EDITORIAL

Dealing with variation in molecular typing of *Mycobacterium tuberculosis*: low-intensity bands and other challenges

"...you should say what you mean," the March Hare went on.
"I do," Alice hastily replied; "at least — at least I mean what I say — that's the same thing, you know."
"Not the same thing a bit!" said the Hatter. [From Alice's Adventures in Wonderland, by Lewis Carroll]

In the last decade considerable progress has been made in molecular strain typing of bacterial pathogens. Various techniques have been established as reproducible and discriminatory. In particular, Southern blot analysis of IS6110 has provided new insights into the transmission and natural history of *Mycobacterium tuberculosis* infection and pulsed-field gel electrophoresis (PFGE) has facilitated epidemiological studies for a range of pathogens [1, 2]. However, one of the major challenges still facing those who perform bacterial strain typing is to provide interpretations that do not leave clinicians and others feeling they have wandered into the Mad Hatter’s tea party.

The central question of molecular typing is deceptively simple: Do two bacterial isolates represent the same strain or different strains? By ‘isolate’ we mean ‘a pure culture of bacteria obtained from a primary isolation plate, presumed to be derived from a single organism’, and a ‘strain’ is one or more isolates that give an indistinguishable result in a particular typing system. There is considerable genetic diversity within most bacterial species and epidemiologically unrelated isolates typically represent different genotypes. Consequently, such isolates will give distinctively different typing results when analysed by a ‘discriminatory’ typing method. In contrast, epidemiologically related isolates representing the same strain are assumed to be recently derived from a common precursor and, consequently, are expected to have indistinguishable, or at least closely related, genotypes (sometimes referred to as ‘clonal’).

The potential for ambiguous strain typing results has been appreciated for years and many studies that have used molecular typing techniques have demonstrated ‘similar’ genotypes in isolates from different sources – that is, demonstrate differences that, intuitively, seem minor relative to the overall diversity within the respective species. There have been at least two common explanations for such observations. First, clinical isolates that are epidemiologically unrelated may, nevertheless, be genetically related, most commonly due to selective pressures associated with virulence factors or antimicrobial resistance. Second, epidemiologically related isolates representing the same strain but from different sources (e.g., different patients or a patient and the environment) may show modest genetic differences consistent with ongoing evolutionary divergence [3].

Recently, two further levels of variation have become apparent. First, examination of multiple isolates cultured from individual hosts has demonstrated that occasionally two distinct genotypes of the same species are present in a single patient. Most invasive infections are monoclonal, i.e., derived from a single organism; nevertheless, ‘polyclonal’ presence appears to be relatively common in a few conditions, e.g., disseminated *M. avium* infection in HIV-infected persons [4] and staphylococcal colonisation of mucosal and cutaneous surfaces [5]. Second, even when a single patient is infected with only one strain, multiple isolates analysed by the same molecular typing technique may demonstrate minor, but distinct and reproducible differences. This level of variation within a single infection, which might be described as ‘microheterogeneity’ [6], has been noted among multiple isolates from one person, either from separate specimens collected over time from an infected site, or collected concurrently from different sites, or even among different colonies from a single culture. Moreover, such variation has been demonstrated by several methods, including Southern blot analysis [6, 7] and PFGE [4, 5], as well as PCR-based techniques. Different typing systems examine different genotypic characteristics which can evolve independently and at different rates, colloquially referred to as ‘different molecular clock speeds’. Consequently, the same set of isolates from an individual patient or outbreak may or may not appear heterogeneous when analysed by different typing systems [6].
These observations have immediate implications for the design and interpretation of molecular epidemiology studies. Reference laboratories that perform typing and susceptibility testing need to consider the provenance of the cultures submitted by the clinical microbiology laboratories that processed the primary specimen. A recent survey of 43 clinical laboratories in the Netherlands indicated that 38 sites sent organisms that represented a direct pool or subculture of multiple colonies, whereas five laboratories prepared subcultures of M. tuberculosis from single colonies [7].

The consequences of such a procedure were documented by De Boer et al. while investigating low-intensity fragments observed in Southern blots probed with IS6110 [7]. Overall, 94 (7.6%) of 1277 cultures had one and occasionally more low-intensity fragments in their IS6110 RFLP patterns. Southern blots prepared from subcultures of individual colonies of these preparations did not have the low-intensity fragment, but instead showed either a normal intensity fragment or no hybridisation signal at all at that position. Of note, no such fragments were observed for any of the isolates from the laboratories that submitted single colony subcultures. In the entire series, there were no composite patterns representing two distinctly different profiles, suggesting that concurrent infection with two or more different M. tuberculosis strains is rare, although some have been reported [8]. Other studies of serial isolates from individual patients with pulmonary tuberculosis have estimated that IS6110 restriction profiles vary in vivo by single band variations with a ‘half-life’ of the order of 3–5 years [9, 10].

Given that molecular typing results may demonstrate even low levels of divergence in ‘real time’, then determining ‘bc’ molecular strain type of an infecting pathogen becomes both a statistical and technical challenge. There are at least three strategies available. (a) Type a pure subculture of a single colony. This is the standard approach for routine bacterial pathogens and, while preparing such subcultures for mycobacteria may require greater effort and time, it effectively eliminates the problem of low-intensity fragments [7]. There is of course a small, but real likelihood that the single colony will itself be a variant, thereby confounding the complete detection of a cluster or outbreak. (b) Type multiple subcultures, each derived from an independent colony. This provides a statistically robust, quantifiable assessment of the level of variation actually present in the culture. Unfortunately, it encumbers the laboratory with substantial additional work and is probably warranted only in the context of specific research studies. (c) Perform typing on a single subculture representing a pool of multiple primary colonies. This is, at least initially, a simple approach. If an unambiguous strain type is observed, then presumably variants are infrequent among the organisms sampled. However, if low-intensity fragments are present, then appreciable effort may be needed to subculture pure genotypes. Alternatively, computer software for analysing restriction profiles (e.g., Bionumerics, Applied Maths, Kortrijk, Belgium) can be used to compare the intensity of each fragment with the overall mean intensity of the profile. Ignoring bands whose intensity is <15% of the mean intensity of bands in a fingerprint pattern effectively edits profiles with low-intensity fragments [7]. However, any such automated interpretation entails a risk of error should there be variation due to unexpected technical factors rather than true genotypic alteration. Therefore, visual checks of the computer interpretation of fingerprints remain of the utmost importance.

As future studies include larger populations and extend over longer periods of time, we can expect to observe genotypic variation more frequently among isolates that are epidemiologically related. Unfortunately, the lexicon of molecular typing remains poorly equipped for describing typing results that are ‘similar’. The most common approaches – e.g., ‘types and subtypes’ or ‘probably or possibly related’ – are applicable primarily to the modest sets of isolates collected over relatively limited time spans [2, 11]. The combination of digital scanning and sophisticated computer programmes greatly facilitates the analysis of larger datasets, but these technologies are themselves associated with new and often unresolved complexities. For example, independent internal standards are needed to determine both mol. wt and reproducibility, and seemingly obscure programme parameters may have significant and unappreciated impact on results. Thus, substantial inaccuracies may lurk behind the apparently precise and objective computer output. New approaches are needed for describing such variation and providing clear molecular and epidemiological interpretations, e.g., standardised algorithms to include low-intensity bands in the computer-assisted analysis, statistical models for analysing restriction profiles [12], or perhaps more quantifiable typing methods such as multilocus sequencing [13]. At this time, restricting epidemiological clusters of tuberculosis cases to sets of isolates with indistinguishable genotypes remains a sound, if conservative, interpretation.

An interesting unresolved question is whether the genotypic variations observed during the course of infection have any pathogenic significance. For example, IS6110 profiles with low-intensity bands representing variant M. tuberculosis genotypes were detected more frequently among cultures from older patients [7], suggesting that latent infection or the transition from latent to active infection may be associated with evolutionary divergence in vivo and, perhaps, adaptation. Potentially, some IS6110 transpositions may alter gene expression and thereby directly facilitate reactivation or confer other selective advantage. Future studies combining molecular strain typing and quantification of gene expression by use of genomic DNA micro arrays [14] will be required to address these issues.
DICK VAN SOOLINGEN and ROBERT D. ARBEIT*
Reference Laboratory for Mycobacteriology,
National Institute of Public Health and the Environment,
PO Box 1, 3720 BA Bliethoven, The Netherlands
and 3Cubist Pharmaceuticals,
Cambridge, MA 02130, USA
Corresponding author: Dr D. van Soolingen
(e-mail: d.van.soolingen@rivm.nl)

References
2. Arbeitz RD. Laboratory procedures for the epidemiologic analysis of microorganisms. In: Murray PR, Baron EJ, Pfaller
5. Masslow JN, Brequer S, Gunn J, Darbin A, Barlow MA, Arbeitz RD. Variation and persistence of methicillin-resistant Staphy-
11. Tenover FC, Arbeitz RD, Goering RV et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-