JJBS

Volume 18, Number 3,September 2025
ISSN 1995-6673
Pages 411 – 419

Jordan Journal of Biological Sciences

Analysis of EsxH Gene Mutations and Homology Modeling in *Mycobacterium tuberculosis* from Clinical Isolates in South Sulawesi

Ummi Chaera[1], Muhammad Nasrum Massi[2*], Doddy Irawan Setyo Utomo[3], Astutiati Nurhasanah[3], Nihayatul Karimah[3], Israini W. Iskandar[1], Najdah Hidayah[4], Firdaus Hamid[2], Yunialthy D. Pertiwi[2], Fadhilah[2],

[1]*Master Program of Biomedical Science Graduate School, Hasanuddin University, Makassar, 90245, South Sulawesi, Indonesia;* [2]*Department of Microbiology, Faculty of Medicine, Hasanuddin University, Makassar, 90245, South Sulawesi, Indonesia;* [3]*Research Center for Vaccines and Drugs, Health Research Organization, National Research and Innovation Agency (BRIN), Banten 15314, Indonesia;* [4]*Graduate School, Hasanuddin University, Makassar, 90245, South Sulawesi, Indonesia.*

Received: August 3, 2024; Revised: January 13, 2024; Accepted: February 2, 2025

Abstract

Introduction: Tuberculosis (TB) is a significant global public health issue, with Indonesia ranking second worldwide for the highest incidence. Despite the widespread use of the Bacillus Calmette-Guérin (BCG) vaccine, its efficacy remains limited, particularly in preventing the reactivation of latent TB infections in adults. Moreover, *Mycobacterium tuberculosis* (*Mtb*), the causative agent of TB, exhibits considerable genetic diversity, which can influence the virulence, transmissibility, and resistance to treatment. The EsxH gene, part of the ESX-3 secretion system, plays a crucial role in the virulence and immune evasion mechanisms of *Mtb*. Methods and Materials: A total of 40 *Mtb*-positive isolates from South Sulawesi, Indonesia, were used. Samples were analyzed using molecular detection techniques in accordance with the standard procedures. The EsxH gene was amplified using specific primers via polymerase chain reaction (PCR), purified, and Sanger sequencing. Nucleotide variations were identified using Unipro UGENE, and phylogenetic analysis was conducted to explore the evolutionary relationship as well as the uniqueness of mutations. Additionally, the structural impact of the identified mutations was analyzed using the SWISS-MODEL server by comparing the three-dimensional (3D) models of the EsxH protein with and without mutations to assess any significant structural changes.

Results: The study identified unique mutations in the EsxH gene of *Mtb* isolates from South Sulawesi, specifically at nucleotide positions 28 G>A and 29 C>T, leading to amino acid changes at position 10 (alanine to threonine and alanine to valine). Phylogenetic analysis indicated that these mutations were distinct from the region, while 3D homology modeling showed no significant structural changes. Conclusion: This study contributes to the understanding of *M. tuberculosis* genetic diversity in South Sulawesi by identifying region-specific mutations in the EsxH gene. While these mutations did not result in major structural alterations to the EsxH protein, further bioinformatic and in vitro studies are needed to explore their functional implications. These findings underscore the importance of continued surveillance and research to improve TB control strategies, as well as to inform the development of more effective vaccines and treatments.

Keywords: *Mycobacterium tuberculosis*, genetic diversity, mutation, EsxH, Homology modeling

1. Introduction

Tuberculosis (TB) is considered a highly significant global emergency. According to World Health Organization (WHO) data in 2021, an estimated 10.6 million people worldwide suffer from the disease (WHO, 2022). Indonesia ranked second among countries with the highest incidence of tuberculosis in the world after India in 2019. The 2021 Global TB Report also reported 824,000 TB cases (Ministry of Health, 2022). Patients with TB infection have a 5-10% lifetime risk of developing the disease, which increases in various immunodeficiency states to a 16% annual risk (Gill et al, 2022).

The most effective approach to controlling TB in humans entails administering the *Bacillus Calmette-Guérin* (BCG) vaccine. It triggers a strong Th1 immune response and supports the formation of T cells specific to mycobacteria, which are both polyfunctional and cytotoxic. However, BCG vaccine is known to be ineffective in protecting adults and may reactivate latent TB infection (Singh et al., 2022)

The infection process of *Mycobacterium tuberculosis* (*Mtb*) relies on the capacity to interact with the immune system in various ways, including evading the innate immune response, enduring the adaptive immune response without leading to symptomatic illness, and inducing disease. A strong inflammatory reaction leads to significant tissue damage, which facilitates transmission (Chandra et al., 2022)

Mycobacteria use a type VII secretion system (T7S) to transport proteins across intricate cell envelopes, of which

* Corresponding author. e-mail: nasrumm@unhas.ac.id.

there are more than five excretion systems from ESX-1 to ESX-5 known to be associated with *Mtb* host-pathogen interactions. This secreted protein is one of the immunodominant antigens best known to the human immune system. Several antigens that have been identified include the Esx family proteins, including EsxA, EsxB, EsxG, EsxH, EsxG, and EsxN. The ESX family proteins, TB10.4 (EsxH), have been found to possess similar characteristics in terms of immune recognition during infection and protective efficacy in TB vaccines. EsxH is known to disrupt the recognition of infected macrophages by CD4+ T cells. The *EsxH* gene encodes the TB10.4 protein, which elicits CD4 and CD8 T cell responses in both humans and mice. It is part of a vaccine that is currently in clinical trials (Darrah et al., 2019).

Studies on mutational variations in the EsxH protein remain limited. The protective efficacy of EsxH subunit vaccines might be compromised on the global stage. In the research conducted by Davilla et al. (2010), no DNA variations were found in 88 samples tested, meaning the EsxH gene is highly conserved. Therefore, the results suggest that the efficacy of TB10.4 vaccines is unlikely to be affected by the genetic diversity of the *M. tuberculosis* population. However, further research is needed that includes a broader range of *M. tuberculosis* strains to validate this finding.

Building upon this, our research aimed to identify sequence variations in the EsxH antigen coding gene in clinical isolates of *Mtb* from South Sulawesi, Indonesia. The potential ability of these mutations to affect amino acid sequences was also evaluated. In addition, phylogenetic analysis was carried out to determine the evolutionary relationships between *Mtb* strains and genetic diversity in certain populations or species. This study used a simple in silico method with a three-dimensional (3D) structural homology model of the EsxH protein through the SWISS-MODEL server. Information about the 3D structure of proteins can facilitate the development of structure-based studies, including the nature and structure, mechanisms of interaction with other compounds through molecular docking, as well as virtual drug screening related to the EsxH protein.

Information regarding the EsxH gene mutations in the *Mtb* population in Indonesia is important to be further investigated, as it can serve as a foundation for consideration in designing an appropriate TB vaccine. We expected to provide valuable insights into the genetic diversity of *Mycobacterium tuberculosis* (*Mtb*) and identify the potential of EsxH as a key target in the development of innovative TB vaccines, by integrating genomic, structural, and phylogenetic data.

2. Methods and Materials

2.1. Genomic DNA Cultivation and Isolation

This study used 40 clinical samples obtained from TB patients at the HUMR laboratory. The aim of isolating *Mtb* genomic DNA was to obtain the EsxH gene. The *Mtb* culture was then used for genome isolation using the gSYNC DNA extraction kit.

2.2. DNA Purification

After genome extraction, the sample was purified using the Sbeadex Pathogen Nucleic Acid Purification Kit. A 100 µl aliquot was combined with 20 µl of Sbeadex particles and 160 µl of binding buffer, then incubated at room temperature with constant shaking for 5 minutes. A magnet was used for 2 minutes to separate the supernatant. Following this, 400 µl of wash buffer BN1 was added, incubated for 5 minutes with shaking, and separated with the magnet. The sample was further washed with 400 µl each of wash buffers TN1 and TN2. Finally, 100 µl of elution buffer was added, incubated for 10 minutes at 60°C with shaking and placed near the magnet for 3 minutes before transferring the eluate to a new tube.

2.3. Amplification of the EsxH Gene of Mycobacterium tuberculosis using PCR.

Mtb DNA amplification was carried out using a PCR mixture that included 6.25 µl of redmix enzyme, 0.5µl of each forward and reverse primers, 2.75µl of ddH2O, as well as 2.5µl of DNA sample. The negative control used nuclease-free water, while the primers used were EscGH-F-Sukkhu (5' GACCGCAACCAAAGAAC-3') and EscGH-R-Sukkhu (5' CCAGCACCCACGGAAAG-3'), targeting a 951 bp region. The PCR procedure consisted of 35 cycles with initial denaturation at 95°C for 1 minute, followed by denaturation at 95°C for 15 seconds, annealing at 55°C for 15 seconds, extension at 72°C for 20 seconds, and a final extension at 72°C for 2 minutes before cooling to 4°C.

2.4. Electrophoresis

Agarose gel electrophoresis was carried out using a 1% agarose gel, prepared by dissolving 1 g of agarose in 50 ml of 0.5% TBE buffer, heating in the microwave for 1 minute, followed by cooling for 30 minutes. About 4µl of SybrSafe DNA Gel Stain was added, mixed, and poured into a gel mold. The gel was left to solidify before being placed in an electrophoresis chamber with 0.5% TBE buffer. Subsequently, wells were loaded with 4 µl of negative control, 2.5µl of DNA marker, and 4µl of the DNA sample. Electrophoresis was then performed at 80 volts for 60 minutes. The results were analyzed using a 100 bp and 1 kb gene ruler, which confirmed the presence of the expected 951 bp amplified gene fragments for EsxH.

2.5. Sanger Sequencing

Sequencing of PCR results was carried out using the Sanger DNA sequencing method. Sample preparation started with DNA Polymerase, Primer, dNTPs, and ddNTPs which were added to the amplified DNA sample. This was followed by heating to a temperature of 90˚C to trigger DNA denaturation and polymerization. Subsequently, the sample was placed into an electrophoresis machine and the waves captured by the detector were analyzed.

2.6. Phylogenetic analysis

The phylogenetic analysis used Mega Software, running in 500 and 1000 bootstrap to observe the comparison between a mutation in this study and other lineages from NCBI.

2.7. Homology analysis using Swiss Model

Homology modeling is a method used to predict the 3D structure of proteins by comparing and aligning the amino acid sequence with others. This sequence information was used to generate a 3D protein structure model using the SWISS-MODEL server (https://swissmodel.expasy.org/).

Protein sequences were transferred from the Unipro UGENE server to SWISS-MODEL in FASTA format.

3. Results

3.1. Gene Amplification

The results of amplification using Polymerase Chain Reaction (PCR) to detect the presence of EsxH genes were read using agar gel electrophoresis which showed the amplified gene fragment with a target size of 951 bp for 40 samples. As shown in Figure 1, the amplification results showed a clear DNA band in each sample, with a corresponding target size of 951 bp. The results indicate that the sample is ready to proceed to the sequencing stage.

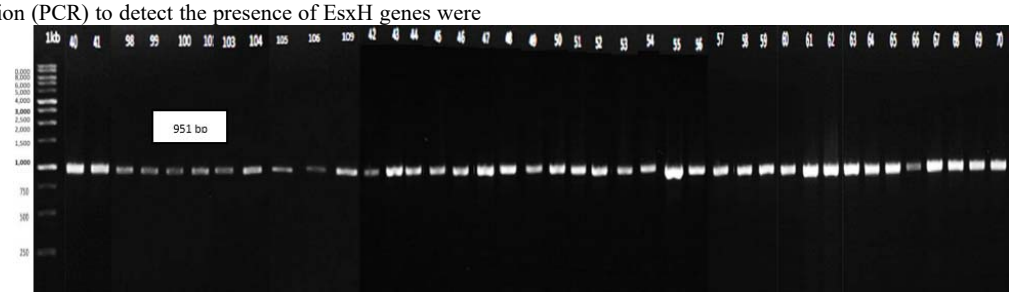

Figure 1. Gene electrophoresis in 40 clinical isolates of *Mtb* EsxH

3.2. Mutation Analysis of Sequencing Sequence Results

After obtaining the amplification results, Sanger sequencing was carried out on 40 samples. The alignment results used the Unipro UGENE application and several DNA variation points were found in the EsxH target gene. As shown in Figure 2, two mutations were found in DNA with sample codes 11X and 30X. These mutations correspond to variations in the EsxH gene with a base change from G to A (Guanine to Adenine) in sample 11X at nucleotide position 28 as well as base C to T (Cytosine to Thymine) in nucleotide position 29 in the 30X sample with a percentage of 2.4% each. The changes in the two samples affected substitutions at the amino acid level, where alterations were found at the 10th amino acid position from Alanine to Threonine in sample code 11X and Alanine to Valine in sample code 30X. (Figure 3)

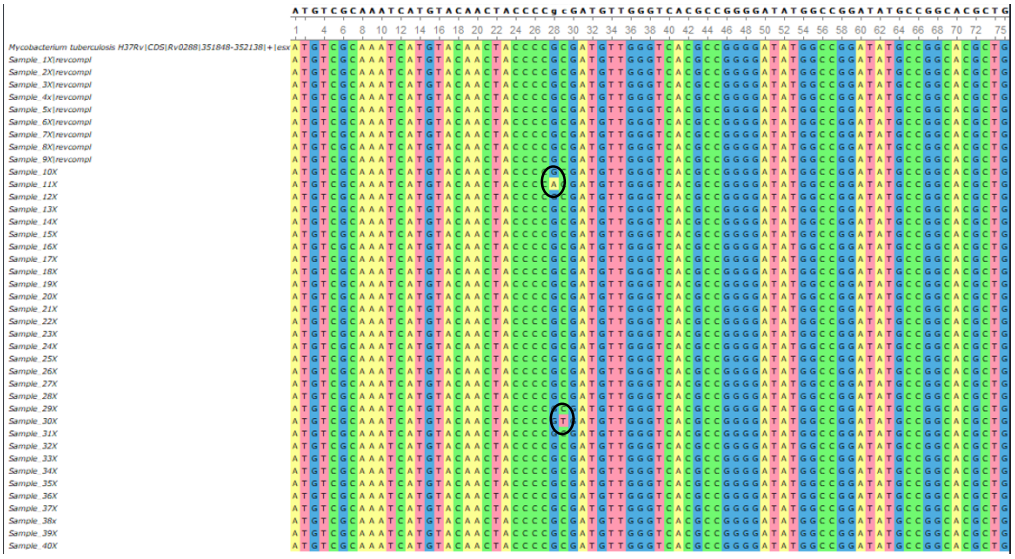

Figure 2. Substitution of nucleotide codon G>A at position 28 and C>T at position 29 of the EsxH gene

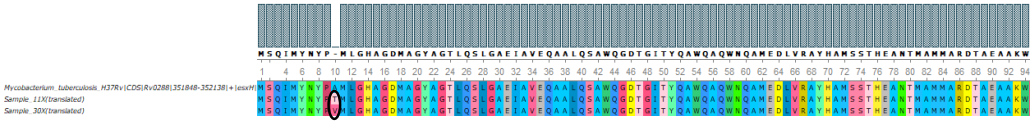

Figure 3. Amino acid substitution at position 10 from Alanine to Threonine with code sample 11X and sample Alanine to Valine sample 30X

The overall variations in the target genes are shown in Table 1. The existence of several variations in these target genes indicates that there are several sequence variations in the *Mtb* clinical isolates.

Table 1. Mutation analysis results of EsxH

Isolate code	H37Rv Nucleotide	Isolate Nucleotide	Mutation sites	Amino acid Change	Percentage
Sample 11X	G	A	28G>A	Ala10Thr	2.4%
Sample 30X	C	Q	29C>T	Ala10Val	2.4%

Ala: Alanine; Thr: Threonine; Val: Valine

3.3. Phylogenetic

In this study, clinical samples of *M. tuberculosis* isolates from around the city of Makassar were screened to obtain genetic information that can be used as reference material for developing TB treatment. A total of 40 samples were obtained for each EsxH gene. The sequencing results showed that the target gene has variations at certain positions, observed phylogenetically to determine the evolutionary relationship. A phylogenetic tree was constructed based on the genetic sequence of the EsxH gene. In this analysis, sequences from clinical samples were compared with reference sequences from strain H37Rv to identify patterns of divergence and evolutionary clusters.

The phylogenetic tree was carried out using the Maximum likelihood (ML) method to ensure consistency of results. The sequences were aligned using the Muscle and ClustalW programs, then a phylogenetic tree was generated with MEGA (Molecular Evolutionary Genetics Analysis) software version 11.0.13. Additionally, bootstrap analysis was performed with 500 and 1000 replications.

Before analyzing the phylogenetic tree, this study first identified the lineage for *Mtb*. Sequences belonging to this lineage were also included in the alignment stage together with samples and reference H37Rv. The data and reference lineage sources for *Mtb* are from the NCBI database.

- L1_1_3→>NZ_CP054014.1 *Mycobacterium tuberculosis* strain FDAARGOS_756 chromosome, complete genome

  (https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP054014.1)
- L1→>NZ_CP040689.1 *Mycobacterium tuberculosis* complex sp. N0072 chromosome, complete genome (https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP040689.1)
- L2→>CP005082.1 *Mycobacterium tuberculosis* str. Beijing/NITR203, complete genome (https://www.ncbi.nlm.nih.gov/nuccore/CP005082.1)
- L3→>NZ_CP028428.1 *Mycobacterium tuberculosis* strain CAS chromosome (https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP028428.1)
- L4→>NC_022350.1 *Mycobacterium tuberculosis* str. Haarlem, complete sequence (https://www.ncbi.nlm.nih.gov/nuccore/NZ_JLDE00000000.1)
- L5→>KK338837.1 Mycobacterium africanum MAL010074 whole genome sequence (https://www.ncbi.nlm.nih.gov/nuccore/KK338837.1)
- L6→>CP010334.1 *Mycobacterium tuberculosis* variant africanum strain 25, complete genome (https://www.ncbi.nlm.nih.gov/nuccore/CP010334.1)

The alignment results for the EsxH gene (Figure 4) showed that, across all lineages, no variation sites were identical to those observed in the 11X and 30X samples at the nucleotide or amino acid level. Furthermore, in the phylogenetic tree (Figure 4.c) with 500 bootstraps, samples "11X" and "30X" were on the same branch as "CP010334.1 africanum strain 25". The bootstrap value for the branch connecting “11X” and “30X” was 37, indicating a relatively low level of confidence in this relationship. In the tree with 1000 bootstraps (Figure 4.d), samples “11X” and “30X” were also on the same branch as “CP010334.1 africanum strain 25”. The bootstrap value for the branch connecting “11X” and “30X” was 58, indicating increased confidence in this relationship compared to a tree with 500 bootstraps.

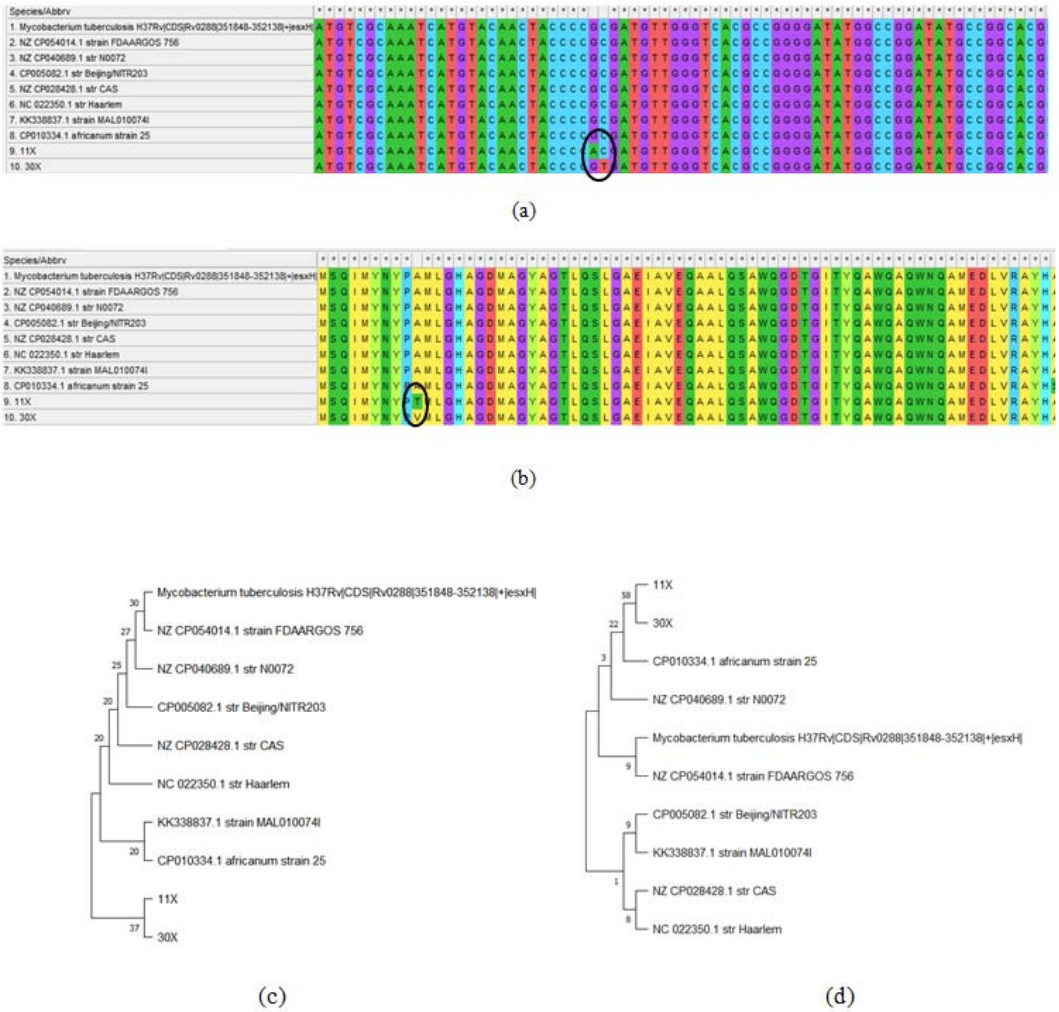

Figure 4. Alignment results of the EsxH gene with the reference gene at the variation point (a) Nucleotide positions 28 and 29 (b) Amino acid position 10 (c) Maximum likelihood phylogenetic tree 500 bootstrap (d) Maximum-likelihood phylogenetic tree 1000 bootstrap.

3.4. Homology Modelling using SWISS-MODEL

The 3D structure of proteins largely determines molecular biochemical properties and functions. In this study, analysis of changes in protein structure was carried out to understand how the mutations identified in the EsxH gene affect the 3D structure. Homology modeling is a technique used to predict the 3D structure of proteins by aligning the amino acid sequence with those of other proteins. Generally, a protein with a recognized 3D structure is called a template. Proteins whose amino acid sequences are related have similar structures. This sequence data was used to create a 3D protein structure model using the SWISS-MODEL server. Copying protein sequences from the Unipro UGENE server to the SWISS-MODEL server was carried out using the FASTA format as in Table 2.

Table 2. Protein target sequence of Esxh

EsxH	MSQIMYNYPAMLGHAGDMAGYAGTLQSLGAEIAVEQAALQSAWQGDTGITYQAWQAQWNQAMEDLVRAYHAMSSTH
Reference	EANTMAMMARDTA EAAKWGGX
11X	MSQIMYNYPTMLGHAGDMAGYAGTLQSLGAEIAVEQAALQSAWQGDTGITYQAWQAQWNQAMEDLVRAYHAMSSTHE ANTMAMMARDTAEAAKWGG
30X	MSQIMYNYPVMLGHAGDMAGYAGTLQSLGAEIAVEQAALQSAWQGDTGITYQAWQAQWNQAMEDLVRAYHAMSSTH EANTMAMMARDTAEAAKWGG

The target protein sequence of EsxH, provided in FASTA format as shown in Figure 1 was submitted to the SWISS-MODEL server (https://swissmodel.expasy.org/). The server then automatically constructed the model using the suitable template. Subsequently, three types of structure models appeared, and the structure with the

coverage value closest to the reference template was selected. In this study, 3D modeling was carried out for samples that had mutations in the EsxH gene. The results showed that model 2 for samples 11X and 30X had the highest coverage among the other two models with the same value, namely 98.96% with the reference template 2kg7.1.BESAT-6-like protein EsxH. (Figure 5)

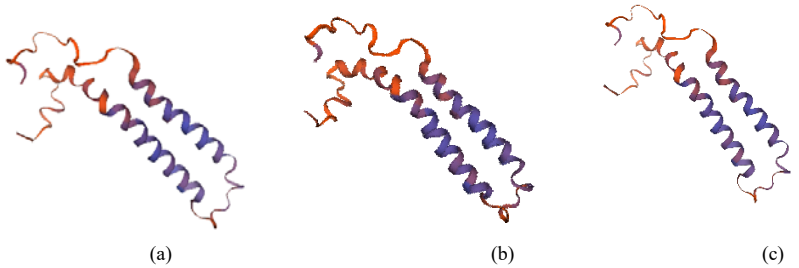

Figure 5. Model structure (a) EsxH reference, (b) 11X (c) 30X

To obtain information regarding the suitability of the model structure that has been obtained, this study also analyzed the protein structure using the Ramachandran Plot to determine the suitability of amino acids position in the protein structure. Figure 6 shows Ramachandran plot results for all EsxH reference protein structures in samples 11X and 30X. The reference structure shows Ramachandran Favored with a value of 89.36% and Ramachandran Outliers. The plot image for each protein structure shows that the amino acids in the protein are in the appropriate area, specifically in the light-dark green zone. The protein structure that has been created in the SWISS-Model has good confidence to be used in the next analysis stage.

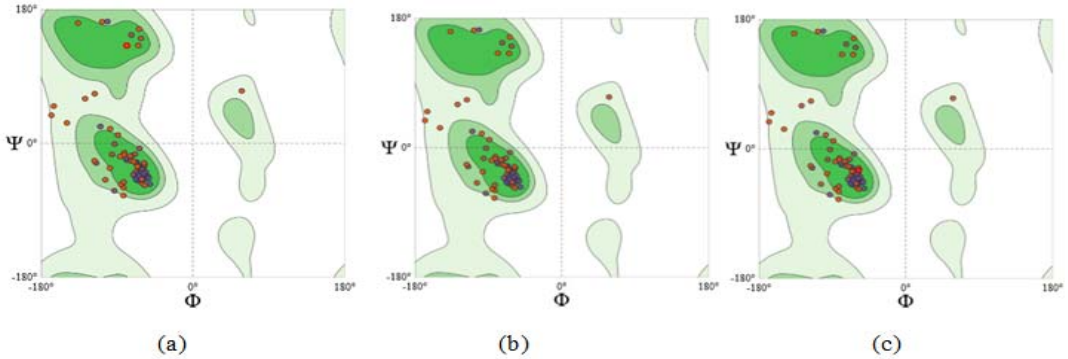

Figure 6. Ramachandran plot on the amino acid protein structure model (a) 11X (b) 30X, (c) Reference EsxH

The variation point in the 10th amino acid sequence did not significantly affect changes in the structure of the protein as a whole when observed from the visualization of the variation position of each amino acid obtained. The first amino acid change of 11X, the reference H37Rv structure is shown in Figure 7.c for the amino acid Alanine (pink).

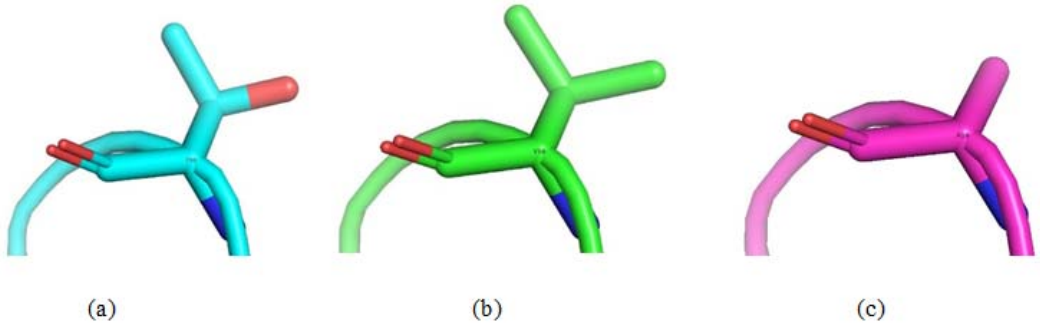

Figure 7. Visualization of the variation position of each amino acid in the protein structure model showing 2 amino acid variation points at (a) 11X (b) 30X, and the reference protein (c) EsxH H37Rv (pink)

4. Discussion

Among the 40 samples isolated, there were nucleotide changes in the EsxH gene with substitutions of G>A (Guanine to Adenine) in the 11X and C>T (Cytosine to Thymine) in the 30X sample with a percentage of 2.4% each. Nucleotide substitution in the two samples was at the amino acid level. The substitution was found at the 10th amino acid position with a change in the amino acid

Alanine to Threonine in the 11X sample and Alanine to Valine in the 30X sample.

Previous studies found that C>T and G>A mutations in clinical isolates of *Mtb* were related to oxidative stress mechanisms. Increased mutation rates were observed in patients with positive smears who showed a higher number of de novo single-nucleotide polymorphisms (SNPs) in *Mtb*, especially C>T mutations, indicating oxidative damage and increased oxidative stress in patients. This study found that increased mutation rate is host-dependent, rather than an inherent characteristic of the *Mtb* strain, linking variation in mutation rates to the host immune response as well as the production and regulation of reactive oxygen species (ROS). Oxidative damage was identified as the main source of mutations, with a significant proportion being C>T transitions. Additionally, *Mtb* samples from HIV-positive hosts had a higher C>T mutation ratio compared to HIV-negative hosts, emphasizing the influence of host immune status on mutation rates(Ly and Liu, 2020)

The balance between ROS and antioxidants may play a significant role in the onset and progression of diseases, particularly those affecting the respiratory system (Crusio and Radeke, 2019). Macrophages are crucial for maintaining tissue homeostasis and responding to pathogenic threats. According to a study, macrophages infected with *Mtb* produce ROS, which might have effects against TB (Shastri et al., 2018). Advancements in antioxidant drug studies could also offer new opportunities to reduce oxidative damage and enhance treatment(Su et al., 2023).

EsxG/EsxH is known to directly interact with host endosomal sorting complexes that are crucial for protein transport, thereby inhibiting phagosome maturation and antigen presentation during *Mtb* infection of macrophages. This gene, secreted through the ESX-3 type VII secretion system, specifically inhibits the endosomal sorting complex required for transport (ESCRT) machinery. Studies on overexpression and loss-of-function show that EsxH impedes the ability of macrophages and dendritic cells to activate CD4+ T cells specific to *Mtb* antigens. The EsxH protein, part of the ESX-3 cluster, plays a critical role in bacteria survival by regulating iron and zinc homeostasis (Portal-Celhay et al., 2016) (Ates et al., 2016)

The exact function of EsxH clusters is not yet completely understood, but studies indicate that EsxG and EsxH may be crucial in mediating interactions with host cells and assisting in nutrient acquisition. The complex is implicated in metal uptake, drug resistance, and evasion of the immune response, making it a promising target for targeted vaccine development (Martinez-Olivares, 2023)

Based on a previous study (Sutiwisesak et al., 2020) *Mtb* natural polymorphism in the EsxH gene disrupts immunodominance by TB10.4-specific CD8 T cell responses. Mice infected with the *Mtb* strain showed changes, where the response of CD8 T cells which normally recognize and fight bacteria was altered. The TB10.44-11 epitope, typically recognized by CD8 T cells, was no longer detected following infection with a strain that has A10T mutation. In contrast, the response to other parts of the bacteria, previously thought to be less strong than the primary response, was increased. This result is quite interesting because although both versions of the TB10.44-11 epitope (A10T and Wild Type versions) tested have the same ability to bind KB proteins and stimulate T cells, the changes affect how the immune response to *Mtb* infection occurs.

Polymorphism A10V in the esxH gene of *Mtb* has been found to play a significant role in immunogenicity. This polymorphism evolved independently three times and was present in one group of four and two unrelated isolates. Although the A10T variant was more abundant, found in 41 isolates, the A10V polymorphism was less. It changes the hierarchy of CD8 T cell responses specific to the antigen, influencing the immune response to *Mtb* infection.

Amino acids are fundamental building blocks in living organisms. Aside from the role in protein synthesis, amino acids are crucial nutrients for host immune cells, aiding in intracellular metabolic processes such as ATP synthesis, nucleotide production, and maintaining redox balance. Amino acids also provide the primary source of nitrogen for *Mtb*, including alanine and valine. Although *Mtb* can absorb all amino acids, it primarily takes up and rapidly metabolizes glutamate (Glu), glutamine (Gln), aspartate (Asp), asparagine (Asn), and alanine (Ala) as the main nitrogen sources (Agapova et al., 2019.) (Amalia et al., 2022).

Alanine dehydrogenase can suppress host immunity, offering valuable insights for the development of effective immunomodulators targeting *Mtb* (Peng et al., 2024). Substitutions from alanine to threonine (A to T), caused by SNPs, occur in various proteins. In some instances, these substitutions lead to the formation of amyloid fibrils or the aggregation of amyloidogenic proteins (Podoly et al., 2010).

Several studies have been conducted on the early functional analysis of EsxH using in silico methods. However, there has been no discussion on the impact of mutations in the EsxH gene on epitope affinity for TB vaccine development. A study by Ramaiah et al. (2019) evaluated the mutation patterns of specific T-cell epitopes in *Mycobacterium tuberculosis* (*Mtb*) accumulated in *Mtb* strains from India. The study identified 64 mutated T-cell epitopes (mTCE) from 79 strains. These mutations involved significant single amino acid changes, most of which affected the binding affinity of the HLA-DR molecule. Mutations that enhanced epitope function exhibited higher HLA-DR binding affinity, while mutations that diminished function resulted in lower affinity. Predictive analysis using IEDB-based tools revealed a strong correlation between changes in binding affinity and the functional impact on immune responses. Functionally, 62% of the tested mTCEs showed significant alterations in IFN-γ and/or IL-2 secretion by CD4 T-cells, with some mutations enhancing the immune response, while others reduced it.

Based on information from identifying variations in genes, a phylogenetic tree was identified to determine the relationship between the variation points found in the sample along with the reference gene H37Rv and several lineages. The phylogenetic tree shows that there were points of variation in the samples tested, but no similar mutations were found in other lineages. Genetic polymorphisms that vary between species and within genomes have important implications for the evolution and conservation of species. According to a previous study, genetic polymorphisms found in a species have been

influenced by demographic history (Ellegren & Galtier, 2016)

Several specific polymorphisms in a species have been studied, such as the substitution of the Ala71Ser amino acid change in EsxH, commonly found in *M. Africanum* strains. The 3D structure of the EsxH complex has been studied using NMR spectroscopy. The contact surface between EsxG and EsxH is essentially hydrophobic (water-repellent), and the Ala71 residue is located at the interface between the molecules in the EsxH complex, near Met72, which, together with Met18, forms the base of the groove. The conservative Ala71Ser substitution in *M. africanum* likely disrupts these hydrophobic interactions (Yruela et al., 2016). However, there has been no further discussion about the mutations occurring in the two isolates, making it a novel finding in vaccine development.

For example, studies on *Mtb*32A, *Mtb*39A, *Mtb*72F, and M72 as vaccine candidates show that these genes have a high degree of identity among different *Mtb* strains. This high identity suggests the genes are relatively stable and do not vary significantly, which is important for vaccine development because the targets tend to persist across strains. The genetic stability suggests vaccines developed from these genes may be effective against multiple *Mtb* variants, reducing the risk of mutations that could cause the vaccine to be less effective. Therefore, genes that show little or no variation between lineage could be promising targets for vaccines (Mortier et al., 2015)

Homology modeling has become a crucial method in structural biology, greatly helping to bridge the gap between known protein sequences and experimentally determined structures. This technique predicts a protein 3D structure by using the amino acid sequence and similarity to other proteins with established structures. SWISS-MODEL is a server used for automatic protein homology modeling (Waterhouse et al., 2018)

The results from the 3D homology modeling visualization showed that structurally, the differences in amino acids in the EsxH gene did not significantly change the overall structure of the protein. The modeling results have also been confirmed regarding the suitability of the position of each amino acid based on Ramachandran calculations. The plot indicates that a protein has high structural quality when there are generally more amino acid residues in the favored regions compared to those in the outlier regions. The results obtained for the EsxH gene have more favorable values than the outlier, suggesting the model can be considered quite good (Park et al., 2023).

5. Conclusion

This study analyzed the genetic diversity of *Mycobacterium tuberculosis* strains from South Sulawesi, Indonesia, focusing on their EsxH gene mutations and implications. The present study showed a unique nucleotide variation at positions 28 G>A and 29 C>T, leading to amino acid substitutions at position 10 (Alanine to Threonine and Alanine to Valine). Phylogenetic analysis of these mutations has shown that this is region-specific and different from previously reported lineages. Structural modelling using SWISS-MODEL server revealed no significant changes in the EsxH protein structure; however the functional implications of these mutations for immune evasion and vaccine design are still to be investigated.

These findings underscore the necessity of further strengthening our efforts in genomic surveillance and molecular studies of diversity in *Mtb* in Indonesia. The comprehensive functional assay and immunological review have yet to be addressed accordingly in these results. Further, bioinformatics analysis and in vitro testing are needed to delineate the effect of these alterations on cell epitopes.

Acknowledgments

The authors are grateful to the National Research and Innovation Agency (Kemenristek/BRIN) and StudiBio for providing computational resources and support during the study.

Source of Funding

This study was funded by the National Research and Innovation Agency (BRIN) under the Research and Innovation for Advanced Indonesia (RIIM) grant, No. 12/II.7/HK/2023, as well as the Health Research Organizations grant for 2023, No. 23/III.9/HK/2023.

Conflict of Interest

The authors declare that there is no conflict of interest or personal relationships.

Ethical Approval

This study protocol adheres to the Declaration of Helsinki and received approval from the Health Research Ethics Committee at the Faculty of Medicine, Hasanuddin University (No. 678/UN4.6.4.5.31/PP36/2023, September 13, 2023). All participants provided written informed consent to take part in the study.

Data Availability

The *Mycobacterium tuberculosis* EsxH gene sequences used in this study have been deposited in the GenBank database and are publicly available at https://www.ncbi.nlm.nih.gov/genbank/ with accession numbers:
BankIt2921642 M.tuberculosis_EsxH_isolate_19x PV074341
BankIt2921642 M.tuberculosis_EsxH_isolate_42x PV074342

References

Agapova, A., Serafini, A., Petridis, M., Hunt, D. M., Garza-Garcia, A., Sohaskey, C. D., & Pedro Só rio de Carvalho, L. 2019. Flexible nitrogen utilization by the metabolic generalist pathogen Mycobacterium tuberculosis. e*life*, **8:**1-22

Amalia, F., Syamsunarno, M. R. A. A., Triatin, R. D., Fatimah, S. N., Chaidir, L., & Achmad, T. H. 2022. The Role of Amino Acids in Tuberculosis Infection: A Literature Review. *Metabolites.,* **12(10)**: 1-13

Ates, L. S., Houben, E. N. G., & Bitter, W. 2016. Type VII Secretion: A Highly Versatile Secretion System. *Microbiol Spectr*. **4(1)**: 1-21

Chandra, P., Grigsby, S. J., & Philips, J. A. 2022. Immune evasion and provocation by Mycobacterium tuberculosis. *Nat Rev Microbiol*. **20(12)**: 750-766.

Crusio, W. E., & Radeke, H. H. 2019. **Advances in Experimental Medicine and Biology** Volume 1304 Series Editors. USA

Darrah, P. A., DiFazio, R. M., Maiello, P., Gideon, H. P., Myers, A. J., Rodgers, M.A., Hackney, J. A., Lindenstrom, T., Evans, T., Scanga, C. A., Prikhodko, V.,Andersen, P., Lin, P. L., Laddy, D., Roederer, M., Seder, R. A., & Flynn, J. A. L.2019. Boosting BCG with proteins or rAd5 does not enhance protection against tuberculosis in rhesus macaques. *NPJ Vaccines*. **4(1)** : 1-13

Davila, J., Zhang, L., Marrs, C. F., Durmaz, R., & Yang, Z. 2010. Assessment of the Genetic Diversity of *Mycobacterium tuberculosis* esxA, esxH, and fbpB Genes among Clinical Isolates and Its Implication for the Future Immunization by New Tuberculosis Subunit Vaccines Ag85B-ESAT-6 and Ag85B-TB10. 4. *Biomed Res Int*. **2010(1)**:1-8.

Ellegren, H., and Galtier, N. 2016. Determinants of genetic diversity. *Nat Rev Genet*. **17(7)**:1-12

Martinez-Olivares, C. E., Hernández-Pando, R., & Mixcoha, E. 2023. In silicoEsxG EsxH rational epitope selection: Candidate epitopes for vaccine designagainst pulmonary tuberculosis. *Plos one*. **18(4)**:1-27

Ly, A., and Liu, J. 2020. Mycobacterial virulence factors: Surface-exposed lipids and secreted proteins. *Int J Mol Sci*. **21(11)**:1-14.

Mortier, M. C., Jongert, E., Mettens, P., & Ruelle, J. L. 2015. Sequence conservation analysis and in silico human leukocyte antigen-peptide binding predictions for the *Mtb*72F and M72 tuberculosis candidate vaccine antigens. *BMC Immunol*. **16(1)**:1-14

Park, S. W., Lee, B. H., Song, S. H., & Kim, M. K. 2023. Revisiting the Ramachandran plot based on statistical analysis of static and dynamic characteristics of protein structures. *J Struct Biol*. **215(1)**:1-10

Peng, C., Cheng, Y., Ma, M., Chen, Q., Duan, Y., Liu, S., Cheng, H., Yang, H.,Huang, J., Bu, W., Shi, C., Wu, X., Chen, J., Zheng, R., Liu, Z., Ji, Z., Wang, J.,Huang, X., Wang, P., Wang, L. 2024. *Mycobacterium tuberculosis* suppresses host antimicrobial peptides by dehydrogenating L-alanine. *Nat Commun*. **15(1)**:1-19

Podoly, E., Hanin, G., & Soreq, H. 2010. Alanine-to-threonine substitutions and amyloid diseases: Butyrylcholinesterase as a case study. *Chem-Biol Interact*. **187(1)**:64–71.

Portal-Celhay, C., Tufariello, J. M., Srivastava, S., Zahra, A., Klevorn, T., Grace, P. S., Mehra, A., Park, H. S., Ernst, J. D., Jacobs, W. R., & Philips, J. A. 2016. *Mycobacterium tuberculosis* EsxH inhibits ESCRT-dependent CD4+ T-cell activation. *Nat Microbiol*. **2(2)**:1-22.

Ramaiah, A., Nayak, S., Rakshit, S., Manson, A. L., Abeel, T., Shanmugam, S., Sahoo, P. N., John, A. J. U. K., Sundaramurthi, J. C., Narayanan, S., D'Souza, G., Von Hoegen, P., Ottenhoff, T. H. M., Swaminathan, S., Earl, A. M., & Vyakarnam, 2019. Evidence for highly variable, region-specific patterns of T-cell epitope mutations accumulating in *Mycobacterium tuberculosis* strains. *Front in immunol*. **10(195)**:1-18

Shastri, M. D., Shukla, S. D., Chong, W. C., Dua, K., Peterson, G. M., Patel, R. P.,Hansbro, P. M., Eri, R., & O'Toole, R. F. 2018. Role of oxidative stress in the pathology and management of human tuberculosis. *Oxid Med Cell Longev.*. **1**:1-10

Singh, S., Saavedra-Avila, N. A., Tiwari, S., & Porcelli, S. A. 2022. A century of BCG vaccination: Immune mechanisms, animal models, non-traditional routes and implications for COVID-19. *Front Immunol*. **13**:1-22

Su, R., Yuan, J., Gao, T., Liu, Y., Shu, W., Wang, Y., Pang, Y., & Li, Q. 2023.Selection and validation of genes related to oxidative stress production and clearance in macrophages infected with Mycobacterium tuberculosis. *Front Cell Infect Microbiol*. **13**:1-13

Sutiwisesak, R., Hicks, N. D., Boyce, S., Murphy, K. C., Papavinasasundaram, K., Carpenter, S. M., Boucau, J., Joshi, N., Le Gall, S., Fortune, S. M., Sassetti, C. M.,and Behar, S. M. 2020. A natural polymorphism of *Mycobacterium tuberculosis* in the esxH gene disrupts immunodomination by the TB10.4-specific CD8 T cell response. *PLoS Pathog*. **16(10)**:1-29

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*., **46**:1-8

World Health Organization. 2022. **Global Tuberculosis Report 2022**. World Health Organization.