoretical contact was in the household.

If the MIRU data had been the only data available, a higher number of the epidemiologically linked pairs would have been found to be consistent with the epidemiological data (8/12, 66.7%). Discrepancies between the information obtained by RFLP and MIRU have also been found in molecular epidemiology studies [15,19,22]. In some of these studies, MIRU-15-defined clusters fit better with the epidemiological data than RFLP-defined ones [19,22]. Our findings in the present study provide additional examples of circumstances where MIRU-15 data could fit better with epidemiological findings.

Other studies had already examined epidemiologically linked cases with no molecular matches, although the more detailed methodological and microbiological approach of our method allowed us to reduce the percentage of true discrepancies. In an extensive US study [10] covering seven sentinel areas over 5 years, 29% of the epidemiologically linked cases did not share matching genotypes. Discrepant microepidemics were also found for the theoretically linked cases in circumstances where the certainty of links is expected to be highest.

The existence of microepidemics not supported by molecular matches could be because TB in the secondary cases is a reactivation of an infection acquired in the past or a recent infection from another unknown source case. In our study, there was no evidence of previous TB in any of the cases involved in the microepidemics not supported by the molecular analysis. Transmission in contexts outside the household has been described, mainly in high-incidence settings [23]. Casual contacts and transmissions in non-conventional settings have been observed and are mainly revealed by universal molecular strategies and in novel epidemiological surveys based on social networks [24,25] or reinterviewing of clustered cases [26].

In summary, our study has succeeded in establishing a molecular match for epidemiologically linked pairs that were not initially supported by fingerprinting, and this allowed us to reduce the percentage of discrepant pairs from 27.3% to 15.9%. Nevertheless, seven pairs were still found to be
mismatched after the refined re-analysis. Our findings should alert us to the need to (i) refine the molecular analysis as much as possible to identify those case pairs whose molecular data truly differed from the epidemiological data, and (ii) broaden the search for epidemiological links including non-conventional contexts outside the household or work/school settings in order to increase the correlation between epidemiology and molecular analysis.

Acknowledgements

We are indebted to T. O’Boyle for proofreading and editing the manuscript.

Transparency Declaration

This study was partially financed by Fondo de Investigaciones Sanitarias (FIS 02/0572, 02/1307, 02/0882, 030654; 060882) and by Instituto de Salud Carlos III (CIBER Enfermedades Respiratorias CB06/06/0058; Spanish Network for the Research in Infectious Diseases REIPI RD06/0008). A.M. is the recipient of a grant from the Comunidad de Madrid, co-financed by the European Social Fund (Order no. 5297/2006). None of the authors declares conflicts of interest.

References