ACIN
ASOCIACIÓN COLOMBIANA DE INFECTOLOGÍA

infectio

ARTÍCULO ORIGINAL

Statistical analysis of 484 *Mycobacterium tuberculosis* genomes reveals an association between single nucleotide polymorphisms on ponA1 gene and LAM and Haarlem lineages

Jhonatan Rabanal[1,2,*]; Mirko Zimic[1,3]

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 (ρ 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 (ρ 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.

Análisis estadístico de 484 genomas de *Mycobacterium tuberculosis* revela una asociación entre los polimorfismos de un solo nucleótido en el gen *ponA1* y los linajes LAM y Haarlem

Resumen

Objetivos: Evaluar la asociación entre la resistencia a rifampicina y la presencia de al menos un SNP en los genes *rpoB* y *ponA1* y los linajes definidos por espoligotipos.

Material y Métodos: Este estudio analizó dos bases de datos de 484 genomas de *M. tuberculosis* de cepas aisladas de pacientes de las ciudades de Lima y Callao, para lo cual se calculó el odds ratio (OR) considerando la pertenencia a determinado linaje definido por espoligotipos como un factor de exposición.

Resultados: No se encontró una asociación estadísticamente significativa (valor de ρ >0.05) entre la presencia de al menos un SNP en el gen *rpoB* y los linajes incluidos en el estudio (LAM, Haarlem, T y Beijing). No obstante, se halló una asociación estadísticamente significativa entre la presencia de al menos un SNP en el gen *ponA1* y los linajes LAM y Haarlem (valor de ρ <0.05). Se encontró una asociación entre el SNP P631S del gen *ponA1* y los linajes LAM y Haarlem; y el SNP A516T, de este mismo gen, presentó una asociación con el linaje LAM. Asimismo, se halló una asociación entre la resistencia a rifampicina y el linaje LAM.

Conclusiones: La presencia de SNPs en el gen ponA1 está asociada con los linajes LAM y Haarlem.

Palabras claves: Espoligotipos; *Mycobacterium tuberculosis*; gen *rpoB*; gen *ponA1*; resistencia a rifampicina.

Introduction

Tuberculosis, which is caused by *Mycobacterium tuberculosis*, is one of the deadliest infectious diseases in the world and represents a public health problem[1]. In recent years, the number of strains of *M. tuberculosis* resistant to first-line drug, such as rifampicin, has been increasing. Rifampicin plays an important role in the first line of tuberculosis treatment due to its powerful bactericidal effect. This drug inhibits the synthesis of messenger RNA by binding to RNA polymerase. Rifampicin resistance is caused by single nucleotide polymorphisms (SNPs) on *rpoB* gene, which encodes the beta subunit of RNA polymerase[2].

1 Universidad Peruana Cayetano Heredia, Lima, Peru
2 https://orcid.org/0000-0002-6534-7406
3 https://orcid.org/0000-0002-7203-8847
* Autor para correspondencia:
Correo electrónico: jhonatan.rabanal@upch.pe

Recibido: 04/07/2021; Aceptado: 11/09/2021

Cómo citar este artículo: J. Rabanal, *et al*. Statistical analysis of 484 *Mycobacterium tuberculosis* genomes reveals an association between single nucleotide polymorphisms on ponA1 gene and LAM and Haarlem lineages. Infectio 2022; 26(2): 168-171

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[3,4]. 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[5,6]. 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[7]. 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[8]. Several studies have reported an association between the presence of certain lineages identified by spoligotyping and resistance to certain drugs[9-11]; 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[12].

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[13,14] 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[13,14].

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 1. Classification of genomes according to established criteria.

	With at least one SNP in an "X" gene	Without any SNP in an "X" gene
Lineage "Y"	A	B
Other lineages different from the "Y" lineage	C	D

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.

Table 2. Classification of genomes according to rifampicin resistance.

	Rifampicin-resistant	Rifampicin-susceptible
Lineage "Y"	A	B
Other lineages different from the "Y" lineage	C	D

"Y" represents any of the lineages included in this study.
The odds ratio (OR) was then calculated for each lineage included in this study. A significance level of 95% was considered and the statistical calculations were carried out using programming language R.

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[12]. In this sense, various studies have found a possible association between certain lineages and resistance to first-line drugs[9-11]. Which has opened a debate on whether these associations are causal or results of chance[12]. 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[15]. Likewise, studies conducted in China and Russia have found a probable association between rifampicin resistance and the Beijing lineage[16,17]. An association between this lineage and multidrug-resistant tuberculosis (MDR-TB) has also been found in India and Vietnam[9,18]. Other lineages have also been associated with rifampicin resistance[12]; for example, a study in Brazil reported the association between the LAM and T lineages and resistance to rifampicin and isoniazid[19]. 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[20]. 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 [21,22,23].

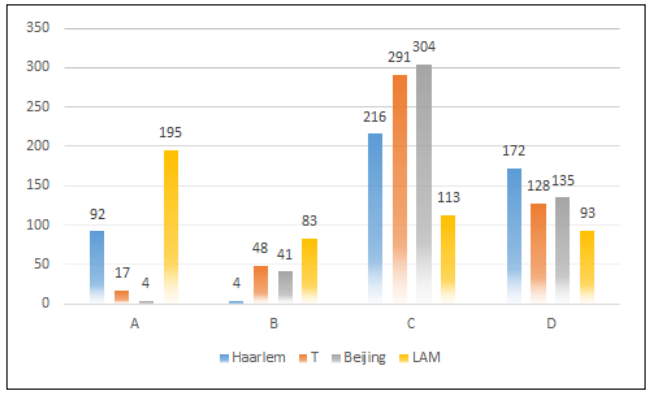

Figure 1. Frequency of spoligotypes defined lineages considering the presence or absence of at least one SNP on *ponA1* gene.

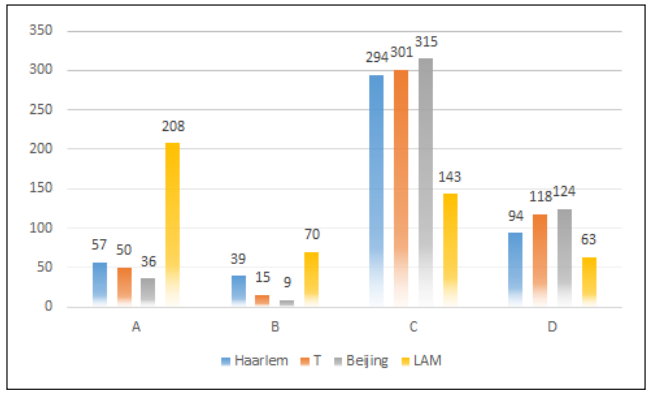

Figure 2. Frequency of spoligotypes defined lineages considering the presence or absence of at least one SNP on *rpoB* gene.

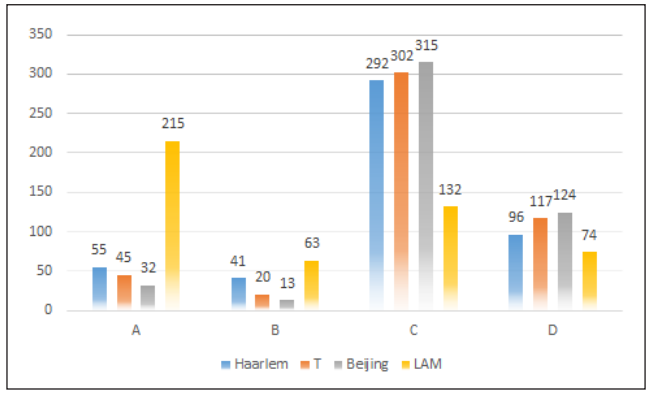

Figure 3. Frequency of spoligotypes defined lineages considering the considering resistance or susceptibility to rifampicin.

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*[5,6]. 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[7]. 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

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

Lineage	Rifampicin-resistant			*rpoB* gene			*ponA1* gene		
	OR	CI	p-value	OR	CI	p-value	OR	CI	p-value
Haarlem	0.4410*	0.2769-0.7025	0.0006	0.4673*	0.2924-0.7468	0.0015	18.3148*	6.5982-50.8373	<0.0001
Beijing	0.969	0.4922-1.9075	0.9274	1.5746	0.7368-3.3650	0.2413	0.0433*	0.0152-0.1234	<0.0001
T	0.8717	0.4938-1.5388	0.6358	1.3068	0.7065-2.4171	0.3939	0.1558*	0.0863-0.2813	<0.0001
LAM	1.9132*	1.2827-2.8536	0.0015	1.3091	0.8762-1.9558	0.1886	1.9336*	1.3278-2.8158	0.0006

*Statistical significance. CI: Confidence Interval.

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

1. MacNeil A, Glaziou P, Sismanidis C, et al. Global Epidemiology of Tuberculosis and Progress Toward Meeting Global Targets - Worldwide, 2018. MMWR Morb Mortal Wkly Rep. 2020; 69(11):281-85.
2. Zaw MT, Emran NA, Lin Z. Mutations inside rifampicin-resistance determining region of rpoB gene associated with rifampicin-resistance in Mycobacterium tuberculosis. J Infect Public Health. 2018; 11(5):605-10.
3. Gao B, Wang J, Huang X, Sha W, et al. The dynamic region of the peptidoglycan synthase gene, Rv0050, induces the growth rate and morphologic heterogeneity in Mycobacteria. Infect Genet Evol. 2019;72: 86-92.
4. Raghavendra T, Patil S, Mukherjee R. Peptidoglycan in Mycobacteria: chemistry, biology and intervention. Glycoconj J.2018; 35(5): 421-32.
5. Farhat MR, Shapiro BJ, Kieser KJ, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. Nat Genet. 2013;45(10):1183-89.
6. Kieser KJ, Boutte C, Kester JC, et al. Phosphorylation of the Peptidoglycan Synthase PonA1 Governs the Rate of Polar Elongation in Mycobacteria. PLoS Pathog. 2015;11(6): e1005010.

Table 4. Statistical analysis of ponA1's SNPs.

Lineage	P631S			A516T		
	OR	CI	p-value	OR	CI	p-value
Haarlem	9.7184	4.5873-20.5892	<0.0001	0.0166	0.0010-0.2701	0.004
LAM	2.2763	1.5673-3.3061	<0.0001	204.8391	12.6271-3322.9507	0.0002
T	0.1363	0.0729-0.2547	<0.0001	0.027	0.0017-0.4409	0.0112
Beijing	0.0054	0.0003-0.0888	0.0003	0.0413	0.0025-0.6765	0.0255

*Statistical significance. CI: Confidence Interval.

7. Rabanal-Sanchez J. Potenciales mutaciones compensatorias asociadas a la resistencia a rifampicina en *Mycobacterium tuberculosis* [Tesis de Maestría], Lima, Universidad Peruana Cayetano Heredia, 2020.
8. Driscoll JR. Spoligotyping for molecular epidemiology of the Mycobacterium tuberculosis complex. Methods Mol Biol. 2009; 551:117-28.
9. Desikan P, Chauhan DS, Sharma P, et al. Clonal diversity and drug resistance in *Mycobacterium tuberculosis* isolated from extra-pulmonary samples in central India – A pilot study. Indian J Med Microbiol. 2014;32(4):434-7.
10. Augustynowicz-Kopec E, Jagielski T, Zwolska Z. Genetic diversity of isoniazid-resistant *Mycobacterium tuberculosis* isolates collected in Poland and assessed by spoligotyping. J Clin Microbiol. 2008;46 (12): 4041-4.
11. Kisa O, Tarhan G, Gunal S, et al. Distribution of spoligotyping defined genotypic lineages among drug-resistant *Mycobacterium tuberculosis* complex clinical isolates in Ankara, Turkey. PLoS One. 2012; 7(1): e30331.
12. Panwalkar N, Chauhan DS, Desikan P. Spoligotype defined lineages of *Mycobacterium tuberculosis* and drug resistance: Merely a casual correlation? Indian J Med Microbiol. 2017; 35(1):27-32.
13. Sheen P, Requena D, Gushiken E, et al. A multiple genome analysis of Mycobacterium tuberculosis reveals specific novel genes and mutations associated with pyrazinamide resistance. BMC Genomics 2017;18(1):769.
14. Grandjean L, Gilman RH, Iwamoto T, et al. Convergent evolution and topologically disruptive polymorphisms among multidrug-resistant tuberculosis in Peru. PLoS One. 2017;12: e0189838.
15. Kumar S & Jena L. Understanding Rifampicin Resistance in Tuberculosis through a Computational Approach. Genomics Inform. 2014;12(4): 276-82.
16. Guo YL, Liu Y, Wang SM, et al. Genotyping and drug resistance patterns of *Mycobacterium tuberculosis* strains in five provinces of China. Int J Tuberc Lung Dis.2011;15(6):789-94.
17. Mäkinen J, Marjamäki M, Haanperä-Heikkinen M, et al. Extremely high prevalence of multidrug resistant tuberculosis in Murmansk, Russia: A population-based study. Eur J Clin Microbiol Infect Dis. 2011; 30(9):1119-26.
18. Buu TN, Huyen MN, van Soolingen D, et al. The *Mycobacterium tuberculosis* Beijing genotype does not affect tuberculosis treatment failure in Vietnam. Clin Infect Dis. 2010;51(8):879-86.
19. Perizzolo PF, Dalla-Costa ER, Ribeiro AW, et al. Characteristics of multidrug-resistant Mycobacterium tuberculosis in Southern Brazil. Tuberculosis (Edinb). 2012; 92(1):56-9.
20. Taype CA, Agapito JC, Accinelli RA, et al. Genetic diversity, population structure and drug resistance of *Mycobacterium tuberculosis* in Peru. Infect Genet Evol. 2012; 12(3):577-85.
21. Garzelli C, Lari N, Cuccu B, et al. Impact of immigration on tuberculosis in a low-incidence area of Italy: A molecular epidemiological approach. Clin Microbiol Infect. 2010;16(11):1691-7.
22. Lopez-Alvarez R, Badillo-Lopez C, Cerna-Cortes JF, et al. First insights into the genetic diversity of *Mycobacterium tuberculosis* isolates from HIV-infected Mexican patients and mutations causing multidrug resistance. BMC microbiol. 2010; 10:82.
23. Bazira J, Asiimwe BB, Joloba ML, et al. *Mycobacterium tuberculosis* spoligotypes and drug susceptibility pattern of isolates from tuberculosis patients in South-Western Uganda. BMC Infect Dis. 2011; 11:81.