



**HAL**  
open science

## IS6110 Copy Number in Multi-Host Mycobacterium bovis Strains Circulating in Bovine Tuberculosis Endemic French Regions

Ciriac Charles, Cyril Conde, Franck Biet, Maria Laura Boschioli, Lorraine Michelet

► **To cite this version:**

Ciriac Charles, Cyril Conde, Franck Biet, Maria Laura Boschioli, Lorraine Michelet. IS6110 Copy Number in Multi-Host Mycobacterium bovis Strains Circulating in Bovine Tuberculosis Endemic French Regions. *Frontiers in Microbiology*, 2022, Evolution and Genomics of the Mycobacterium tuberculosis complex, 23 (13), pp.891902. 10.3389/fmicb.2022.891902 . hal-03738004

**HAL Id: hal-03738004**

**<https://hal.inrae.fr/hal-03738004v1>**

Submitted on 25 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# IS6110 Copy Number in Multi-Host *Mycobacterium bovis* Strains Circulating in Bovine Tuberculosis Endemic French Regions

Ciriac Charles<sup>1,2</sup>, Cyril Conde<sup>2</sup>, Franck Biet<sup>2†</sup>, Maria Laura Boschirolì<sup>1†</sup> and Lorraine Michelet<sup>1\*†</sup>

<sup>1</sup> ANSES, Animal Health Laboratory, National Reference Laboratory for Tuberculosis, Paris-Est University, Paris, France,

<sup>2</sup> INRAE, ISP, Université de Tours, Nouzilly, France

IS6110 is an insertion sequence found in the *Mycobacterium tuberculosis* complex, to which *Mycobacterium bovis* belongs, which can play a role in genome plasticity and in bacterial evolution. In this study, the abundance and location of IS6110 on *M. bovis* genomic data of French animal field strains were studied. A first analysis was performed on a panel of 81 strains that reflect the national *M. bovis* population's genetic diversity. The results show that more than one-third of them are IS6110 multicopy and that 10% have IS6110 in a high copy number (more than 6 copies). Multicopy strains are those circulating in the regions where prevalence was above the national average. Further study of 93 such strains, with an IS6110 copy number of 10-12, showed stability of IS6110 copy number and genome location over time and between host species. The correlation between *M. bovis* multicopy strains and high bovine tuberculosis (bTB) prevalence leads us to consider whether their epidemiological success could be partly due to genetic changes originated by IS6110 transposition.

**Keywords:** *Mycobacterium bovis*, bovine tuberculosis (bTB), insertion sequence (IS), IS6110, France

## OPEN ACCESS

### Edited by:

Adrian Allen,  
Queen's University Belfast,  
United Kingdom

### Reviewed by:

Stephen V. Gordon,  
University College Dublin, Ireland  
Maxime Godfroid,  
Collège de France, France

### \*Correspondence:

Lorraine Michelet  
lorraine.michelet@anses.fr

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 08 March 2022

**Accepted:** 03 May 2022

**Published:** 23 June 2022

### Citation:

Charles C, Conde C, Biet F,  
Boschirolì ML and Michelet L (2022)  
IS6110 Copy Number in Multi-Host  
*Mycobacterium bovis* Strains  
Circulating in Bovine Tuberculosis  
Endemic French Regions.  
*Front. Microbiol.* 13:891902.  
doi: 10.3389/fmicb.2022.891902

## INTRODUCTION

IS6110 is a specific insertion sequence of the *Mycobacterium tuberculosis* complex (MTBC) historically used for genotyping techniques, mainly restriction fragment length polymorphism (RFLP) IS6110, the former gold standard to identify epidemiologically linked isolates of the MTBC species (van Soolingen et al., 1994). Easier and more discriminatory methods such as spoligotyping, mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR), and more recently, whole-genome single nucleotide polymorphism (wgSNP) have been developed (Kamerbeek et al., 1997; Supply et al., 2006; Comas et al., 2009; Driscoll, 2009; Hauer et al., 2019).

IS6110 is of particular interest due to its role in genome plasticity and bacterial evolution. In fact, this insertion sequence can promote gene inactivation and structural variation (insertion, inversion, or deletion) and can act as a mobile promoter (Soto et al., 2004; Alonso et al., 2013). As an example, the insertion of IS6110 in the *phoP* promoter has been described as the cause of gene upregulation and virulence increase of the *Mycobacterium bovis* (*M. bovis*) B strain, which caused severe human multidrug resistant tuberculosis outbreaks in Spain (Samper et al., 1997; Soto et al., 2004). In addition, the correlation between the high amount of IS6110 in *M. tuberculosis* Beijing lineage and its increased virulence, antibiotic resistance, and the ability of these strains to better adapt to the environment have been described (Kremer et al., 2004; McEvoy et al., 2007; Merker et al., 2013). The

number of IS6110 copies varies among MTBC members: strains belonging to the *M. tuberculosis* Beijing lineage have the most important number of IS6110 copies (15 on average), while *M. bovis* strains present only one or few copies of IS6110 (Gonzalo-Asensio et al., 2018). Nonetheless, the recent publication of a new complete genome sequence of a French field strain, *M. bovis* Mb3601, showed that this particular strain possess 11 IS6110 copies (Branger et al., 2020). This strain, which belongs to the European 3 (Eu3) complex clonal, is found in an endemic bovine tuberculosis (bTB) region (Côte d'Or in Burgundy, Central-East France) and belongs to one of the most abundant genotypes in France in the last years (Hauer et al., 2015).

Bovine tuberculosis is an old worldwide chronic zoonotic disease due to *M. bovis*. The main maintenance host species of *M. bovis* is domestic cattle but this bacterium can circulate in multi-host systems that include not only domestic but also wild animals and their environment, therefore, explaining the persistence of the pathogen in some areas (Réveillaud et al., 2018). France obtained the officially bTB free (OTF) status in 2001 (Benet et al., 2006). However, this OTF status is threatened by a steady rise of bTB cattle outbreaks over the past 15 years, partly due to more efficient surveillance in cattle and also due to the presence of wildlife infection which contributes to the persistence of the disease at regional levels.

A previous study shows a strong regionalization of bTB in France caused by specific *M. bovis* genotypes in different endemic regions (Hauer et al., 2015). Totally, 80% of bTB outbreaks in France in the last years are due to 3 major *M. bovis* genotypes (Delavenne et al., 2021): SB0120-DHV genotype (SB0120-VNTR 5 3 5 3 9 4 5 6) found in Dordogne and Haute-Vienne departments (Nouvelle Aquitaine region); SB0120-CO genotype (SB0120-VNTR 5 5 4 3 11 4 5 6) found in Côte d'Or department (Burgundy region); Cluster A/F4 family genotypes (Hauer et al., 2019), especially due to SB0821/F007 and SB0832/F015 genotypes in the Atlantic Pyrenees department (Nouvelle Aquitaine region) and SB0840/F001 genotype in Corsica.

The persistence of these main genotypes could be due to the ability of the strains to adapt to the environment. In fact, these strains seem to be maintained and expand geographically despite the implementation of control measures adapted to each context (Boschiroli et al., 2015; Hauer et al., 2015; Delavenne et al., 2021).

The objectives of this study were (i) to establish using a whole-genome sequencing (WGS) approach the abundance of IS6110 copy and their chromosomal distribution in 80 strains representative of French *M. bovis* genotypes diversity and (ii) to evaluate the level of the IS6110 sequence stability and their orthology in a panel of 92 sympatric strains of SB0120-CO genotype evolving in a multi-host system.

## MATERIALS AND METHODS

### Genomes of *Mycobacterium bovis* French Strains

The first panel (panel 1) constituted genomes of 80 *M. bovis* strains isolated from French bovine tuberculosis outbreaks between 1983 and 2011 which were selected from a previous

study (Supplementary Figure 1 and Supplementary Table 1) (Hauer et al., 2019). DNA from bacterial clones was extracted using the phenol-chloroform method detailed in a previous study and sequenced with Illumina HiSeq technology [paired-end (2 × 100 bp)] (Hauer et al., 2019). The second panel (panel 2) constituted 92 SB0120-CO strains that have been selected from another study (Michelet et al., 2018). Bacterial thermolysates without any further purification steps were used as DNA sources. DNA sequencing was carried out with Illumina HiSeq technology [paired-end (2 × 250 bp)] (Michelet et al., 2018). These strains, isolated from the bTB endemic region in Côte d'Or (Central East France), were obtained from different animal hosts in the framework of the French bTB control campaign between 2009 and 2014 (Supplementary Figure 1 and Supplementary Table 1).

The short-read quality was evaluated using FastQC (version 0.11.9 with default parameters), and reads were trimmed with Sickle<sup>1</sup> (version 1.33 with default parameters) using a quality phred-score of Q20 (Xia et al., 2016; Wingett and Andrews, 2018). The SPAdes (version 3.15.2 with careful option) tool was used to assemble short reads (Bankevich et al., 2012), and Prokka (version 1.14.6 with default parameters) was used on the genomes' assemblies (Seemann, 2014) to check the absence of contamination of the short-read sequencing. Strains of panel 2 with more than 3 rRNA or with a size between 4,200,000 and 4,600,000 base pairs were considered for further analyses.

### Phylogenetic Analysis

Single nucleotide polymorphisms were obtained using the Bionumerics software, version 7.6 (AppliedMath, Belgium). Identified SNPs were selected according to strict criteria of the wgSNP module as follows: they had to be present on at least five reads in both forward and reverse directions, twelve base pairs had to separate them, they were not present in repetitive regions of the genome, and ambiguous SNPs [at least one unreliable (N) base, ambiguous (non-ATCG) base, or gap] were not included. The evolutionary trees were inferred on Mega (Kumar et al., 2016) using the maximum likelihood method (Hasegawa–Kishino–Yano model) based on concatenated and validated SNPs (panel 1, 8,981 SNPs for 81 genomes; panel 2, 124 SNPs for 93 genomes). The trees were drawn to scale, with branch lengths measured in the number of substitutions per site.

### IS6110 Distribution Analysis

ISMMapper (version 2.0.1) pipeline (Hawkey et al., 2015) was carried out in this study for IS6110 identification on genome short-read sequences using Mb3601 reference genome (Branger et al., 2020). The employed script is available on GitHub.<sup>2</sup> To achieve 100% detection with high confidence, average genome-wide read depths of < 75 × were excluded for panel 2 strains (Hawkey et al., 2015). Only 93 genomes (with Mb3601 reference

<sup>1</sup><https://github.com/najoshi/sickle>

<sup>2</sup>[https://github.com/jhawkey/IS\\_mapper](https://github.com/jhawkey/IS_mapper)

strains) out of the initial 147 passed the quality selection described in this and the previous section.

For uncertain IS6110 presence and/or localization (37 and 9 uncertain insertion sites in panel 1 and panel 2, respectively) due to low read depth in the IS insertion site, IS6110 presence was manually checked with Integrative Genomics Viewer (Robinson et al., 2011) using the bam file generated by ISMapper.

## Orthology Analysis

Genomic positions were deduced through ISMapper analysis. Surrounding genes (other than IS6110) of these insertion sites were identified based on Mb3601 reference to determine orthologous genomic sites of the IS. Insertion sites present in the same coding sequence (CDS) are grouped in a common locus.

## IS6110 Site Gene Ontology Enrichment Analysis

Gene ontology (GO) analysis was performed on the upstream and downstream genes of IS6110 using their protein sequence. Functional annotation was performed using eggno3 5.0 (Huerta-Cepas et al., 2019). Specific colors were applied to the gene according to their Cluster of Orthologous Gene (COG) functional categories.

## RESULTS

### IS6110 Abundance in *Mycobacterium bovis* French Diversity

Among the 81 *M. bovis* strains representing the French genotypic diversity of panel 1, 65% (53/81) have only one IS6110 copy, 25% (20/81) are IS6110 low copy number (2–5 IS6110 copies), and 10% (8/81) present a high copy number with more than six copies (Figure 1 and Supplementary Table 1) (Fomukong et al., 1998; Gonzalo-Asensio et al., 2018). Among the different clusters described by Hauer and collaborators (Hauer et al., 2019), strains of Cluster C (SB0134) and those of the European 1 clonal complex (Smith et al., 2011) are mostly single copies, except for one strain in each cluster with two and three copies (B7 and D6, respectively). F9 family/cluster G strains are all IS6110 single copies. The IS6110 copy number seems to be correlated with the phylogenetic group (cluster or subcluster). All strains in Cluster A/F4 family group are IS6110 multicopy with three common insertions, which are present in most of the strains belonging to this cluster (8/10). Among the 14 strains that belong to the European 2 (Eu2) clonal complex (Rodriguez-Campos et al., 2012), 43% of them (6/14), which are localized on two specific branches of the phylogenetic tree, possess more than one IS6110 copy (3–5 copies). Notably, 31% of Eu3 clonal complex strains (9/29), including the Mb3601, are IS6110 multicopy. This group includes the strain with the highest IS6110 copy number, i.e., 16 copies, strain A5. Interestingly, these high copy number strains (7–16 copies) are all present in a specific monophyletic clade (8/8). Strains that have caused the majority of outbreaks in France in the last 10 years belong to strain groups, which are multicopy, with a low copy number of 2–5 copies for Cluster

A/F4 family and a high copy number of 11–13 copies for SB0120-DHV and SB0120-CO.

In the 81 strains of panel 1, 58 insertion sites have been identified (insertion sites 1 to 58 related to position on the genome Mb3601) and are summarized in Supplementary Table 2. Almost 50% of these insertion sites (28/58) are only observed in a unique strain. Almost all these strains possess an IS6110 insertion in the Direct Repeat (DR) locus (position 43). However, this insertion is absent in two Cluster A/F4 family strains: D5, with an SB0840 spoligotype and D12 with a SB0928 spoligotype. Strain D12 has an IS6110 insertion in position 42, also observed in strain D3 which is phylogenetically close to strain D12. In addition to the number of copies, which seems to be related to certain strain groups, the localization of these insertions also seems to be conserved. For instance, Cluster A/F4 family strains have 3 common IS6110 insertion sites (insertion sites 1, 33, and 43). Eight of these insertion sites are shared between several *M. bovis* groups. For example, insertion site 33 is common between Cluster A/F4 family (10 strains) and SB0120-CO genotype (2 strains), and insertion sites 19 and 50 are both shared by 7 strains of three different clonal groups.

SB0120-DHV genotype strains (strains H2 and C10) and the SB0120-CO genotype (strains C12 and Mb3601) strains have each 10 IS6110 conserved insertion sites, but only two of them (insertion sites 43 and 50) are common to the four strains. Strain A5, phylogenetically closely to SB0120-DHV, has also these ten IS6110 insertion sites but also shares a common insertion site (insertion site 19) with strains belonging to the monophyletic clade of IS6110 high multicopy strains (SB0120-CO, F1, and B9 strains).

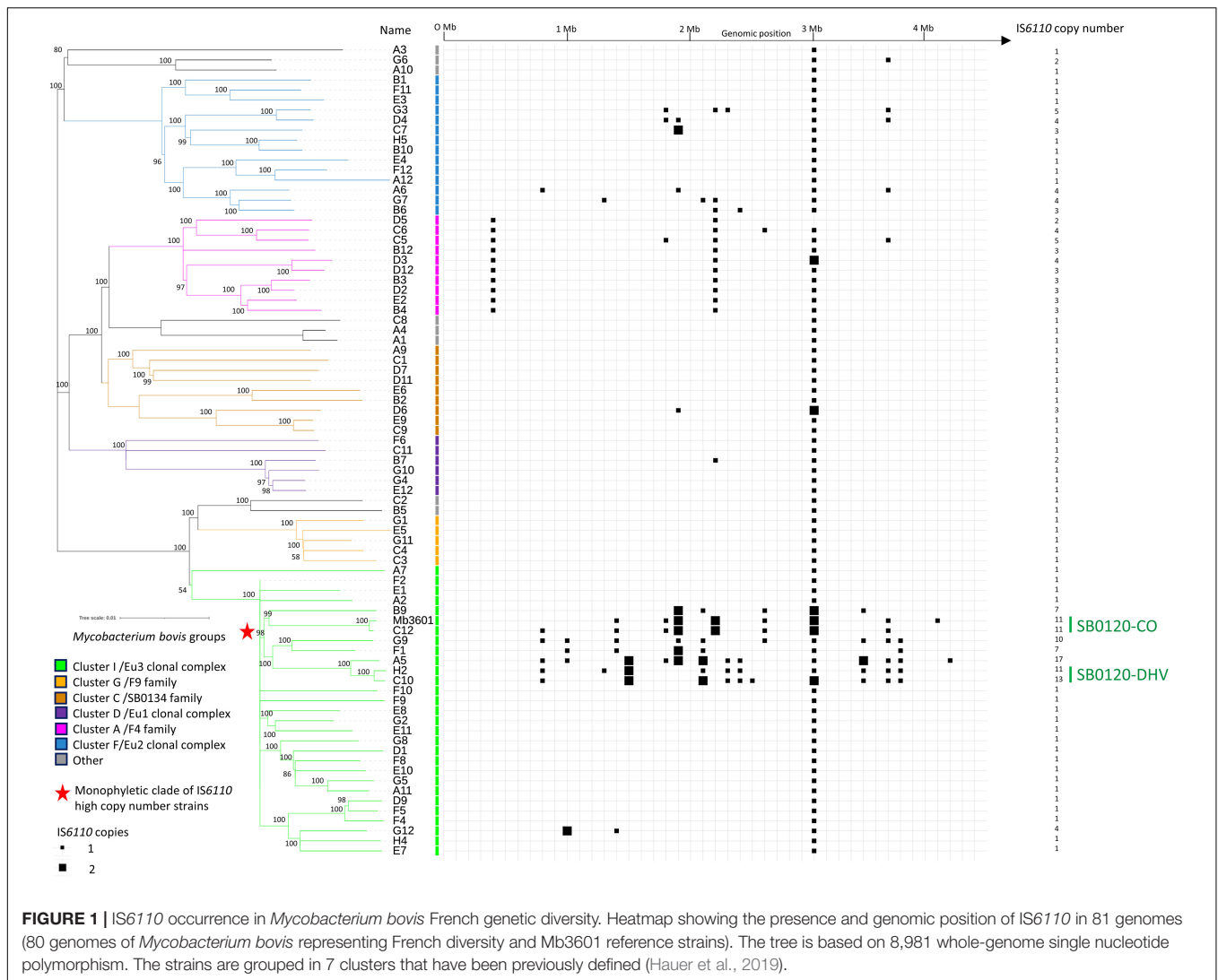
### Longitudinal Study of IS6110 in Sympatric Strains

To further study the IS6110 high copy number phenomenon in *M. bovis* strains, IS6110 sites were searched in a panel of multicopy strains of the SB0120-CO genotype. The studied panel, panel 2, is composed of strains isolated over a period of 6 years and from four different species [cattle, badger (*Meles meles*), wild boar (*Sus scrofa*), and fox (*Vulpes vulpes*)]. SB0120-CO strains possess between 10 and 12 IS6110 copies (Figure 2 and Supplementary Table 3). These insertion sites are stable over time; in fact, ten of them are found in all SB0120-CO strains of panel 1 and panel 2 (Figures 1, 2). Furthermore, IS6110 genomic positions are also conserved independently of the animal species from which the strains were isolated. Eight other IS6110 genomic positions exist but are unique to a specific strain. One of them is present in 3 very close SB0120-CO strains. Besides, SB0120-CO strains with an additional insertion are present on different branches of the tree. These insertions are neither recurrent in our sampling nor conserved over time or correlated to host adaptation.

### Genetic Impact of the IS6110 Insertions

The analysis of IS6110 genomic positions in the *M. bovis* French diversity (panel 1) revealed 58 insertion loci and shows that



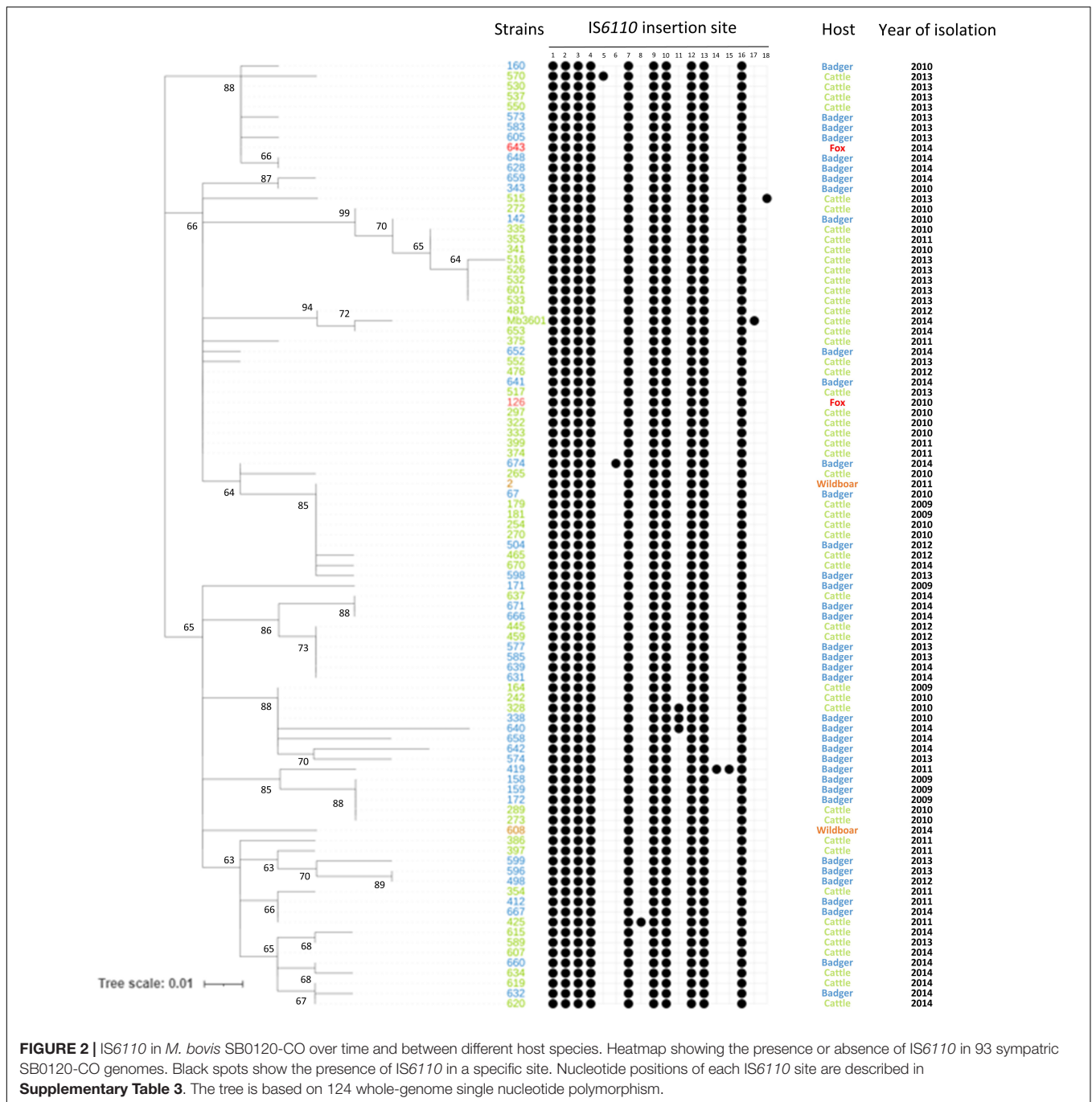


62% of them are intragenic (36/58) (**Supplementary Table 2**). Of the 36 intragenic insertions, 7 are present in Mb3601, and their identities are confirmed with blast alignment on AF2122/97 reference genome (data not shown). However, the proportion of intragenic IS6110 sites is different in several groups according to different IS6110 copy numbers.

A schematic representation of insertion sites (**Supplementary Table 4**) shows that some IS6110 insertion sites can be localized in different positions and orientations in the same insertion region such as in the Clustered Regularly Interspaced Short Palindromic Repeats CRISPR-associated (CRISPR-Cas) region, near the MBS3601\_RS09105 locus or near the MBS3601\_RS05055 locus.

The intragenic insertions are mostly found in Cluster A/F4 family, Cluster F/Eu2 clonal complex, and Cluster I/Eu3 clonal complex strains which can be explained by the presence of several copies of IS6110 in these groups (**Supplementary Table 2**). Analyses of the genes interrupted by IS6110 in the 3 most representative French *M. bovis* types defined previously, i.e.,

SB0120-CO, SB0120-DHV, and Cluster A/F4 family, show genes associated with 10 gene ontology pathways. Most genes interrupted by IS6110 in Cluster A/F4 family and SB0120-CO strains have unknown, and “Replication, recombination and repair” functions but also “Coenzyme transport and metabolism,” “Defense mechanisms,” “Energy production and conversion,” and “Cell wall/membrane/envelope biogenesis” function in SB0120-CO. These functions are not found in SB0120-DHV, where genes involved in “Cell motility” and “Transcription” functions are those presenting the most interruptions. These insertion sites in specific genes can be conserved in the same *M. bovis* cluster such as *rpfD*, *moeY*, and *cas1* for SB0120-CO, MB3601\_RS02050 (hyaluronidase/chondrosulfatase) for Cluster A/F4 family, or MBS3601\_RS11780 (*LysR* family transcriptional regulator) and MBS3601\_RS16315 (PPE family protein) for SB0120-DHV. Some other genes can be interrupted in several *M. bovis* clusters such as *radD* (12 strains from 2 different clusters) or *plcD* (7 strains from 3 different clusters). However, further study on these insertion sites shows the difference in genomic

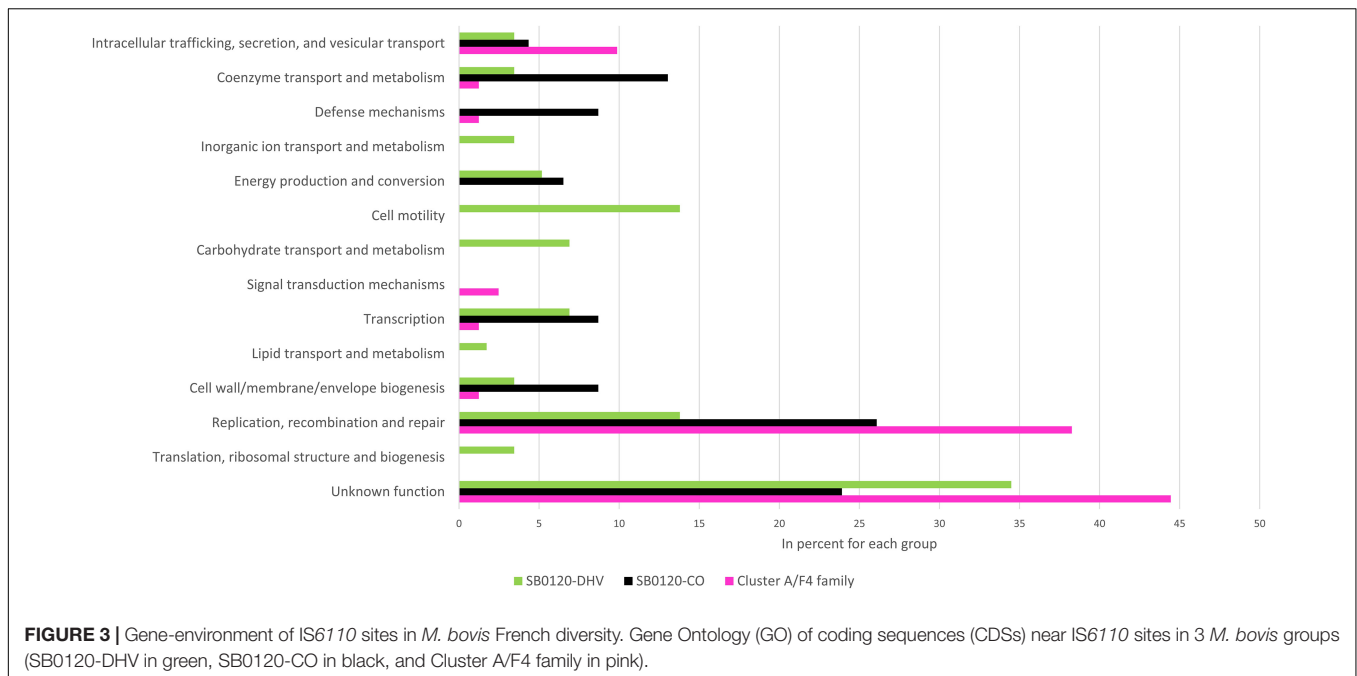


position and/or in IS orientation among *M. bovis* clusters (**Supplementary Tables 2, 4**).

A total of 13 GO pathways present in upstream or downstream of IS6110 insertion loci in SB0120-CO, SB0120-DHV, and Cluster A/F4 family could be highlighted (**Figure 3**). Most of these genes are associated with an unknown function GO category. The “Replication, recombination and repair” GO category is the most frequently represented category with a determined function in the three groups, with 38% (31/81) in Cluster A/F4 family, 26% (12/46) in SB0120-CO, and 14% (8/58) SB0120-DHV.

## DISCUSSION

*Mycobacterium bovis* strains are commonly considered as possessing one or few copies of IS6110 (Gonzalo-Asensio et al., 2018), and only few strains were described as IS6110 multicopy (Soto et al., 2004; Allix et al., 2006; Gonzalo-Asensio et al., 2018; Branger et al., 2020). However, our study focusing on French *M. bovis* diversity shows that more than one-third of our samples are multicopy. Specifically, strains of Cluster A/F4 family, SB0120-CO and SB0120-DHV genotypes, which are



those circulating in the most bTB prevalent regions in the past 10 years (Delavenne et al., 2021), are IS6110 multicopy. This observation leads us to consider whether IS6110 transposition could be partly responsible for their epidemiological success, although other explanations could support this success, notably the drop in strain diversity (Boschiroli et al., 2015; Hauer et al., 2015; Delavenne et al., 2021) and the associated bottleneck following the establishment of control programs. Moreover, other genetic events can also contribute to the success of these strains (indel, SNP). Further analyses are needed to combine SNP and deletion events as well as IS6110 transposition events.

Strains with the highest IS6110 copy numbers belong to the Eu3 clonal complex and in particular to a specific monophyletic clade in the phylogenetic tree to which the two monophyletic types SB0120-CO and SB0120-DHV belong. *In silico* results describing the presence of IS6110 high copy number strains in the same monophyletic clade are sustained by the description of the 11 IS6110 copies in the Mb3601 complete genome (Branger et al., 2020) and that of strain F1 (7 IS6110 copies). F1 is characterized by spoligotype SB0162 (Supplementary Table 1) and found not only in France but also in Belgium and previously described by IS6110 RFLP analysis as presenting 8–11 copies, which is close to our results (Allix et al., 2006).

According to the results on SB0120-CO sympatric strains, IS6110 copy number and insertion site seem to be stable over time among this type of strain, with an average of 10 copies and recurrence in their genomic positions. Furthermore, a specific host adaptation through IS6110-linked changes does not seem to have taken place as suggested for highly epidemic *M. tuberculosis* strains (Beijing lineage) (Gonzalo-Asensio et al., 2018), given that no strain variability is observed, neither on

the copy number nor in IS6110 insertion site, independently of the different animal hosts from which the strains were isolated from. This high stability over time and independently of the affected animal species demonstrates that *M. bovis* strains do not seem to evolve *via* IS6110 transposition for specific host adaptation as suggested by Gonzalo-Asensio et al. (2018). Both this stability and independence of the animal host were also confirmed in IS6110 multicopy SB0120-DHV and Cluster A/F4 family strains of SB0821 and SB0832 spoligotypes, and also in SB0134 single-copy strains infecting livestock and wildlife in several regions in France (data not shown). Furthermore, the stability of the 3 common IS6110 sites in Cluster A/F4 family strains suggests the evolution of these strains from a common ancestor. However, if the genotypes SB0120-CO and SB0120-DHV have each 10 conserved insertion sites, only 2 are common between the two types. As expected, one of them is present in the DR locus (insertion 43). The second one (insertion 50) is present in an IS6110 Preferential Locus (ipl) region which is described to be a hotspot of IS6110 (Fang and Forbes, 1997; Roychowdhury et al., 2015). In fact, this insertion is also present in strains of the other groups such as Cluster A/F4 family and Eu2 clonal complex.

As expected, almost all strains in our panel (79 out of 81) have the IS6110 in the DR locus. This insertion site is common to most MTBC species and is defined as the ancestral insertion of a common ancestor (Hermans et al., 1990; Dale, 1995; Philipp et al., 1996; Thorne et al., 2011; Gonzalo-Asensio et al., 2018). MTBC that lacks the DR IS6110 has previously been described (Refrégier et al., 2020) and may be explained by recombination events between two IS6110 and IS transpositions (Gonzalo-Asensio et al., 2018; Shitikov et al., 2019; Refrégier et al., 2020).

The specific insertion in several genes such as *moeY* in SB0120-CO strains or *radD* in Cluster A/F4 family and SB0120-DHV strains shows that their expression is not essential for bacterial virulence and transmission given that these genotypes are prevalent in France (Delavenne et al., 2021).

IS6110 is often inserted in the environment of genes implied in “replication, recombination and repair,” “transcription,” “cell wall/membrane/envelope biogenesis,” and “coenzyme transport and metabolism pathways” GO categories according to the literature (Reyes et al., 2012). These four categories are also the most frequently identified in our analysis. In fact, 21% (78/369) of genes around IS6110 insertion sites of multicopy strains are involved in the “replication, recombination and repair” GO category. This category with a determined function is the most frequent in our analysis. Genes encoding replication, recombination, and DNA repair functions seem to play an important role in the evolution of highly clonal bacteria such as *M. bovis* (Vultos et al., 2008; Mestre et al., 2011; Reis et al., 2021). Other frequent categories are “Intracellular trafficking, secretion, and vesicular transport” (7%, 26/369), “Cell wall/membrane/envelope biogenesis” (6%, 22/369), and “Transcription” (6%, 22/369) or “Cell motility” (5%, 20/369). IS6110 insertions in or near these genes could play a role in *M. bovis* genome plasticity.

New IS6110 insertion sites in *M. bovis* genomes have been described in this study, especially in SB0120-CO, SB0120-DHV, Cluster A/F4 family, and Cluster F/Eu2 clonal complexes such as insertion site 15 interrupting *cmr* gene or insertion site 24 interrupting *cyp144* gene. Some IS6110 insertion sites have already been described in other MTBCs, suggesting that an *ipl* region could be common to several MTBC species. For example, the ancestral IS6110 is present in almost all MTBC strains, and some other IS6110 can be found in the CRISPR-Cas region (Hermans et al., 1990; Thorne et al., 2011; Refrégier et al., 2020). Moreover, the 3 insertion sites highlighted in our study that are common to several clusters have already been described as *ipl* in previous studies: the *plcD* region (insertion site 19) (Gordon et al., 1999; Sampson et al., 1999; Viana-Niero, 2006; Roychowdhury et al., 2015), insertion site 33 around *radD* (Chen et al., 2015), and insertion site 50 around IS1547 (Fang et al., 1999) (Supplementary Table 2). The result gene regulation of IS6110 insertion in these hot spots can be different depending on the IS6110 orientations and/or insertion sites in the same genomic region and can lead to gene interruption or expression regulation (Sampson et al., 1999; Warren et al., 2000; Vera-Cabrera et al., 2001; Soto et al., 2004; Kim et al., 2010).

As expected, given that our strains are issued from well-established infected animals, IS6110 insertion sites were not identified in genes previously described as essential for virulence (Sasseti and Rubin, 2003; Forrellad et al., 2013). However, some genes shown in our study as being interrupted by IS6110, such as *plcD*, were described as conferring an increased virulence for thoracic tuberculosis and a role in bacterial persistence within macrophages (Raynaud et al., 2002; Kong et al., 2005; Forrellad et al., 2013). Nonetheless, and in line with the fact that the strains in our study were infectious to animals from which they were

isolated, other authors suggested that PLCs could play a less important role in bacterial virulence (Le Chevalier et al., 2015). However, genes such as *plcD* and *rpfD* have homologous genes, and their inactivation could be overcome by other homolog genes as reported by Kana and collaborators for *rpf* genes (Raynaud et al., 2002; Kana et al., 2008; Shafipour et al., 2021).

IS6110 can also lead to gene regulation *via* its strong promoter as observed in the B strain on *phoP*. This insertion upstream of this essential gene for virulence probably led the strain to adapt to the human host and provoked an important multidrug-resistant (MDR) tuberculosis outbreak due to *M. bovis* (Soto et al., 2004). No such IS6110 insertion leading to *phoP* regulation was found in our study. Some studies suggest that IS6110 inserted no more than 400 bp upstream of a gene and in the same transcriptional orientation may upregulate its expression (Safi et al., 2004; Soto et al., 2004; Alonso et al., 2013). Such type of insertions has been detected in our analyses (data not shown). However, as they deserve complementary *in vivo* or *in vitro* studies with the corresponding strains, we cannot assess if they play any role on virulence or other survival traits.

## CONCLUSION

Contrary to what was expected from the literature, this study shows that 35% of the strains representative of the French diversity possess more than one IS6110. In addition, 10% that belong to a specific monophyletic clade of the Eu3 clonal complex present a high IS6110 copy number. These multicopy strains are those circulating in French regions where bTB is most prevalent. IS6110 insertion sites appear to be stable within specific genotypes over time and between host species, suggesting that IS6110 transposition is not an evolutionary driver for modern French *M. bovis* strains at least over a 15-year period. Moreover, the correlation between the epidemiological success of *M. bovis* strains and multicopies of IS6110 leads us to consider whether it could be the consequence of an increase in their fitness due to the genetic changes originated by IS6110 transposition which could have occurred during ancient evolutionary diversification events. Any phenotypical consequence of IS6110 insertions should need confirmation by *in vitro* or *in vivo* experiment.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

LM, MLB, and FB supervised this study. CCh performed bioinformatics work and wrote the first draft. CCo provided bioinformatics support for data analysis. All authors contributed to this study correction, discussed the results and their interpretation, and approved the submitted version.



## FUNDING

CCh is a doctoral fellow financed by the ANSES and the One Health European Joint Programme (OHEJP), Grant number 773830.

## REFERENCES

- Allix, C., Walravens, K., Saegerman, C., Godfroid, J., Supply, P., and Fauville-Dufaux, M. (2006). Evaluation of the Epidemiological Relevance of Variable-Number Tandem-Repeat Genotyping of *Mycobacterium bovis* and Comparison of the Method with IS6110 Restriction Fragment Length Polymorphism Analysis and Spoligotyping. *J. Clin. Microbiol.* 44, 1951–1962. doi: 10.1128/JCM.01775-05
- Alonso, H., Samper, S., Martín, C., and Otal, I. (2013). Mapping IS6110 in high-copy number *Mycobacterium tuberculosis* strains shows specific insertion points in the Beijing genotype. *BMC Genomics* 14:422. doi: 10.1186/1471-2164-14-422
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Benet, J.-J., Boschioli, M.-L., Dufour, B., and Garin-Bastuji, B. (2006). Lutte contre la tuberculose bovine en France de 1954 à 2004: Analyse de la pertinence épidémiologique de l'évolution de la réglementation. *Epidémiol et santé anim.* 1, 127–143.
- Boschioli, M.-L., Michelet, L., Hauer, A., De Cruz, K., Courcou, A., Hénault, S., et al. (2015). Tuberculose bovine en France : cartographie des souches de *Mycobacterium bovis* entre 2000-2013. *Bull. Épidémiol. santé animale alimentation.* 70, 2–8.
- Branger, M., Loux, V., Cochard, T., Boschioli, M. L., Biet, F., and Michelet, L. (2020). The complete genome sequence of *Mycobacterium bovis* Mb3601, a SB0120 spoligotype strain representative of a new clonal group. *Infect. Gene. Evol.* 1:104309. doi: 10.1016/j.meegid.2020.104309
- Chen, S. H., Byrne, R. T., Wood, E. A., and Cox, M. M. (2015). *Escherichia coli* radD (yejH) gene: a novel function involved in radiation resistance and double-strand break repair. *Mol. Microbiol.* 95, 754–768. doi: 10.1111/mmi.12885
- Comas, I., Homolka, S., Niemann, S., and Gagneux, S. (2009). Genotyping of Genetically Monomorphic Bacteria: DNA Sequencing in *Mycobacterium tuberculosis* Highlights the Limitations of Current Methodologies. *PLoS One* 4:e7815. doi: 10.1371/journal.pone.0007815
- Dale, J. W. (1995). Mobile genetic elements in mycobacteria. *Eur. Respir. J. Suppl.* 20, 633s–648s.
- Delavenne, C., Desvaux, S., Boschioli, M.-L., Carles, S., Durand, B., and Forfait, C. (2021). Surveillance de la tuberculose due à *Mycobacterium bovis* en France métropolitaine en 2019: Résultats et indicateurs de fonctionnement. *Bull. Épidémiol. Santé Animale Alimentation* 94, 1–23.
- Driscoll, J. R. (2009). “Spoligotyping for Molecular Epidemiology of the *Mycobacterium tuberculosis* Complex,” in *Molecular Epidemiology of Microorganisms: Methods and Protocols*, ed. D. A. Caugant (Totowa, NJ: Humana Press), 117–128. doi: 10.1007/978-1-60327-999-4\_10
- Fang, Z., Doig, C., Morrison, N., Watt, B., and Forbes, K. J. (1999). Characterization of IS1547, a New Member of the IS900 Family in the *Mycobacterium tuberculosis* Complex, and Its Association with IS6110. *J. Bacteriol.* 181, 1021–1024. doi: 10.1128/JB.181.3.1021-1024.1999
- Fang, Z., and Forbes, K. J. (1997). A *Mycobacterium tuberculosis* IS6110 preferential locus (ipl) for insertion into the genome. *J. Clin. Microbiol.* 35, 479–81. doi: 10.1128/jcm.35.2.479-481.1997
- Fomukong, N., Beggs, M., El Hajj, H., Templeton, G., Eisenach, K., and Cave, M. D. (1998). Differences in the prevalence of IS6110 insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS6110. *Tubercle Lung Dis.* 78, 109–116. doi: 10.1016/s0962-8479(98)80003-8
- Forrellad, M. A., Klepp, L. I., Gioffré, A., Garcia, J. S. Y., Morbidoni, H. R., Santangelo, M., et al. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* 4, 3–66.
- Gonzalo-Asensio, J., Pérez, I., Aguiló, N., Uranga, S., Picó, A., Lampreave, C., et al. (2018). New insights into the transposition mechanisms of IS6110 and its dynamic distribution between *Mycobacterium tuberculosis* Complex lineages. *PLoS Genet.* 14:e1007282. doi: 10.1371/journal.pgen.1007282
- Gordon, S. V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K., and Cole, S. T. (1999). Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol. Microbiol.* 32, 643–655. doi: 10.1046/j.1365-2958.1999.01383.x
- Hauer, A., De Cruz, K., Cochard, T., Godreuil, S., Karoui, C., Henault, S., et al. (2015). Genetic Evolution of *Mycobacterium bovis* Causing Tuberculosis in Livestock and Wildlife in France since 1978. *PLoS One* 10:e0117103. doi: 10.1371/journal.pone.0117103
- Hauer, A., Michelet, L., Cochard, T., Branger, M., Nunez, J., Boschioli, M.-L., et al. (2019). Accurate Phylogenetic Relationships Among *Mycobacterium bovis* Strains Circulating in France Based on Whole Genome Sequencing and Single Nucleotide Polymorphism Analysis. *Front Microbiol.* 10:955. doi: 10.3389/fmicb.2019.00955
- Hawkey, J., Hamidian, M., Wick, R. R., Edwards, D. J., Billman-Jacobe, H., Hall, R. M., et al. (2015). ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC Genomics* 16:667. doi: 10.1186/s12864-015-1860-2
- Hermans, P. W., van Soolingen, D., Dale, J. W., Schuitema, A. R., McAdam, R. A., Catty, D., et al. (1990). Insertion element IS986 from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. *J. Clin. Microbiol.* 28, 2051–2058. doi: 10.1128/jcm.28.9.2051-2058.1990
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085
- Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., et al. (1997). Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* 35, 907–914. doi: 10.1128/jcm.35.4.907-914.1997
- Kana, B. D., Gordhan, B. G., Downing, K. J., Sung, N., Vostroktunova, G., Machowski, E. E., et al. (2008). The resuscitation-promoting factors of *Mycobacterium tuberculosis* are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro. *Mol. Microbiol.* 67, 672–684. doi: 10.1111/j.1365-2958.2007.06078.x
- Kim, E. Y., Nahid, P., Hopewell, P. C., and Kato-Maeda, M. (2010). Novel Hot Spot of IS6110 Insertion in *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 48, 1422–1424. doi: 10.1128/JCM.02210-09
- Kong, Y., Cave, M. D., Yang, D., Zhang, L., Marrs, C. F., Foxman, B., et al. (2005). 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.* 43, 6048–6053. doi: 10.1128/JCM.43.12.6048-6053.2005
- Kremer, K., Glynn, J. R., Lillebaek, T., Niemann, S., Kurepina, N. E., Kreiswirth, B. N., et al. (2004). Definition of the Beijing/W Lineage of *Mycobacterium tuberculosis* on the Basis of Genetic Markers. *J. Clin. Microbiol.* 42, 4040–4049. doi: 10.1128/JCM.42.9.4040-4049.2004
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Le Chevalier, F., Cascioferro, A., Frigui, W., Pawlik, A., Boritsch, E. C., Bottai, D., et al. (2015). Revisiting the role of phospholipases C in virulence and the lifecycle of *Mycobacterium tuberculosis*. *Sci. Rep.* 5:16918. doi: 10.1038/srep16918
- McEvoy, C. R. E., Falmer, A. A., van Pittius, N. C. G., Victor, T. C., van Helden, P. D., and Warren, R. M. (2007). The role of IS6110 in the evolution of

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.891902/full#supplementary-material>

- Mycobacterium tuberculosis*. *Tuberculosis* 87, 393–404. doi: 10.1016/j.tube.2007.05.010
- Merker, M., Kohl, T. A., Roetzer, A., Truebe, L., Richter, E., Rüscher-Gerdes, S., et al. (2013). Whole Genome Sequencing Reveals Complex Evolution Patterns of Multidrug-Resistant *Mycobacterium tuberculosis* Beijing Strains in Patients. *PLoS One* 8:e82551. doi: 10.1371/journal.pone.0082551.
- Mestre, O., Luo, T., Vultos, T. D., Kremer, K., Murray, A., Namouchi, A., et al. (2011). Phylogeny of *Mycobacterium tuberculosis* Beijing Strains Constructed from Polymorphisms in Genes Involved in DNA Replication, Recombination and Repair. *PLoS One* 6:e16020. doi: 10.1371/journal.pone.0016020
- Michelet, L., Courcou, A., Durand, B., and Boschirol, M. L. (2018). Reconstitution of *Mycobacterium bovis* transmission chain between cattle and wild animals. *Épidémiologie .Santé Animale* 2018, 81–90.
- Philipp, W. J., Poulet, S., Eiglmeier, K., Pascopella, L., Balasubramanian, V., Heym, B., et al. (1996). An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *PNAS* 93, 3132–3137. doi: 10.1073/pnas.93.7.3132
- Raynaud, C., Guillhot, C., Rauzier, J., Bordat, Y., Pelicic, V., Manganeli, R., et al. (2002). Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*. *Mol. Microbiol.* 45, 203–217. doi: 10.1046/j.1365-2958.2002.03009.x
- Refrégier, G., Sola, C., and Guyeux, C. (2020). Unexpected diversity of CRISPR unveils some evolutionary patterns of repeated sequences in *Mycobacterium tuberculosis*. *BMC Genom.* 21:841. doi: 10.1186/s12864-020-07178-6
- Reis, A., Salvador, L., Robbe-Austerman, S., Tenreiro, R., Botelho, A., Albuquerque, T., et al. (2021). Whole Genome Sequencing Refines Knowledge on the Population Structure of *Mycobacterium bovis* from a Multi-Host Tuberculosis System. *Microorganisms* 26:9. doi: 10.3390/microorganisms9081585
- Réveillaud, É., Desvaux, S., Boschirol, M.-L., Hars, J., Faure, É., Fediaevsky, A., et al. (2018). Infection of Wildlife by *Mycobacterium bovis* in France Assessment Through a National Surveillance System, Sylvatub. *Front Vet Sci.* 5:262. doi: 10.3389/fvets.2018.00262
- Reyes, A., Sandoval, A., Cubillos-Ruiz, A., Varley, K. E., Hernández-Neuta, I., Samper, S., et al. (2012). IS-seq: a novel high throughput survey of in vivo IS6110 transposition in multiple *Mycobacterium tuberculosis* genomes. *BMC Genomics.* 13:249. doi: 10.1186/1471-2164-13-249
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
- Rodríguez-Campos, S., Schürch, A. C., Dale, J., Lohan, A. J., Cunha, M. V., Botelho, A., et al. (2012). European 2 – A clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. *Infect. Genet. Evol.* 12, 866–872. doi: 10.1016/j.meegid.2011.09.004
- Roychowdhury, T., Mandal, S., and Bhattacharya, A. (2015). Analysis of IS6110 insertion sites provide a glimpse into genome evolution of *Mycobacterium tuberculosis*. *Sci. Rep.* 5:12567 doi: 10.1038/srep12567
- Safi, H., Barnes, P. F., Lakey, D. L., Shams, H., Samten, B., Vankayalapati, R., et al. (2004). IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. *Mol. Microbiol.* 52, 999–1012. doi: 10.1111/j.1365-2958.2004.04037.x
- Samper, S., Martín, C., Pinedo, A., Rivero, A., Blázquez, J., Baquero, F., et al. (1997). Transmission between HIV-infected patients of multidrug-resistant tuberculosis caused by *Mycobacterium bovis*. *AIDS* 11, 1237–42. doi: 10.1097/00002030-199710000-00006
- Sampson, S. L., Warren, R. M., Richardson, M., van der Spuy, G. D., and van Helden, P. D. (1999). Disruption of coding regions by IS6110 insertion in *Mycobacterium tuberculosis*. *Tuber Lung. Dis.* 79, 349–359. doi: 10.1054/tuld.1999.0218
- Sassetti, C. M., and Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *PNAS* 100, 12989–12994.
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Shafipour, M., Shirzad-Aski, H., Kochaksaraei, M. B., Sohrabi, A., Taziki, M., Mahghani, G. A., et al. (2021). The Prevalence of plcD Gene and Evaluation of IS6110 Insertion Status in This Gene in Some Clinical *Mycobacterium tuberculosis* Isolates. *Mol. Gene. Microbiol. Virol.* 36.111–118
- Shitikov, E., Guliaev, A., Bespyatykh, J., Malakhova, M., Kolchenko, S., Smirnov, G., et al. (2019). The role of IS6110 in micro- and macroevolution of *Mycobacterium tuberculosis* lineage 2. *Mol. Phylogenetics Evol.* 139:106559. doi: 10.1016/j.ympev.2019.106559
- Smith, N. H., Berg, S., Dale, J., Allen, A., Rodriguez, S., Romero, B., et al. (2011). European 1: A globally important clonal complex of *Mycobacterium bovis*. *Infect. Genet. Evol.* 11, 1340–1351. doi: 10.1016/j.meegid.2011.04.027
- Soto, C. Y., Menendez, M. C., Perez, E., Samper, S., Gomez, A. B., Garcia, M. J., et al. (2004). IS6110 Mediates Increased Transcription of the phoP Virulence Gene in a Multidrug-Resistant Clinical Isolate Responsible for Tuberculosis Outbreaks. *J. Clin. Microbiol.* 42, 212–219. doi: 10.1128/JCM.42.1.212-219.2004
- Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rusch-Gerdes, S., Willery, E., et al. (2006). Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 44, 4498–4510.
- Thorne, N., Borrell, S., Evans, J., and Magee, J. (2011). Garcia de Viedma D, Bishop C, et al. IS6110-based global phylogeny of *Mycobacterium tuberculosis*. *Infect. Genetics Evol.* 11, 132–138. doi: 10.1016/j.meegid.2010.09.011
- van Soolingen, D., de Haas, P. E., Haagsma, J., Eger, T., Hermans, P. W., Ritacco, V., et al. (1994). Use of various genetic markers in differentiation of *Mycobacterium bovis* strains from animals and humans and for studying epidemiology of bovine tuberculosis. *J. Clin. Microbiol.* 32, 2425–2433. doi: 10.1128/jcm.32.10.2425-2433.1994
- Vera-Cabrera, L., Hernández-Vera, M. A., Welsh, O., Johnson, W. M., and Castro-Garza, J. (2001). Phospholipase Region of *Mycobacterium tuberculosis* Is a Preferential Locus for IS6110 Transposition. *J. Clin. Microbiol.* 39, 3499–3504. doi: 10.1128/JCM.39.10.3499-3504.2001
- Viana-Niero, C. (2006). Identification of an IS6110 insertion site in plcD, the unique phospholipase C gene of *Mycobacterium bovis*. *J. Med. Microbiol.* 55, 451–457. doi: 10.1099/jmm.0.46364-0
- Vultos, T. D., Mestre, O., Rauzier, J., Golec, M., Rastogi, N., Rasolof, V., et al. (2008). Evolution and Diversity of Clonal Bacteria: The Paradigm of *Mycobacterium tuberculosis*. *PLoS One* 3:e1538. doi: 10.1371/journal.pone.0001538
- Warren, R. M., Sampson, S. L., Richardson, M., Spuy, G. D. V. D., Lombard, C. J., Victor, T. C., et al. (2000). Mapping of IS6110 flanking regions in clinical isolates of *Mycobacterium tuberculosis* demonstrates genome plasticity. *Mol. Microbiol.* 37, 1405–1416. doi: 10.1046/j.1365-2958.2000.02090.x
- Wingett, S. W., and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* 7:1338 doi: 10.12688/f1000research.15931.2
- Xia, E., Teo, Y.-Y., and Ong, R. T.-H. (2016). SpoTyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. *Genome Med.* 8:19. doi: 10.1186/s13073-016-0270-7

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

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Charles, Conde, Biet, Boschirol and Michelet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.