MOLECULAR CHARACTERISTICS OF THE M. TUBERCULOSIS LAM-RUS FAMILY PREVALENT IN CENTRAL RUSSIA.

NOTE

Svetlana Dubiley¹*, Eugene Kirillov ¹, Anna Ignatova¹, Valentina Stepanshina², Igor Shemyakin².

¹Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, 16/10 Miklukho-Maklaya, Moscow 117997, Russia and ²State Research Center for Applied Microbiology and Biotechnology, Obolensk, Moscow region, Russia.

Running title: M. tuberculosis LAM-RUS family

*Author for correspondence: Svetlana Dubiley, IBCH RAS, 16/10 Miklukho-Maklaya, Moscow 117997, Russia; phone: +7(495)429-8269; fax: +7(495)335-7103; e-mail: lana@ibch.ru.

Key words: Mycobacterium tuberculosis, strain, epidemiology, genotyping, LAM-RUS
ABSTRACT

We analyzed IS6110-associated polymorphisms in the phospholipase C genes of 107 isolates of *Mycobacterium tuberculosis* selected to be representative of isolates circulating in central Russia. We found that the majority LAM family strains contained an insertion in a unique position in the *plcA* gene, suggesting a common ancestor. This insertion can serve as a specific genetic marker for this group, which we designate LAM-RUS.
Russia is a country with a high prevalence of tuberculosis (TB) with an estimated 119 new cases per 100,000 population in 2005 (18). Although the incidence rate has begun to slowly decrease during past few years, the number of TB-related deaths in Russia remains substantial. It is generally accepted that different *Mycobacterium tuberculosis* strains have distinctive epidemiological and clinical characteristics: virulence, clinical presentation, and behavior in animals appear to be strain-dependent (7). Some *M. tuberculosis* strains are noted for their wide dissemination and acquisition of drug resistance (4). As a result, there has been increased attention paid to *M. tuberculosis* strain identification recently.

Phospholipases C (PLC) have been shown to be involved in *M. tuberculosis* virulence (9). Four genes, encoded by genes *plcA*, *plcB*, *plcC*, and *plcD*, encode phospholipase C proteins. It has been shown that *plc* genes in clinical isolates are often interrupted with IS6110 insertion elements (15; 17). The *plcD* gene was suggested to be a hot spot for IS6110 integration; insertions into the *plcABC* are less common (15). In this study, we analyzed the frequency of IS6110 insertions into the phospholipase C operon (*plcABC* genes) among clinical isolates of *M. tuberculosis* from Central Russia.

A set of strains was selected from a collection of *M. tuberculosis* clinical isolates maintained at the State Research Center for Applied Microbiology and Biotechnology, Obolensk, Russia. The collection contains more than 1200 isolates obtained during 1998-2006 from patients in Moscow city and in the Moscow, Tula, and Kaluga regions. All strains with a unique spoligotype (57 of a total 72 spoligotypes) were included into the sample. Among the strains with shared
spoligotypes, two to eight epidemiologically unlinked strains were chosen for analysis. The resulting panel of 107 strains should be representative of the diversity of *M. tuberculosis* strains circulating in the region.

IS6110 insertions in the *plcABC* locus were detected by measuring the sizes of six overlapping PCR-amplicons that cover the entire locus. Of the 107 isolates analyzed, 47 contained an IS6110 insertion. In 45 of the 47 isolates, the IS6110 element was inserted at the same position and in the same orientation within the *plcA* gene as determined by sequencing. One additional isolate of the 107 contained an insertion in the *plcC* gene (Table 1).

The genetic relatedness of the isolates in the sample was assessed using five genotyping methods: IS6110-RFLP (16), spoligotyping (6), MIRU-VNTR typing (14), SNP-based identification of principal genetic groups, PGGs (13), and SNP clustered groups, SCGs (3). Genetic distance calculations and phylogenetic trees construction were performed using Mathlab 6.5 Statistics and Bioinformatics Toolboxes. The genetic characteristics of the strains are listed in Supplementary Figure S1.

The distribution of isolates with *plcABC::IS6110* insertions between the PGGs and SCGs groups was uneven (Table 1). Both of the isolates with the Ins1 or Ins3 IS6110 insertions were members of SCG6/PGG3. All but two isolates from the SCG5/PGG2 group contained the Ins2 insertion in the *plcA*.

A dendrogram of the 105 unique IS6110-fingerprint patterns found in the 107 isolates was constructed (Supplementary Figure S1). Ins2-containing strains showed highly similar IS6110-fingerprint patterns and were grouped together on the dendrogram. An unrooted phylogenetic
tree of 58 MIRU-VNTR genotypes was constructed using the neighbour-joining (NJ) method. The tree is shown in Figure 1 and the groups are colored according to their SCG/PGG genotype. The Ins2-containing isolates form a compact, discrete group on the dendrogram. In addition, the SCG5 isolates with uninterrupted plcA gene showed high similarity to some of the SCG6 strains.

The small genetic distance across the Ins2-containing group and the presence of an IS6110 element at a unique site of integration suggest that the strains are derived from the common progenitor. The combination of IS6110-RFLP, spoligotyping, and SCG/PGG genotyping methods allowed recognizing this group as being related to the Latin American–Mediterranean (LAM) family (1). The M. tuberculosis LAM family was initially described as a clade on a NJ tree of spoligotypes (12). However, as data accumulated, the genetic heterogeneity of this family became evident which suggested the presence of a number of distinct LAM groups in various geographical regions (2; 8; 10). To reflect the high prevalence of this type of strain in Central Russia as well as its epidemiological importance, we propose to name the group of Ins2-containing strains as the LAM-RUS family and to use Ins2 as a specific genetic marker for the group.

To identify members of the LAM-RUS family, a multiplex PCR assay was developed. In brief, two pairs of primers were used in the PCR reaction: U (universal plcA-specific forward primer, 5’gaagttgattcgcgccgtt), R (universal plcA-specific reverse primer, 5’gctgggagtcccgcggacg), H (hybrid forward primer, specific to the site of Ins2 IS6110 insertion in the plcA gene, 5’ccaactcagaaaccaactgaacc), and I (reverse IS6110-specific primer, 5’ctcgaatctgctgaccgc). The amplification reaction was performed in a total volume of 30 µl containing ~1 ng DNA template, 200 µM of each dNTPs, 0.3 µM of each primer, and 2 U Taq DNA polymerase.
(Fermentas, Lithuania) in the buffer recommended by the manufacturer. LAM-RUS strains produced two amplicons (163 and 410 bp), while strains without an insertion produced a product of 268 bp. Strains with the other insertions/deletions in the analyzed region produced PCR products of various lengths.

Although further study is needed to determine the actual prevalence of the LAM-RUS family, the high degree of clustering, high frequency of being MDR, and increased risk of outbreaks associated with these strains have already been reported. About 45% and 30% of \( M. \text{tuberculosis} \) isolates collected in prison hospitals in Tula and Moscow regions respectively were members of the newly defined LAM-RUS family (5; 11). More than 80% of these isolates were MDR and 68% of them were resistant to four or five drugs tested (rifampicin, isoniazid, streptomycin, and ethambutol or kanamycin). Wide dissemination of these strains in two remote penitentiaries implies that we may face an outbreak of XDR TB in the region before long.

Molecular methods are gradually establishing their role in studies of TB epidemiology. The availability of convenient strain- and lineage-specific PCR markers simplifies strain classification, speeds up the identification of clinically relevant strains, and may give a hint about the nature of the disease before microbiological and drug resistance tests can be completed. Epidemiological studies benefit from efficient tools for unambiguous strain differentiation and tracking. We believe that description of the specific genetic characteristics of LAM-RUS \( M. \text{tuberculosis} \) family will contribute to the TB control programs in the region and worldwide.

This work was supported by BTEP63/ISTC2628 and CRDF2718. We are grateful to Dr. Thomas M. Shinnick from CDC, Atlanta, GA and Dr. Richard O’Brien from Foundation for Innovative New Diagnostics for critical reading of the manuscript.
REFERENCES


### Table 1. Prevalence of \( \text{plcABC}::\text{IS6110} \) genotype in different phylogenetic groups.

<table>
<thead>
<tr>
<th>Phylogenetic groups</th>
<th>Total isolate no.</th>
<th>( \text{plcABC}::\text{IS6110} ) genotype</th>
<th>Position of insertion ( ^c )</th>
<th>Adjacent sequence ( ^d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGG1(^a)/SCG2(^b)</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PGG2/SCG3</td>
<td>30</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PGG2/SCG5</td>
<td>47</td>
<td>45 (Ins2, ( \text{plcA} ))</td>
<td>2630571</td>
<td>cgggtgtggttgtttc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 (Ins1, ( \text{plcA} ))</td>
<td>2630703</td>
<td>gcaacggggggtggtgctc</td>
</tr>
<tr>
<td>PGG3/SCG6</td>
<td>20</td>
<td>1 (Ins3, ( \text{plcC} ))</td>
<td>2628680</td>
<td>ttagccagccaggaat</td>
</tr>
</tbody>
</table>

\(^a\)Principal Genetic Groups; \(^b\)SNP clustered groups; \(^c\)positions are numbered according to \( M. \text{tuberculosis} \) H37Rv genome; \(^d\)duplicated nucleotides are shown in bold.
FIGURE LEGEND.

Figure 1. Distribution of Ins2-containing strains on phylogenetic tree based on their MIRU-VNTR patterns. PGGs and SCGs are color-coded. LAM-RUS strains are circled. *Strains containing Ins2 IS6110 insertion.

SUPPLEMENTARY FIGURE LEGEND.

Figure S1. Genetic characteristics of M. tuberculosis clinical isolates. Isolates are grouped according to the dendrogram of IS6110-RFLP patterns shown on the left. The dendrogram was constructed using the unweighted pair group method with arithmetic averages algorithm and Jaccard’s distance matrix with 1% band position tolerance. PGG/SCG groups are highlighted with color. Two isolates without IS6110 RFLP patterns are listed in the bottom.