Tuberculosis 103 (2017) 28–36

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

Single nucleotide polymorphisms may explain the contrasting phenotypes of two variants of a multidrug-resistant *Mycobacterium tuberculosis* strain

María Mercedes Bigi [a], Beatriz Lopez [b], Federico Carlos Blanco [c], María del Carmen Sasiain [d], Silvia De la Barrera [d], Marcelo A. Marti [e], Ezequiel Jorge Sosa [f], Darío Augusto Fernández Do Porto [f], Viviana Ritacco [b], Fabiana Bigi [c, *], Marcelo Abel Soria [a, **]

[a] Universidad de Buenos Aires, Facultad de Agronomía, Cátedra de Microbiología Agrícola.INBA-CONICET, Av. San Martín 4453, C1417DSE, Buenos Aires, Argentina
[b] Instituto Nacional de Enfermedades Infecciosas-ANLIS Carlos Malbrán, Av. Vélez Sarsfield 563, C1282AFF, Buenos Aires, Argentina
[c] Instituto de Biotecnología, Instituto Nacional de Tecnología Agropecuaria (INTA), N. Repetto and De los Reseros, Hurlingham, 1686, Buenos Aires, Argentina
[d] IMEX-CONICET, Academia Nacional de Medicina, José Andrés Pacheco de Melo 3081, C1425AUM, Buenos Aires, Argentina
[e] Departamento de Química Biológica, e IQUIBICEN-CONICET, FCEyN, UBA, Intendente Güiraldes 2160, C1428EGA, Buenos Aires, Argentina
[f] Plataforma de Bioinformática Argentina, Instituto de Cálculo, FCEyN, UBA, Intendente Güiraldes 2160, C1428EGA, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 4 September 2016
Received in revised form
27 December 2016
Accepted 30 December 2016

Keywords:
Mycobacterium tuberculosis
MDR
Haarlem
Genome sequencing
Polymorphisms

ABSTRACT

Globally, about 4.5% of new tuberculosis (TB) cases are multi-drug-resistant (MDR), i.e. resistant to the two most powerful first-line anti-TB drugs. Indeed, 480,000 people developed MDR-TB in 2015 and 190,000 people died because of MDR-TB. The MDR *Mycobacterium tuberculosis* M family, which belongs to the Haarlem lineage, is highly prosperous in Argentina and capable of building up further drug resistance without impairing its ability to spread. In this study, we sequenced the whole genomes of a highly prosperous M-family strain (Mp) and its contemporary variant, strain 410, which produced only one recorded tuberculosis case in the last two decades. Previous reports have demonstrated that Mp induced dysfunctional $CD8^+$ cytotoxic T cell activity, suggesting that this strain has the ability to evade the immune response against *M. tuberculosis*. Comparative analysis of Mp and 410 genomes revealed non-synonymous polymorphisms in eleven genes and five intergenic regions with polymorphisms between both strains. Some of these genes and promoter regions are involved in the metabolism of cell wall components, others in drug resistance and a SNP in *Rv1861, a gene* encoding a putative transglycosylase that produces a truncated protein in Mp. The mutation in Rv3787c, a putative S-adenosyl-L-methionine-dependent methyltransferase, is conserved in all of the other prosperous M strains here analysed and absent in non-prosperous M strains. Remarkably, three polymorphic promoter regions displayed differential transcriptional activity between Mp and 410. We speculate that the observed mutations/polymorphisms are associated with the reported higher capacity of Mp for modulating the host's immune response.

1. Introduction

TB development depends on the host's natural resistance/susceptibility to *Mycobacterium tuberculosis* (*Mtb*) infection and differences in transmissibility, virulence, and immunogenicity among *Mtb* strains. In turn, these latter bacterial factors are determined by the genetic background of the organisms. In this line, certain strains of *Mtb* with special transmission potential are able to manipulate

* Corresponding author. N. Repetto and De los Reseros, Hurlingham, 1686, Buenos Aires, Argentina.
** Corresponding author. Av. San Martín 4453, C1417DSE, Buenos Aires, Argentina.
E-mail addresses: mebigi@hotmail.com (M.M. Bigi), bealopez@anlis.gov.ar (B. Lopez), blanco.federico@inta.gob.ar (F.C. Blanco), mariadelcarmen.sasiain@hotmail.com (M.C. Sasiain), sildelabar@gmail.com (S. De la Barrera), marti.marcelo@gmail.com (M.A. Marti), ezequieljsosa@gmail.com (E.J. Sosa), dariofd@gmail.com (D.A. Fernández Do Porto), vivianaritacco@gmail.com (V. Ritacco), bigi.fabiana@inta.gob.ar (F. Bigi), soria@agro.uba.ar (M.A. Soria).

http://dx.doi.org/10.1016/j.tube.2016.12.007
1472-9792/

host immunity, which could impact on the evolution and/or outcome of the disease [1,2]. Multidrug-resistant tuberculosis (MDR-TB) poses a threat to the control and elimination of TB, which, despite being a curable and preventable disease, is still a major public health problem.

In Argentina, 9605 new cases of TB were reported in 2014, with an incidence of 22.5 per 100,000 inhabitants and MDR-TB was documented in 101 cases. MDR-TB outbreaks emerged in the early 1990s in AIDS patients and thereafter disseminated to immunocompetent individuals [3]. Epidemiological, bacteriological, and genotyping data allowed the identification of certain MDR *Mtb* outbreak strains, such as the M strains of the Haarlem family and the Ra strain of Latin America and the Mediterranean (LAM). Each of these two strains managed to perpetuate in their geographical niches, the metropolitan areas of Buenos Aires and Rosario cities, respectively [4]. In particular, the M strains are still prosperous in the country and can build up further drug resistance without impairing their ability to spread [5]. In contrast, its clonal variant 410 has caused a single tuberculosis case since the onset of the outbreak [6].

Previous studies have demonstrated that Mp, a highly prosperous M family member, and 410 strains modulate the host immune response in different ways [7,8]. Particularly, Mp induced dysfunctional $CD8^+$ cytotoxic T cell (CTL) activity, while 410 elicited a CTL response similar to that of the *M. tuberculosis* reference strain, H37Rv [9,10]. CTL activity has been associated with lysis of *Mtb*-infected macrophages [11] and with reduction in *Mtb* viability [12]. Therefore, the reduced CTL activity of Mp could be considered as part of an evasion mechanism to leave the bacterial niche intact, thus allowing the persistence and its successful spreading to the community. Despite the increased cost in public health importance of MDR-TB, the bacterial factors involved in the immune evasion mechanisms and the consequent ability to cause outbreaks are still unknown.

Relevant pathogenic differences between *Mtb* strains may be revealed by genomic comparison. With this premises in mind, we performed whole genome sequencing of two variants of the M strains: the highly prosperous Mp strain, and a non-prosperous close relative, the 410 strain.

The sequence alignment of both genomes against the reference Haarlem strain allowed us to identify genomic differences between Mp and 410 strains. These differences could explain the contrasting immune phenotypes that both strains induce *in vitro* and likely the higher capacity of Mp to perpetuate in the community, compared to 410.

2. Results and discussion

2.1. Genome sequencing and assembly

Whole-genome sequence reads from Mp and 410 were assembled using as reference the genomic sequence of Haarlem strain (see Materials and Methods for accession numbers). Specifically, we generated the genome assembly of strain 410 by aligning 7.91 million of reads producing a 169× average mapping coverage with a fraction on non-mapped reads of 6.14%. With respect to strain Mp, we generated its genome assembly by aligning approximately 4.67 million reads, with an average coverage of 105× and 7.31% of non-mapped reads.

Comparative analysis of the assembled genomes to that of reference Haarlem strain showed 345 and 362 single nucleotide mismatches in 410 and Mp, respectively. In addition, 48 insertions and deletions (INDELs) in 410 and 46 INDELs in Mp were detected in the open reading frames (ORFs). Besides, assembly gaps were also present in Mp and 410 (Fig. 1). This finding suggests deletions, as previously demonstrated in the Haarlem lineage [13].

We identified 4067 and 4066 genes for Mp and 410 strains, respectively. In Mp strain the annotation pipeline recognized a small region of 171 nt that bears similarity to gene *rv2819c* (1128 nt) and annotated the smaller fragment as a full gene. This annotation error might explain the extra gene assigned to Mp compared to 410.

Two additional contigs for each genome were built *de novo* with the reads that did not map to Haarlem strain. One of these contigs, carrying 12 genes, was very similar in Mp and 410, with eleven genes having a high level of similarity. These genes, which are present in other *M. tuberculosis* complex genomes, encode possible transposases and proteins similar to Esat6 and PPE protein families.

2.2. Comparative analysis of predicted proteins and promoter regions in MDR Mtb strains

Comparing to the Haarlem reference strain, we found that sixteen genes showed one SNP in either Mp, 410 or both. We identified additional SNPs but they mapped in regions with abundant gaps and undetermined sequences and were discarded from this analysis.

Table 1 shows that three out of the 16 relevant polymorphisms were synonymous SNPs and 13 were non-synonymous (NS-SNPs). These polymorphisms were distributed in 11 genes, and confirmed by PCR with specific primers (SI 3).

Mp and 410 showed two different SNPs in *rpoB* (*rv0667*) and *pncA* (*rv2043c*). In total, five amino acid changes were present in Mp and five in 410: a nucleotide deletion produced a predicted reading frameshifting for the Rv1861 protein of Mp and the insertion of three nucleotides added a proline in Rv0668 of Mp. Finally, a SNP was present in the 16S RNA of Mp strain and this SNP was associated to kanamycin resistance.

The analysis of promoter regions showed five intergenic regions with at least one SNP exclusive of Mp compared to the Haarlem genome sequence (Fig. 2). We identified −10 and −35 boxes of σ^{70} recognition sites in four of the five non-coding intergenic regions using the BPROM online service.

A SNP was identified upstream of *rv0010c*, *eccA3* (*rv0282*) and *rv1682*. In addition, the intergenic regions between *rv2172c-idsA2* (*rv2173*) showed two consecutive SNPs, while that of *rv3253c-rv3254* carries a 12-bp duplication.

To assess the importance of the SNPs in activity of the promoter regions, we compared the transcriptional level of the genes surrounding all polymorphic promoter regions. Fig. 3 shows the expression of *rv0010, rv2172c, idsA2* and *rv3254* was upregulated in 410 strain as compared to Mp strain. The rest of the tested genes did not show significant difference in their expression between strains under the experimental conditions used in this study.

2.3. Polymorphisms in other isolates of the M strain family

We then extended the analysis of the polymorphic proteins and promoter regions (between Mp and 410) to other 16 M genomes obtained from the European Nucleotide Archive public database [5]. These 16 strains were MDR isolated from patients in Argentina from 1998 to 2008 [5], eight were prosperous (caused outbreaks) and eight were non-prosperous (caused isolated cases). Polymorphisms in the antibiotic resistant-related genes *rpoB*, *pncA*, *embB* (*rv3795*), and *16S-rRNA* gene were conserved in all prosperous strains but absent in non-prosperous strains (Fig. 4). Interestingly, the mutation in *rpoB* of 410 strain differs from that of the other non-prosperous M strains. This result suggests that this mutation may be responsible for the slower replication rate observed *in vitro* for this strain (data not shown).

Regarding to the genes not reported as involved in antibiotic