

## RESEARCH ARTICLE

## Open Access

# Mapping IS6110 in high-copy number *Mycobacterium tuberculosis* strains shows specific insertion points in the Beijing genotype

Henar Alonso<sup>1,3,4</sup>, Sofia Samper<sup>1,2,3</sup>, Carlos Martín<sup>1,2,3</sup> and Isabel Otal<sup>1,3\*</sup>**Abstract**

**Background:** *Mycobacterium tuberculosis* Beijing strains are characterized by a large number of IS6110 copies, suggesting the potential implication of this element in the virulence and capacity for rapid dissemination characteristic of this family. This work studies the insertion points of IS6110 in high-copy clinical isolates specifically focusing on the Beijing genotype.

**Results:** In the present work we mapped the insertion points of IS6110 in all the Beijing strains available in the literature and in the DNA sequence databases. We generated a representative primer collection of the IS6110 locations, which was used to analyse 61 high-copy clinical isolates. A total of 440 points of insertion were identified and analysis of their flanking regions determined the exact location, the direct repeats (DRs), the orientation and the distance to neighboring genes of each copy of IS6110. We identified specific points of insertion in Beijing strains that enabled us to obtain a dendrogram that groups the Beijing genotype.

**Conclusions:** This work presents a detailed analysis of locations of IS6110 in high-copy clinical isolates, showing points of insertion present with high frequency in the Beijing family and absent in other strains.

**Keywords:** IS6110, Insertion points, *M. tuberculosis*, Beijing genotype

**Background**

The insertion sequence (IS) 6110 is specific for the *Mycobacterium tuberculosis* complex (MTBC) [1]. *M. tuberculosis* strains typically contain multiple copies of IS6110 (up to 25 per genome) [2], although strains with only a single copy or no copies have also been identified [3-5]. In contrast, *M. bovis* strains are characterized by having a low copy number, with *M. bovis* BCG substrains having either one or two copies [6]. The high variability in copy number and location of IS6110, as well as its stability over time, renders IS6110 a useful diagnostic and epidemiological tool. Moreover, in some cases, the location of a copy that is specific for a strain can be used for the rapid identification and differentiation of that particular strain from other isolates [7].

Without a known insertion target, IS6110 has been found within ORFs and intergenic regions [2,8]. The IS6110 locations along the genome are not equally distributed, being found more often in some regions, while completely absent in others [9-11]. In single-copy strains of *M. tuberculosis* and *M. bovis*, IS6110 is always present in the conserved 36-bp array, designated the Direct Repeat region (DR region) [12] and, with minor exceptions, all members of the MTBC carry a copy integrated in this locus. General hot-spots of IS6110 include IS1547 in *iplA-iplB* region [13,14], the phospholipase C region (locus *plcABC* and the *plcD* gene) [15,16], members of the PPE gene family [17] and the origin of replication (*oriC*) [18,19].

One of the most relevant and better studied mechanisms of genome change is IS-mediated. It is considered that about 5% to 15% of the spontaneous mutations in the bacterial genome are due to changes in IS locations [20]. Transposition is one of the more common mechanisms used by IS to move along genome, whereas recombination between two IS copies can lead to genome

\* Correspondence: [otali@unizar.es](mailto:otali@unizar.es)

<sup>1</sup>Grupo de Genética de Micobacterias. Departamento de Microbiología, Medicina Preventiva y Salud Pública, Universidad de Zaragoza, C/ Domingo Miral sn. 50009, Zaragoza, Spain

<sup>3</sup>CIBER Enfermedades Respiratorias (CIBERES), Madrid, Spain

Full list of author information is available at the end of the article

deletions [21]. Hence, an IS could either play a role as a potential enemy or a helpful ally affecting the fitness of the bacterium. *IS6110* insertions, genetic reorganizations and deletions are some of the mechanisms proposed to be responsible for differences in the virulence phenotypes among *M. tuberculosis* strains. *IS6110* has been associated with participation in adaptation to a particular host [22], activation of genes during infection [23], evolution [21] and in activation of downstream genes with an orientation-dependent activity promoter [10,23,24].

The Beijing strains are considered one of the most successful families in tuberculosis transmission. Three features that characterize the Beijing family are related to the *IS6110* insertion sequence. These include: i) the presence of a copy in the *oriC* known as insertion A1, ii) the deletion of the region of difference (RD) 207, and iii) a similar *IS6110*-RFLP multiband pattern profile [25]. Members of this genotype are characterized by a high-copy number of *IS6110* [26], suggesting the potential implication of this element in the special characteristics of this family related to virulence and capacity for rapid dissemination.

In the present work, we conducted an in-depth study of the points of insertion of *IS6110* in selected Beijing clinical isolates and in the Beijing strains available in the literature and in the DNA sequence databases. With the obtained locations, we generated a representative primer collection of Beijing-*IS6110* points of insertion, which was used to analyse 61 high-copy clinical isolates. We established the specific and shared copy locations in the studied strains and constructed a dendrogram allowing the accurate classification of Beijing genotype strains.

## Results and discussion

### Generation of a primer collection of *IS6110* insertion points

Firstly, we studied the points of insertion of *IS6110* in eight representative Beijing strains (NHN5, HM77, HM903, HM764, 990172, W4, N4 and CAM22) by LMPCR. Twenty-two new genomic insertion points were obtained by this technique. Primers for the identified new *IS6110* locations were designed and added to primers used in a previous work for localization of *IS6110* copies in the Beijing strain GC1237 [10]. Through LMPCR and PCR, a total of 106 (45 different) *IS6110* insertion points were obtained and plotted on the H37Rv genome map. The shared insertion points by these eight strains and the three reference Beijing strains (GC1237, 210 and W [10,31]) were obtained (Figure 1 and Additional file 1: Tables S4 and S5).

In addition, we analysed the *IS6110* insertion points in 43 reference sequenced strains in the GeneBank comparing their sequenced fragments with H37Rv genome. We obtained a total of 486 points of insertion (Additional file 1: Table S3), which were plotted on the H37Rv

genome map and were grouped in two concentric circles: one corresponding to non-Beijing strains (green circle) and another to Beijing strains (red circle) (Figure 2).

In addition to the GeneBank analysis, we reviewed the available literature [10,11,13,16-18,20,22-24,27-35] for new locations of *IS6110* in Beijing strains, for which new pairs of primers were designed. In addition, to detect whether the preferred insertion site of *IS6110*, DK1 region [28], in low-copy number strains (LCS) (< 7 copies) is present or absent in high-copy number strains (HCS), a specific pair of primers was designed. Moreover, the primers used to amplify locations of *IS6110* in *M. bovis* human isolates [22] were included, as the host in both cases is the same. With all these pairs of oligonucleotides we generated a primer collection (238 oligonucleotides) (Additional file 1: Table S2). Figure 2 depicts the amplified regions with the primer collection (first concentric blue circle) showing quite a homogeneous distribution along the *M. tuberculosis* genome.

### Locations of *IS6110* in Beijing genotype

All the Beijing strains (8 representative strains, GeneBank and literature) presented the three characteristic *IS6110* locations of Beijing family: the insertion A1, between Rv1754c-Rv1765c genes, which corresponds to deleted RD152, and in the DR region (Figure 1). Of note, two copies of *IS6110*, both in the same orientation, were detected in the DR region in CAM22 strain.

Interestingly, when we compared the eight representative strains, we observed six sites of insertion present with high frequency in the genes Rv1371, *ctpD* (Rv1469c), Rv2016 and *idsB* (Rv3383c), between the two *IS1532* and in *esxR-esxS* region. These 6 insertion points are also present in the three reference Beijing strains (GC1237, 210 y W) [10] (Figure 1 and Additional file 1: Table S5).

According to some authors, it is frequent to find at least one copy of *IS6110* in the NTF region [36] in strains belonging to Beijing family and some members of this family may have a second insertion within this locus such as the MDR W strain [37,38]. It has been proposed that the absence of *IS6110* in the NTF locus may be associated with ancestral Beijing genotype sublineages [39] and in the course of evolution some strains ("modern" sublineages) have acquired the insertion of *IS6110* in this region [37]. The eight analyzed strains do not present *IS6110* in this region, which according to some authors, would be classified as ancestral.

### Distribution of *IS6110* among Beijing and non-Beijing strains

We studied the distribution of *IS6110* in the eight analysed Beijing strains and in the sequenced *M. tuberculosis* genomes (from the GeneBank and in the literature) and observed that the locations were random (Figure 1 and

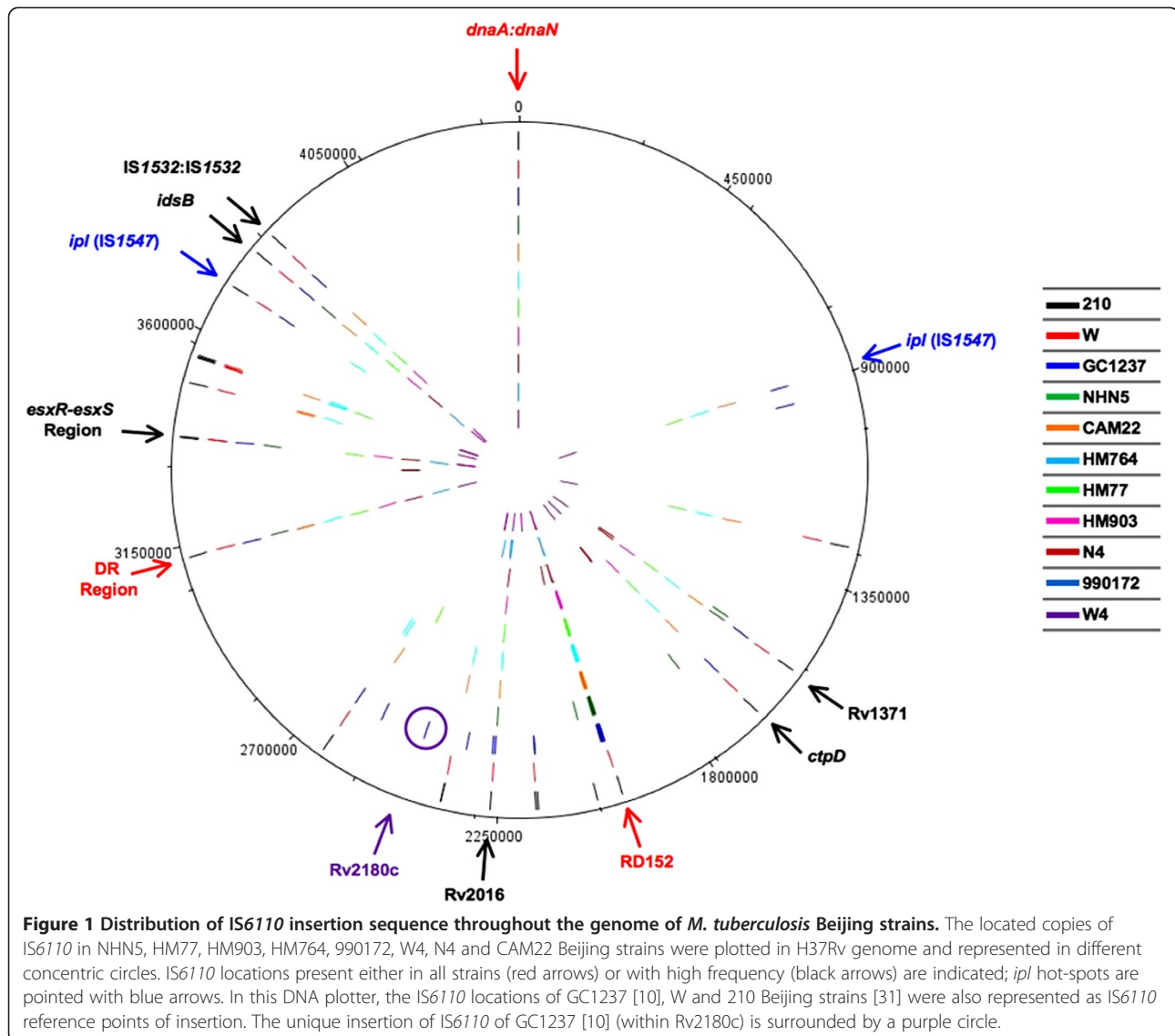
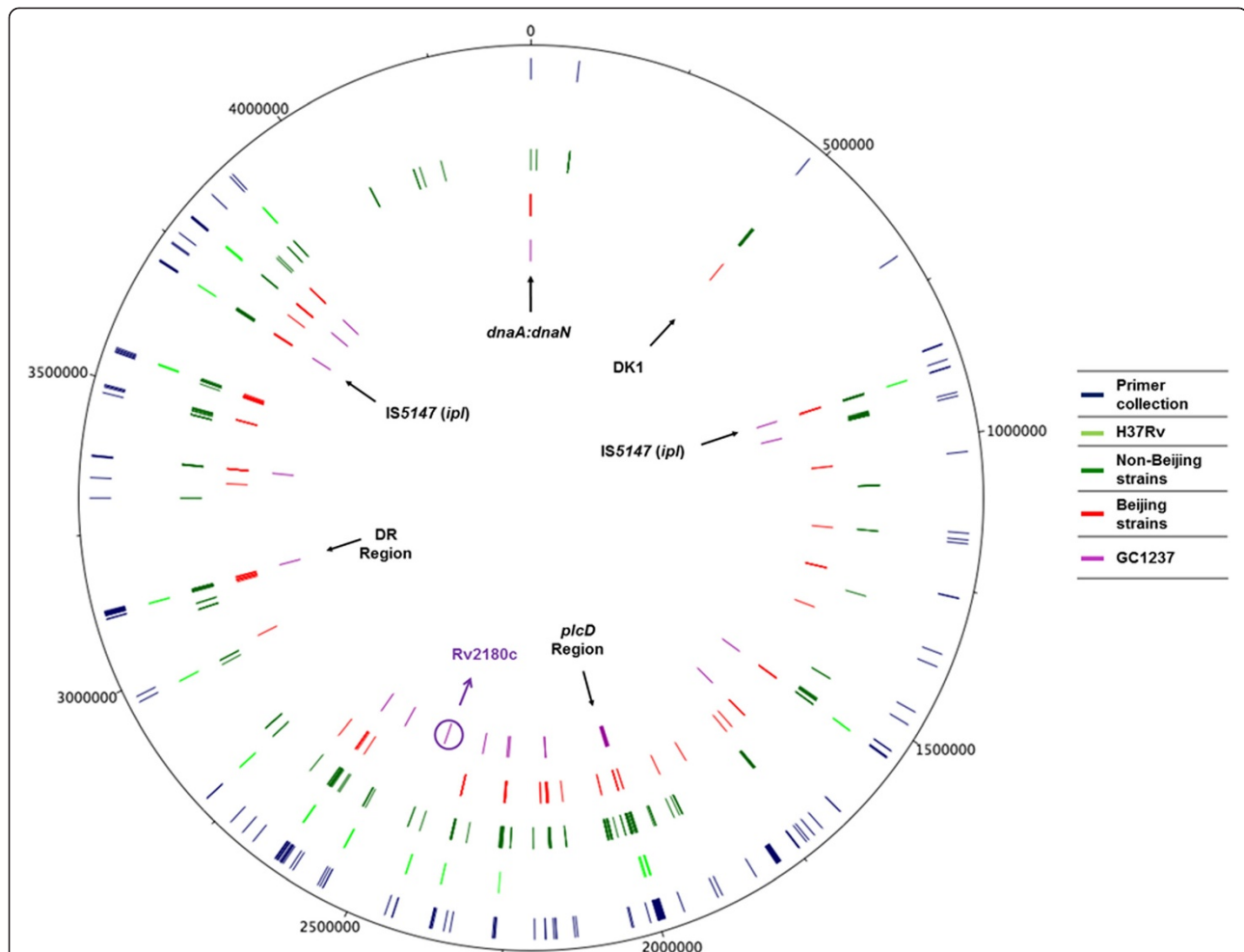


Figure 2). However, the presence of numerous preferential integration loci of IS6110 were also detected (some examples are indicated by arrows in Figure 2) in Beijing and non-Beijing strains, corroborating favored integration regions common among *M. tuberculosis* strains. In all strains, this element was found within the DR region. In the non-Beijing strain SUMu003, a copy was observed in a different point other than A1 in the *dnaA:dnaN* region, indicating that the presence of an IS6110 in this region is not exclusive to the Beijing family (Figure 2). This result is in agreement with studies which show different locations of IS6110 in the *dnaA:dnaN* region in non-Beijing strains [18,19,40]. The amplification of this entire intergenic region is a useful tool to detect Beijing isolates, but another genomic feature of this genotype is also necessary to avoid potential false positive results.

IS6110 was found inserted more often in some genomic regions (e.g., 1800000 bp - 2700000 bp) than in others that could be more abundant in essential genes (e.g., the region near *oriC*) (Figure 2). In fact, if an IS6110 transposes in essential regions, the outcome of this event would not be observed. These findings are in agreement with previous studies of chromosomal distribution of IS6110 [11,27,30,32].

#### Mapping IS6110 in 61 HCS

With the aim of studying the IS6110 insertion points and compare them between Beijing and non-Beijing HCS strains, we analysed 61 *M. tuberculosis* clinical isolates selected for their high copy number. The isolates comprised 44 non-Beijing and 17 Beijing strains, including the eight previously studied in this work.

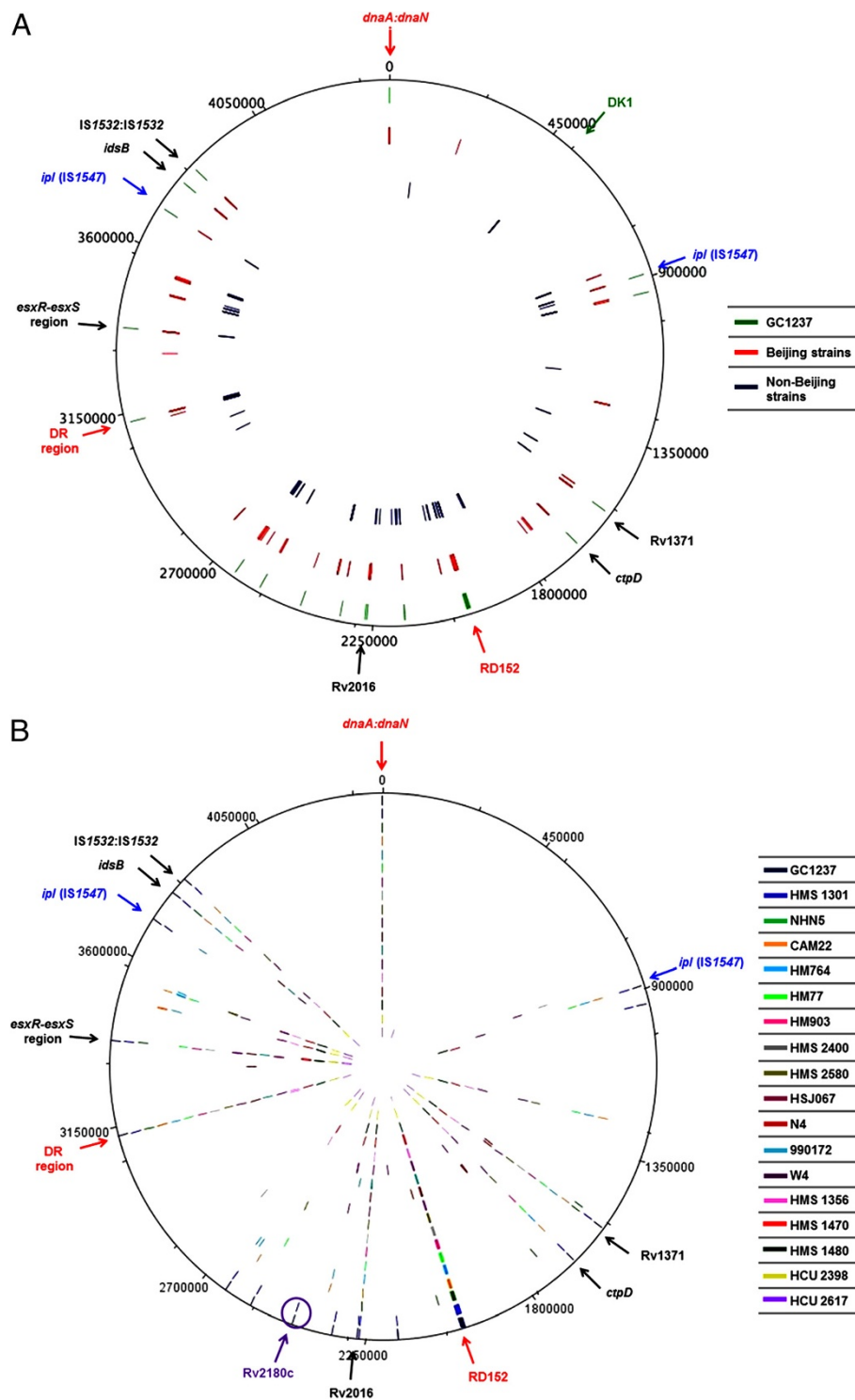


**Figure 2 Amplified regions with primer collection and distribution of points of insertion of IS6110 of *M. tuberculosis* sequenced genomes (GeneBank).** The first circle (blue) represents the distribution of the regions which can be amplified with the primer collection generated in this work. The second concentric circle (fluorescent green) represents the genomic locations of IS6110 in the reference strain H37Rv. The third and fourth circles represent the distribution of this sequence in non-Beijing strains (green) and in Beijing strains (red) respectively and the fifth concentric circle (purple) correspond to the genomic points of insertion of IS6110 in GC1237 used as another reference strain. All the points of insertion and amplified regions are plotted in *M. tuberculosis* H37Rv genome. The hot-spots: IS1547, *plcD* region, DR region, *dnaA:dnaN* region and DK1 region are indicated by arrows. The unique insertion of IS6110 of GC1237 [10] (within Rv2180c) is surrounded by a purple circle.

The study was carried out by PCR using the entire primer collection (138 reactions for each clinical isolate) (Additional file 1: Table S2) and when the amplified fragment indicated that the IS6110 was present, the product was sequenced with IS61 and IS62 primers (Additional file 1: Table S2). By this method we obtained the two flanking regions of all the copies. By analysis of the sequences, we obtained the insertion site, the flanking 3–4 bp direct repeats (DRs), the orientation and the distance to neighboring genes of each copy of IS6110 (Additional file 1: Tables S4 and S5). A total of 440 (160 different) insertion points were obtained (Additional file 1: Tables S4 and S5) and plotted on the genomic map of H37Rv and represented with DNA plotter in two different circles, red (insertions in Beijing strains) and blue (non-Beijing

strains) (Figure 3A). The locations of IS6110 in the reference Beijing strain GC1237 were included in a separated circle (green) as reference points of insertion.

From the 61 analysed clinical isolates, we localized a lower number of insertion points in non-Beijing strains (1 to 4) than in Beijing strains (at least 10). This finding is probably because the primer collection is based mainly in points of IS6110 from Beijing strains. All strains presented one copy of IS6110 in the DR region. Of the non-Beijing strains, 90.9% presented the same insertion site in this region. The remaining non-Beijing strains presented a copy in different points within this region (Additional file 1: Table S4), which is in agreement with other authors who have observed 16 different locations within the DR region [11].



**Figure 3** Distribution of IS6110 through the *M. tuberculosis* genome of the studied strains. **(A)** The obtained locations of IS6110 of each strain were plotted in *M. tuberculosis* H37Rv genome and represented in three concentric circles: the blue one corresponds to the IS6110 locations in non-Beijing strains. The red one represents the IS-locations in Beijing strains and the green one corresponds to the locations of IS6110 in GC1237 strain. The general hot-spots of *M. tuberculosis* are indicated by blue/red arrows, the DK1 region is indicated by green arrow and possible specific hot-spots of Beijing genotype are indicated with black arrows. **(B)** The IS6110 locations obtained in the 17 Beijing strains were plotted in *M. tuberculosis* H37Rv genome and represented each one in a concentric circle. The IS6110 locations of GC1237 were included as control. The general hot-spots are indicated by blue/red arrows and with black arrows the IS6110-Beijing-hot-spots.

Other special locations that we studied in HCS were the DK1 insertion point and the sites of IS6110 in *M. bovis* clinical isolates [22]. According to Fomukong et al. [28], the DK1 (*mmpS1* gene) insertion point is highly preferred in LCS and the authors defend the idea that its prevalence decreases in HCS, suggesting a separate lineage for HCS and LCS [28,29]. No IS6110 was detected in the DK1 locus in all 17 Beijing strains analysed (Figure 3A). Six of the 44 studied non-Beijing strains presented an IS6110 in the exact DK1 site (Additional file 1: Table S4) (Figure 3B). The high-copy number in these strains could be the result of transposition of another IS6110. Additionally, we found that none of the studied *M. tuberculosis* isolates have an equal insertion point of IS6110 as the *M. bovis* clinical isolates, supporting that *M. bovis* and *M. tuberculosis* evolved separately from a common precursor at an early stage.

The insertion of IS6110 in the characteristic Beijing spot RD152 was also observed in non-Beijing strains (region between Rv1754c-Rv1765c), indicating that this region is not exclusive to this family (Figure 3A and Additional file 1: Table S4). However, comparing the insertion points in Beijing and non-Beijing strains, only Beijing presented the deletion of RD152 generated by reorganization of two IS6110. Nonetheless, locations within Rv1371, Rv2016, *ctpD*, *idsB* genes and between the two IS1532 were only observed in Beijing strains (Figure 3A and Figure 3B, and Additional file 1: Table S4 and S5) as observed with the 8 Beijing isolates earlier in this work. Due to the high frequency of all of these IS6110 locations, these regions seem to be specific hotspots in Beijing strains. These results agree with Thorne et al. that showed the presence of some insertion points conserved within genetic lineages [41]. With the exception of the characteristic A1 point, to our knowledge, this is the first time that specific insertion points are described for Beijing strains, allowing us to speculate their possible relation with the fitness advantages of this family.

On the other hand unique locations were observed in *papA4* and Rv2957 genes (N4 strain), interrupting *mez* and *PPE49* genes (CAM22 strain) and in the intergenic region of Rv1542c:Rv1543 (W4 strain) (Additional file 1: Table S5). We corroborated the uniqueness of these sites, after analyzing the literature [10,11,13,16-18,20,22-24,27-35], as was the case for the copy located upstream of Rv2180c gene in GC1237 [10] (Figure 1 and Figure 2). The data suggests that although IS6110 has preferential genomic regions, its insertion is sufficiently random that could generate differences among strains of the same family.

#### IS6110 distribution in the eleven functional categories of *M. tuberculosis*

In the analyzed 61 strains, we observed that 58% of the IS6110 insertions (94 of 160 points) occurred in ORFs

(Additional file 1: Table S4). The interruption of coding regions can be seen as a naturally occurring knock-out assay and could provide information on non-essential genes for mycobacteria capacity of infection in human host, as all the studied strains are clinical isolates. Given that the ORFs represent 91% of *M. tuberculosis* genome [9], our data suggest that transposition is relatively more frequent in intergenic than in intragenic regions. The higher number of intergenic events is likely due to selection, as they are less probable to be deleterious than those that occur within genes. Such events could increment the probability of insertion of IS6110 in possible promoter regions, which could influence the expression of neighbouring genes.

Our results agree with other studies which found that 58% of discrete IS6110 insertion sites occurred within coding regions in *M. tuberculosis* [32] and in *M. bovis* strains [22].

We observed that IS6110 does not interrupt ORFs of 3 of the 11 functional categories: the stable RNAs, information pathways and IS and phages (<http://tuberculist.epfl.ch/>). Only one gene, Rv2103c, of the virulence category was disrupted (Additional file 1: Table S4). The distribution of the interrupted genes of the rest of the categories was markedly similar to the percentage of ORFs in *M. tuberculosis* genome with high frequency of disrupted genes in categories of hypothetical proteins and the intermediary metabolism and respiration. The details of the distribution in the 11 categories are indicated in Additional file 1: Table S4.

Several studies have indicated that it is frequent to find IS6110 inserted in PPE/PE genes [17,21,22,27,31,35]. These genes are associated with antigenic variation in *M. tuberculosis* [42] but it has been suggested that the disruption of a member of this family would not be expected to produce severe disadvantages as it can be compensated by another member [27]. We observed that the interrupted PPE genes were PPE16, PPE34, PPE38, PPE40 and PPE49 in several points and in both orientations and there were no disrupted PE genes (Additional file 1: Table S4). PPE40 has been suggested as an essential gene [43], but one of the studied strains (clinical isolate) presents a copy in it.

#### Analysis of DRs flanking IS6110

After analysing the flanking sequences of each copy of IS6110 in the studied strains, we observed that 80% of the different insertion points (128 of 160) were flanked by DRs of 3–4 bp (Additional file 1: Table S4), indicating that these were the result of transposition events. The other 32 copies without detected DRs showed genomic reorganizations or loss of genomic regions. The deletions of RD152 and RD207 are examples of recombination between two adjacent copies of IS6110 [44]. The

IS6110 without DRs localized between Rv0794c:Rv0797, is in the opposite orientation to the IS6110 in H37Rv and in identical position as the copy in the reference Beijing GC1237 [10], and it is associated with genomic reorganization of this region. Due to the high number of copies of IS6110 per strain, it was possible to observe copies flanked by DRs and other copies without DRs suggesting that the probability of rearrangement between copies rises when the number of those increases, producing more variability among strains. This finding is in agreement with different studies indicating that strains with a high number of IS6110 copies have lost genomic regions more often than strains with only few copies [10,22]. Although in some of the studied HCS the number of the located IS6110 copies was less than 5, some of them were observed without DRs, corroborating the idea that the probability of rearrangement processes between copies rises when the number increases. This is in agreement with other studies which localized IS6110 in LCS and observed that all copies were flanked by DRs [22,45]. In one copy of IS6110 we detected DRs of 5 nucleotides (Additional file 1: Table S4).

#### **Twenty percent of the located copies of IS6110 in the studied strains could act as mobile promoter**

As we reported previously in this work, IS6110 is relatively more frequent in intergenic regions, increasing its probability of being inserted in promoter regions, influencing the expression of neighbouring genes. Different studies have indicated that when IS6110 is inserted in the same orientation as, and close enough to, a downstream gene could potentially function as a promoter [10,23,24]. The orientation of the 440 located copies of IS6110 in the 61 strains and the distance to the close genes were analysed in order to test the promoter function of this element. Thirty-two locations of IS6110 were located close enough to (less than 400 bp) and in the same orientation as the neighboring gene (Additional file 1: Table S4). Of the 32 candidate locations to act as a mobile promoter, 4 of them were observed at a frequency of 3.2% and 23 at 1.6%. The remaining 5 locations were observed in a higher number of strains with 3 of these locations observed in non-Beijing strains and 2 in Beijing strains (Additional file 1: Table S4). Of note, one of the two frequent locations among Beijing strains corresponds to the copy located in *ctpD* gene and upstream Rv1468c gene and it has already been demonstrated that this copy is acting as a promoter inside monocytes [23]. The fact that this location is quite frequent in Beijing genotype and even is specific of this family, could be one of the special features of the Beijing family in terms of virulence and transmission.

Two of the 61 strains presented a copy of IS6110 in the promoter region of *phoP* gene (Additional file 1:

Table S4). The multidrug resistant (MDR) strain *M. bovis* B or MBZ responsible for large tuberculosis outbreaks in Spain has a copy of IS6110 located 75 bp upstream the *phoP* gene [24]. Soto *et al.* demonstrated that this IS6110 causes an increment in the transcription of *phoP* [24], which could have an important consequence as the product of this gene is an important transcriptional regulator [46]. One of the two strains (non-Beijing strain HMS 2405) presents a copy of IS6110 196 bp upstream and in the same orientation as *phoP* gene. In this case, the *phoP* promoter region is not interrupted, as IS6110 is inserted upstream of transcription starting points, *tsp1* and *tsp2*, of this gene and could also provide an additional *tsp*. Although this point is different from that in MBZ, HMS 2405 could be an interesting candidate for studying the effect of IS6110 in *phoP* gene in a drug-susceptible *M. tuberculosis* strain.

#### **Generation of an accurate dendrogram based on IS6110 points of insertion**

After observing that the 17 studied Beijing strains shared a high number of locations not present in non-Beijing strains, we decided to group the 61 clinical isolates based on their insertion points. The obtained dendrogram, using Bionumerics program, (Figure 4A) shows that Beijing strains were grouped better than when using the IS6110-RFLP classification system (Figure 4B). Furthermore, some non-Beijing strains were also grouped in families despite having very few located copies of IS6110. This fact could indicate that probably each family has their preferential sites of insertion. However, the grouping based on IS6110 locations in non-Beijing families was not perfect due to the lack of information on all their points of insertion.

The classification technique based on IS6110 insertion points developed in this study allows grouping the strains in families as spoligotyping and provides information on the clone or strain as IS6110-RFLP. As Figure 4A shows, although IS6110-RFLP groups genotypes, if the IS6110-RFLP of a Beijing strain is quite different from other Beijing strains, the strain is classified as distant from the group. Based on the results obtained in this study, the development of a 96-well plate with pairs of primers which amplify copies of IS6110 characteristics of each family could allow to obtain a dendrogram in one day. This PCR-based method is quick, accurate and economic and possible to complete in one day.

#### **Conclusions**

This study provides a detailed analysis of the locations of IS6110 in *M. tuberculosis* Beijing genotype compared to other HCS, including the exact insertion sites, the flanking DRs, the orientation and the distance of IS6110 to neighbouring genes. In general, the insertion of





common linker primer SALGD (Additional file 1: Table S2). PCR products were purified using GFX PCR DNA gel band purification kit (Amersham Pharmacia Biotech) and the restriction enzyme ExoSAP-IT® (Affymetrix). The amplified products were sequenced with the corresponding oligonucleotides and when a match was found, additional primers (Additional file 1: Table S2) were designed and used with the 8 strains to verify whether the point of insertion was present in other of the 8 strains. These primers amplify the completed sequence of IS6110 and approximately 300 bp of both flanking sequences (Additional file 1: Table S2).

Secondly, PCRs were performed with specific primers (Additional file 1: Table S2), designed in a previous study for amplifying the locations of IS6110 in 210, W [31] and GC1237 strains [10]. PCRs were carried out in a total volume of 25 µl, containing 50 ng of DNA, 2.5 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 to 2 min depending on the amplified product. After the last cycle, the samples were incubated at 72°C for 10 min.

The genomes of H37Rv and GC1237 were used as control in both cases.

#### Sequence analysis of points of insertion of IS6110 in the DNA database and in the available literature

The points of insertion of IS6110 of the reference sequenced strains: Beijing (210, 02\_1987, 94\_M4241A, HN878, R1207, T85, X-122, W and W-148 ) and non-Beijing (98-R604 INH-RIF-EM, BTB05-552, BTB05-559, C, CDC1551, CDC1551A, CPHL\_A, EAS054, F11, GM 1503, K85, KZN 605, KZN R506, KZN V2475, KZN 1435, KZN 4207, NCGM 2209, str.Haarlem, S96-129, SUMu001, SUMu002, SUMu003, SUMu004, SUMu005, SUMu006, SUMu007, SUMu008, SUMu009, SUMu010, SUMu011, SUMu012, T17, T46 and T92) were obtained comparing the flanking regions of each IS6110 in the genome sequences with the reference strain H37Rv using NCBI genetic sequence database (GeneBank) (<http://www.ncbi.nlm.nih.gov/genome/166>).

Moreover, available literature was analyzed to identify any further point of insertion of this sequence not described in the DNA database [10,11,13,16-18,22-24,27-35].

After this sequence analysis, when a point of insertion of IS6110 of a Beijing strain (of GeneBank or available literature) was found outside of the amplified regions of our primer collection, additional primers were designed and included in the collection (Additional file 1: Table S2). In addition, flanking primers of DK regions [28] were designed to detect whether these frequent locations of

IS6110 in LCS were also present in the HCS selected in this study. Moreover, the primers designed for the study of locations of IS6110 in clinical isolates of *M. bovis* with human host [22] were also included to study whether preferred locations of IS6110 in these *M. bovis* strains were also present in the selected *M. tuberculosis* strains.

#### Localization and analysis of copies of IS6110 in the sixty-one selected clinical isolates

The localization of copies of IS6110 in the 61 clinical isolates was carried out by PCR as previously describe in this work with all the pairs of oligonucleotides of the generated primer collection (Additional file 1: Table S2). H37Rv and GC1237 were used as external controls and HMS 1301 as internal control as this strain presents the same IS6110-RFLP as GC1237. The 8 representative Beijing strains studied before in this work by LMPCR and specific PCR were again included in this part of the study. The PCR products which might include an IS6110 were sequenced with IS61 and IS62 primers (Additional file 1: Table S2).

#### Determination of direct repeats (DR) and analysis of the flanking regions of each copy of IS6110 in the genomes

The DRs generated by the mechanism of the transposition of IS6110 were determined with the sequence analysis of the flanking regions of each copy of IS6110 in the genomes.

#### Dendrogram based on the points of insertion of IS6110

The informatic analysis of spoligotyping results is carried out by Bionumerics program and is based in presence/absence or numerical analysis 1/0 of a specific sequence of *M. tuberculosis*. Based on this idea, all the obtained IS-locations were arranged in columns on an excel sheet and the 61 strains (also the two controls GC1237 and H37Rv) were arranged in rows. When a strain presented a point of insertion, it was given the number 1 and if not, it was assigned the number 0. After that, the data was introduced in SpolDB4 database as a new informatic event, included to the IS6110-RFLP and the spoligotyping of each strain and analyzed by Bionumerics program.

#### Additional file

**Additional file 1: Supplementary informations.**

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

Conceived and designed the experiments: HA, IO. Performed the experiments HA. Analyzed the data HA, SS, CM, IO. Wrote the paper: HA, SS, CM, IO. All authors read and approved the final manuscript.

## Acknowledgements

We thank Dessislava Marinova for her kind help in English revision and for her valuable comments on the manuscript. We would like to thank to Alberto Cebollada, Carlos Lampreave and Carmen Lafoz for their technical support. Thanks to the network working in TB in Aragón, specially the micobacterial laboratories of the Hospital Universitario Miguel Servet (HMS), Hospital Clínico Universitario Lozano-Blesa (HCU) from Zaragoza and Hospital General San Jorge from Huesca (Spain) for providing the well characterized strains. This work is part of the TB-VIR Project supported by the European Commission under the Health Cooperation of the 7<sup>th</sup> Framework Programme 200973, Project of the Ministerio de Economía y Competitividad BIO2011/23555 and partially supported by FIS09/051. Henar Alonso was financially supported by Gobierno de Aragón (B034/09).

## Author details

<sup>1</sup>Grupo de Genética de Micobacterias. Departamento de Microbiología, Medicina Preventiva y Salud Pública, Universidad de Zaragoza, C/ Domingo Miral sn. 50009, Zaragoza, Spain. <sup>2</sup>Laboratorio de Investigación Molecular, Hospital Universitario Miguel Servet, IIS Aragón, Zaragoza, Spain. <sup>3</sup>CIBER Enfermedades Respiratorias (CIBERES), Madrid, Spain. <sup>4</sup>Institut de Pharmacologie et de Biologie Structurale, UMR5089 CNRS, 205 Route de Narbonne, BP 64182, 31077 Toulouse, France.

Received: 27 February 2013 Accepted: 19 June 2013

Published: 25 June 2013

## References

1. Thierry D, Brisson-Noel A, Vincent-Levy-Frebault V, Nguyen S, Guesdon JL, Gicquel B: **Characterization of a Mycobacterium tuberculosis insertion sequence, IS6110, and its application in diagnosis.** *J Clin Microbiol* 1990, **28**:2668–2673.
2. Brosch RG, Gordon SV, Eiglmeier K, Garnier T, Tekaija F, Yeramian E, Cole ST: **Genomics, biology, and evolution of the Mycobacterium tuberculosis complex.** In *Molecular Genetics of Mycobacteria*. Edited by Hatfull GF & Jacobs WR Jr. Washington DC, USA: American Society for Microbiology Press; 2000:19–36. ISBN ISBN 1-55581-191-4.
3. Fomukong NG, Tang TH, al-Maamary S, Ibrahim WA, Ramayah S, Yates M, Zainuddin ZF, Dale JW: **Insertion sequence typing of Mycobacterium tuberculosis: characterization of a widespread subtype with a single copy of IS6110.** *Tuber Lung Dis* 1994, **75**:435–440.
4. Lok KH, Benjamin WH Jr, Kimerling ME, Pruitt V, Lathan M, Razeq J, Hooper N, Cronin W, Dunlap NE: **Molecular differentiation of Mycobacterium tuberculosis strains without IS6110 insertions.** *Emerg Infect Dis* 2002, **8**:1310–1313.
5. Steensels D, Fauville-Dufaux M, Boie J, De Beenhouwer H: **Failure of PCR-Based IS6110 Analysis To Detect Vertebral Spondylodiscitis Caused by Mycobacterium bovis.** *J Clin Microbiol* 2012, **51**:366–368.
6. Fomukong NG, Dale JW, Osborn TW, Grange JM: **Use of gene probes based on the insertion sequence IS986 to differentiate between BCG vaccine strains.** *J Appl Bacteriol* 1992, **72**:126–133.
7. Millan-Lou MI, Alonso H, Gavin P, Hernandez-Febles M, Campos-Herrero MI, Copado R, Canas F, Kremer K, Caminero JA, Martin C, Samper S: **Rapid test for identification of a highly transmissible Mycobacterium tuberculosis Beijing strain of sub-Saharan origin.** *J Clin Microbiol* 2012, **50**:516–518.
8. Fang Z, Doig C, Kenna DT, Smittipat N, Palittapongarnpim P, Watt B, Forbes KJ: **IS6110-mediated deletions of wild-type chromosomes of Mycobacterium tuberculosis.** *J Bacteriol* 1999, **181**:1014–1020.
9. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, et al: **Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.** *Nature* 1998, **393**:537–544.
10. Alonso H, Aguilo JI, Samper S, Caminero JA, Campos-Herrero MI, Gicquel B, Brosch R, Martin C, Otal I: **Deciphering the role of IS6110 in a highly transmissible Mycobacterium tuberculosis Beijing strain, GC1237.** *Tuberculosis (Edinb)* 2011, **91**:117–126.
11. Reyes A, Sandoval A, Cubillos-Ruiz A, Varley KE, Hernandez-Neuta I, Samper S, Martin C, Garcia MJ, Ritacco V, Lopez L, et al: **IS-seq: a novel high throughput survey of in vivo IS6110 transposition in multiple Mycobacterium tuberculosis genomes.** *BMC Genomics* 2012, **13**:249.
12. Hermans PW, Van Soolingen D, Bik EM, De Haas PE, Dale JW, Van Embden JD: **Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains.** *Infect Immun* 1991, **59**:2695–2705.
13. Fang Z, Forbes KJ: **A Mycobacterium tuberculosis IS6110 preferential locus (ipl) for insertion into the genome.** *J Clin Microbiol* 1997, **35**:479–481.
14. Fang Z, Kenna DT, Doig C, Smittipat DN, Palittapongarnpim P, Watt B, Forbes KJ: **Molecular evidence for independent occurrence of IS6110 insertions at the same sites of the genome of Mycobacterium tuberculosis in different clinical isolates.** *J Bacteriol* 2001, **183**:5279–5284.
15. Iøerger TR, Feng Y, Ganesula K, Chen X, Dobos KM, Fortune S, Jacobs WR Jr, Mizrahi V, Parish T, Rubin E, et al: **Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories.** *J Bacteriol* 2010, **192**:3645–3653.
16. Vera-Cabrera L, Hernandez-Vera MA, Welsh O, Johnson WM, Castro-Garza J: **Phospholipase region of Mycobacterium tuberculosis is a preferential locus for IS6110 transposition.** *J Clin Microbiol* 2001, **39**:3499–3504.
17. McEvoy CR, Warren RM, van Helden PD, Gey van Pittius NC: **Multiple, independent, identical IS6110 insertions in Mycobacterium tuberculosis PPE genes.** *Tuberculosis (Edinb)* 2009, **89**:439–442.
18. Kurepina NE, Sreevatsan S, Plikaytis BB, Bifani PJ, Connell ND, Donnelly RJ, Van Soolingen D, Musser JM, Kreiswirth BN: **Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in Mycobacterium tuberculosis: non-random integration in the dnaA-dnaN region.** *Tuber Lung Dis* 1998, **79**:31–42.
19. Turcios L, Casart Y, Florez I, De Waard J, Salazar L: **Characterization of IS6110 insertions in the dnaA-dnaN intergenic region of Mycobacterium tuberculosis clinical isolates.** *Clin Microbiol Infect* 2009, **15**:200–203.
20. Cardona PJ: **Undestanding tuberculosis - Deciphering the secret life of the Bacilli. Chapter 3: IS6110 the Double-Edged Passenger.** *INTECH* 2012, **1**:69–98.
21. McEvoy CR, Falmer AA, Warren RM, van Helden PD, Victor TC, Gey van Pittius NC: **The role of IS6110 in the evolution of Mycobacterium tuberculosis.** *Tuberculosis (Edinb)* 2007, **87**:393–404.
22. Otal I, Gomez AB, Kremer K, De Haas P, Garcia MJ, Martin C, Van Soolingen D: **Mapping of IS6110 insertion sites in Mycobacterium bovis isolates in relation to adaptation from the animal to human host.** *Vet Microbiol* 2008, **129**:333–341.
23. Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, Howard ST: **IS6110 functions as a mobile, monocyte-activated promoter in Mycobacterium tuberculosis.** *Mol Microbiol* 2004, **52**:999–1012.
24. Soto CY, Menendez MC, Perez E, Samper S, Gomez AB, Garcia MJ, Martin C: **IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks.** *J Clin Microbiol* 2004, **42**:212–219.
25. Hanekom M, Van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM: **Mycobacterium tuberculosis Beijing genotype: a template for success.** *Tuberculosis (Edinb)* 2011, **91**:510–523.
26. Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, Musser JM: **Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites.** *J Infect Dis* 2006, **193**:121–128.
27. Sampson SL, Warren RM, Richardson M, van der Spuy GD, Van Helden PD: **Disruption of coding regions by IS6110 insertion in Mycobacterium tuberculosis.** *Tuber Lung Dis* 1999, **79**:349–359.
28. Fomukong N, Beggs M, El Hajj H, Templeton G, Eisenach K, Cave MD: **Differences in the prevalence of IS6110 insertion sites in Mycobacterium tuberculosis strains: low and high copy number of IS6110.** *Tuber Lung Dis* 1997, **78**:109–116.
29. Dale JW, Al-Ghusein H, Al-Hashmi S, Butcher P, Dickens AL, Drobniowski F, Forbes KJ, Gillespie SH, Lamprecht D, McHugh TD, et al: **Evolutionary relationships among strains of Mycobacterium tuberculosis with few copies of IS6110.** *J Bacteriol* 2003, **185**:2555–2562.
30. Warren RM, Sampson SL, Richardson M, Van Der Spuy GD, Lombard CJ, Victor TC, Van Helden PD: **Mapping of IS6110 flanking regions in clinical isolates of Mycobacterium tuberculosis demonstrates genome plasticity.** *Mol Microbiol* 2000, **37**:1405–1416.
31. Beggs ML, Eisenach KD, Cave MD: **Mapping of IS6110 insertion sites in two epidemic strains of Mycobacterium tuberculosis.** *J Clin Microbiol* 2000, **38**:2923–2928.
32. Sampson S, Warren R, Richardson M, van der Spuy G, Van Helden P: **IS6110 insertions in Mycobacterium tuberculosis: predominantly into coding regions.** *J Clin Microbiol* 2001, **39**:3423–3424.

33. Kong Y, Cave MD, Yang D, Zhang L, Marrs CF, Foxman B, Bates JH, Wilson F, Mukasa LN, Yang ZH: **Distribution of insertion- and deletion-associated genetic polymorphisms among four *Mycobacterium tuberculosis* phospholipase C genes and associations with extrathoracic tuberculosis: a population-based study.** *J Clin Microbiol* 2005, **43**:6048–6053.
34. Viana-Niero C, Rodriguez CA, Bigi F, Zanini MS, Ferreira-Neto JS, Cataldi A, Leao SC: **Identification of an IS6110 insertion site in *plcD*, the unique phospholipase C gene of *Mycobacterium bovis*.** *J Med Microbiol* 2006, **55**:451–457.
35. McEvoy CR, Van Helden PD, Warren RM, Gey van Pittius NC: **Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic *Mycobacterium tuberculosis* PPE38 gene region.** *BMC Evol Biol* 2009, **9**:237.
36. Plikaytis BB, Marden JL, Crawford JT, Woodley CL, Butler WR, Shinnick TM: **Multiplex PCR assay specific for the multidrug-resistant strain W of *Mycobacterium tuberculosis*.** *J Clin Microbiol* 1994, **32**:1542–1546.
37. Kurepina N, Likhoshvay E, Shashkina E, Mathema B, Kremer K, Van Soolingen D, Bifani P, Kreiswirth BN: **Targeted hybridization of IS6110 fingerprints identifies the W-Beijing *Mycobacterium tuberculosis* strains among clinical isolates.** *J Clin Microbiol* 2005, **43**:2148–2154.
38. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN: **Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains.** *Trends Microbiol* 2002, **10**:45–52.
39. Kremer K, Glynn JR, Lillebaek T, Niemann S, Kurepina NE, Kreiswirth BN, Bifani PJ, Van Soolingen D: **Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers.** *J Clin Microbiol* 2004, **42**:4040–4049.
40. Casart Y, Turcios L, Florez I, Jaspe R, Guerrero E, De Waard J, Aguilar D, Hernandez-Pando R, Salazar L: **IS6110 in *oriC* affects the morphology and growth of *Mycobacterium tuberculosis* and attenuates virulence in mice.** *Tuberculosis (Edinb)* 2008, **88**:545–552.
41. Arnold C, Gharbia S, Gonzalez-Martin J, Bishop C, GarciadeViedma D, Magee J, Evans J, Borrell S, Thorne N: **IS6110-based global phylogeny of *Mycobacterium tuberculosis*.** *Infect Genet Evol* 2011, **11**:132–138.
42. Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, Cole ST: **Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens?** *Mol Microbiol* 2002, **44**:9–19.
43. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sasseti CM: **High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism.** *PLoS Pathog* 2011, **7**:1002251.
44. Tsolaki AG, Gagneux S, Pym AS, Small PM, Van Soolingen D, Kreiswirth BN, Goguet de la Salmoniere YO: **Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*.** *J Clin Microbiol* 2005, **43**:3185–3191.
45. Mendiola MV, Martin C, Otal I, Gicquel B: **Analysis of the regions responsible for IS6110 RFLP in a single *Mycobacterium tuberculosis* strain.** *Res Microbiol* 1992, **143**:767–772.
46. Gonzalo-Asensio J, Mostowy S, Harders-Westerveen J, Huygen K, Hernandez-Pando R, Thole J, Behr M, Gicquel B, Martin C: **PhoP: a missing piece in the intricate puzzle of *Mycobacterium tuberculosis* virulence.** *PLoS One* 2008, **3**:e3496.
47. Van Soolingen D, De Haas PE, Hermans PW, Van Embden JD: **DNA fingerprinting of *Mycobacterium tuberculosis*.** *Methods Enzymol* 1994, **235**:196–205.
48. Prod'homme G, Lagier B, Pelicic V, Hance AJ, Gicquel B, Guillhot C: **A reliable amplification technique for the characterization of genomic DNA sequences flanking insertion sequences.** *FEMS Microbiol Lett* 1998, **158**:75–81.

doi:10.1186/1471-2164-14-422

**Cite this article as:** Alonso et al.: Mapping IS6110 in high-copy number *Mycobacterium tuberculosis* strains shows specific insertion points in the Beijing genotype. *BMC Genomics* 2013 **14**:422.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

