



OPEN

## Analysis of *Mycobacterium africanum* in the last 17 years in Aragon identifies a specific location of IS6110 in Lineage 6

Jessica Comín<sup>1,2</sup>, María Luisa Monforte<sup>3</sup>, Sofía Samper<sup>1,2,4,6</sup>✉, Aragonese Working Group on Molecular Epidemiology of Tuberculosis (EPIMOLA)\* & Isabel Ota<sup>2,4,5</sup>

The purpose of this study was to increase our knowledge about *Mycobacterium africanum* and report the incidence and characteristics of tuberculosis (TB) due to their lineages in Aragon, Spain, over the period 2003–2019. The study includes all the cases in our region, where all the *M. tuberculosis* complex isolates are systematically characterised. We detected 31 cases of *M. africanum* among 2598 cases of TB in the period studied. TB caused by *M. africanum* is rare (1.19%) in our population, and it affects mainly men of economically productive age coming from West African countries. Among the isolates, Lineage (L) 6 was more frequent than L5. The genotyping of these strains identified five clusters and 13 strains with a unique pattern. The isolates' characterisation identified a copy of IS6110 within the *moaX* gene, which turned out to be specific for L6. It will allow the differentiation of this lineage from the rest of MTBC with a simple PCR reaction. It remains to be established whether this polymorphism may limit *M. africanum* transmission. Furthermore, a mutation in the *mutT2* promoter was found as specific for L6 strains, which could be related to the high variability found for L6 compared to L5.

Tuberculosis (TB) is still an important cause of death, especially in developing countries<sup>1</sup>. All members of the *Mycobacterium tuberculosis* complex (MTBC) can cause TB, with *M. tuberculosis* being the most important. However, the strains from *M. africanum* lineages (L) are responsible for almost half of the TB cases in West Africa<sup>2</sup>. *M. africanum* was first described in 1968 from TB patients in Senegal<sup>3</sup>, showing intermediate characteristics between *M. tuberculosis* and *M. bovis*<sup>4,5</sup>. There are three hypotheses being considered of why *M. africanum* is almost restricted to West Africa: it is not able to compete with modern *M. tuberculosis* lineages<sup>6</sup>; it is adapted to the African population<sup>7,8</sup>; and there could be an animal reservoir, being a zoonotic disease<sup>5,9–12</sup>.

Regarding phylogeny, both *M. africanum* L5 and L6 have a common ancestor with the region of difference (RD) 9 deleted<sup>13,14</sup>. Besides, L5 split from the common phylogenetic branch before L6, and the latter also has deleted RD7, RD8, and RD10 regions. Recent studies subdivided L5 into two sub-lineages, L5.1 and L5.2, discriminated by RD711<sup>15</sup>.

The prevalence of *M. africanum* is rare in industrialised countries, such as in Spain; a retrospective study over 10 years (2000–2010) was published<sup>16</sup>. They analysed 36 cases due to *M. africanum* and concluded that most of them were immigrants from Africa, and only four cases were Spaniards. However, not all cases were searched exhaustively, considering that they did not systematically genotype the isolates.

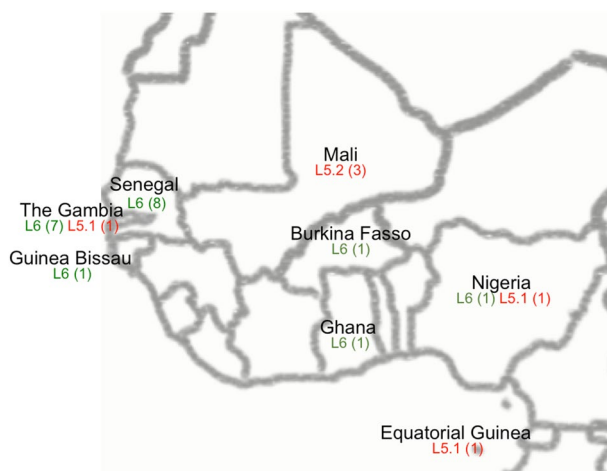
The insertion sequence (IS) 6110, specific to MTBC, has been used to genotype these strains since 1993<sup>17</sup>. Some studies analysed the location of the IS6110 copies trying to clarify its role in the bacteria's physiology<sup>19,20</sup> or revealing that some of the IS6110 locations are characteristic of some specific families<sup>21,22</sup>. However, to our knowledge, the location of IS6110 in the genome of *M. africanum* strains has not been studied until now.

In this work, we used an epidemiological and molecular perspective to investigate the presence of *M. africanum* for the last 17 years in Aragon, Spain. The molecular analysis of these strains allowed us to identify specific polymorphisms described for the first time in these lineages.

<sup>1</sup>Instituto Aragonés de Ciencias de la Salud, Zaragoza, Spain. <sup>2</sup>Fundación IIS Aragón, Zaragoza, Spain. <sup>3</sup>Hospital Royo Villanova, Zaragoza, Spain. <sup>4</sup>CIBER de Enfermedades Respiratorias, Zaragoza, Spain. <sup>5</sup>Universidad de Zaragoza, Zaragoza, Spain. <sup>6</sup>Laboratorio de Investigación Molecular-UIT, Hospital Universitario Miguel Servet, PO Isabel la Católica 1-3, planta calle, 50009 Zaragoza, Aragón, Spain. \*A list of authors and their affiliations appears at the end of the paper. ✉email: ssamper.iaacs@aragon.es

	N = 31	%
<b>Sex</b>		
Male	24	77.41
Female	7	22.59
<b>Age group</b>		
15–24	6	19.35
25–34	14	45.16
35–44	8	28.80
45–64	2	6.45
Unknown	1	3.22
<b>Origin</b>		
Immigrant	28 <sup>a</sup>	90.32
Autochthonous	3	9.67
<b>Location of the disease</b>		
Pulmonar only	16	51.61
Extrapulmonary	15	48.38
<b>Municipality</b>		
Urban	20	64.51
Rural	11	35.48

**Table 1.** Patients' characteristics whose isolates were identified as *M. africanum* in Aragon from 2003 to 2019. <sup>a</sup>27 immigrants came from West African countries and one came from Bulgaria.

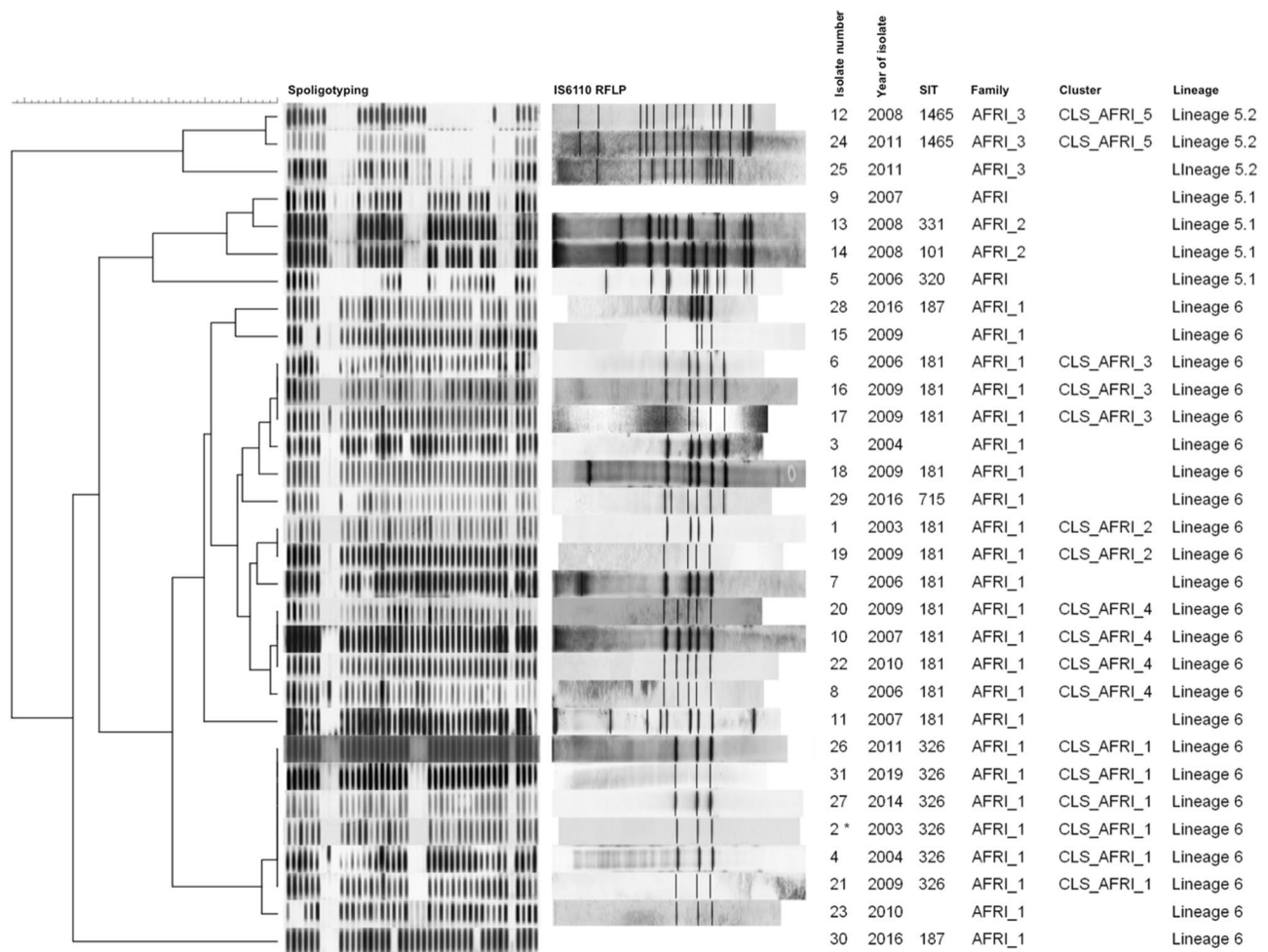


**Figure 1.** Map drawn with Adobe Photoshop CS6 ([www.adobe.com](http://www.adobe.com)) of the West African countries of origin for 26 of the cases. African countries where the patients came from are marked. The red colour indicates an L5 case of TB, and the green colour an L6 case. The number of cases is presented in brackets.

## Results

A retrospective descriptive study of the TB cases caused by isolates identified as *M. africanum* in the Autonomous Community of Aragon was carried out. All TB cases with positive culture in Aragon between 2003 and 2019 were genotyped. Also, spoligotyping supplied the ability to distinguish among the different members of the complex, such as *M. tuberculosis*, *M. bovis*, and *M. africanum*. Out of 2598 MTBC cases of TB over this 17-year period, 31 cases (1.19%) have been caused by an African spoligo-family strain.

**Social and clinical characterisation of the TB cases.** The characteristics of the cases due to *M. africanum* are detailed in Table 1. Of the 31 patients studied, 77.41% were male, the age range was between 18 and 62 years, with the largest in the 25–34-year age-group (45.16%). No cases were detected in the youngest and the eldest age groups. Regarding their origin, 27 patients were born in West African countries (87.09%), three were Spaniards (9.67%), and one patient was born in Bulgaria. The African countries of origin are detailed in Fig. 1. According to the geographical location, the number of patients who lived in an urban area was superior to those who lived in a rural area (20, 64.51% vs 11, 35.48%). At least 12 of the cases lived in our country for less than

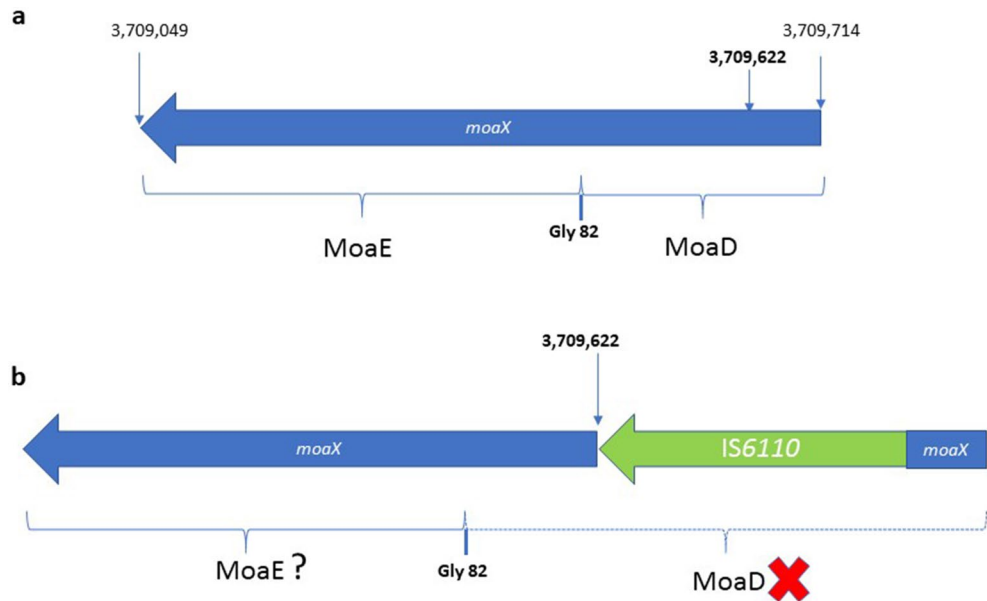


**Figure 2.** Dendrogram based on spoligotype patterns. The data shown are spoligotypes, IS6110-RFLP types, isolate number, year of isolation, SIT, family under SITVIT definition, and lineage of each *M. africanum* isolate in our population from 2003 to 2019. For two isolates, RFLP-type was not available, both showing a unique spoligotype belonged to AFRI and AFRI1 families. \*Isolate 2 in CLS\_AFRI\_1 has a different location of one copy of IS6110 despite sharing the RFLP pattern with the other five isolates included in its cluster.

5 years when TB was diagnosed. Attending to the location of the disease, the samples studied were: 16 sputum, four bone biopsies, six abscesses from different locations, two lymphadenopathies, and three other specimens. Near 50% of the cases (15) presented extrapulmonary disease (Table 1). For all cases, TB bacillus were susceptible to the treatment.

**Genotypic characterisation of *M. africanum* isolates.** The molecular analysis based on IS6110-RFLP and Spoligotyping of the *M. africanum* isolates showed five different clusters, including from two to six cases, and 13 isolates with a unique pattern (Fig. 2). Spoligotyping showed 13 different patterns, three were detected more than once (SIT 181 in 13 isolates, SIT 326 in six isolates, and SIT 1465 in two isolates) and distributed in the AFRI\_1, AFRI\_2, AFRI\_3 or AFRI families according to the SITVIT definition<sup>23</sup>. To confirm the *M. africanum* lineages, the specific differential regions TbD1, RD9, RD702, and RD711 were analysed. TbD1 was present, and RD9 absent in all the isolates. Based on the study of RD702 and RD711, we could classify them into the two existent lineages of *M. africanum*. Twenty-four isolates belonged to L6 and seven to L5. There was a total concordance in the classifications obtained by spoligotyping and the RD analysis. The isolates classified as AFRI1 had the RD702 region deleted, and therefore corresponded to L6 isolates<sup>2</sup>. AFRI, AFRI\_2, and AFRI\_3 strains had RD702 present and, therefore, were considered L5. Two AFRI and two AFRI\_2 spoligotype isolates had deleted RD711 and were sub-classified as L5.1. Meanwhile, the three AFRI\_3 isolates had RD711 present, classifying them as L5.2<sup>15</sup>. The IS6110-RFLP showed a low number of IS6110 copies ( $\leq 6$ ) in the L6 isolates, except one which showed eight copies. However, the six L5 isolates with available RFLP-pattern presented more than 10 IS6110 copies (Fig. 2).

Following the strain characterisation and in the context of a new assay performed in our laboratory to rapidly identify resistances and lineages, 32 isolates of our DNA collection, including different MTBC lineages, were analysed using AmpliSeq-based methodology. Two of the 32 isolates belonged to L6 (isolates 15 and 27) and three

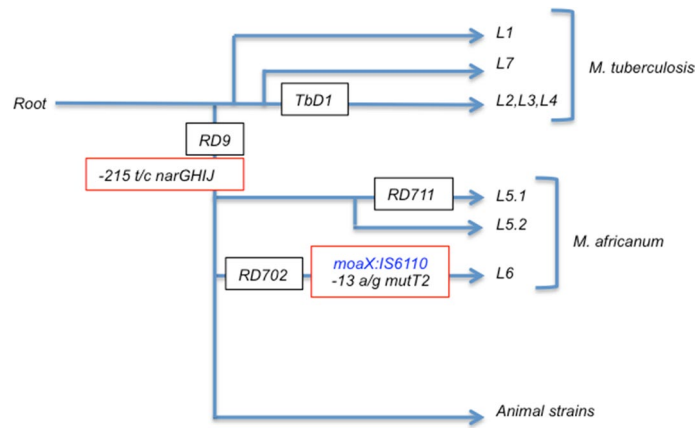


**Figure 3.** Schematic representation of *moaX* (*Rv3323c*) gene, coding for MoaX protein. The cleavage by Gly82 residue is required for the functionality of this MPT synthase. (a) Schema of H37Rv. (b) Schema of *M. africanum* L6 with IS6110 inserted in *moaX* gene. The effect of IS6110 is unknown for MoaE, but MoaD is going to be unfunctional. Numbers are referred to the position of the nucleotide in the genome of *M. tuberculosis* H37Rv.

to L5.1 (isolates 5, 13, and 14). The sequence of the amplicons obtained showed five specific SNPs for L6 isolates located in *rpoB* (1163c/t<sup>24</sup> and 1917a/c mutations, non-synonymous SNPs), *inhA* (233t/c, non-synonymous SNP), *katG* (609c/t, synonymous SNP), and *Rv0309* (474g/a, synonymous SNP) genes. These SNPs were reviewed and confirmed in NCBI L6 complete genomes. In the three L5 isolates studied, a SNP was present in *gyrA* (2265c/t, synonymous SNP). Finally, one specific SNP was detected in *leuB* (550t/c, non-synonymous SNP) in the isolates of both lineages L5.1 and L6 and absent in the rest of the isolates of the MTBC studied by AmpliSeq.

**Specific IS6110 location in L6 strains.** We studied the location of IS6110 in three L6 strains (isolates 2, 15, and 21) using ligation-mediated PCR (LM-PCR) within a study of MTBC strains with a low copy number of this IS. In addition to the copy located in the DR area, three locations were detected in *Rv0963c*, *lipX:mshB* and *moaX* genes. In the three strains, one of the IS6110 copies was located in the *moaX* gene and at identical point for all three cases (Fig. 3). Based on the results obtained, the primers *moaXr* (ccagtcgacgcggttgggg) and *moaXd* (atcgggtcattaccggcggc) were designed to verify the point of insertion of IS6110. The expected PCR products were 2128 bp if IS6110 was present and 788 bp if IS6110 was absent from the site of amplification. We sequenced the amplified fragment noting that IS6110 was inserted at nucleotide 3709622, referred to H37Rv reference genome, flanked by three bp direct repeats (gac), as a consequence of the transposition, and located 90 nucleotides from the beginning of the *moaX* gene (*Rv3323c*) and in its same direction (Fig. 3). Further analysis showed that this IS6110 copy was present in all our collection strains of *M. africanum* belonging to the L6 but never in our L5 strains. In addition, we observed that this location was absent in 42 isolates of low copy number studied by our group. We have also analysed this insertion point in the strains belonging to *M. africanum* L6, whose genomes are available in the NCBI (CP010334.1 and FR878060.1), verifying the presence of IS6110 in the *moaX* gene (Fig. 4). On the other hand, to investigate the intergenic IS6110 insertion in *lipX:mshB* as a possible specific location for *M. africanum*, we amplified the region with the primers *LipX-F* (gccgttcccaatcgaatc) and *LipX-R* (gctcaggctctcatcgtcg). The expected fragment was 264 bp if the IS was absent and 1591 bp if it was present. The PCR results revealed the insertion in five out of nine L6 isolates tested and never in L5 isolates, which means it was not specific but frequent in L6 strains. IS6110 was inserted at nucleotide 1300194, flanked by two bp direct repeats (tt), in all the isolates at the same point, including those in the NCBI database. However, the location of IS6110 in *Rv0963c* was not detected in any other of the *M. africanum* isolates analysed.

**Nitrate reduction activity absent in *M. africanum* strains.** Due to the location of IS6110 interrupting the *moaX* gene, which codes for the enzymes involved in the synthesis of molybdenum cofactor (MoCo), necessary for the activity of the nitrate reductase (NR) enzyme<sup>25</sup>, we wanted to investigate whether this fact would be reflected in a difference in NR activity between the L5 and L6 strains. We analysed the reduction of nitrates of the *M. africanum* isolate 5 (L5) and isolate 11 (L6), and H37Rv and BCG as positive and negative controls, respectively. However, both L5 and L6 strains showed a negative result of NR activity. Both positive and negative controls were in line with expectations. Then, we analysed the sequences of *narG* and *narI* genes, which did not present any mutation to explain the L5 strain result. Nevertheless, the analysis of the *narGHIIJ* promoter



**Figure 4.** Partial evolutionary scenario of MTBC detailing *M. africanum* lineages. IS6110 is involved in the natural evolution of MTBC, and the seemingly random transposition may have contributed to the differentiation of MTBC. Mutations and specific insertion of IS6110 in the *moaX* gene found in this work are red-framed.

in four L5 isolates and two L6 isolates showed an identical mutation in  $-215$  (t/c), which was also present in the NCBI complete genomes of *M. africanum* (L6) and *M. bovis*. Additionally, the study of the sequenced promoter region of the *narGHIJ* operon showed a mutation in the  $-13$  (a/g) *mutT2* gene, upstream *narG* gene, in the two L6 isolates, which was absent in the four L5 isolates tested. This was also observed in the *M. africanum* L6 genomes included in the NCBI database.

## Discussion

This work was carried out to understand the epidemiological situation in Aragon, Spain, related to TB cases caused by *M. africanum*. This study used the data set on TB cases linked to the genotypes of the clinical isolates. The findings from this study indicate that *M. africanum* is a rare cause of TB in our region and represents 1.19% of the cases with available genotype data reported during the 2003–2019 period. A previous study regarding this causal agent<sup>16</sup> reviewed the percentage detected in other countries, as Brazil, Australia and Portugal, where it represented less than 1% of the TB isolates. In their work, the authors collected information on 36 TB cases of *M. africanum* over a 10-year period in Spain. Nineteen of these isolates were from our region and therefore included in this study. The fact that they did not systematically identify the *M. africanum* isolates leads us to believe that there was an underestimation of the TB cases caused by L5 and L6 in our country. Nevertheless, our study was exhaustively conducted since 2003, identifying all *M. africanum* cases. We consider that the incidence of *M. africanum* in our country should be low given the results observed in this study, even though in higher African migration areas it could be slightly different. We hypothesised that the African lineages that are rare in our population are not adapted to transmit.

The descriptive analysis of the TB cases caused by *M. africanum* showed that most were male (77.41%) and in the 25–34 age group (45.16%). All of them were in the labour force, which could be related to being the most abundant age group among immigrants. It was more likely to occur in foreign-born people coming from West African countries (87%), being that only three cases (9.67%) were of Spanish-born people. In Esteban's study<sup>16</sup> performed in Spain, few Spaniards (7%), in contrast to immigrants, presented TB caused by *M. africanum*. The slight difference obtained by us may be because we did an exhaustive genotyping of all the cases. In our study, the patients came from different rural areas in a higher percentage (35%) than the common TB caused by *M. tuberculosis*, which occurred around 80% in urban areas<sup>26</sup>. These associations suggest that the epidemiology of *M. africanum* in our region is driven primarily by the migration of people from West Africa. The TB in Spaniards suggests that transmission of *M. africanum* might occur in Spain, but the possibility of TB acquisition during a trip (e.g., to West Africa) cannot be excluded, as one of the Spaniards presented a unique genotype strain. It would be of interest to continue the study in the coming years to check if the *M. africanum* strains of this work are maintained or are displaced by other MTBC strains. In previous reports, a lower transmission of *M. africanum* in comparison to L4 was observed. Nevertheless, the proportion of L5 and L6 is maintained over time, suggesting that other factors may be responsible for its continued presence in Africa<sup>27</sup>.

The presentation of the disease was in half of the cases restricted to pulmonary location. The extrapulmonary type of the disease (48.38%) was identified in a higher percentage than for the TB case notifications in Spain in 2017 (27.5%)<sup>28</sup>. Some studies showed a high proportion of extrapulmonary TB caused by L5 strains, suggesting that these strains might show a different ability to cause pulmonary disease than *M. tuberculosis* sensu stricto strains<sup>15</sup>. Curiously, while the extra respiratory location of the TB was high among the African cases, the three Spaniards presented only respiratory disease. The differential HLA distribution among the Mali population has been studied, and it was concluded that it might be at least partially responsible for the geographical restriction of *M. africanum* infections to West Africa<sup>29</sup>. The possibility that HLA could also affect the clinical presentation of the

disease would explain these differences. Other studies support the hypothesis that *M. africanum* has a low degree of virulence that may be related to dissemination, rather than lung damage, during the early stages of infection<sup>30</sup>.

We could detect a higher percentage of L6 strains (77.41%) than L5 (22.58%), including both L5.1 and L5.2 sub-lineage cases. Other studies in Mali and Gambia also showed a higher identification of L6 among their cases<sup>2,31</sup>. A systematic review of current knowledge about MTBC strain diversity and geographical distribution in African regions showed a different prevalence of *M. africanum* in West African countries. It represented 8% in Nigeria, 19.75% in Ghana, 20% of the isolates in Burkina Fasso, 3.3% in Guinea, 47.10% in Guinea Bissau, and 38.4% in the Gambia<sup>32</sup>. Previous results published about Gambia reflected that half of their isolates were *M. africanum*, and nearly all of them belonged to L6 with SIT 181 as the most prevalent pattern. According to our results, seven out of eight *M. africanum* isolates of patients coming from Gambia were L6 isolates, and five of them had SIT 181<sup>2</sup>. Traore et al.<sup>31</sup> determined 27.8% of the cases as *M. africanum* in Mali, and almost all of them (94.2% of the strains) were MAF2 (L6). However, the three cases detected from Mali in our study were L5.2.

For all cases, TB bacillus were susceptible to the treatment, although the Ampliseq method applied in seven *M. africanum* isolates detected some mutations in genes related to resistance. These polymorphisms could be specific evolutionary characteristics of the respective lineages. These results indicated that we must be cautious when reporting resistant genotypes, such as the mutations found in this study, which do not confer a resistance phenotype. Nevertheless, they could be assessed as specific for L6 (*rpoB*, *inhA*, and *katG* genes) and L5 (*gyrA* gene).

RFLP showed a substantial difference in the number of IS6110 copies between the L5 and L6 strains. L6 strains carried a lower number of copies in contrast to L5 strains. Spoligotyping and IS6110-RFLP allowed us to detect five clusters, including 17 cases. Although each technique has low discriminatory power separately, especially among low copy number strains, it increases when considered together. On the other hand, the location of some of the IS6110 insertion points adds differentiation capacity to the RFLP, as indicated in other publications where it has been described that RFLP analysis can underestimate the real copy number for the IS6110 element<sup>33,34</sup>. In this work, the isolates 2 and 21 present three bands that seem identical when observing their RFLP pattern. However, they share two locations (DR region and *moaX*) but isolate 21 has an IS6110 in the *Rv0963* gene, which does not share isolate 2 (Fig. 2). This indicates that it can happen in some cases that the coincidence of a band in the RFLP pattern does not imply that the IS6110 insertion point is the same. The explanation for this would be that a small difference between the lengths of the restriction fragment generated for two different locations of IS6110 is not appreciable in the RFLP pattern. Despite this, transmission was not considered in this study as it could overestimate the recent transmission rate.

In the context of a study of the IS6110 location in low copy number strains, we discovered an insertion within the *moaX* gene for the L6 strains analysed, and later we verified its presence in all L6 strains but never in L5 or other MTBC families studied. A previous work studied the insertion points of IS6110 in high-copy clinical isolates, specifically focusing on the Beijing genotype and revealed that its location in *moaX* gene was not characteristic of Beijing family<sup>22</sup>. Also, we found that in a previous work where the locations of IS6110 were studied in 579 MTBC strains representatives of the major lineages circulating in Europe and Latin America, the location of IS6110 in *moaX* was not detected in any case<sup>21</sup>. In all L6 strains included in our collection and the strains whose genomes are available online, the insertion point was the same. Altogether, it strongly suggests that this location is specific for L6, allowing us to differentiate this lineage from the rest of the strains of the MTBC. Within the scheme of the evolutionary stage of the tubercle bacillus, proposed by Brosch et al.<sup>13</sup>, we suggest the transposition of IS6110 into the *moaX* gene when L6 is separated from the rest of the lineages (Fig. 4). Besides, the location in *lipX:mshB* was frequent in L6 strains. These results agree with previous observations, indicating that each family has preferential insertion sites<sup>21,22,35</sup>, which is probably related to their evolutionary relationship.

The *moaX* gene encodes a molybdopterin (MPT) synthase with *moaD* and *moaE* activity that contributes to molybdenum cofactor (MoCo) synthesis in MTBC<sup>25</sup>. It has been shown that there is functional interchangeability between the MPT synthase subunits of *M. tuberculosis*, and in the case of MoaX, post-translational cleavage at the Gly82 residue is required for the functionality of this enzyme<sup>36</sup>. According to that, the IS6110 inserted in *moaX* gene of L6 strains is interrupting the MoaD subunit (Fig. 3). It has been described that some mutants in genes involved in molybdopterin biosynthesis had lost their ability to resist phagosomal acidification<sup>37</sup>. In most molybdenum-containing enzymes, the metal is coordinated to the dithiolene group of MPT to form MoCo. Enzymes that utilise MoCo harness the redox properties of molybdenum to catalyse redox reactions in carbon, nitrogen, and sulfur metabolism and to reduce terminal electron acceptors for anaerobic respiration<sup>25</sup>. One of these enzymes is NarG, a membrane-bound respiratory NR, suggesting a potentially important role for MoCo in the metabolism of *M. tuberculosis* in vivo. In an anaerobic environment, many bacteria can use nitrate as a final electron acceptor. Historically, *M. tuberculosis* has been differentiated from *M. bovis* because only *M. tuberculosis* can reduce significant amounts of nitrate (NO<sub>3</sub><sup>-</sup>) to nitrite (NO<sub>2</sub><sup>-</sup>). NR activity occurs at a low level during the aerobic growth of *M. tuberculosis* and increases significantly upon entry into the microaerobic stage. When we discovered the IS6110 insertion in *moaX* for L6 strains, we expected to find differences in NR activity between L5 and L6, but none showed NR activity. This indicates that the disruption of the MoaD subunit from MoaX in the L6 strains is not the only one responsible for the lack of activity observed in vitro. This result supports the hypothesis that homologous genes could compensate for any adaptive disadvantage of the bacteria due to the natural knockouts created by IS6110 insertion or other mutations<sup>25</sup>. Looking for another explanation for this result, we analysed the operon *narGHII* implicated in NR activity. The first mutation described<sup>38</sup> that prevented NR activity was -215 (t/c) SNP in the promoter of *narGHII* operon for *M. bovis*. *M. africanum* L5 and L6 have this mutation, but also *M. canetti*, which has NR activity<sup>39</sup>. There is another region responsible for NR activity, the *narK2* operon. A mutation in -10 promoter elements of the *narK2* operon reduced NR activity in BCG<sup>40,41</sup>. We found this mutation in the L6 strains available in NCBI but not in *M. canetti*, which had the same genotype as H37Rv. It seems that the presence of both mutations could explain the lack of NR activity we observed for *M.*

*africanum* L5 and L6. However, in latent anaerobiosis, BCG overexpressed the *narX* gene, a fused NR<sup>42</sup>. Thus, a similar enzyme could play this role for *M. africanum*.

Surprisingly, the search for mutations in the *narGHJ* promoter led us to the location of a mutation in –13 (a/g) *mutT2* gene, upstream of this operon, in the L6 strains analysed and in the NCBI complete genomes of *M. africanum* (L6), but not in L5 strains analysed nor in other TB genomes available in NCBI. This gene was studied in the Beijing lineage as a possible cause of a major number of SNPs related to resistance<sup>43</sup>. It has been observed that L6 has a higher variability in its genome in comparison to L5, which could be related to a higher mutation rate<sup>44,45</sup>. *MutT2* is involved in DNA repair, therefore the mutation detected in the *mutT2* promoter could increase the polymorphisms in L6 strains<sup>46</sup>.

A possible limitation of this work is that the number of strains studied was low. Nevertheless, all the isolates have been exhaustively and systematically characterised in a continuous period of 17 years. Consequently, the results objectively reflect the incidence of *M. africanum* in our region. On the other hand, genotyping methods do not discriminate enough to analyse transmission, so that whole-genome sequencing of the isolates would be more informative.

In summary, the results of this study indicate that TB caused by *M. africanum* is rare in Aragon, and the majority of the cases were in immigrants from West Africa. L6 was more prevalent, with few cases of L5. As far as we know, this is the first time that IS6110 locations have been determined in *M. africanum* strains, which has allowed us to detect the presence of a copy of IS6110 in the *moaX* gene in all L6 strains. Further studies on the implication of interruption of MPT synthase subunit-encoding genes in the physiology of L6 strains and its possible relationship with lower virulence would be of interest. The analysis of this location showed that it is a specific characteristic of the L6 strains, which allows us to distinguish this lineage of *M. africanum* from the rest of MTBC in a simple and fast way, using a PCR-based test.

## Material and methods

**Origin of clinical isolates.** In Aragon, a north region in Spain, all MTBC isolates are genotyped for surveillance purposes routinely since 2004, but 2003 isolates are also registered in the context of a previous study. In this work, we selected all patients with a microbiological diagnosis of TB caused by *M. africanum* between 1 Jan 2003 and 31 Dec 2019. The demographic (age, sex, country of birth, years since entry to Spain) and clinical (location of disease, sputum smear status, and previous diagnosis of TB) characteristics of the patients were retrospectively reviewed.

**Genotyping.** Genomic DNA was isolated using the cetyltrimethylammonium bromide (CTAB) method<sup>47</sup>. DNA was frozen at –80 °C and used in the different molecular techniques in this study. All strains were systematically genotyped by restriction fragment length polymorphism (RFLP) based on IS6110 and Spoligotyping. RFLP was performed as described by van Embden et al.<sup>17</sup>. Spoligotyping used a commercial membrane (Mapmygenome India Limited) to hybridise with the amplicons of the direct repeats region of each isolate. The procedure was previously described<sup>48</sup>. The genetic patterns were analysed by Bionumerics v7.6 software (Applied Maths, Kortrijk, Belgium) and introduced into the Database of the University of Zaragoza. TB cases caused by *M. africanum* were selected retrospectively by their spoligotype, a specific intermediate pattern between those of *M. tuberculosis* and *M. bovis*, according to the SITVIT definition<sup>23</sup>. Isolates were considered in cluster if they carried an identical IS6110-RFLP pattern and the same spoligotype if they had less than five copies of IS6110.

**Study of differential regions.** Subsequently, the presence or absence of the differential regions RD9, TbD1, RD702, and RD711 were analysed and used to classify the isolates into the different lineages of *M. africanum* L5, its sub-lineages L5.1 and L5.2, and L6<sup>15</sup>. The PCR were performed using the following primers: TBD1fla1-F (ctacctatctccgggtcca) and TBD1fla1-R (catagatcccggacatggtg) 2637/484 pb; RD9-flankF (gtgtag-gtcagcccatcc) and RD9-flankR (gcccaacagctcgacatc) 2484/72 pb<sup>13</sup>; RD702-F (ccgcaactcagtagctctt) and RD702-R (gttgggttctggtccat), and RD711-F (ggccgcctgctaagaacct) and RD711-R (cctaggccggcgacgaagt)<sup>14</sup>.

**Study of single polymorphisms.** A panel of primers focused on genes related to resistance in MTBC, and SNPs for lineage differentiation was analysed by AmpliSeq-based methodology using next-generation sequencing. This panel of primers was designed to amplify the *gyrA* gene from 7302 to 9818 (2516 pb), *rpoB* from 759,807 to 763,325 (3518 pb), *rpsL* from 781,560 to 781,934 (374 pb), *inhA* promoter from 1,673,303 to 1,673,440 (137 pb), *inhA* from 1,674,102 to 1,675,011 (909 pb), *katG* from 2,153,889 to 2,156,211 (2322 pb), *pncA* 2,288,681 to 2,289,341 (660 pb), *eis* from 2,714,124 to 2,715,432 (1308 pb), and *embB* from 4,246,514 to 4,249,810 (3296 pb). In addition, other hotspots to identify the lineages, specifically to identify *M. africanum* L5 (SNP in point 1377185, *Rv1234*) and L6 (SNP in point 378404, *Rv0309*), were amplified and analysed. Besides, the polymorphisms previously described for *M. bovis* in the *narGHJ* operon were analysed by amplification of the different regions. Primers used were the following: mutT2F-2 (tccggatgatattacctcc) and mutT2R-2 (tccgccgggtcggg-gac)<sup>43</sup>; narG-Fw (gccagctttgacacatcg) and NarG-Rv (gccagatgacgtttccgac); NarI-Fw (tggctaccactcggaatgac) and NarI-Rv (acgatgaggccggaacg). The detailed points are referred to as NC\_000962\_3.

**Location of IS6110 insertion sites.** To study IS6110 insertion sites, a ligation mediated PCR was used as described by Prod'homme et al.<sup>49</sup> to amplify one or both ends of each copy of IS6110 and its flanking sequence. Briefly, genomic DNA was digested with *SalI* enzyme and ligated to a linker containing a *SalI* restriction site. The resulting template was then digested by *SalI*. PCR was performed using ISA1 or ISA3, specific primers for IS6110 and directed outwards from this element<sup>50</sup>, and the linker primer Salgd. The template was initially denatured by incubation at 95 °C for 9 min and amplified by 35 cycles of PCR (95 °C for 30 s, 70 °C, and 72 °C for 90 s)

followed by a final extension at 72 °C for 10 min. Amplified products were separated by standard horizontal gel electrophoresis in a 1.5% agarose gel in tris–borate–EDTA buffer (90 mM tris, 90 mM boric acid, 2 mM EDTA) and stained with ethidium bromide. PCR products were purified, using GFX PCR DNA and Gel Band Purification Kit (Amersham Pharmacia Biotech) followed by ExoSAP-IT PCR Product Cleanup Reagent (Affymetrix), sequenced and analysed for homology with Tuberculist (<http://genolist.pasteur.fr/TubercuList>).

**Enzymatic assay of NR.** The NR activity test was performed with actively growing cultures, which were inoculated directly into phosphate buffer supplemented with nitrate and incubated for 2 h at 37 °C. The mycobacteria were cultured on 7H10 agar supplemented with 0.2% glycerol and 10% albumin/dextrose/catalase (ADC). One L5 strain and one L6 were inoculated into phosphate buffer supplemented with 10 mM nitrate. Following 2 h of incubation at 37 °C, naphthylamide and sulfanilic acid reagents were added, and the colour was then observed<sup>51</sup>.

**Computer analysis.** The sequences generated were aligned and compared with the sequences of *M. tuberculosis* H37Rv (<http://genolist.pasteur.fr/TubercuList>) and *M. africanum* complete genomes, NC\_015758.1 and CP010334.1 (<http://blast.ncbi.nlm.nih.gov>), using the Basic Local Alignment Search Tool (BLAST).

**Ethics declarations.** The permission to take informed consent was formally waived by the Comité de Ética de la Investigación de la Comunidad Autónoma de Aragón (CEICA), Spain, CI.PI18/068. No human tissues were used in the study. Once received the bacterial isolate, it was coded (NSTRAIN). The epidemiological data of the cases were sent by fax and were anonymised keeping only the code given to track the analysis of the clinical characteristics, to follow the Helsinki ethical principles for medical research involving human data. The experiment protocol followed was revised and approved by the CEICA and is in line with the Declaration of Helsinki, as revised in 2013.

### Data availability

Sequences data reported in the present study were deposited in GenBank under accession numbers MW987573MW987574 and MW987575.

Received: 24 February 2021; Accepted: 23 April 2021

Published online: 14 May 2021

### References

- World Health Organization. Global Tuberculosis Report 2020. Licence: CC BY-NC-SA 3.0 IGO. Geneva. <https://doi.org/10.3201/eid1910.121023> (2020).
- Gehre, F. *et al.* Immunogenic *Mycobacterium africanum* strains associated with ongoing transmission in the Gambia. *Emerg. Infect. Dis.* **19**, 1599–1605. <https://doi.org/10.3201/eid1910.121023> (2013).
- Castets, M., Boisvert, H., Grumbach, E., Brunel, M. & Rist, N. Les bacilles tuberculeux de type Africain: Note préliminaire. *Rev. Tuberc. Pneumol.* **32**, 179–184 (1968).
- Meyer, L. & David, H. L. Evaluation de l'active urease et de l'activité  $\beta$ -glucosidase pour l'identification pratique des mycobactéries. *Ann. Microbiol.* **130B**, 323–332 (1979).
- Thorel, M. F. Isolation of *Mycobacterium africanum* from monkeys. *Tubercle* **61**, 101–104. [https://doi.org/10.1016/0041-3879\(80\)90018-5](https://doi.org/10.1016/0041-3879(80)90018-5) (1980).
- Bold, T. D. *et al.* Impaired fitness of *Mycobacterium africanum* despite secretion of ESAT-6. *J. Infect. Dis.* **205**, 984–990. <https://doi.org/10.1093/infdis/jir883> (2012).
- Asante-Poku, A. *et al.* Molecular epidemiology of *Mycobacterium africanum* in Ghana. *BMC Infect. Dis.* **16**, 385. <https://doi.org/10.1186/s12879-016-1725-6> (2016).
- Asante-Poku, A. *et al.* *Mycobacterium africanum* is associated with patient ethnicity in Ghana. *PLoS Negl. Trop. Dis.* **9**, e3370. <https://doi.org/10.1371/journal.pntd.0003370> (2015).
- Coscolla, M. *et al.* Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. *Emerg. Infect. Dis.* **19**, 969–976. <https://doi.org/10.3201/eid1906.121012> (2013).
- Rahim, Z. *et al.* Characterization of *Mycobacterium africanum* subtype I among cows in a dairy farm in Bangladesh using spoligotyping. *Southeast Asian J. Trop. Med. Public Health.* **38**, 706–713 (2007).
- Alfredsen, S. & Saxegaard, F. An outbreak of tuberculosis in pigs and cattle caused by *Mycobacterium africanum*. *Vet. Rec.* **131**, 51–53. <https://doi.org/10.1136/vr.131.3.51> (1992).
- Gudan, A. *et al.* Disseminated tuberculosis in hyrax (*Procavia capensis*) caused by *Mycobacterium africanum*. *J. Zoo Wildl. Med.* **39**, 386–391. <https://doi.org/10.1638/06-041.1> (2008).
- Brosch, R. *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA* **99**, 3684–3689. <https://doi.org/10.1073/pnas.052548299> (2002).
- Gagneux, S. *et al.* Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **103**, 2869–2873. <https://doi.org/10.1073/pnas.0511240103> (2006).
- Ates, L. S. *et al.* Unexpected genomic and phenotypic diversity of *Mycobacterium africanum* lineage 5 affects drug resistance, protein secretion, and immunogenicity. *Genome Biol. Evol.* **10**, 1858–1874. <https://doi.org/10.1093/gbe/evy145> (2018).
- Isea-Peña, M. C. *et al.* *Mycobacterium africanum*, an emerging disease in high-income countries?. *Int. J. Tuberc. Lung Dis.* **16**, 1400–1404. <https://doi.org/10.5588/ijtld.12.0142> (2012).
- Van Embden, J. D. A. *et al.* Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: Recommendations for a standardized methodology. *J. Clin. Microbiol.* **31**, 406–409 (1993).
- Soto, C. Y. *et al.* 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. <https://doi.org/10.1128/JCM.42.1.212-219.2004> (2004).
- Alonso, H. *et al.* Deciphering the role of IS6110 in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237. *Tuberculosis* **91**, 117–126. <https://doi.org/10.1016/j.tube.2010.12.007> (2011).
- Gonzalo-Asensio, J. *et al.* New insights into the transposition mechanisms of IS6110 and its dynamic distribution between *Mycobacterium tuberculosis* complex lineages. *PLoS Genet.* **14**, e1007282. <https://doi.org/10.1371/journal.pgen.1007282> (2018).



21. Reyes, A. *et al.* IS-seq: A novel high throughput survey of in vivo IS6110 transposition in multiple *Mycobacterium tuberculosis* genomes. *BMC Genom.* **13**, 249. <https://doi.org/10.1186/1471-2164-13-249> (2012).
22. Alonso, H., Samper, S., Martín, C. & OtaI, I. Mapping IS6110 in high-copy number *Mycobacterium tuberculosis* strains shows specific insertion points in the Beijing genotype. *BMC Genom.* **14**, 422. <https://doi.org/10.1186/1471-2164-14-422> (2013).
23. Demay, C. *et al.* SITVITWEB—A publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect. Genet. Evol.* **12**, 755–766. <https://doi.org/10.1016/j.meegid.2012.02.004> (2012).
24. Huard, R. C. *et al.* Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. *J. Bacteriol.* **188**, 4271–4287. <https://doi.org/10.1128/JB.01783-05> (2006).
25. Williams, M. J., Kana, B. D. & Mizrahi, V. Functional analysis of molybdopterin biosynthesis in mycobacteria identifies a fused molybdopterin synthase in *Mycobacterium tuberculosis*. *J. Bacteriol.* **193**, 98–106. <https://doi.org/10.1128/JB.00774-10> (2011).
26. López-Calleja, A. I. *et al.* Genotyping of *Mycobacterium tuberculosis* over two periods: A changing scenario for tuberculosis transmission. *Int. J. Tuberc. Lung Dis.* **11**, 1080–1086 (2007).
27. Asare, P. *et al.* Reduced transmission of *Mycobacterium africanum* compared to *Mycobacterium tuberculosis* in urban West Africa. *Int. J. Infect. Dis.* **73**, 30–42. <https://doi.org/10.1016/j.ijid.2018.05.014> (2018).
28. WHO. *WHO Global tuberculosis report 2019*. Geneva: World Health Organization; Licence: CC BY-NC-SA 3.0 IGO (2019). <https://doi.org/10.1037/0033-2909.126.1.78>.
29. Kone, A. *et al.* Differential HLA allele frequency in *Mycobacterium africanum* vs *Mycobacterium tuberculosis* in Mali. *HLA*. **93**, 24–31. <https://doi.org/10.1111/tan.13448> (2019).
30. Cá, B. *et al.* Experimental evidence for limited in vivo virulence of *Mycobacterium africanum*. *Front. Microbiol.* **10**, 2102. <https://doi.org/10.3389/fmicb.2019.02102> (2019).
31. Traore, B. *et al.* Molecular strain typing of *Mycobacterium tuberculosis* complex in Bamako, Mali. *Int. J. Tuberc. Lung Dis.* **16**, 911–916. <https://doi.org/10.5588/ijtld.11.0397> (2012).
32. Kone, B. *et al.* Exploring the usefulness of molecular epidemiology of tuberculosis in Africa: A systematic review. *Int. J. Mol. Epidemiol. Genet.* **11**, 1–15 (2020).
33. OtaI, I. *et al.* Mapping of IS6110 insertion sites in *Mycobacterium bovis* isolates in relation to adaptation from the animal to human host. *Vet. Microbiol.* **135**, 406. <https://doi.org/10.1016/j.vetmic.2007.11.038> (2008).
34. Millán-Lou, M. I. *et al.* In Vivo IS6110 profile changes in a *Mycobacterium tuberculosis* strain as determined by tracking over 14 years. *J. Clin. Microbiol.* **53**, 2359–2361. <https://doi.org/10.1128/JCM.00607-15> (2015).
35. Comín, J. *et al.* A whole-genome sequencing study of an X-family tuberculosis outbreak focus on transmission chain along 25 years. *Tuberculosis* **126**, 102022. <https://doi.org/10.1016/j.tube.2020.102022> (2021).
36. Narrandes, N. C., Machowski, E. E., Mizrahi, V. & Kana, B. D. Cleavage of the moaX-encoded fused molybdopterin synthase from *Mycobacterium tuberculosis* is necessary for activity. *BMC Microbiol.* **15**, 22. <https://doi.org/10.1186/s12866-015-0355-2> (2015).
37. Brodin, P. *et al.* High content phenotypic cell-based visual screen identifies *Mycobacterium tuberculosis* acyltrehalose-containing glycolipids involved in phagosome remodeling. *PLoS Pathog.* **6**, e1001100 (2010).
38. Stermann, M., Sedlacek, L., Maass, S. & Bange, F. C. A promoter mutation causes differential nitrate reductase activity of *Mycobacterium tuberculosis* and *Mycobacterium bovis*. *J. Bacteriol.* **186**, 2856–2861. <https://doi.org/10.1128/JB.186.9.2856-2861.2004> (2004).
39. Goh, K. S., Rastogi, N., Berchel, M., Huard, R. C. & Sola, C. Molecular evolutionary history of tubercle bacilli assessed by study of the polymorphic nucleotide within the nitrate reductase (narGHIJ) operon promoter. *J. Clin. Microbiol.* **43**, 4010–4014. <https://doi.org/10.1128/JCM.43.8.4010-4014.2005> (2005).
40. Honaker, R. W. *et al.* *Mycobacterium bovis* BCG vaccine strains lack narK2 and narX induction and exhibit altered phenotypes during dormancy. *Infect. Immun.* **76**, 2587–2593. <https://doi.org/10.1128/IAI.01235-07> (2008).
41. Chauhan, S., Singh, A. & Tyagi, J. S. A single-nucleotide mutation in the -10 promoter region inactivates the narK2X promoter in *Mycobacterium bovis* and *Mycobacterium bovis* BCG and has an application in diagnosis. *FEMS Microbiol. Lett.* **303**, 190–196. <https://doi.org/10.1111/j.1574-6968.2009.01876.x> (2010).
42. Hutter, B. & Dick, T. Up-regulation of narX, encoding a putative ‘fused nitrate reductase’ in anaerobic dormant *Mycobacterium bovis* BCG. *FEMS Microbiol. Lett.* **178**, 63–69. [https://doi.org/10.1016/S0378-1097\(99\)00348-1](https://doi.org/10.1016/S0378-1097(99)00348-1) (1999).
43. Ebrahimi-Rad, M. *et al.* Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerg. Infect. Dis.* **9**, 838–845. <https://doi.org/10.3201/eid0907.020803> (2003).
44. Oterche, I. D. *et al.* Comparative genomics of *Mycobacterium africanum* Lineage 5 and Lineage 6 from Ghana suggests distinct ecological niches. *Sci. Rep.* **8**, 11269. <https://doi.org/10.1038/s41598-018-29620-2> (2018).
45. Coscolla, M. *et al.* Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microb. Genom.* <https://doi.org/10.1099/mgen.0.000477> (2021).
46. Dos Vultos, T., Blázquez, J., Raugier, J., Matic, I. & Gicquel, B. Identification of nudix hydrolase family members with an antimutator role in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *J. Bacteriol.* **188**, 3159–3161. <https://doi.org/10.1128/JB.188.8.3159-3161.2006> (2006).
47. van Soolingen, D., de Haas, P. E., Hermans, P. W. & van Embden, J. D. DNA fingerprinting of *Mycobacterium tuberculosis*. *Methods Enzymol.* **235**, 196–205 (1994).
48. Kamerbeek, J. *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**, 907–914. <https://doi.org/10.1128/jcm.35.4.907-914.1997> (1997).
49. Prod'homme, G. *et al.* A reliable amplification technique for the characterization of genomic DNA sequences flanking insertion sequences. *FEMS Microbiol. Lett.* **158**, 75–81. [https://doi.org/10.1016/S0378-1097\(97\)00503-X](https://doi.org/10.1016/S0378-1097(97)00503-X) (1998).
50. Mendiola, M. V., Martín, C., OtaI, I. & Gicquel, B. Analysis of the regions responsible for IS6110 RFLP in a single *Mycobacterium tuberculosis* strain. *Res. Microbiol.* **143**, 767–772. [https://doi.org/10.1016/0923-2508\(92\)90104-V](https://doi.org/10.1016/0923-2508(92)90104-V) (1992).
51. Organización Panamericana de la salud. Manual para el diagnóstico bacteriológico de la tuberculosis. Normas y guía técnica. Parte II Cultivo. Washington, D.C. <https://iris.paho.org/handle/10665.2/18616> (2008).

## Acknowledgements

Authors would like to acknowledge the use of Servicio General de Apoyo a la Investigación-SAI, Universidad de Zaragoza (Servicio de Análisis Microbiológico), and Servicios Científico Técnicos, IACS (Servicio de Secuenciación y Genómica Funcional and Servicio de Biocomputación). We would like to thank the Aragon Health Department for their constant support in this work.

## Author contributions

I.O., and S.S., participated similarly in the coordination of the work. J.C. performed the laboratory techniques to confirm results. They three wrote the manuscript. M.L.M. analysed the epidemiological data. J.V., J.S. and L.T. worked in the mycobacterial laboratories and are participants in the Working Group on Tuberculosis in Aragon

(EPIMOLA), they diagnosed, and completed minimal information on the TB cases and performed the genotyping. M.J.I., D.I., C.L. and C.M. work in the molecular epidemiology studies of TB in Aragon.

### Funding

This work was supported by the Carlos III Health Institute in the context of two consecutive Grants FIS18/0336 and J.C. was awarded a scholarship by the Government of Aragon/European Social Fund, “Building Europe from Aragon”.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to S.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

---

## Aragonese Working Group on Molecular Epidemiology of Tuberculosis (EPIMOLA)

María José Iglesias<sup>5</sup>, Daniel Ibarz<sup>5</sup>, Jesús Viñuelas<sup>7</sup>, Luis Torres<sup>8</sup>, Juan Sahagún<sup>9</sup>, María Carmen Lafoz<sup>5</sup> & María Carmen Malo<sup>10</sup>

<sup>7</sup>Hospital Universitario Miguel Servet, Zaragoza, Spain. <sup>8</sup>Hospital General Universitario San Jorge, Huesca, Spain.

<sup>9</sup>Hospital Universitario Lozano Blesa, Zaragoza, Spain. <sup>10</sup>Salud Pública, Gobierno de Aragón, Zaragoza, Spain.