Abstract
Objectives: Evaluate the association between rifampicin resistance and the presence of at least one SNP in the rpoB and ponA1 genes and the spoligotype defined lineages.

Material and Methods: This study analyzed two databases of 484 genomes of M. tuberculosis from strains isolated from patients in the cities of Lima and Callao, for which the odds ratio (OR) was calculated considering belonging to a certain spoligotype defined lineages as an exposure factor.

Results: No statistically significant association (p value > 0.05) was found between the presence of at least one SNP in the rpoB gene and the lineages included in the study (LAM, Haarlem, T and Beijing). However, a statistically significant association was found between the presence of at least one SNP in the ponA1 gene and the LAM and Haarlem lineages (p value < 0.05). An association was found between the P631S SNP in the ponA1 gene and the LAM and Haarlem lineages; and the A516T SNP, of this same gene, presented an association with the LAM lineage. Likewise, an association was found between rifampicin resistance and the LAM lineage.

Conclusions: The presence of SNPs in the ponA1 gene is associated with the LAM and Haarlem lineages.

Keywords: Spoligotypes; Mycobacterium tuberculosis; rpoB gene; ponA1 gene; rifampicin resistance.
On the other hand, ponA1 gene encodes the penicillin binding protein PonA1, which participates in peptidoglycan biosynthesis and regulates the enzymatic activity of other proteins involved in bacterial morphology and growth. Some SNPs (T34D and Q365H) have been reported to alter the tolerance of M. tuberculosis to rifampicin and increase the minimum inhibitory concentration (MIC) almost twice. Likewise, a recent study found that some SNPs (P631S and A516T) are associated with rifampicin resistance and suggested these SNPs could alter the rifampicin tolerance of M. tuberculosis. Nevertheless, a causal association between rifampicin resistance and the presence of these SNPs could not be established. For these reasons, this study included the ponA1 gene in the analysis since we consider that it plays a role in rifampicin resistance, and could be associated with some spoligotypes defined lineages.

At the end of the last century, a new method was developed for the genotyping of M. tuberculosis strains, which was called spacer oligonucleotide typing or spoligotyping, based on the presence or absence of certain spacer sequences in the direct repeat region (DR). The DR region is a DNA segment made up of a 36 bp repeat sequence and several 31 to 41 bp non-repeating DNA segments called spacer sequences. 43 spacer sequences were identified in M. tuberculosis H37Rv and made it possible to group the strains of this bacteria into certain lineages. Several studies have reported an association between the presence of certain lineages identified by spoligotyping and resistance to certain drugs, for example, most of the Beijing lineage M. tuberculosis strains that have been studied presented drug resistance and this frequency was higher compared to strains of other lineages. Therefore, a probable causal association between the presence of certain lineages and drug resistance has been suggested; however, this probable causal association could not be confirmed.

Our study hypothesizes that some lineages would be associated with the presence of SNPs in genes associated with drug resistance such as rpoB gene and genes involved in some drug tolerance mechanism such as ponA1 gene. In this sense, this study analyzed two secondary database of 484 M. tuberculosis genomes in order to identify an association between the presence of SNPs on rpoB and ponA1 genes and the presence of certain spoligotypes defined lineages.

### Materials and methods

Our study performed a statistical analysis of a secondary database, which comes from two previous studies of 484 genomes obtained from strains of M. tuberculosis, which were isolated from patients with active tuberculosis in the cities of Lima and Callao. The genomic DNA extracted from M. tuberculosis was sequenced using Illumina HiSeq2000. High quality sequencing reads were assembled taking into account the genome of the reference strain M. tuberculosis H37Rv using NextGENe software. The alignment of the sequences with this reference genome made it possible to identify SNPs, insertions and deletions using NextGENe software. These studies identified the lineage (clade and SIT code) of each strain.

The rpoB and ponA1 genes were classified considering each lineage separately as an exposure variable and the presence or absence of at least one SNP on rpoB and ponA1 genes, both genes were studied separately, as outcome variables. This analysis was performed to determine the association between lineages defined by spoligotypes and the presence of at least one SNP on rpoB and ponA1 genes or the absence of SNP in these genes (Table 1).

<table>
<thead>
<tr>
<th>Lineage “Y”</th>
<th>With at least one SNP in an “X” gene</th>
<th>Without any SNP in an “X” gene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Other lineages different from the “Y” lineage</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

Where “X” can be rpoB gene or ponA1 gene. “Y” represents any of the lineages included in this study.

Our study also carried out an analysis where it considered rifampicin resistance as an outcome variable (table 2). Furthermore, we included in our study SNPs with a frequency greater than 5% of the ponA1 gene, for which SNPs were considered separately as outcome variables.

### Results

Approximately 63.64% of the genomes analyzed in this study presented at least one SNP in ponA1 gene and 72.52% presented at least one SNP in rpoB gene. Therefore, it could be inferred that the SNPs of these genes are relatively frequent in M. tuberculosis strains isolated from patients with active tuberculosis in the cities of Lima and Callao. Furthermore, 71.69% of these genomes come from rifampicin-resistant strains.

Our study found that the analyzed genomes was grouped into 4 lineages (LAM, Haarlem, T and Beijing). If we consider the classifications described in the methodology, we can observe that the LAM lineage contained the highest number of genomes in the three proposed classifications. However, the Haarlem lineage was presented in approximately 29.9% of genomes with at least one SNP on ponA1 gene (figure 1). Also 16.2% and 15.9% of genomes with at least on SNP on rpoB gene (figure 2) and genomes derived from rifampicin-resistant strains (figure 3), respectively; were found in the Haarlem.
lineage. Only 4.2% of genomes of the Haarlem lineage did not present at least one SNP on ponA1 gene. Hence this fact could be suggesting that there is an association between the presence of SNPs on ponA1 gene and the Haarlem lineage.

This study found a statistical association between the presence of at least one SNP on ponA1 gene and the Haarlem and LAM lineages (p-value<0.0001). Likewise, the absence of SNPs on ponA1 gene is associated with the T and Beijing lineages (p-value<0.001) (table 3). Furthermore, the rifampicin resistance is associated with the LAM lineage (p-value<0.01) (table 3).

Our study found an association between the presence of P631S SNP and the Haarlem and LAM lineages (p-value <0.0001). Likewise, SNP A516T SNP presented an association with the LAM lineage (p-value <0.001) (table 4). Other SNPs of the ponA1 gene were not included in the analysis since their frequency was less than 5% of the total SNPs.

Discussion

Since the advent of spoligotyping, it has been possible to analyze the spacer sequences of the DR region of M. tuberculosis strains, all over the world. Which has provided considerable information on the distribution of the lineages defined by spoligotypes. In this sense, various studies have found a possible association between certain lineages and resistance to first-line drugs. Which has opened a debate on whether these associations are causal or results of chance. Our study evaluated whether the presence of at least one SNPs on rpoB and ponA1 genes or rifampicin resistance are associated with some spoligotype defined lineages.

More than 96% of rifampicin resistant strains have been reported to have SNPs in an 81bp region of rpoB gene. Likewise, studies conducted in China and Russia have found a probable association between rifampicin resistance and the Beijing lineage. An association between this lineage and multidrug-resistant tuberculosis (MDR-TB) has also been found in India and Vietnam. Other lineages have also been associated with rifampicin resistance; for example, a study in Brazil reported the association between the LAM and T lineages and resistance to rifampicin and isoniazid. Our study found that LAM lineage is associated with rifampicin resistance and the absence of SNPs on rpoB gene is associated with the Haarlem lineage. We did not find a statistically significant association between the presence of at least one SNP in rpoB gene and the lineages included in the study. This could be due to the fact that there is no causal relationship between rifampicin resistance and the most frequent lineages reported in Peru. A study carried out in Peru did not find an association between resistance to drugs, including rifampin, and the lineages found in this country. However, we found an association between rifampicin resistance and LAM lineage. This could be due to the fact that not all SNPs in the rpoB gene cause resistance to rifampin and there is almost 5% of cases of resistance to this drug that are not associated with the rpoB gene. Other studies, conducted in Mexico, Uganda and Italy also found no association between drug resistance and spoligotype defined lineages.

On the other hand, in vitro studies have reported that some SNPs (T34D and Q365H) on ponA1 gene increase the MIC of rifampicin to almost double, compared to strains of wild-type M. tuberculosis. Likewise, a study analyzed 914 genomes of M. tuberculosis from Peruvian strains and found that SNPs P631S and A516T are associated with resistance to this drug; this study found that SNP A516T is more common in rifampicin-resistant strains. Our study found that the presence of at least one SNP on ponA1 gene is associated with the LAM and Haarlem lineages and the absence of SNPs on this gene is associated with the Beijing and T lineages. Furthermore, we found that the most common SNP in ponA1 gene (P631S) is associated with LAM and Haarlem lineages. The A516T SNP also is associated with...
LAM lineage. Our study suggests that this could be due to the fact that the strains of this lineage developed in a medium that allowed the selection of SNPs on *ponA1* gene, especially SNPs that alter bacterial morphology and growth. It has been suggested that the SNP P631S could generate these alterations; likewise, it was the most frequent SNP found in Peru\(^7\).

On the other hand, associating the absence of a SNP to a group, as this study did, might not be appropriate since we could be dealing with a conserved gene or with low evolutionary pressure. However, the genes evaluated present a high frequency of SNPs both in our study and in previous studies.

Most of the genomes included in our study come from MDR-TB strains; hence, our results could be due to resistance to other drugs or related to genes associated with other drugs. However, we present the first evidence that suggests an association between the presence of at least one SNP on *ponA1* gene and the LAM and Haarlem lineages.

**Ethical disclosures**

**Protection of human and animal subjects.** The authors declare that no experiments were performed on humans or animals for this investigation.

**Right to privacy and informed consent.** The authors declare that no data that enables identification of the patients appears in this article.

**Conflict of interest.** The authors declare that the revision was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**References**


---

### Table 3. Statistical analysis of rifampicin resistance, rpoB and *ponA1* genes.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Rifampicin-resistant</th>
<th>rpoB gene</th>
<th><em>ponA1</em> gene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>CI</td>
<td>p-value</td>
</tr>
<tr>
<td>Haarlem</td>
<td>0.4410*</td>
<td>0.2769-0.7025</td>
<td>0.0006</td>
</tr>
<tr>
<td>Beijing</td>
<td>0.969</td>
<td>0.4922-1.9075</td>
<td>0.9274</td>
</tr>
<tr>
<td>T</td>
<td>0.8717</td>
<td>0.4938-1.5388</td>
<td>0.6358</td>
</tr>
<tr>
<td>LAM</td>
<td>1.9132*</td>
<td>1.2827-2.8536</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

*Statistical significance. CI: Confidence Interval.

---

### Table 4. Statistical analysis of *ponA1*’s SNPs.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>P631S</th>
<th>A516T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>CI</td>
</tr>
<tr>
<td>Haarlem</td>
<td>9.7184</td>
<td>4.5873-20.5892</td>
</tr>
<tr>
<td>LAM</td>
<td>2.2763</td>
<td>1.5673-3.3061</td>
</tr>
<tr>
<td>T</td>
<td>0.1363</td>
<td>0.0729-0.2547</td>
</tr>
<tr>
<td>Beijing</td>
<td>0.0054</td>
<td>0.0003-0.0888</td>
</tr>
</tbody>
</table>

*Statistical significance. CI: Confidence Interval.

---


