Deciphering the role of IS6110 in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237

Henar Alonso, Juan Ignacio Aguiló, Sofia Samper, Jose Antonio Caminero, María Isolina Campos-Herrero, Brigitte Gicquel, Roland Brosch, Carlos Martín, Isabel Otal

**MOLECULAR ASPECTS**

**SUMMARY**

The capacity of infection and the ability of *Mycobacterium tuberculosis* strains belonging to the Beijing family to spread rapidly probably result from genetic advantages and unidentified mechanisms of virulence not yet thoroughly investigated. Among the mechanisms proposed to be responsible for the varying virulence phenotypes of *M. tuberculosis* strains we find IS6110 insertions, genetic reorganizations and deletions, which have strong influences on fitness.

Beijing family is one of the lineages with the highest number of copies of IS6110. By studying genetic markers characteristic for this lineage, here we have characterized the clinical isolate *M. tuberculosis* GC1237 strain responsible for important epidemic outbreaks in the Gran Canary Island. We have identified and analyzed each point of insertion of IS6110 using a bacterial artificial chromosome (BAC) library of this strain, in addition to the use of other approximations.

Nineteen copies of IS6110 have been localized in GC1237 genome of which, four copies of IS6110 can act as a promoter and we have focused in the characterization of one copy located 31 bp upstream of the essential gene Rv2179c and compared to the reference strain H37Rv.

1. Introduction

Strains of the Beijing genotype were first described in China and neighboring countries in 1995. They are widespread in many regions of the world and frequently cause epidemic outbreaks. Different studies have indicated that one-third of global tuberculosis (TB) cases is caused by Beijing family strains assigning this lineage to one of the most successful mycobacterial families in terms of morbidity and mortality. The predominance of the Beijing lineage probably results from genetic advantages, including unidentified virulence factors and the modulation of specific host responses not yet thoroughly investigated. There are some studies that relate hypervirulence of W-Beijing strains with production of phenolic glycolipid PGL, which is a putative virulence factor that attenuates the host’s innate immune response and ability to control infection. The *pks15* locus, described to be polymorphic among members of the *Mycobacterium tuberculosis* complex, is involved in the biosynthesis of the glycolipid PGL and although it is characteristic of Beijing strains to have an intact *pks15*, not all members of this family are producers of PGL.

Nowadays, Beijing strains are currently attracting considerable worldwide attention because they display important pathogenic features. These strains are often associated with multi-drug resistance as the well-known case in New York in the 1990s caused by the *W* strain. The clinical isolate *M. tuberculosis* GC1237 which belongs to the Beijing family has been responsible for epidemic outbreaks in the Gran Canary Island since its first introduction in the community in 1993. Nowadays, this strain continues being predominant in the area due to its rapid and successful dissemination within the community. The increased capacity of infection and the high success rate of the Beijing family to spread rapidly could be a consequence of genetic advantages and unidentified mechanisms of virulence not yet thoroughly investigated.
IS6110 insertions, genetic reorganizations and deletions are some of the mechanisms proposed to be responsible for differences in the virulence phenotypes of M. tuberculosis.

It has been demonstrated that IS6110 may increase the expression of neighboring virulence genes by generating new promoter sequences capable of driving their expression. IS6110 can upregulate downstream genes through an outward-directed promoter in its 3' end. This activity has been demonstrated for upregulation of the two-component system PhoP/PhoR. Promoter activity was orientation dependent and was localized within 110-bp fragment adjacent to the right terminal inverted repeat. The fact that the Beijing lineage contains a larger number of IS6110 copies than other lineages could be related with the special characteristics of this family in terms of virulence and capacity for rapid dissemination.

In this study we have classified the M. tuberculosis GC1237 within the Beijing family. We also identified the locations of IS6110 insertions in this strain and compared them to the IS6110 insertions in two Beijing strains, 210 and W. We studied the orientation and distance to neighboring genes of each copy focusing our study in one copy of IS6110 that is acting as a promoter located 31 bp upstream of the hypothetical essential gene Rv2179c.

2. Material and methods

2.1. Bacterial strains, culture media, and growth conditions

The clinical isolate M. tuberculosis GC1237 and the reference M. tuberculosis H37Rv strains were used. Mycobacterial strains were grown in Middlebrook 7H9 broth supplemented with albumin-dextrose-catalase and 0.05% Tween 80 or in Middlebrook 7H10 medium Bacto agar supplemented with oleic acid-albumin-dextrose-catalase (Difco Laboratories, Detroit, Mich.) and 0.05% Tween 80. Liquid cultures were grown to logarithmic phase to be used for macrophage infection in vitro and for mycobacterial RNA extraction. Escherichia coli DH10B cultures were grown in Luria-Bertani (LB) medium supplemented with chloramphenicol (12.5 µg/ml) in order to isolate the BACs. All of the strains were grown at 37 °C pBeloBAC11 plasmid was used for the construction of the BAC library of M. tuberculosis GC1237, pPFV27-int, derived from pPFV27, was used for the construction of the GFP strains.

2.2. Cell culture and infections

MH-S murine alveolar macrophages were obtained from HPA culture collections. Cells were grown in Dulbecco’s modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 4 mM L-glutamine. Infections were performed during 4 h at a multiplicity of infection (MOI) of 10 bacteria per cell. After incubating with bacteria, cells were washed three times with PBS, and cultured in complete medium during the time indicated for each experiment.

2.3. Isolation of genomic DNA

Genomic DNA of mycobacterial strains was isolated using the CTAB method as previously described by van Soolingen et al.

2.4. Construction and characterization of BAC library of M. tuberculosis GC1237 strain

The construction of BACs was carried out as previously described by Brosch et al. Briefly, the M. tuberculosis GC1237 library was constructed by ligation of partially digested HindIII fragments (50–125 Kb) into pBeloBAC11 plasmid. From almost 10,000 clones obtained 2000 were placed into 96-well plates and stored at –80 °C. Plasmid preparations of recombinant clones for sequencing reactions were prepared in 96-well plates containing an overnight culture in 250 µl of 2X yeast-tryptone medium with 12.5 µg/ml of chloramphenicol.

BAC DNA extraction was done as previously described by Birnboim et al. with minor modifications. Briefly, 100 ml of BAC-transformed E. coli was prepared in LB medium containing 12.5 µg/ml chloramphenicol, and the cultures were grown overnight at 37 °C with vigorous and continuous agitation. Then, the bacterial cells were collected by centrifugation. The bacterial pellet was softly resuspended in a solution of 5 ml of 50 mM glucose 10 mM EDTA, 25 mM Tris pH 8, 200 mg lysozyme was added to the tube and a solution of 4 ml of ice-cold NaAc pH 4.8 was added to the mixture. The tube was placed on ice and the precipitated debris was removed by centrifugation. Then, 14 ml of chloroform/isoamyl alcohol (24/1) were added to the supernatant and after centrifugation the aqueous phase was transferred to a new microfuge tube. DNA of the BACs was precipitated adding isopropanol and finally obtained by centrifugation.

End-sequencing reactions were performed with Taq DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems) by using a mixture of 13 µl of BAC DNA solution, 2 µl of primer SP6-BAC1 or T7-BAC1 (supplementary material), 2.5 µl of Big Dye, and 2.5 µl of a 5X buffer (50 mM MgCl2, 50 mM Tris). Thermal cycling was performed on a thermocycler (MJ Research Inc.) with an initial denaturation step of 60 s at 95 °C, followed by 26 cycles of 15 s at 96 °C, 15 s at 56 °C, and 4 min at 60 °C. DNA was precipitated with 70 µl of 70% ethanol, centrifuged, rinsed with 70% ethanol, dried and resuspended in 2 ml of formamide-EDTA buffer. The sequencing was performed on a model 373A automatic DNA sequencer (Applied Biosystems) for 12–16 h. Sequence data were transferred to Digital workstations and edited with TED software from the Staden package. Edited sequences were compared using BLAST programs to the M. tuberculosis H37Rv sequence database.

2.5. Identification of genomic deletions and analysis of pks15/1 region

The study of the genomic deletions of the regions of difference RD105, RD142, RD150, RD181 and RD207 in M. tuberculosis GC1237 strain, which identify and phylogenetically sub-classify the Beijing lineage, was performed by PCR. Other genomic deletions (RD108, RD110a, RD127, RD139BW, RD149, RD152, RD165, RD166 and RD182a) also were analyzed. The primers used in these amplifications were as described elsewhere. The PCR was carried out in a total volume of 50 µl, containing 0.5 µg of DNA, 5 µl of 10X PCR buffer, 200 µM dNTPs, 12.5 pmol of each primer and 1 U Taq Gold polymerase (Roche). Before the amplification, the template was initially denatured by incubation at 94 °C for 9 min then the amplification was performed for 35 cycles of 94 °C for 30 s, corresponding annealing temperature for 30 s, and 72 °C for 1–2 min depending on the amplified product. After the last cycle, the samples were incubated at 72 °C for 10 min. The RD deletions were confirmed by DNA sequencing using H37Rv as reference genome.

The pks15/1 polymorphism in this strain was determined by PCR and sequencing as previously described in Ref.

2.6. Location of the copies of IS6110 insertion sequence

The study of the presence of IS6110 insertion sequence in GC1237 BACs library was carried out by PCR as previously described in this study with the specific primers of this sequence, Gab 1 and Gab 2 (supplementary material).

We used Ligation-mediated PCR (LMPCR) to locate the copies of IS6110 in GC1237 strain as described by Prod’hom et al. This
technique amplifies both ends of each copy of IS6110. Briefly, genomic DNA and BACs were digested with Sall or with Smal and ligated to a linker containing a Sall restriction site or a Smal restriction site. The resulting template was then digested by Sall or with Smal. PCR was performed using ISA1 and ISA3, specific primers for IS6110 directed outwards and the common linker primer Salg (supplementary material). PCR products were purified using GFX PCR DNA gel band purification kit (Amersham Pharmacia Biotech) and the restriction enzyme EcoSAP-IT (Affymetrix).

The amplified products were sequenced with the corresponding oligonucleotides and when a match was found the amplification products after 35 cycles of PCR and the integrity of the DNA from the different strains was checked by gel electrophoresis on a 1% agarose gel. Purified DNA was kept at –80 °C until further use.

2.9.2. Intracellular conditions

Intracellular mycobacteria RNA extraction was performed with a modified protocol based on the one described by Fontan et al. Briefly, 12.10⁶ MH-S cells per flask were seeded in 150 cm² flasks. An appropriate volume of liquid mycobacteria culture was suspended in complete medium and added to each flask containing cells. After incubation, cells were washed with phosphate buffered saline (PBS) to remove extracellular bacteria, and cultured in complete medium for 0 and 48 h post-infection, when they were treated to extract intracellular mycobacteria. With this aim, cells were lysed and homogenized during 5 min adding 10 ml per flask of GTC buffer (25 mM sodium citrate, 4 M guanidine thiocyanate, 0.5% N-lauryl sarcosine, 0.125 M 2-mercaptoethanol and 0.5% Tween 80, pH 7.0). Next, samples were collected in 15-ml centrifuge tubes and centrifuged for 1 h at 4000 rpm. Bacterial pellets were washed with GTC buffer and centrifuged again for 1 h at 4000 rpm. After that, dry bacterial pellets were treated as described above to extract mycobacterial RNA.

2.10. RT-PCR

Reverse transcription-PCR (RT-PCR) was carried out in two steps. RT was carried out with Expand Reverse Transcriptase (Roche) using 1 μg RNA as the template and the appropriate reverse primer. Reaction mixtures were incubated at 42 °C for 90 min. RT products were the subjected to PCR amplification, using Taq Gold polymerase (Roche) as we previously described with the appropriate primers. Samples were analyzed by electrophoresis on a 1% agarose gel.

2.11. qRT-PCR

cDNA libraries from M. tuberculosis H37Rv and GC1237 strains were constructed as follows. One μm of RNA was mixed with 25 pmol of random hexanucleotides primers (Sigma) and 50 units of Expand Reverse Transcriptase (Roche) in a final volume of 20 μl. Reaction mixtures were incubated at 65 °C for 10 min and then at 42 °C for 90 min. Reaction of Rv2179c was measured and normalized with respect to the levels of rrrnPl mRNA by quantitative real-time PCR (qRT-PCR). qRT-PCR was carried out in a StepOne Plus (Applied Biosystems) instrument, using the cDNA generated by RT from 25 ng of RNA as a template, 1X Power SYBER green PCR master mix (Applied Biosystems), and the appropriated primers (supplementary material), each at a concentration of 250 nM. The PCR program involved an initial denaturation step for 10 min at 95 °C, followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. The specificity of the PCR products was confirmed by the loss of fluorescence at a single temperature, when the double-stranded DNA melted to single-stranded DNA.
3. Results

3.1. Analysis and representativity of the pBeloBAC11 library of M. tuberculosis GC1237 strain

We have constructed a BAC library of GC1237 strain. To characterize the BAC clones, we systematically subjected them to end-sequencing reactions. The sequence of the clones provided us the exact BAC positions on the physical map of the reference chromosome H37Rv. A total of 59 BACs named A1–12 to H1–12 were selected covering different regions of the genome of GC1237. These BACs represented around 75% of the chromosome of the studied strain (Figure 1A).

3.2. Characterization of GC1237 strain: study of RD and pks15/1 region and analysis of extra regions

We studied the regions RD105, RD207, RD181, RD150 and RD142 in order to classify GC1237 strain within one of the four subgroups of the Beijing family. The results indicated that GC1237 has deleted RD181 and RD105, a robust marker for strains pertaining to the Beijing family, and RD207 concerning the Direct Repeat Locus. The findings regarding these deletions indicate that the studied strain belongs to subgroup four of the seven groups of M. tuberculosis described by Kong et al. We further found the deletions of RD149 and RD152, associated with mobile genetic elements such as IS6110 (Figure 1B). GC1237 presents a copy of IS6110 in place of RD152. Finally, the studied strain presented intact regions in the rest of the analyzed regions including RD110a, RD127, RD129, RD139BW, RD165, RD166 and RD182a.

We observed two extra regions in GC1237 absent in H37Rv. When we compared the findings in Bovilist database, we found that the extra regions corresponded to Mb3356-Mb3359c and Mb2047c-Mb2048c. A copy of IS6110 was detected between the genes corresponding to Mb2047c-Mb2048c (Figure 1B).

The study of pks15/1 region revealed an intact gene in GC1237 (data not shown) indicating that this strain could retain the ability to produce functional glycolipid PGL in contrast to the reference strain H37Rv.

3.3. Different strategies to localize all copies of IS6110 insertion sequence in GC1237 strain

We localized 19 different copies of IS6110 in GC1237 applying two main strategies: LMPCR technique with BAC library and genomic DNA of the studied strain and PCR with designed primers based on the known points of the insertion sequence in the two reference Beijing strains 210 and W. The presence of IS6110 in the representative GC1237 BAC library was detected using PCR with the specific oligonucleotides Gab 1 and Gab 2 and ten of total of 59 BACs were positive for IS6110 (data not shown). Using LMPCR the ten different copies were localized in GC1237 and this technique was further utilized with genomic DNA to detect the rest of the copies present in GC1237 by amplifying one or both ends of each copy of IS6110 and its flanking sequences. As a second strategy, based on genome known locations of IS6110 in 210 and W strains, we designed primers to amplify the different regions that might include IS6110 in the studied strain. This design was realized using M. tuberculosis H37Rv as genome reference. As a result of the two different strategies followed in this study we determined the exact location of the 19 copies of IS6110 in GC1237 strain (Table 1). The three Beijing strains share nine IS6110 positions, these include the two insertions characteristic of the Beijing family, between dnaA and dnaN, and in the Direct Repeat Locus (Figure 1C).

GC1237 strain contains eight copies of IS6110 in intergenic regions and eleven in intragenic regions (Table 1). Of the eleven interrupted genes, pip, ctpD, idsB and Rv2180c are involved in bacterial metabolism and respiration, four correspond to genes with no predicted function and three are members of the PE and PPE family. We investigated the location of IS6110 within Rv2180c in other M. tuberculosis strains including other Beijing strains (data not shown) and this point of insertion is unique to GC1237 strain. Two IS6110 within the PPE34 were consecutive and in the same orientation but the second one lacks 60 bp at the 5′ end. One of the

Figure 1. Circular plot of GC1237 genome prepared with DNAplotter. Graphic representation of M. tuberculosis GC1237 genome. (A) BACs regions. (B) Genomic regions absent in M. tuberculosis H37Rv (RvDs), lacking regions of difference (RD) and extra regions in GC1237 strain. (C) Location of IS6110 in different Beijing strains GC1237, 210 and W. The DR locus and dnaA-dnaN region are marked.

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS6110 in a highly transmissible Mycobacterium tuberculosis Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007
Table 1
Insertion sites of IS6110 in M. tuberculosis GC1237, 210 and W.

<table>
<thead>
<tr>
<th>GENES</th>
<th>GC1237</th>
<th>210</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GC1237 genome</td>
<td>BAC region</td>
<td></td>
</tr>
<tr>
<td>Rv0001 (dnaA):Rv0002 (dnaN)</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Rv0794c:Rv0797 (IS1547)</td>
<td>+</td>
<td>B4</td>
<td></td>
</tr>
<tr>
<td>Rv0840c (pip)</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rv1135c</td>
<td>+</td>
<td>B11</td>
<td>+</td>
</tr>
<tr>
<td>Rv1371</td>
<td>+</td>
<td>C7</td>
<td>+</td>
</tr>
<tr>
<td>*Rv1469 (cpdP)</td>
<td>+</td>
<td>C10</td>
<td>+</td>
</tr>
<tr>
<td>Rv1754c</td>
<td>+</td>
<td>D4</td>
<td>+</td>
</tr>
<tr>
<td>Rv1754c-Rv1765c (RD152)</td>
<td>+</td>
<td>D4</td>
<td></td>
</tr>
<tr>
<td>Rv1798:Rv1799 (ipfT)</td>
<td>+</td>
<td>D6</td>
<td>+</td>
</tr>
<tr>
<td>Rv1917c (PPE34)</td>
<td>+</td>
<td>D9</td>
<td></td>
</tr>
<tr>
<td>Rv1917c (PPE34)</td>
<td>++</td>
<td>D9</td>
<td></td>
</tr>
<tr>
<td>Rv2016</td>
<td>+</td>
<td>D12</td>
<td>+</td>
</tr>
<tr>
<td>MT2080 (Mb2047c):MT2081 (Mb2048c)</td>
<td>+</td>
<td>D12</td>
<td></td>
</tr>
<tr>
<td>Rv2077A:Rv2078</td>
<td>+</td>
<td>E1</td>
<td></td>
</tr>
<tr>
<td>Rv2104c:Rv2107 (PE22)</td>
<td>+</td>
<td>E1</td>
<td>+</td>
</tr>
<tr>
<td>Rv2104c:Rv2107 (PE22)</td>
<td>E1</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Rv2107 (PE22):Rv2108 (PPE36)</td>
<td>E1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>*Rv2180c</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Rv2286c</td>
<td>+</td>
<td>E7</td>
<td></td>
</tr>
<tr>
<td>Rv2352c (PPE38)</td>
<td>E8</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rv2353c (PPE39):Rv2356c (PPE40)</td>
<td>+</td>
<td>E8</td>
<td>+</td>
</tr>
<tr>
<td>Rv2813-Rv2820c (RD207)</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>*Rv3019A (PE27A):Rv3019c (esxR)</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rv3019c (esxR):Rv3020c (esxS)</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rv3128c</td>
<td>G1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rv3128c:Rv3129</td>
<td>G1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rv3178c:Rv3179</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rv3179c:Rv3180c</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Rv3184c:Rv3187</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rv3324A:Rv3327 (IS1547)</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Rv3383c (idb8)</td>
<td>+</td>
<td>G7</td>
<td>+</td>
</tr>
<tr>
<td>*Rv3427c (IS1532):Rv3428c (IS1532)</td>
<td>+</td>
<td>G8</td>
<td>+</td>
</tr>
</tbody>
</table>

Candidate locations to act as a promoter.

Gene names come from the M. tuberculosis H37Rv genome map. One gene is listed when the location of insertion is in that gene and two genes are listed when the insertion of IS6110 is between both.

Sites are in order of appearance on the M. tuberculosis H37Rv genome map.

Name of BAC.

The plus sign indicates that the M. tuberculosis 210 and W contain a copy of IS6110 in those regions.

Table 2
Direct repeats of IS6110 insertion sequence in GC1237 strain.

<table>
<thead>
<tr>
<th>Genes of IS6110 in GC1237 genome</th>
<th>Direct Repeat (DR)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv0001 (dnaA):Rv0002 (dnaN)</td>
<td>ATT</td>
</tr>
<tr>
<td>Rv0794c:Rv0797 (IS1547)</td>
<td></td>
</tr>
<tr>
<td>Rv0840c (pip)</td>
<td>ACG</td>
</tr>
<tr>
<td>Rv1371</td>
<td>GAGG</td>
</tr>
<tr>
<td>Rv1469 (cpdP)</td>
<td>CCT</td>
</tr>
<tr>
<td>Rv1754c-Rv1765c (RD152)</td>
<td></td>
</tr>
<tr>
<td>Rv1917c (PPE34)</td>
<td>TTA</td>
</tr>
<tr>
<td>Rv1917c (PPE34)</td>
<td>TTA</td>
</tr>
<tr>
<td>Rv2016</td>
<td></td>
</tr>
<tr>
<td>MT2080 (Mb2047c):MT2081 (Mb2048c)</td>
<td></td>
</tr>
<tr>
<td>Rv2077A:Rv2078</td>
<td></td>
</tr>
<tr>
<td>Rv2179c:Rv2180c</td>
<td></td>
</tr>
<tr>
<td>Rv2286c</td>
<td></td>
</tr>
<tr>
<td>Rv2353 (PPE39):Rv2356c (PPE40)</td>
<td></td>
</tr>
<tr>
<td>Rv2813-Rv2820c (RD207)</td>
<td></td>
</tr>
<tr>
<td>Rv3019A (PE27A):Rv3019c (esxR)</td>
<td></td>
</tr>
<tr>
<td>Rv3324A:Rv3327 (IS1547)</td>
<td></td>
</tr>
<tr>
<td>Rv3383c (idb8)</td>
<td></td>
</tr>
<tr>
<td>Rv3427c (IS1532):Rv3428c (IS1532)</td>
<td></td>
</tr>
</tbody>
</table>

* The absence of Direct Repeat is indicated by minus.

H. Alonso et al. / Tuberculosis xxx (2011) 1–10

3.4. Analysis of direct repeats of each copy of IS6110 in GC1237 genome

We analyzed the flanking regions of each of the nineteen copies of IS6110 in GC1237 genome and detected the presence of direct repeats (DR) of 3–4 nucleotides at the extremities of sixteen IS6110 sequences as result of IS6110 transposition (Table 2). The other three copies analyzed were not flanked by DRs. Genomic reorganization was observed in two of these locations (RD152 and RD207), and this is probably a consequence of recombination between two adjacent copies of IS6110. The last of the three IS6110 elements without DR, which was localized between Rv0794c:Rv0797 was in the opposite orientation with respect to the IS6110 in the reference genome.

3.5. Study of the orientation and distance of IS6110 to the neighboring genes

Different studies have indicated that IS6110 could upregulate the expression of downstream genes. Previous studies have shown that when IS6110 is inserted in the same orientation as, and close enough to, a downstream gene, IS6110 could potentially function as a promoter. We analyzed the orientation of each copy of the IS6110 insertion sequence in GC1237 strain and the distance to the close gene in order to test the promoter function of IS6110 and we observed that IS6110 could act as a mobile promoter in four locations (Table 1). One of them is located 297 bp upstream of the Rv1468c gene, another 138 bp upstream of PE27A gene, the third one 41 bp upstream Rv3427c gene and the last one is located 31 bp upstream of the essential gene Rv2179c. We focused our study in the last one as this location is specific for GC1237 and the distance to the downstream gene is the shortest.

3.6. Detection of a transcript from IS6110 to aroG in GC1237 strain and from Rv2180c to aroG in H37Rv strain

To determine if a transcript extended from IS6110 or Rv2180c into downstream genes Rv2179c and aroG, we extracted RNA from broth cultures of GC1237 and H37Rv strains and RT-PCR was performed using the corresponding oligonucleotides as previously described in this work. We obtained amplification products from Rv2180c to aroG in H37Rv (Figure 2A and C). These results indicate that the three genes are cotranscribed in an operon. In the case of GC1237, amplification products were obtained from 3’ end of IS6110 to aroG suggesting that the 3’ end of IS6110, Rv2179c and aroG genes are cotranscribed (Figure 2B and D). All the RT-PCR products were sequenced and the Blast results verified their specificity. Our results indicated that the three genes are cotranscribed in an intergenic copies in GC1237, localized between PPE39 and PPE40 genes, is also present in H37Rv but the region is longer in the studied strain. Within the two ipl regions two copies of IS6110 were located: one between Rv0794c and Rv0797, which is in the opposite orientation in H37Rv and other studied Beijing strains (data not shown), and the second one was located between Rv3324A:Rv3327 in the same orientation as reference H37Rv.

The NTF region of GC1237 does not contain any copies of IS6110 which is an indication of being an ancestral strain different to 210 and W.

3.4. Analysis of direct repeats of each copy of IS6110 in GC1237 genome

We analyzed the flanking regions of each of the nineteen copies of IS6110 in GC1237 genome and detected the presence of direct repeats (DR) of 3–4 nucleotides at the extremities of sixteen IS6110 sequences as result of IS6110 transposition (Table 2). The other three copies analyzed were not flanked by DRs. Genomic reorganization was observed in two of these locations (RD152 and RD207), and this is probably a consequence of recombination between two adjacent copies of IS6110. The last of the three IS6110 elements without DR, which was localized between Rv0794c:Rv0797 was in the opposite orientation with respect to the IS6110 in the reference genome.
operon, but these results do not exclude the possibility of independent promoters for each gene.

3.7. IS6110 is acting as a promoter of Rv2179c gene both under extracellular and intracellular conditions

To verify that IS6110 located upstream Rv2179c in GC1237 genome was functional as a promoter, we constructed the recombinant plasmids pFPV79c and pFPVIS79c containing the promoterless gfp gene preceded by the region upstream Rv2179c in H37Rv strain or in GC1237 strain, respectively. They were transformed into H37Rv obtaining FPV79c and FPVIS79c strains to compare GFP expression of both constructions in the same genetic background. As control, we used H37Rv transformed with pFPV27 which contains gfp gene without promoter (FPV27 strain). GFP expression was measured both by fluorometry and flow cytometry. Results of both techniques showed that the fluorescence of FPVIS79c strain is about 5-fold higher than the observed in FPV79c strain (Figure 3A). Moreover, data obtained by fluorometry during the growth curve indicated that the effect of IS6110 on the regulation of GFP in this case was independent of bacterial growth phase (Figure 3B). Therefore, this result suggests that in GC1237, IS6110 is acting as a promoter of Rv2179c gene.

We decided to study the promoter activity of IS6110 upstream Rv2179c under intracellular conditions. For that issue, we infected murine immortalized alveolar macrophages (MH-S cell line) with the different GFP expressing strains and infected cells FMI was measured by FACS. Thus, to exclude non-infected cells, we also infected with non-fluorescent H37Rv, considering their FMI as negative fluorescence level (R1 in Figure 4A). On the other hand, cells contained in the region with a higher fluorescence level were established as the infected ones (R2 in Figure 4A). Moreover, data were referred to annexinV-negative cells, in order to analyze only live cells. As it is observed in Figure 4B cells infected with FPVIS79c strain showed a clear increase of their FMI values regarding the control at all times studied. These results are in agreement with the obtained data under extracellular conditions. Furthermore, GFP expression of FPVIS79c-infected cells tended to increase during the experiment, being this rise more dramatic at 72 h. In contrast, in the absence of IS6110 this event failed to happen (Figure 4B).

Figure 2. RT-PCR analysis of Rv2179c region from M. tuberculosis H37Rv and GC1237 strains. (A) Schematic diagram of Rv2179c region in H37Rv strain. (B) Schematic diagram of Rv2179c region in GC1237 strain. In both diagrams, the primers used for RT-PCR (Table 1) and the sizes of the fragments obtained with each pair of primers are indicated. The direction of transcription for Rv2179c is indicated by arrows. (C) RT-PCR analysis of this region in H37Rv. (D) RT-PCR analysis of this region in GC1237. In both RT-PCR the combination of primers is indicated above each set of reaction. Each set of the three reaction consists of a positive control PCR assay with genomic DNA as the template (+), an RT-PCR (–), and a negative control assay without reverse transcriptase (−).

Figure 3. GFP expression in FPV79c and FPVIS79c strains. (A) Fluorescence median intensity (FMI) of FPV27, FPV79c and FPVIS79c strains was measured by flow cytometry at logarithmic phase. (B) Time course the evolution of the fluorescence of FPV27, FPV79c and FPVIS79c strains during the bacterial growth normalized with O.D600.

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS6110 in a highly transmissible Mycobacterium tuberculosis Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007
3.8. Rv2179c gene expression is increased in GC1237 strain both in broth and intracellular conditions

As it had been previously described that IS6110 can act as a promoter of downstream genes in *M. tuberculosis*,14 we compared the expression of Rv2179c gene in GC1237 and H37Rv strains. Thus, both strains were cultured to mid-exponential growth phase and RNAs were extracted to generate cDNAs. The expression of Rv2179c was measured by qRT-PCR and normalized to *rrnAP1* expression levels. Results demonstrated that Rv2179c expression was about 6-fold higher in GC1237 than in H37Rv (Figure 5A), which suggests that IS6110 in this location is up-regulating gene expression.

To determine the effect of IS6110 on Rv2179c expression during infection, bacterial RNA was obtained both from intracellular and broth growth conditions. The results showed that Rv2179c expression rose dramatically in GC1237 infected macrophages, being 6 and 10-fold higher at 0 h and 48 h post-infection, respectively (Figure 5B). On the contrary, lower increase was observed in the case of H37Rv, confirming that the presence of IS6110 correlated with higher expression of Rv2179c in the studied strain both under intracellular and extracellular conditions.

As part of the three genes forming an operon we analyzed the expression of *aroG* by qRT-PCR, but we could not observed significant increment of this gene (data no shown). This result does not exclude the possibility of another regulation of *aroG*.

4. Discussion

The *M. tuberculosis* GC1237 strain has demonstrated an enhanced capacity to spread and transmit within a community. This strain belongs to one of the most virulent families, Beijing. The capacity of infection and the ability of this family to spread rapidly probably result from genetic advantages and unidentified mechanisms of virulence not yet thoroughly investigated.

Several reports have strongly suggested that the severity and clinical manifestations of TB depend on the immunogenicity and pathogenicity of the infecting *M. tuberculosis* strain; these properties vary considerably among strains.31 The notion that some *M. tuberculosis* lineages, such as the W-Beijing genotype, are more virulent than others is supported by results obtained with animal and cellular models. Studies in both show that among the Beijing family strains there are hyper-virulent phenotypes that are associated with relapse and treatment failure in humans and distinct immune responses in animal models.32 Furthermore, different studies have demonstrated that some W-Beijing strains can escape the immunoprotective efficacy of the current TB vaccine and thus act as “escape variants” of *Mycobacterium bovis* BCG vaccination.33 We decided to study most extensively the GC1237 strain because it has been responsible for a high number of TB cases since its introduction in the Gran Canary Island in 1990s.

Beijing strains are genetically homogeneous and conserved. They share some genetic markers, such as similar spoligotypes1,29 characterized by the deletions of spacers 1–34 in the Direct Repeat Locus of *M. tuberculosis*; a copy of IS6110 between dnaA: dnaN genes; and it is frequent to find at least a copy of IS6110 in the NTF region in the strains belonging to Beijing family34 and some members of this family may have a second insertion within this locus.35 However, the absence of IS6110 in NTF locus is characteristic of ancestral sub-lineages within the Beijing genotype.36 Due to the course of evolution, some strains (“modern” sublineages) have acquired the insertion of IS6110 in this region.35 The clinical strain GC1237 presents an intact NTF locus indicating that GC1237 is an ancestral Beijing strain.

Comparative whole-genome hybridization of Beijing strains showed large sequence polymorphisms (LSPs) also known as regions of difference (RD), which subdivide the Beijing family into at least four phylogenetic subgroups, raising the possibility that there are phenotypic differences within the Beijing family. The subdivisions

![Figure 4](image-url). GFP expression in FPV79c and FPVIS79c strains in murine macrophages. (A) Histograms of H37Rv and FPVIS79c-infected cells. R1 indicates the non-fluorescent cells and R2 represents the fluorescent cells. (B) Time course of FMI of cells gated in R2, infected with FPV79c and FPVIS79c strains. The percentage of increment was calculated with respect to FMI values of FPV27 strain. Figure shows a representative experiment of three performed.
are made on the basis of RD105 and RD207, and the variable appearance of RD181, RD150 and RD142 deletions. Some authors have proposed an evolutionary pathway of the Beijing lineages based on the RD deletions and on the IS6110 insertions in the NTF region indicating that ancient strains have neither deletion of RD181 nor insertion of IS6110 in NTF region. However, our results are in disagreement with this classification because GC1237 presents RD181 deleted but lacks IS6110 in NTF locus. A recent study supports the view that genomic deletion of RD181 has occurred during the evolutionary process from ancient to modern strains proposing four genetic sublineages. According to this view, GC1237 falls in the subgroup the most predominant in Japan. When we analyzed the appearance of RD152 and RD149, these deletions are consequence of transposition. In our study, because of the high number of different combinations of these regions, present in other sequenced strains of M. tuberculosis and other copies without DR suggesting that the probability of transposing a copy of IS6110 insertion sequence the probability to localize all the copies by LMPCR technique is low. This technique was useful to determine the points of insertion of this sequence in M. bovis. However, the number of copies of IS6110 in M. bovis is very much minor (1–5) than the observed in M. tuberculosis. In this regard, the construction of BAC library of GC1237 strain has been a successful tool. The use of the different molecular strategies has allowed us to localize all IS6110 copies in a strain of M. tuberculosis containing a high number of this insertion sequence. Previous studies supported the idea that some insertion sites were prevalent in the low-copy number strains and suggested a separate evolutionary lineage for the low and the high-copy number strains of M. tuberculosis. According to this, GC1237 does not have any copy of IS6110 inserted at the same genomic position described for low-copy M. tuberculosis strains.

In sixteen of the nineteen copies, we observed DR of 3–4 nucleotides in the flanking regions. These DR indicate that the presence of IS6110 was due to transposition events and the absence of DR in the other copies suggests rearrangements between IS6110 elements. There are very few studies where authors report DR flanking IS6110. In these studies the number of copies is low (<6) and all of them were flanked by DR, indicating that these IS6110 are consequence of transposition. In our study, because of the high number of copies of IS6110 we observed some copies flanked by DR and other copies without DR suggesting that the probability of rearrangement process between copies rises when the number of those increases producing more variability among strains.

Numerous studies indicate that IS6110 insertion sequence has some preferential points of insertion (hot-spots) in the genome. The IS1547 (ipl locus) is a site of intensive insertion of IS6110 and the studied strain corroborates this hot-spot with the insertion of a copy of IS6110 in each ipl locus. On the other hand, mapping of IS6110 insertion points in M. tuberculosis GC1237 has shown that no consensus target sequence was detected in the immediate vicinity of the analyzed copies. Obviously, the DR flanking IS6110 also were different except for the sequence ATC that was found in two cases. All these support the lack of transposition specificity for IS6110, according with other authors. The open reading frames represent 91% of M. tuberculosis genome but in GC1237 the insertion of IS6110 into coding regions occurred in 42% of the cases, suggesting that the transposition is...
relatively more frequent in intergenic regions. Our results agreed with other studies that found that 58% of discrete IS6110 insertion sites occurred within coding regions in *M. tuberculosis* and in *M. bovis* strains.\(^{40}\) In this context, because of the insertion of IS6110 in possible promoter regions the probability of IS6110 influencing the expression of the neighboring genes could be increased. It is necessary to take into account that the locations of IS6110 observed in the different genomes are the result of transposition events and the natural selection of viable mycobacteria because when insertion occurs in a necessary gene it could affect survival and we cannot observe it. In case of the interruption of a non-essential gene, the phenotype of the mycobacteria could be affected. It has been demonstrated the potential for intra- and inter-genic IS6110 insertions to add plasticity to *M. tuberculosis* genome and influence fitness of the mycobacterium.

With the exception of the copy of IS6110 located between *dnaA* and *dnaB* genes, we have not found any copy of IS6110 in the quarter of *M. tuberculosis* GC1237 circular genome surrounding these genes (Figure 1C). Also, we observed the same in the two Beijing strains 210 and W. This result is in agreement with other authors when they have characterized IS6110 integration loci.\(^{47}\) This finding could indicate that in this part of the genome there are no specific sequences or regions recognized by the sequence or that these genes have an important role in *M. tuberculosis* and the interruption or modulation of their expression could result in no advantage for the bacteria.

The IS6110 insertion sequence varies in number and position within the genome generating a high level of DNA polymorphism among strains. It is known that the insertion of IS6110 can alter bacterial gene expression. Depending on the location, IS6110 may cause loss of gene activity or can also upregulate the expression of downstream genes by acting as a mobile promoter. When IS6110 is inserted in the same orientation as, and close enough to, a downstream gene it could potentially function as a promoter.\(^{14,15}\) According to these previous studies we analyzed the orientation and the distance to the closest gene of each copy of IS6110 in GC1237 strain. Thus, we obtained eight genes with an IS6110 inserted upstream, in the right orientation and at a reasonable distance to work as a promoter these genes. However, when we compared the expression of these genes in GC1237 with reference to H37Rv strain, we only observed a clear increase in the case of the gene Rv2179c, whose IS6110 is located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it. To discern that Rv2179c upregulation found a clear increase in the case of the gene Rv2179c, whose IS6110 in GC1237 strain was really due to the presence of IS6110 located 31 bp upstream of it.

When we compared Rv2179c expression in both strains in broth culture, we observed higher expression in GC1237 than in H37Rv. These results confirm the promoter activity of IS6110 and suggest that this activity is upregulated inside macrophages.

On the other hand, the RT-PCR results indicated that in H37Rv strain the Rv2180c, Rv2179c and Rv2178c (*aroG*) genes are cotranscribed. Differently, in GC1237 we could not detect the cotranscript between Rv2180c and Rv2179c genes due to the insertion of IS6110 in Rv2180c, but we detected the cotranscript from the 3' end of IS6110 to *aroG* indicating that this sequence acts as a new promoter for the Rv2179c gene and also for *aroG* gene which is an essential gene involved in biosynthesis of chorismate precursor of the three aromatic amino acids in *M. tuberculosis*.\(^{48}\) There are no available data about the function of the protein encoded by Rv2179c gene. In fact, according to Tuberculist database, the product of this gene is a conserved essential hypothetical protein and it is conserved in different mycobacterial strains.\(^{49}\) From our results we can hypothesize that the over-expression of this gene could be advantageous for GC1237 at least in certain environments such as infecting macrophages but we do not rule out the possibility of not having an effect. Providing that this insertion is unique in this strain we do not discard its implication in specific fitness advantages in GC1237 and on the other hand, this point of insertion can be a useful tool in epidemiological studies because it allows us to identify and differentiate this strain among Beijing strains.

In summary, this study facilitates the location of all the copies of IS6110 in GC1237 which will enable the search of the common locations of IS6110 in Beijing strains that would be characteristic of this family. Knowledge on distribution of IS insertion sites in Beijing strains sheds more light in evolutionary processes involving this family. The location of all the copies will allow studying the implication of each one in the different strain properties. Some of the detected copies are interrupting genes, others are implicated in recombination events and we have demonstrated that IS6110 is acting as a promoter in Rv2179c.

Acknowledgments

We thank Dessislava Marinova for her kind help in English revision and for her valuable comments on the manuscript. Henar Alonso is financially supported by Gobierno de Aragón (B034/09).

This work is part of the TB-VIR Project supported by the European Commission under the Health Cooperation of the 7th Framework Programme 200973 and Project of the Ministerio de Ciencia e Innovación B02008-01561.

Appendix. Supplementary material

The supplementary data associated with this article can be found in the on-line version at doi:10.1016/j.tube.2010.12.007.

Funding: None.

Competing interests: None Declared.

Ethical approval: Not required.

References


