Genomic Characterization of IS*6110* Insertions in *Mycobacterium orygis*

Evolutionary Bioinformatics
Volume 20: 1–7

DOI: 10.1177/11769343241240558

Ahmed Kabir Refaya[1], Umashankar Vetrivel[2] and Kannan Palaniyandi[1]

[1]Department of Immunology, ICMR-National Institute for Research in Tuberculosis, Chetpet, Chennai, India. [2]Department of Virology & Biotechnology/Bioinformatics Division, ICMR-National Institute for Research in Tuberculosis, Chetpet, Chennai, India.

ABSTRACT: *Mycobacterium orygis*, a subspecies of the *Mycobacterium tuberculosis* complex (MTBC), has emerged as a significant concern in the context of One Health, with implications for zoonosis or zooanthroponosis or both. MTBC strains are characterized by the unique insertion element IS*6110*, which is widely used as a diagnostic marker. IS*6110* transposition drives genetic modifications in MTBC, imparting genome plasticity and profound biological consequences. While IS*6110* insertions are customarily found in the MTBC genomes, the evolutionary trajectory of strains seems to correlate with the number of IS*6110* copies, indicating enhanced adaptability with increasing copy numbers. Here, we present a comprehensive analysis of IS*6110* insertions in the *M. orygis* genome, utilizing ISMapper, and elucidate their genetic consequences in promoting successful host adaptation. Our study encompasses a panel of 67 paired-end reads, comprising 11 isolates from our laboratory and 56 sequences downloaded from public databases. Among these sequences, 91% exhibited high-copy, 4.5% low-copy, and 4.5% lacked IS*6110* insertions. We identified 255 insertion loci, including 141 intragenic and 114 intergenic insertions. Most of these loci were either unique or shared among a limited number of isolates, potentially influencing strain behavior. Furthermore, we conducted gene ontology and pathway analysis, using eggNOG-mapper 5.0, on the protein sequences disrupted by IS*6110* insertions, revealing 63 genes involved in diverse functions of Gene Ontology and 45 genes participating in various KEGG pathways. Our findings offer novel insights into IS*6110* insertions, their preferential insertion regions, and their impact on metabolic processes and pathways, providing valuable knowledge on the genetic changes underpinning IS*6110* transposition in *M. orygis*.

KEYWORDS: *Mycobacterium orygis*, IS*6110*, bovine tuberculosis, zoonosis, *Mycobacterium tuberculosis* complex, One Health

RECEIVED: September 1, 2023. **ACCEPTED:** March 4, 2024.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is funded by the Indian Council of Medical Research (Award No. 2020-6334)

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Kannan Palaniyandi, Department of Immunology, ICMR-National Institute for Research in Tuberculosis, #1, Mayor Sathiyamoorthy Road, Chetpet, Chennai 600 031, India. Emails: palaniyandi.k@icmr.gov.in; kannanvet@rediffmail.com

Ahmed Kabir Refaya, Department of Immunology, ICMR-National Institute for Research in Tuberculosis, #1, Mayor Sathiyamoorthy Road, Chetpet, Chennai, 600 031, India. Email: refayanasar@gmail.com

Introduction

The insertion sequence IS*6110* is found explicitly in *Mycobacterium tuberculosis* complex (MTBC), the causative agent of tuberculosis (TB) in humans and other mammals, including wildlife and farm animals, which are responsible for zoonotic or zooanthroponotic TB transmission. This unique trait made IS*6110* a historical epidemiological marker for diagnosing MTBC in biological samples, enabling the detection of TB outbreaks and transmission chains, although currently being replaced by robust sequencing methods.[1-3] The transposition of IS*6110* nurtures various genetic modifications in MTBC strains, conferring genomic plasticity and significant biological implications.[4] Insertions of IS*6110* amid the coding region might render the gene inactive or act as a mobile promoter controlling gene activation, leading to unusual consequences in the host-pathogen evolution.[5,6] The members of the MTBC are classified into high (>6) and low (<7) copy strains based on the content of IS*6110*, although no clear evidence accounts for its phenotypic consequences in bacterial physiology and pathogenesis.[7] However, *M. tuberculosis* Beijing/W lineage (L2) possessing a high content of 15 IS*6110* insertions is linked with high virulence and colossal spread of drug-resistant strains.[8] It has been demonstrated that a reasonable number of IS*6110* might render strain-specific phenotypes catering to selective advantages during its course of infection rather than excessive accumulation resulting in inactivation or deletion of essential genetic regions, being detrimental to the bacterium.[5] Among the animal-adapted MTBC, *Mycobacterium bovis (M. bovis)* is known to possess only 1 or few copies of IS*6110*,[4] whereas *M. orygis* possess high copy numbers ranging from 17 to 20, however, the exact position of insertions are yet unknown[9] and not much information is available on other members which possess a zoonotic risk. *M. orygis*, has been isolated from various species, including oryx, waterbucks, cattle, deer, rhinoceros, monkeys, and humans and the exact extent of its host range still remains unknown.[9-16] A recent study in India identified 7 *M. orygis* isolates among 940 clinical isolates from humans.[17] Our laboratory has frequently been isolating *M. orygis* from cattle and wild ungulates recently, and various reports on this species are also being reported in South Asia.[13,14]

The advent of Whole Genome Sequencing (WGS) has facilitated to localize the IS*6110* position in the genomes of MTBC and has been recently used to identify its role in bovine adaptation.[18,19] This study investigates the distribution of

IS*6110* among *M. orygis* strains and its effect on the genes by *in silico* characterization. We have manifested the use of WGS to localize the IS*6110* copies and their chromosomal distribution using the sequences of 67 *M. orygis* strains, comprising 56 WGS paired reads downloaded from the National Centre for Biotechnology Information-Sequence Read Archive (NCBI-SRA) and 11 WGS paired reads sequenced from our lab and to evaluate the level of IS*6110* sequence stability and their orthology evolving in a multi-host system. Ever since the publication of the complete genome of *M. orygis* 51145,[20] it has been used well as a reference genome (GenBank accession No. CP063804) for analysis of *M. orygis* strains.[13] Although the complete genome has been sequenced, the genes are not functionally annotated, and the mycobrowser (https://mycobrowser.epfl.ch/) identifies these genes as unknown, thereby demanding the use of orthologous comparative methods to identify its function. Hence, to overcome this, we additionally used *M. tuberculosis* H37Rv (NC000962.3) as the reference genome.

Materials and Methods

Data acquisition and analysis

To compare the sequence of our study isolates with existing sequences available in the database, NCBI SRA was searched with the term "*Mycobacterium orygis*." All the sequences branded as *M. orygis* or *M. tuberculosis var. orygis* were downloaded from SRA with the fasterq-dump tool from the SRA toolkit (version 2.9.6) available at https://ncbi.github.io/sra-tools/. Only the sequences which downloaded as paired reads, and the sequences which possessed the species-specific Region of Difference (RDs) and single nucleotide polymorphisms (SNPs) were used for further analysis (Table S6). A total of 56 paired-end sequences of *M. orygis* strains isolated from human and animal sources were used, along with 11 study isolates. The raw sequences were filtered using trimmomatic (version 0.39) with a framework of minimum phred quality score and read length set to 30 and 80, respectively. The filtered sequences were further analyzed through vSNP https://github.com/USDA-VS/vSNP using *M. orygis* 51145 (CP063804) as the reference genome to generate SNPs. The phylogenetic tree was constructed based on the aligned whole-genome SNP sequences under a GTRCATI model of substitution and maximum-likelihood algorithm with a bootstrap replication of 1000[21] and visualized using the interactive tree of life (iTOL).[22]

Analysis of IS6110 distribution

ISMapper (version 2.0)[23] pipeline https://github.com/jhawkey/IS_mapper was used to localize IS*6110* on short-read sequences with *M. tuberulosis* H37Rv (NC000962.3) and *M. orygis* 51145 (CP063804) as reference genomes. The presence of IS*6110* was visualized manually with Integrative Genomics Viewer[24,25] using the bam file generated by ISMapper 2.0.

Analysis of orthologous genes

The genomic positions of the IS*6110* were deduced through ISMapper analysis. The genes surrounding the insertion sites (intergenic) and genes possessing the insertion sites within them (intragenic) were identified based on the respective reference genomes to determine orthologous genomic sites of the IS.

Gene ontology enrichment and KEGG pathway analysis

The protein sequences of the upstream and downstream genes of IS*6110* and the protein sequence of the genes which possessed the IS*6110* insertions were used to perform functional annotation using eggNOG-mapper (version emapper—2.1.9)[26] based on eggNOG orthology data.[27] Sequence searches were performed using DIAMOND (version 2.0.11)[28] to mine the orthologs.

Results

Distibution of IS6110 among M. orygis strains

In our analysis of IS*6110* localization among the 67 sequences from both reference genomes, we observed that 4.5% (3/67) of the strains in the human cluster did not possess any IS*6110* insertions and were classified as no-copy strains. We found that 91% (61/67) of the strains were high-copy, while 4.5% (3/67) were classified as low-copy. The number of IS*6110* loci varied within the samples. In the human cluster, the loci ranged from 2 to 30 (with a median of 25) using the reference CP063804. In the animal cluster, the loci ranged from 20 to 42 (with a median of 26.5) using the same reference. However, when using the reference NC000962.3, the number of loci ranged from 2 to 43 (with a median of 32) in the human cluster and from 26 to 54 (with a median of 35) in the animal cluster. Overall, we identified a total of 24 common insertion loci. The specific regions of these insertion sites are tabulated and schematically represented, along with the phylogenetic tree generated using the maximum likelihood approach with all 67 *M. orygis* sequences (Figure 1 and Table 1).

Unique and novel IS6110 insertions in M. orygis strains

In addition to the previously mentioned typical and known IS*6110* insertions, we also identified novel IS*6110* insertions using ISMapper. When mapped against CP063804, we identified 107 novel insertion loci, of which, 61.7% (66/107) were intragenic and 38.3% (41/107) were intergenic. Approximately 58% (62/107) of these insertion loci were unique or shared between 2 and 3 strains, while 15 loci were shared in at least 4 strains, with only one being intergenic. Of the 15 common insertion sites, 11 were exclusively found in the animal cluster, 3 in both the human and animal clusters, and 1 insertion site

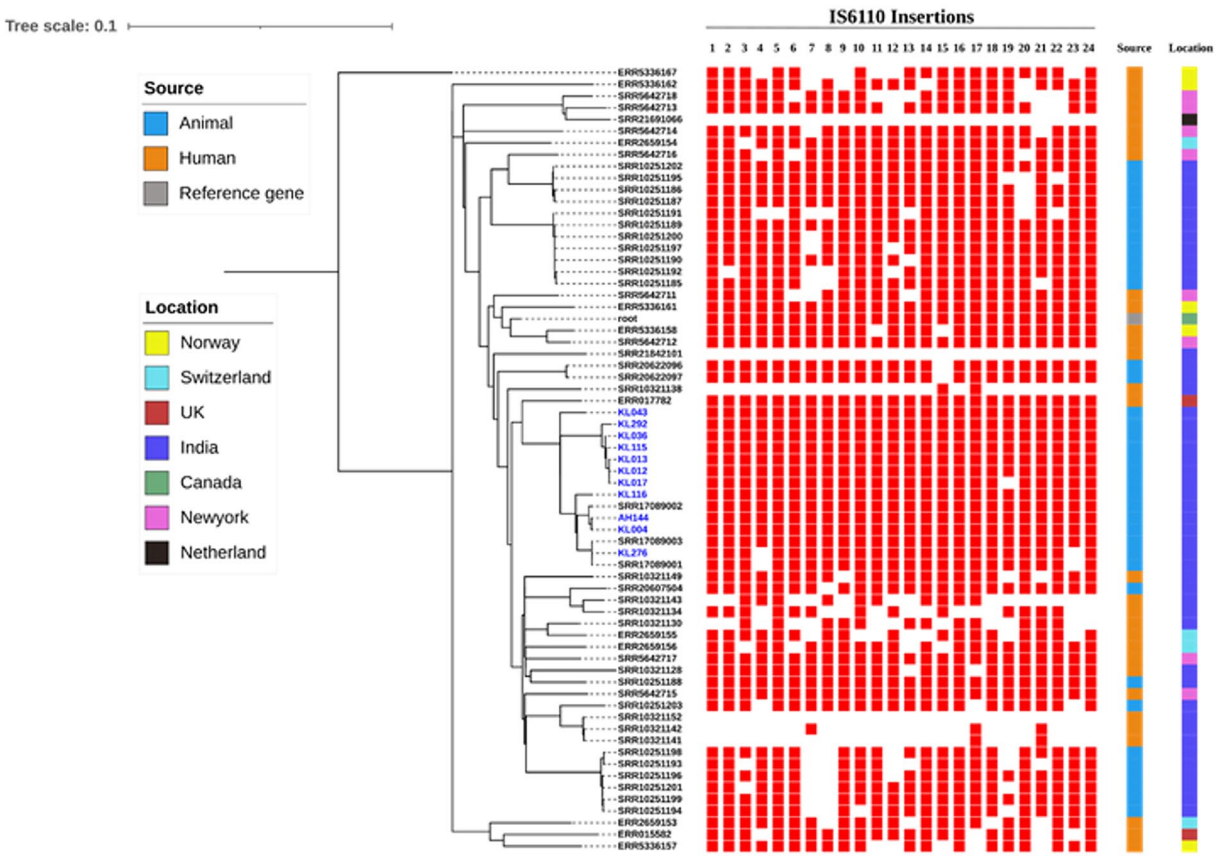

Figure 1. Contextual Phylogenetic tree combining the 11 study isolates along with 56 *M. orygis* isolates around the globe: Constructed using whole genome SNPs obtained after alignment with *M. orygis* reference genome CP063804. The heatmap shows the presence or absence of IS*6110* in 67 *M. orygis* genomes. Red squares indicate the presence of IS*6110* in a specific site. Descriptions of each IS*6110* site are described in Table 1. The source and the geographic location of the isolates are represented in the respective bands. The study isolates are labeled in blue.

was found only in the human cluster. Almost 124 novel insertion loci were identified when mapped against NC000962.3, out of which 39.5% (49/124) were intergenic and 60.5% (75/124) intragenic. Only 14 insertion sites were identified in both clusters, of which only 2 were intergenic. Notably, we observed variations in the localization and orientation of IS*6110* insertions among the strains despite being present in the same genomic region. These findings shed light on the diversity and distribution of novel IS*6110* insertions, providing valuable insights into the genetic landscape of the studied clusters. The genes on either side of IS*6110* and those that contained IS*6110* in both the H37Rv and 51145 reference genomes were combined. We identified 227 genes, among which 85 were from *M. tuberculosis*, 64 from *M. orygis* and 79 genes were found to be common in both *M. orygis* and *M. tuberculosis* (Supplemental Table S1). Using eggNOG Mapper, the protein sequences of these 227 genes were subjected to functional annotation.

Genetic significances of IS6110 insertions

Approximately 204/227 genes were scanned by eggNOG 5.0, and 82 were identified to be functionally involved in Gene Ontology (GO) and KEGG pathway (KP) enrichment analysis. Almost 37/82 genes were exclusively linked to various processes in GO and 19/82 genes were mapped in numerous KP. Additionally, we also identified 26 genes that were involved in both GO and KP. A total of 63 genes were encompassed in 669 GOs comprising 455 Biological processes (BP), 161 Molecular Functions (MF) and 53 Cellular components (CC), and the distribution of the genes in various functions are represented in Figure 2. Among the 63 genes involved in GO, 32 were found only in the human cluster, 15 in the animal cluster, and 16 in both clusters. The intergenic insertion sequence upstream of Rv0742 (Hypothetical protein) and Rv2965c/RJtmp_003059 *(kdtB/coaD)* were found in almost 61 and 60 isolates, respectively. The intragenic insertion sequences present within the genes Rv1266/Rjtmp_001332 (*pknH*), Rv2565 (NTE family protein), Rv1753 *(PPE24)*, Rv0402/Rjtmp_000420/Rjtmp_000422 *(mmpL1)* and Rv1367c/Rjtmp_001448 (Hypothetical protein) were also found in almost 55 to 61 isolates. All these genes were involved in various GO processes (Supplemental Tables S2 and S3). Other insertion regions in the human cluster were unique and found in only 1 strain or at least 2 strains. A similar pattern was observed in the animal cluster except for 1 intergenic region between Rv2338c/Rjtmp_002419 (*moeW/thiF* family adenyltransferase) and Rv2339/RJtmp_002420 (*mmpL9*) and an

Table 1. Details of IS*6110* insertions among all *M. orygis* isolates.

IS NO.	ORIENTATION	LOCI	POSITION	NO. OF ISOLATES
1	R	RJtmp_000364: RJtmp_000365	418095-419370	61
2	R	RJtmp_000420: RJtmp_000421	481559-482854	59
3	R	RJtmp_000782: RJtmp_000783	835711-837062	60
4	R	RJtmp_000841: Rjtmp_000842	892501-893790	52
5	R	RJtmp_001232: RJtmp_001233	1304672-1305974	60
6	F	RJtmp_001332: RJtmp_001333	1423001-1424355	55
7	R	RJtmp_001435: RJtmp_001436	1536637-1537898	44
8	R	RJtmp_001448: RJtmp_001449	1550641-1551955	58
9	F	RJtmp_001738: RJtmp_001739	1891878-1893148	56
10	R	RJtmp_001757: RJtmp_001758	1909881-1911200	59
11	F	RJtmp_001833: RJtmp_001834	198611-1987961	55
12	R	RJtmp_001837: RJtmp_001838	1994072-1995329	44
13	F	RJtmp_001872: RJtmp_001873	2035374-2036722	50
14	F	RJtmp_001988: RJtmp_001989	2151738-2152992	58
15	F	RJtmp_002433: RJtmp_002434	2602192-2603492	56
16	R	RJtmp_002654: RJtmp_002655	2842275-2843648	60
17	F	RJtmp_002901: RJtmp_002902	3068456-3069736	59
18	F	RJtmp_003056: RJtmp_003057	3264524-3265824	57
19	R	RJtmp_003106: RJtmp_003107	3314904-3316281	51
20	R	RJtmp_003217: RJtmp_003218	3429234-3430489	52
21	F	RJtmp_003255: RJtmp_003256	3470593-3471871	53
22	F	RJtmp_003278: RJtmp_003279	3493266-3494534	55
23	R	RJtmp_003430: RJtmp_003431	3653243-3654495	52
24	R	RJtmp_003855: RJtmp_003856	4137024-4138299	59

intragenic insertion in Rv2430/RJtmp_002513 *(PPE41)* which were found in at least 6 and 7 strains, respectively.

Almost 62 KPs were identified in which nearly 45 genes were involved, including 20 from the human cluster, 14 from the animal cluster and 11 genes from both human and animal clusters. Here too, we found that the genes from both clusters mainly were unique or shared between a couple of strains. The majority of the genes (23/45) were involved in metabolic pathways (map01100) and 13/45 genes were involved in the biosynthesis of secondary metabolites (map01110). In addition to metabolic pathways, genes such as Rv0742, Rv2964/Rjtmp_003056 (*purU*), Rv2965c/RJtmp_003059 (*kdtB*/*coaD*) were found in almost 60 strains and these genes were involved in glycerolipid metabolism (map00561), Glycolate and dicarboxylate metabolism (map00630) & Pantothenate and CoA biosynthesis (map00770) respectively. The gene Rv1368/RJtmp_001450 (*lprF*), found in 7 isolates (3 in human and 4 in animal cluster) was involved in the TB pathway (map05152), and RJtmp_000891 (*pip*) found only in 14 isolates from the animal cluster was involved in arginine and proline metabolism (map00330). Most of the genes *dnaN, gap, icd, lpdB, proA, trpE, glgC, plcC*, and Rjtmp_002367 were involved in at least 5 or more pathways but were found only in 1 or 2 isolates (Supplemental Tables S4 and S5).

IS6110 insertions and genetic significance among our study isolates

All our study isolates belonged to the animal cluster and were high-copy strains possessing more than 25 IS*6110* insertions. In addition to the previously mentioned insertion regions, we also identified an IS*6110* intergenic insertion between Rv1470/Rjtmp_001552 (*trxA*) and Rv1471/Rjtmp_001553 (*trxB1*) in 2 of our study isolates (KL013, KL017). *TrxA* was involved in

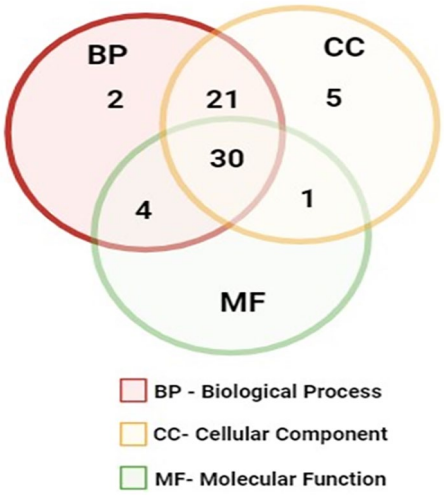

Figure 2. Distribution of 63 genes in various GO process.

seleno compound metabolism (map00450), NOD-like receptor signaling pathway (map04621), and Fluid sheer stress and atherosclerosis pathway (map05418). Another intragenic insertion within RJtmp_001745 was exclusively found in 6 of our isolates (KL012, KL013, KL017, KL115, KL036, and KL292). Strain KL012 and KL292 possess a unique insertion in the genes Rjtmp_000047 (*marR* transcriptional regulator) and Rjtmp_002096 (*csm5*), respectively and the gene *csm5* was identified to be involved in 11 GOs. A total of 7 unique insertions were found in strain KL036, among which 3 were intergenic and 4 intragenic. The intragenic insertions were found in genes Rv1364/Rjtmp_001464 (*carB*), Rv1631/RJtmp_001705 (dephospho-coA kinase), Rv2379/RJtmp_002451 (*mbtF*), and Rv1609/RJtmp_001681 (anthranilate synthase component). Among these genes, *carB* was involved in almost 50 GOs and *mbtF* is involved in the Biosynthesis of siderophore group non-ribosomal peptides (map01053). Intergenic insertions were found between Rv1187/Rjtmp_001253 (*rocA*/*pruA*) & Rv1188/Rjtmp_001254 (proline dehydrogenase/*putA*), Rv1523/RJtmp_001610 (methyltransferase) and Rv1524/RJtmp_001611 (glycosyltransferase), and Rv3800/RJtmp_003912 (*pks13*) and Rv3801/RJtmp_003913 (*fadD32*). Among these genes, *pruA* and *putA* are involved in arginine & proline metabolism along with the *pip* gene and *fadD32* is involved in nearly 47 different GOs.

Discussion

The fact that members of the MTBC exhibit non-specific host preferences have now been well established, as demonstrated by the isolation of *M. tuberculosis* from cattle and the identification of *M. bovis* and *M. orygis* in humans. The concept of host tropism within MTBC species reflects the interplay between host-driven and pathogen-driven processes, where variations in the host immune response and alterations in MTBC phenotypes contribute to differences in infection, persistence, disease development, and transmission.[29] The adaptation of the pathogen to the host is a highly complex process, as the mechanisms underlying the transition from latent TB infection (LTBI) to active TB disease are still not fully understood. In this context, the MTBC species possess a distinctive insertion sequence called IS*6110*, which belongs to the IS3 family and plays a significant role in transposition. The transposition of IS*6110* follows a copy-out-paste-in mechanism.[30-32] This implicit trait offers an outstanding chromosomal polymorphism to study TB outbreaks and its unquestionable role as a clinical epidemiological marker.

The perseverance of this insertion element in MTBC might also steer phenotypic changes and affect the fitness of the strain. Various reports claim these transposition events are instigated by external stress conditions leading to genetic variability.[33,34] IS*6110* insertions are well known to disturb gene expression by intruding protein-coding genes, by facilitating recombination events resulting in deletions and inversions, or by up-regulating the expression of adjacent genes due to its presence within the promoter region.[18,33,35,36] Recently Charles et al[19] have made use of the WGS approach to study the abundance and chromosomal distribution of IS*6110* copy on *M. bovis* genomic data of French animal field strains and stated that these insertions seem to be stable within specific genotypes over time and between host species, signifying that the transposition of IS*6110* is not an evolutionary driver for modern French *M. bovis* strains at least over a 15-year period. The effect of IS*6110* on genome function prompted us to take a deeper look at the distribution and patterns of IS*6110* insertions among the recently circulating strain *M. orygis*, the emerging pathogen in India.

M. tuberculosis as a reference genome has been well utilized and is also well known to possess 16 IS*6110* insertions and all the possible regions of insertions have been vastly explored.[37] Meanwhile, in the case of *M. orygis*, van Ingen et al had previously identified these strains as high copy strains possessing 17 to 20 IS*6110* copies. However, the exact regions of insertions remained unknown.[9] Here, we identified 25 IS*6110* insertions in the reference strain and any additional insertions other than the ones found in the reference were identified as novel insertions.

We found that most of the strains in this study were high copy strains, but also identified strains with no copy (SRR10321152, SRR21691066, SRR21842108) and low copy (SRR10321141, SRR10321142, SRR10321138) IS*6110* insertions although insignificant. No copy and low copy strains were found only in the human cluster, whereas the animal cluster consisted of only high copy strains. Among all the isolates, KL036 possessed the highest number of IS*6110* insertions with 42 insertions against CP063804 and 54 against NC000962.3. Most novel insertions were highly confined to a single strain or found in 2 or 3 strains in both human and animal clusters, indicating the potential role of IS*6110* transposition.

The 25 known IS*6110* were present in all the high-copy strains but were found to be localized in different positions and orientations within the same region and these insertions were mostly intergenic. Nearly all the high copy strains possessed IS*6110* insertions downstream to the genes *mosR, lipx, pknH, EccA5, MgtC*/*sapB* family protein, and ExeA family protein. A few other

genes include oxidoreductase, phospholipase, transposase and hypothetical proteins. We found 2 genes for each of *mmpL1* (RJtmp_000420 (912 bp), RJtmp_000422 (1971 bp)), *lipX* (RJtmp_001232 (177 bp), RJtmp_001234 (129 bp)) and *mmpL10* (RJtmp_001247 (1743 bp), RJtmp_001249 (1269 bp)) in *M. orygis* genome compared to Rv0402c (2877 bp), Rv1169c (303 bp), and Rv1183 (3009 bp), respectively in *M.tuberculosis* genome. A closer observation showed the presence of IS*6110* insertion sequence RJtmp_000421, RJtmp_001233, and RJtmp_001248 between these genes, which leads to the speculation that the existence of 2 genes might be due to the transposition of the insertion element. Most novel and unique insertion sequences were found within the hypothetical protein, PPE/PE family proteins or PPE-domain-containing proteins. Few transcriptional regulators, such as *marR*, *gntR*, *LuxR*, and *TetR* had IS*6110* insertions upstream of the gene but were confined to a single strain. We found no insertion sequences upstream of the previously reported genes *PhoP*, *esxS*, or *ctpD*. However, we found an insertion element in the upstream region of *dnaN* in 1 strain from the human cluster.[18]

Functional annotation by Gene ontology and KEGG pathway analysis identified numerous functions and pathways in which the genes affected by IS*6110* were involved. However, only 7 insertion sequences were found in all these isolates, whereas the rest were only present in a single or a couple of strains. The effect of the insertion sequence between *pknH* and *embR* is unknown. However, the genes *pknH* and *embR* are non-essential regulatory proteins found to be involved in various processes, including biological processes, cellular components and molecular function by GO analysis. Similarly, another insertion sequence existing between *purU* and *coaD*/*kdtB* was also observed in all the isolates, and the gene *kdtB* is involved in the cell wall and cell process.

In the case of the novel and unique insertions, none of the loci were found among our isolates except for Rv2357 (*glyS*) and Rv1368/RJtmp_001450 (*lprF*). The former was found in 6 isolates, 3 from each cluster including 2 study isolates (KL043 and KL036) and the latter was found in 7 isolates in the animal cluster, including 3 study isolates (KL004, KL115, and AH144). Another insertion loci within Rv0804/RJtmp_000891 (*pip*) was found in all 11 of our study isolates and 3 more in the animal cluster. The gene *glyS* known as glycyl-tRNA synthetase, catalyzes the synthesis of glycyl-tRNA, which is required to insert glycine into protein.[38] The role of *LprF* in the tuberculosis pathway is unclear, and we speculate its possibility to display TLR2 agonist activity as most of the Lipoproteins like LprG, LprA and LpqH are well known for the same.[39-41] Similarly, in the arginine and proline metabolism, proline iminopeptidase (Pip) explicitly catalyzes the removal of N-terminal proline residues from peptides.[42] Gamma-glutamyl phosphate reductase (proA) catalyzes the NADPH-specific reduction of L-gamma-glutamyl phosphate into L-glutamate-5 semialdehyde, which spontaneously cyclizes to form pyrroline-5-carboxylate.[43] This intermediate product of proline biosynthesis is converted into glutamate by pyrroline-5-carboxylate dehydrogenase (PruA/RocA).[44]

To the best of our knowledge, this is the first study to report the identification and localization of IS*6110* insertion sites in the *M. orygis* genome. Some of the insertion sites identified in this study have previously been recognized as insertion hotspots of MTBC. Only a few isolates in this study had IS*6110* insertions in the hotspots region like Rv1758 (*cut1*), Rv1777 (*cyp144*), Rv3183, Rv3327, Rv1371, Rv1765c.[45,46] Nearly all the isolates had an insertion element in Rv1682 and Rv1169, the hotspot regions for L5 and L6 lineages of MTBC, respectively.[46] Gene interruption or regulation of genes in these hotspot regions might differ depending on the orientations and/or insertion sites of IS*6110* in the same genomic region. Fewer studies have suggested that the location of IS*6110* in the same transcriptional orientation and not more than 400 bp upstream of a gene may lead to upregulation of the gene.[6,45] Although we have also detected similar insertions in our analysis, it warrants an additional *in vivo* or *in vitro* study to assess their role in virulence and other survival traits. The study is the beginning, and we presume that the genetic changes observed due to IS*6110* transposition could be partly responsible for the fitness of *M. orygis* in establishing infection in various hosts. At the same time, other explanations could also support this success. The sequences used in this study were limited to fewer numbers and most of the sequences obtained were from India. An in-depth analysis involving more sequences from various other geographic locations will provide a clear understanding regarding the fitness of the organism. Further studies are warranted to correlate the effect of IS*6110* transposition events to the successful adaptation of the pathogen.

Conclusion

In conclusion, our findings provide novel information about IS*6110* insertions and the preferential insertion regions in the *M. orygis* genome. We have also elucidated the genomic impact of these insertions and their effect on various metabolic processes and pathways suggestive of their pertinence. Henceforward, as and when MTBC strains are being vastly sequenced, it is ideal to locate the IS*6110* insertion sites, which will provide a better understanding regarding its biological role in the organism's adaptation, evolution, and fitness.

Author Contributions

Ahmed Kabir Refaya: Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Umashankar Vetrivel**: Investigation, Resources, Writing – review & editing. **Kannan Palaniyandi**: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. The authors have read and approved the manuscript.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Thierry D, Cave MD, Eisenach KD, et al. IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Res*. 1990;18:188.
2. Brisson-Noël A, Nguyen S, Aznar C, et al. Diagnosis of tuberculosis by DNA amplification in clinical practice evaluation. *Lancet*. 1991;338:364-366.
3. Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco – a population-based study using conventional and molecular methods. *New Engl J Med*. 1994;330:1703-1709.
4. Gonzalo-Asensio J, Pérez I, Aguiló N, et al. New insights into the transposition mechanisms of IS*6110* and its dynamic distribution between *Mycobacterium tuberculosis* complex lineages. *PLoS Genet*. 2018;14:e1007282.
5. McEvoy CR, Falmer AA, Gey van Pittius NC, et al. The role of IS*6110* in the evolution of *Mycobacterium tuberculosis*. *Tuberculosis*. 2007;87:393-404.
6. Safi H, Barnes PF, Lakey DL, et al. IS*6110* functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. *Mol Microbiol*. 2004;52:999-1012.
7. Fomukong N, Beggs M, el Hajj H, et al. Differences in the prevalence of IS*6110* insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS*6110*. *Tuber Lung Dis*. 1997;78:109-116.
8. Kremer K, Glynn JR, Lillebaek T, et al. Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. *J Clin Microbiol*. 2004;42:4040-4049.
9. van Ingen J, Rahim Z, Mulder A, et al. Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerg Infect Dis*. 2012;18:653-655.
10. Rahim Z, Thapa J, Fukushima Y, et al. Tuberculosis caused by *Mycobacterium orygis* in dairy cattle and captured monkeys in Bangladesh: a new scenario of tuberculosis in South Asia. *Transbound Emerg Dis*. 2017;64:1965-1969.
11. Dawson KL, Bell A, Kawakami RP, et al. Transmission of *Mycobacterium orygis* (*M. Tuberculosis* complex species) from a tuberculosis patient to a dairy cow in New Zealand. *J Clin Microbiol*. 2012;50:3136-3138.
12. Thapa J, Paudel S, Sadaula A, et al. *Mycobacterium orygis*-associated tuberculosis in free-ranging rhinoceros, Nepal, 2015. *Emerg Infect Dis*. 2016;22:570-572.
13. Refaya AK, Ramanujam H, Ramalingam M, et al. Tuberculosis caused by *Mycobacterium orygis* in wild ungulates in Chennai, South India. *Transbound Emerg Dis*. 2022;69:e3327-e3333.
14. Refaya AK, Kumar N, Raj D, et al. Whole-genome sequencing of a *Mycobacterium orygis* strain isolated from cattle in Chennai, India. *Microbiol Resour Announc*. 2019;8:1-3.
15. Marcos LA, Spitzer ED, Mahapatra R, et al. *Mycobacterium orygis* lymphadenitis in New York, USA. *Emerg Infect Dis*. 2017;23:1749-1751.
16. Islam MR, Sharma MK, KhunKhun R, et al. Whole genome sequencing-based identification of human tuberculosis caused by animal-lineage *Mycobacterium orygis*. *J Clin Microbiol*. 2023;61:1-15.
17. Duffy SC, Srinivasan S, Schilling MA, et al. Reconsidering *Mycobacterium bovis* as a proxy for zoonotic tuberculosis: a molecular epidemiological surveillance study. *Lancet Microbe*. 2020;1:e66-e73.
18. Xiong X, Wang R, Deng D, et al. Comparative genomics of a bovine *Mycobacterium tuberculosis* isolate and other strains reveals its potential mechanism of bovine adaptation. *Front Microbiol*. 2017;8:2500-2512.
19. Charles C, Conde C, Biet F, Boschiroli ML, Michelet L. IS*6110* copy number in multi-host *Mycobacterium bovis* strains circulating in bovine tuberculosis endemic French regions. *Front Microbiol*. 2022;13:891902-891909.
20. Rufai SB, McIntosh F, Poojary I, et al. Complete genome sequence of *Mycobacterium orygis* strain 51145. *Microbiol Resour Announc*. 2021;10:2.
21. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312-1313.
22. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47:W256-W259.
23. Hawkey J, Hamidian M, Wick RR, et al. ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC Genomics*. 2015;16:1-11.
24. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24-26.
25. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. *Cancer Res*. 2017;77:e31-e34.
26. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38:5825-5829.
27. Huerta-Cepas J, Szklarczyk D, Heller D, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47:D309-D314.
28. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18:366-368.
29. Malone KM, Gordon SV. *Mycobacterium tuberculosis* complex members adapted to wild and domestic animals. *Adv Exp Med Biol*. 2017;1019:135-154.
30. Polard P, Chandler M. An in vivo transposase-catalyzed single-stranded DNA circularization reaction. *Genes Dev*. 1995;9:2846-2858.
31. Ton-Hoang B, Polard P, Chandler M. Efficient transposition of IS911 circles in vitro. *EMBO J*. 1998;17:1169-1181.
32. Duval-Valentin G, Marty-Cointin B, Chandler M. Requirement of IS911 replication before integration defines a new bacterial transposition pathway. *EMBO J*. 2004;23:3897-3906.
33. Cubillos-Ruiz A, Morales J, Zambrano MM. Analysis of the genetic variation in *Mycobacterium tuberculosis* strains by multiple genome alignments. *BMC Res Notes*. 2008;1:110.
34. Ghanekar K, McBride A, Dellagostin O, et al. Stimulation of transposition of the *Mycobacterium tuberculosis* insertion sequence IS*6110* by exposure to a microaerobic environment. *Mol Microbiol*. 1999;33:982-993.
35. Soto CY, Menéndez MC, Pérez E, et al. Iss *6110* mediates increased transcription of the *phop*p virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. *J Clin Microbiol*. 2004;42:212-219.
36. Warren RM, Sampson SL, Richardson M, et al. Mapping of IS*6110* flanking regions in clinical isolates of *Mycobacterium tuberculosis* demonstrates genome plasticity. *Mol Microbiol*. 2000;37:1405-1416.
37. Ioerger TR, Feng Y, Ganesula K, et al. Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories. *J Bacteriol*. 2010;192:3645-3653.
38. Freist W, Logan DT, Gauss DH. Glycyl-trna synthetase. *Biol Chem Hoppe Seyler*. 1996;377:343-356.
39. Drage MG, Tsai HC, Pecora ND, et al. *Mycobacterium tuberculosis* lipoprotein LprG (Rv1411c) binds triacylated glycolipid agonists of toll-like receptor 2. *Nat Struct Mol Biol*. 2010;17:1088-1095.
40. Pecora ND, Gehring AJ, Canaday DH, Boom WH, Harding CV. *Mycobacterium tuberculosis* LprA is a lipoprotein agonist of TLR2 that regulates innate immunity and APC function. *J Immunol*. 2006;177:422-429.
41. Pai RK, Pennini ME, Tobian AA, et al. Prolonged toll-like receptor signaling by *Mycobacterium tuberculosis* and its 19-kilodalton lipoprotein inhibits gamma interferon-induced regulation of selected genes in macrophages. *Infect Immun*. 2004;72:6603-6614.
42. Cunningham DF, O'Connor B. Proline specific peptidases. *Biochim Biophys Acta Protein Struct Mol Enzymol*. 1997;1343:160-186.
43. Seddon AP, Zhao KY, Meister A. Activation of glutamate by gamma-glutamate kinase: formation of gamma-cis-cycloglutamyl phosphate, an analog of gamma-glutamyl phosphate. *J Biol Chem*. 1989;264:11326-11335.
44. Qamar A, Mysore KS, Senthil-Kumar M. Role of proline and pyrroline-5-carboxylate metabolism in plant defense against invading pathogens. *Front Plant Sci*. 2015;6:503-509.
45. Alonso H, Samper S, Martín C, Otal I. Mapping IS*6110* in high-copy number *Mycobacterium tuberculosis* strains shows specific insertion points in the Beijing genotype. *BMC Genomics*. 2013;14:422.
46. Roychowdhury T, Mandal S, Bhattacharya A. Analysis of IS*6110* insertion sites provide a glimpse into genome evolution of *Mycobacterium tuberculosis*. *Sci Rep*. 2015;5:12567.