



Article

Decoding Ecuadorian *Mycobacterium tuberculosis* Isolates: Unveiling Lineage-Associated Signatures in Beta-Lactamase Resistance via Pangenome Analysis

Gabriel Morey-León ^{1,2,*} , Juan Carlos Fernández-Cadena ³ , Derly Andrade-Molina ^{1,4} and Luisa Berná ^{5,6,7,*}

¹ Facultad de Ciencias de la Salud, Universidad Espíritu Santo, Samborondón 0901952, Ecuador

² Facultad de Ciencias, Universidad de la República, Montevideo 11400, Uruguay

³ African Genome Center, University Mohammed VI Polytechnic (UM6P), Ben Guerir 43150, Morocco; juan.fernandez@um6p.ma

⁴ Laboratorio de Ciencias Ómicas, Universidad Espíritu Santo, Samborondón 0901952, Ecuador; dmandrademolina@uees.edu.ec

⁵ Laboratorio de Interacciones Hospedero-Patógeno, Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo 11400, Uruguay

⁶ Unidad de Genómica Evolutiva, Facultad de Ciencias, Universidad de la República, Montevideo 11400, Uruguay

⁷ Institut Pasteur de Montevideo, Montevideo 11400, Uruguay

* Correspondence: garielmory@uees.edu.ec (G.M.-L.); lberna@pasteur.edu.uy (L.B.)

Abstract: Background: Tuberculosis is the second largest public health threat caused by pathogens. Understanding *Mycobacterium tuberculosis*'s transmission, virulence, and resistance profile is crucial for outbreak control. This study aimed to investigate the pangenome composition of *Mycobacterium tuberculosis* clinical isolates classified as L4 derived from Ecuador. **Methods:** We analyzed 88 clinical isolates of *Mycobacterium tuberculosis* by whole-genome sequencing (WGS) and bioinformatic tools for Lineage, Drug-resistance and Pangenome analysis. **Results:** In our analysis, we identified the dominance of the LAM lineage (44.3%). The pangenomic analysis revealed a core genome of approximately 3200 genes and a pangenome that differed in accessory and unique genes. According to the COG database, metabolism-related genes were the most representative of all partitions. However, differences were found within all lineages analyzed in the metabolic pathways described by KEGG. Isolates from Ecuador showed variations in genomic regions associated with beta-lactamase susceptibility, potentially leading to epistatic resistance to other drugs commonly used in TB treatment, warranting further investigation. **Conclusions:** Our findings provide valuable insights into the genetic diversity of *Mycobacterium tuberculosis* populations in Ecuador. These insights may be associated with increasing adaptation within host heterogeneity, variable latency periods, and reduced host damage, collectively contributing to disease spread. The application of WGS is essential to elucidating the epidemiology of TB in the country

Keywords: pangenome; core-genome; tuberculosis; beta-lactamase; Ecuador; surveillance



Academic Editor: Vishwanath Venketaraman

Received: 26 December 2024

Revised: 19 January 2025

Accepted: 21 January 2025

Published: 28 January 2025

Citation: Morey-León, G.; Fernández-Cadena, J.C.; Andrade-Molina, D.; Berná, L. Decoding Ecuadorian *Mycobacterium tuberculosis* Isolates: Unveiling Lineage-Associated Signatures in Beta-Lactamase Resistance via Pangenome Analysis. *Biomedicines* **2025**, *13*, 313. <https://doi.org/10.3390/biomedicines13020313>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (*Mtb*) is a significant global health concern, with an estimated 10.8 million infections and 1.25 million deaths in 2023 [1]. *Mtb* has encompassed complex conformed of genetically related bacteria including *M. africanum* [2,3], which, within *Mtb*, are the typical human pathogens, and other lineages that infect animals (*M. bovis*, *M. canetti*, *M. microti*, *M. caprae*, *M. pinnipedii*, and

M. orygis) [4–7], which have evolved from common ancestors over centuries from different geographic areas, and adapting to external conditions [8–10]. Some studies have demonstrated the relationship between clinical isolates and specific genotypes and geographical regions as well as their remarkable capacity for increased dissemination that sets them apart from ancestral lineages [11–14]. This is particularly evident due to their higher virulence and shorter latency periods to change the active form.

Furthermore, it has been noted that even within modern lineages, which are known for their heightened virulence and rapid dissemination, not all isolates display identical characteristics [15]. This diversity in behavior is influenced by various external factors, including drug resistance, host heterogeneity, demographic factors, and the presence of dominant lineages that confer advantages in terms of dissemination and impact on the host [10,16]. In Ecuador, Lineage Euro-American (L4) is the most predominant, and molecular epidemiology studies have shown that the population structure of *Mtb* is composed of LAM, X-type, and Haarlem sublineages [17–21].

Genomic approaches, such as WGS of *Mtb* strains, have improved the understanding of composition analysis and provided invaluable information on gain or loss genes, evolutionary markers, and polymorphisms related to drug resistance, virulence, and sub-lineage patterns [22–28]. With a large amount of information on genes from sequence isolates, the pangenome-based approach is more convenient for discerning a complete analysis in search of unique, accessory, and core genes, estimating the diversity of genes and novel marker genes, especially local-distribution-associated [29,30], mostly associated with virulence and drug resistance.

The pangenome represents the complete set of genes present in a species. It comprises the core genome, consisting of genes shared by all species members, the accessory genome, which includes genes present in some but not all members, and the unique genome, which includes genes present in only one species member [31]. Additionally, the cloud genome refers to a subset of genes in the accessory genome that are not universally distributed but may be present in certain subgroups or populations under specific environmental conditions or selective pressures [32]. Pangenome studies have been performed on different isolates from Peru [33], Mexico [34], Brazil, Argentina, Paraguay, and Colombia [35] from clinical samples of *Mtb* to understand the variation in terms of unique sequences among them and to identify the importance of genes related to metabolism to adapt to external conditions and latency periods. However, no extensive pangenome studies have been performed in Ecuador to reveal the composition of the circulating lineages in the country. A unique WGS study identified the presence of 4.3.2/3 (LAM) and 4.1.2 (Haarlem) sublineages in a small dataset of *Mtb* strains from Ecuador [36]. In this study, we aimed to investigate the genomic composition of 88 *Mtb* strains classified as L4 derived from Ecuador using pangenome analysis. In addition, we explored the genetic variation in genomic regions associated with β -lactamase susceptibility using bioinformatics-based inference methods. Our findings provide valuable insights into the genetic diversity of *Mtb* in Ecuador and suggest an association of lineage with specific mutations related to β -lactamase resistance in TB, which drives the development of more effective TB control strategies tailored to the specific characteristics of the local *Mtb* population in the region.

2. Materials and Methods

2.1. *Mtb* Samples, Assembly, Functional Annotation, and Pangenome Analysis

This study included the raw reads from 88 clinical isolates of *Mtb* collected from different provinces in Ecuador between 2019 to 2021, which are available in the Sequence Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>, accessed on 15 June 2024) under the PRJNA827129 BioProject. Of the total isolates, 81.8% were from Guayaquil, while

the remaining isolates were distributed as follows: Babahoyo (5.6%), El Empalme (3.4%), and Quito (2.3%). This distribution highlights the geographical spread and prevalence of TB in these areas. A small proportion of isolates came from Chone, Durán, Guaranda, Machala, and Nueva Loja, with each contributing 1.1% of the total. Raw reads were quality-checked using FastQC v0.11.9 [37] and improved by Trimmomatic v0.38 [38]. Specie confirmation and contamination screening was performed by Kraken v2 [39]. A Unicycler assembly pipeline [40] was used to assemble high-quality reads with a minimum contig size of 300 bp and polished discordance using Pilon v1.24 [41]. The quality of the assemblies was evaluated using *Mtb* strain HR37v (NC_000962.3) as a reference genome in Quast v5.0.2. [42]. Structural and functional annotation was performed using Rapid Annotation with the Subsystem Technology tool kit (RASTtk) in the Pathosystems Resource Integration Center (PATRIC) (<https://www.patricbrc.org/>, accessed on 1 July 2024). Pangenome analysis was performed using Roary v3.11.2 [43] and BPGA (Bacterial Pan-Genome Analysis tool) pipeline v1.3.0 [44]. Roary was performed with a minimum identity percentage for blastp of 95%, and gene detection in 99% of the isolates was recorded as a core gene. Roary_stats, a custom R tool v4.4.0, was used to process the outputs. In the BPGA pipeline, the COG and KEGG databases were used to define the functionality of the characterized gene. The COG database is embedded in the BPGA pipeline to facilitate the functional annotation of genes, helping researchers to understand the functional landscape of the bacterial genomes whilst the KEGG database is integrated into the BPGA pipeline to facilitate the annotation of genes involved in metabolic and signaling pathways. This helps researchers gain a deeper understanding of the biological processes, metabolic capabilities, and functional diversity of bacterial genomes in their pangenomic analysis. The pangenome was divided into three partitions: core, accessory, and cloud/unique. Core genes were present in 100% of the isolates, accessory in 3–99%, and cloud/unique in 1–2%.

2.2. Variant Calling, Lineage Classification, and Drug-Resistance Genes

Genetic relatedness among 88 *Mtb* genomes was processed using the MTBseq pipeline v1.1.0 with standard input parameters [45]. Briefly, BWA-mem and SAMtools algorithms were used to map the reads to the *Mtb* reference genome (NC_000962.3). GATK v3 was applied for base call recalibration and realignment of reads around InDels, and samtools mpileup for variant calling. Good-quality genomes presented a minimum mean coverage > 20X, read depth DP < 5, and reference genome covered >95%. The sub-lineage classification was performed using the strain module to perform lineage classification based on a set of phylogenetic SNPs, and a genetic distance matrix among transmission groups was created from the MTBseq result of the TBgroups module, which infers related isolates based on pairwise distance between distinct SNP positions. The TB-Profiler v4 pipeline was used to predict canonical mutations in the genes associated with resistance to first- and second-line drugs. We used BCFtool as the caller algorithm within the TB-profiler. For beta-lactamase susceptibility analysis, 46 genes were selected due to being especially related to beta-lactamase resistance and cell wall biosynthesis.

3. Results

3.1. Assembly, Functional Annotation, and Pangenome Analysis

In this study, we conducted a comparative genome analysis of 88 *Mtb* isolates from Ecuador. The results reveal an average genome size of 4.32 Mb (range: 4.21–4.33) and a high GC content of 65.2–65.5%. The coverage depth was approximately $\sim 60.9X \pm 23.6$. Analysis of single nucleotide variants (SNVs) revealed an average of approximately 788 ± 111 SNPs, along with approximately 51 ± 8 insertions and 47 ± 7 deletions. Notably, the majority of the InDels were short (≤ 20 nt, as shown in Table 1). The distribution of the

SNPs encompassed approximately 105 ± 18 intergenic, 372 ± 47 non-synonymous, and 232 ± 36 synonymous variations. Genome annotation highlighted the existence of approximately 4294 ± 14 genes, consisting of 4254 coding sequence (CDS) and 44 transfer RNA (tRNA) genes, with functional assignments identified in approximately 3378 ± 12 cases. Approximately 1061 ± 4 proteins were annotated with Enzyme Commission (EC) numbers, whereas approximately 918 ± 4 exhibited Gene Ontology (GO) assignments. Moreover, approximately 814 ± 3 proteins were successfully linked to the KEGG pathways. The assignment of roughly 1961 ± 19 genes was carried out within subsystems, which represent clusters of proteins collaborating to execute specific biological processes or structural complexes (Table S1).

Table 1. Summary of detected genomic variants in the lineages of 88 *Mtb* isolates identified in this study.

Lineages	SNP	Deletions		Insertions	
		More than 20 nt	Less than 20 nt	More than 20 nt	Less than 20 nt
Euro-American	900	55	4	55	4
Haarlem	877	45	3	55	3
LAM	760	43	4	45	2
mainly T	494	33	2	32	1
S-type	760	37	4	47	3
X-type	863	45	6	55	3

SNP: Single nucleotide polymorphism.

Pangenome analysis of the 88 *Mtb* isolates by Roary revealed a pangenome consisting of 6703 gene families, which comprised 3032 core genes, 2073 accessory genes, and 1598 cloud genes (Figure 1A–D). The ratio between core and pangenome size was 0.45, indicating that the genome comprised almost half of the pangenome (45.2%), which is indicative of the highest variability. Pathway analysis revealed that 19.2% of the annotated protein-coding genes were associated with amino acid metabolism, followed by genes related to carbohydrates (16.2%), lipid metabolism (13.6%), and xenobiotics, biodegradation, and metabolism (12.5%). Genes involved in amino acid biosynthesis were found to be conserved and essential for pathogenicity in bacteria, including *Mtb* [46]. Through the BPGA pipeline, our analysis revealed that the pangenome consisted of an average of 4397 genes. These genes were categorized into a core genome comprising 3104 genes, ~659 accessory genes, and 767 unique genes. The sensitive isolates had a higher number in the core genome whilst in the resistant isolates, the accessory genome comprised more genes than the sensitive (Figure 1E,F). According to the pangenome evaluation, the “b” value of 0.0861522 in the power-law regression model suggests that the pangenome of *Mtb* is near to close.

Notably, the three *Mtb* isolates with the highest number of cloud genes belonged to the X-type (S1454 [n = 261] and S1477 [n = 200]) and LAM (S1453 [n = 177]) sub-lineages (Figure 1C), and a similar pattern was identified by the BPGA pipeline in these isolates (S1454 [n = 211], S1477 [n = 163]), and S1453 [n = 144]). Only 6.3% (40 of 638) of the cloud genes in these isolates were assigned to COG (Clusters of Orthologous Genes), encompassing lipid transport and metabolism, secondary metabolite biosynthesis, transport and catabolism, amino acid transport and metabolism, and energy production and conversion categories. Notably, 80% of the genes associated with lipid transport and metabolism were found in isolate S1454. In comparison, 75% of the genes associated with energy production and conversion were found in isolate S1477.

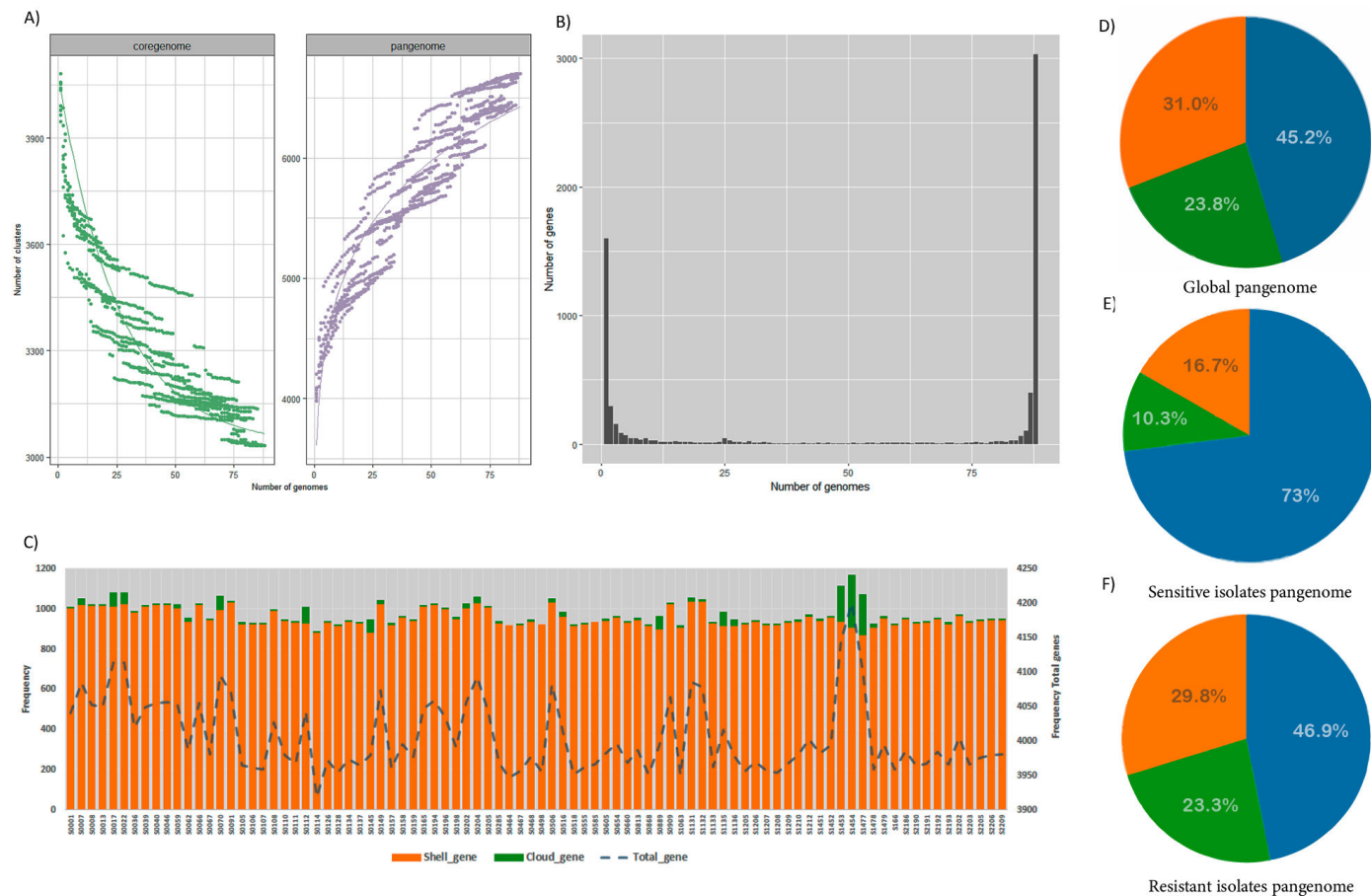


Figure 1. The pan-genome of the 88 *Mtb* isolates from Ecuador. (A) Rarefaction curves of the core genome (green) and the pangenome (light purple). (B) Cluster frequency in the core genome. Each bar represents the core genome composition of an individual isolate, with the global core genome depicted by the final bar after analysis by Roary for all isolates. (C) Distribution of accessory, cloud, and total genes across all isolates analyzed in this study, highlighting the variability in pangenome components among different strains. (D) Global pangenome composition, where the core genome constitutes almost half of the total gene content. (E) Pangenome composition for sensitive isolates, with the core genome comprising nearly 75% of all genes. (F) Pangenome composition for resistant isolates, where the accessory genes increase, potentially reflecting specific characteristics associated with resistance. Colors in pies: core gene (blue), accessory gene (orange), and cloud gene (green). The dotted line in panel C represents the total genes of each isolate.

A functional analysis among different pangenome partitions based on COG categories globally showed that genes associated with metabolism (45.45% in the core genes to 35.27% in the accessory genes), as well as genes related to cellular processes and signaling (24.97% in the unique gene to 9.79% in the core genes), present significant global variations. Additionally, we determined that genes with general function predictions [R] (over 10% in all partitions) are the most recurrent, followed by genes associated with secondary metabolites biosynthesis, transport, and catabolism [Q] (10% in accessory and unique genes). Finally, genes associated with cell motility [N] were predominantly in the accessory gene category.

Furthermore, according to functional assignments based on the KEGG pathways database, our genomes globally exhibited mostly genes associated with metabolism (core and accessory genes), whereas genes related to environmental information processing were mostly identified as unique genes. Within the core-genome gene distributions, amino acid metabolism (15.0%) had the highest representation, followed by carbohydrate metabolism (14.6%), overview (11.8%), and xenobiotics biodegradation and metabolism (8.9%). Among accessory genes, carbohydrate metabolism (13.9%) and overview (10.0%) were the most prevalent. In the unique gene category, genes associated with the cellular community, digestive system, energy metabolism, immune system, infectious diseases, and signaling molecules and interactions had a similar distribution, each accounting for 8.7%.

3.2. Pangenome Diversity on Lineage of *Mtb*

When we analyzed the core genome composition based on the drug resistance pattern, the sensitive isolate (73.0% by Roary and 83.0% by BPGA) had a more closed pangenome than the resistant isolates (46.8% by Roary and 60.7% by BPGA) (Figure 1E,F). By lineage, the LAM and X-type isolates had a higher number of cloud genes (1073 and 874, respectively); however, by BPGA, the results differed slightly in the core genome, and the largest changes were found mostly in accessory and unique genes (Table 2).

Table 2. Pangenome composition according to drug resistance and lineage features of 88 *Mtb* isolates from Ecuador.

	n	Core Genes	Roary Pipeline			Pangenome	Core Genes	BPGA Pipeline			Pangenome
			Accessory Gene	Cloud Gene				Accessory Gene	Cloud Gene		
Total	88	3032	2073	1598	6703		3104	659	767		4397
Genomic resistance											
Resistant_isolate	64	3054	1940	1518	6522		3123	646	756		5145
Sensitive_isolate	24	3573	817	504	4894		3477	255	198		4188
Lineages											
LAM	39	3281	1385	1073	5739		3330	443	463		4649
X_Type	21	3306	1281	874	5461		3352	392	563		4655
S_type	10	3684	573	269	4526		3589	150	239		4095
Haarlem	9	3674	593	224	4491		3540	209	190		4146
mainly_T	5	3754	394	290	4438		3618	112	194		3990
Euro-American	4	3760	334	268	4362		3616	144	118		3985

A closer analysis of the distribution within the COG database and KEEG pathways revealed that the diversity in lineage comes mainly from accessory and unique genes. When analyzing the functional annotation of COG major categories for accessory genes, we found that Euro-American isolates had the highest percentage of information storage and processing genes (29.14%), while the S-type lineage had a higher percentage of cellular processes and signaling genes (31.16%), and the X-type lineage exhibited the highest percentages of metabolism and poorly characterized genes (38.29% and 20.0%, respectively). In terms of unique genes, the S-type lineage showed the highest representation of information storage and processing genes (33.24%), followed by Euro-American with cellular processes and signaling genes (30.30%); metabolism had the highest representation in the LAM lineage

(48.13%), and the X-type lineage had the highest unique poorly characterized (20.33%). Reviewing the distribution of COG functional categories, we found that genes related to [Q] secondary metabolites biosynthesis, transport, and catabolism genes (found in all six lineages), and [N] cell motility genes (presented predominantly in Haarlem, LAM, mainly-t, S-type, and X-type) were the most prevalent, accounting for more than 10% in each lineage. Additionally, [R] general function prediction genes were exclusively present in the LAM, S-type, and X-type lineages with the highest percentages (Figure 2 and Table S2).

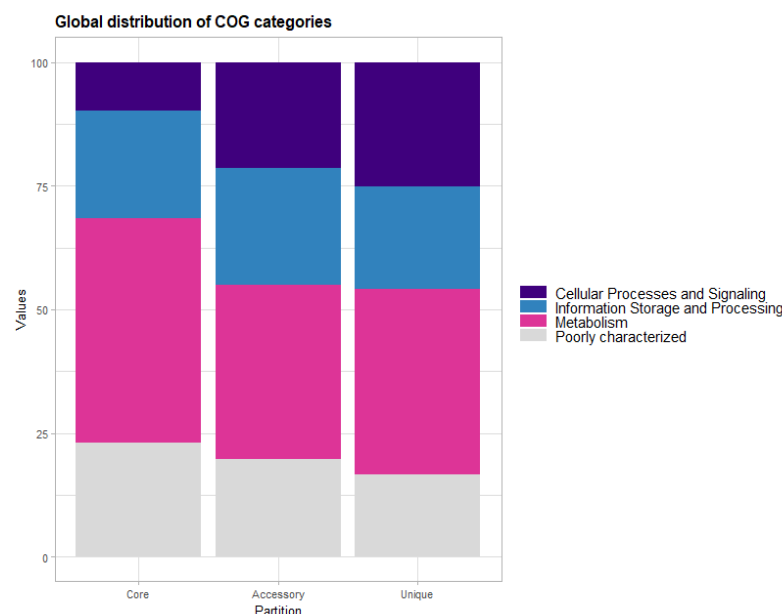


Figure 2. Distribution of the major COG categories. Stacked bars represent the percentage distribution of the functional COG annotations among the genes highly conserved in the core, accessory, and unique partitions from all lineages analyzed in this study.

Among the unique genes, the Euro-American isolate exhibited greater diversity, primarily composed of genes related to Cell cycle control, cell division, chromosome partitioning, cell motility, intracellular trafficking, secretion, vesicular transport, energy production, and conversion (12.9% each). Additionally, a significant portion of their unique genes were associated with secondary metabolites biosynthesis, transport, and catabolism (21.65%). In contrast, the X-type lineage had a notable representation of genes involved in carbohydrate transport and metabolism, as well as secondary metabolites biosynthesis, transport, and catabolism (14.79% each). Furthermore, a proportion of their unique genes fell into the category of function unknown (11.09%) (Figure 3 and Table S3).

From Global Metabolic KEEG pathways, the mainly-T isolate significantly contributed to the accessory partition. It had the highest percentage of genes related to environmental information processing (33.3%), followed by organismal systems (25.0%) and cellular processes (12.5%). Metabolism had the highest representation, with percentages of 57.96% in the X-type lineage, 54.14% in LAM, and 53.57% in Haarlem. Human diseases accounted for 13.64% of genes in the S-type lineage, and genetic information processing represented 11.05% in LAM. Regarding unique genes, all lineages exhibited a significant presence of genes related to environmental information processing, human diseases, and organismal systems, followed by cellular processes (Euro-American, Haarlem, LAM, and S-type). Metabolism genes were notably represented in the LAM, mainly-T, and S-type lineages. However, only the X-type lineage showed the highest percentage of genes associated with genetic information processing (15.38%) (Figure 4 and Table S2).

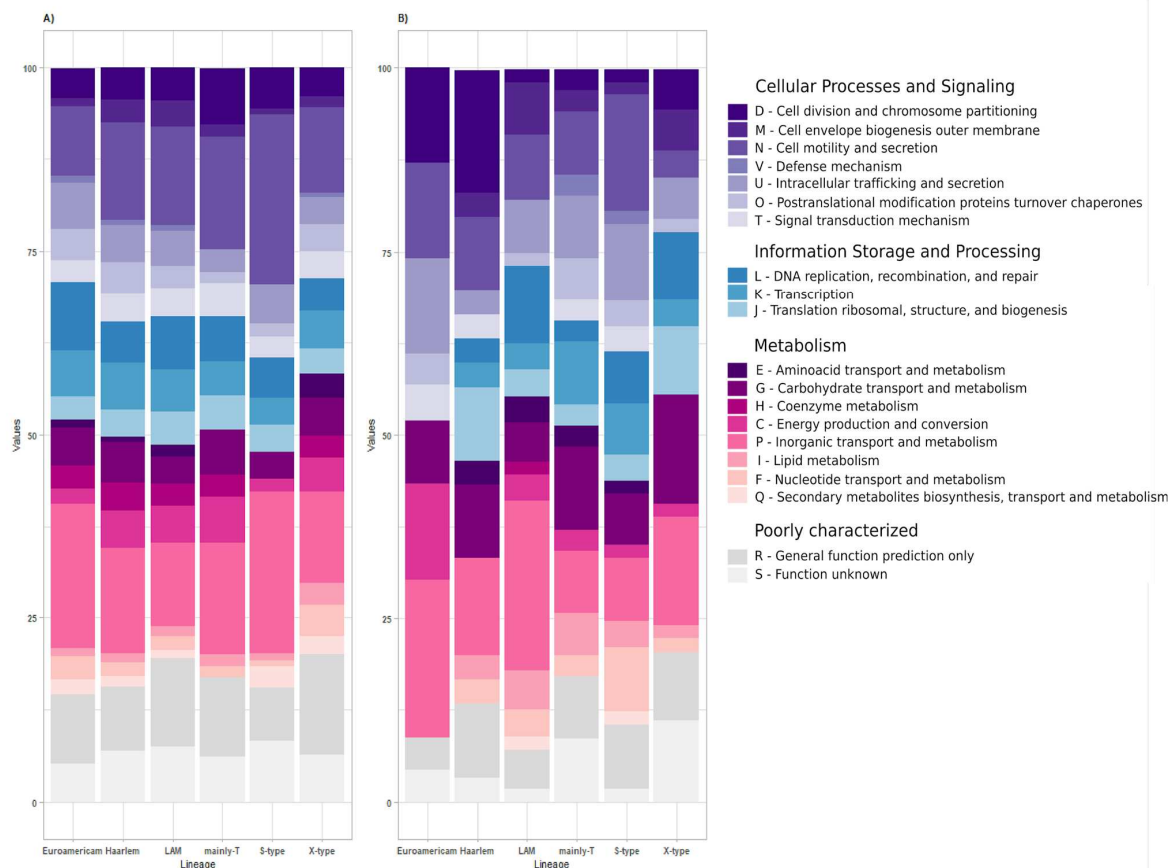


Figure 3. Distribution of primary COG categories. Stacked bars represent the percentage distribution of the functional annotations among the genes highly conserved on (A) the accessory partition; (B) the unique genes from all lineages analyzed in this study. **Functional Annotation Groups.** *Cellular Processes and Signaling.* D: cell division and chromosome partitioning; M: cell envelope biogenesis outer membrane; N: cell motility and secretion; V: defense mechanism; U: intracellular trafficking and secretion; O: posttranslational modifications proteins turnover chaperones; T: signal transduction mechanism. *Information Storage and Processing.* L: DNA replication, recombination, and repair; K: transcription; J: translation ribosomal, structure, and biogenesis. *Metabolism.* E: amino acid transport and metabolism; G: carbohydrate transport and metabolism; H: coenzyme metabolism; C: energy production and conversion; P: inorganic transport and metabolism; I: lipid metabolism; F: nucleotide transport and metabolism; Q: secondary metabolites biosynthesis, transport and metabolism. *Poorly characterized.* S: function unknown; R: general function prediction only.

However, a more detailed analysis of accessory genes, based on the KEGG pathways distribution, reveals that carbohydrate metabolism genes exhibited a higher level of diversity. The Haarlem lineage represented 14.3% of carbohydrate metabolism genes, followed by LAM (12.7%) and X-type (12.10%). Similarly, signal transduction genes displayed significant variability, with the mainly-T isolate having 20.8%, S-type with 13.6%, and Euro-American with 11.5%. Notably, the mainly-T- and S-type isolates lacked genes associated with carbohydrate metabolism (Figure 5A).

From the unique genes, infectious diseases were prevalent across almost all lineages (except mainly-T) followed by the cellular community, digestive system, immune system, signaling molecules and interaction, and signal transduction. The Euro-American and Haarlem isolates represented 11.3% of these categories, while the LAM lineage had 11.11%, and the S-type lineage had 11.43% (increasing to 14.29% in signal transduction genes). It is worth noting that several pathways in the case of accessory genes did not involve any genes (Figure 5B and Table S4).

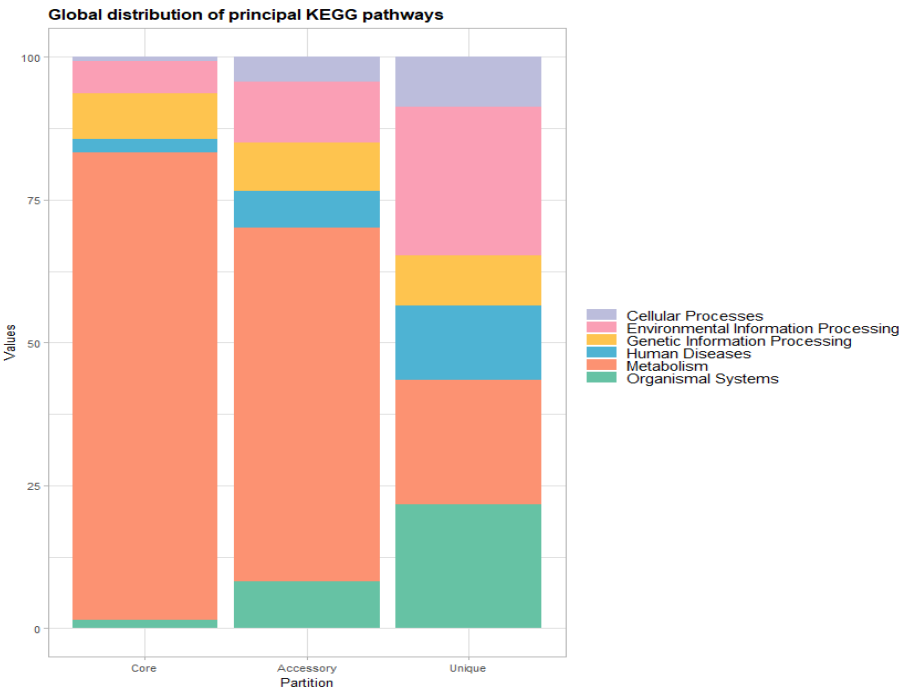


Figure 4. General distribution of KEGG pathways. Differences in the percentage of functional KEGG pathways annotated in the core, accessory, and unique partitions from all lineages analyzed in this study.

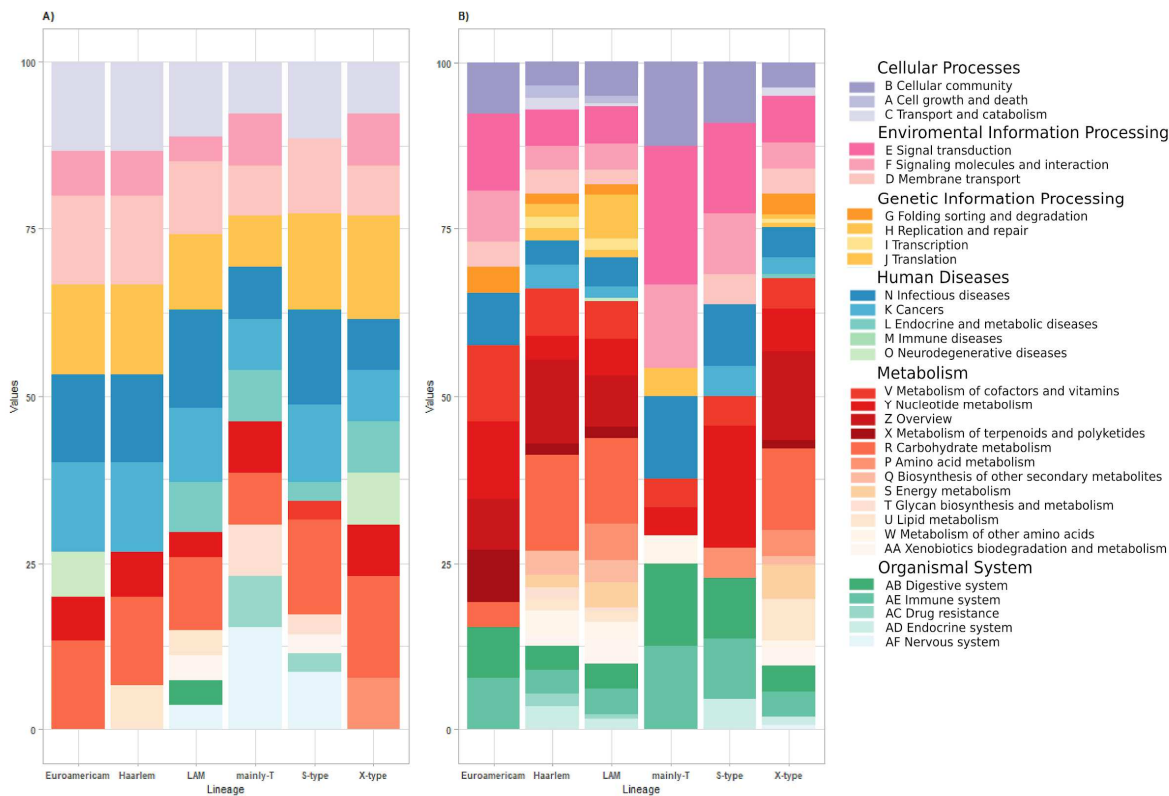


Figure 5. Distribution of primary KEGG pathways. (A) The percentage of functional annotations among the genes highly conserved among the unique genes from all lineages analyzed; (B) the percentage of functional annotations among the genes highly conserved among the accessory genes from all lineages analyzed in this study. Alphabetical labels correspond to standard KEGG pathway categories, with full names included to ensure clarity and consistency with KEGG conventions, facilitating the interpretation of genes according to their biological roles.

3.3. Genetic Diversity of the Beta-Lactamase Resistance-Associated Genes

To identify the existence of mutations that may affect genes related to beta-lactamase susceptibility, we analyzed the 88 Ecuadorian *Mtb* genomes to identify the SNPs present in 46 genomic regions commonly identified as beta-lactamase resistance-associated and cell wall biosynthesis genes of MTBC [47]. In all the isolates, 107 SNPs were identified in 46 genomic regions. *murD* (132 SNPs), *ddlA* (86 SNPs), *rpfC* (30 SNPs), *rpfE* (99 SNPs), *ponA1* (95 SNPs), *dacB2* (29 SNPs), *ftsK* (51 SNPs), *Rv0192* (93 SNPs), and *Rv0008c* (104 SNPs) genes showed the highest accumulation of SNPs distributed into various substitutions among the characterized resistance profiles (Table 3 and Table S5). These genes were primarily associated with peptidoglycan biosynthesis. Typically, the MDR isolates presented the highest number of mutations. The lineages that frequently exhibited a higher number of SNPs were LAM and S-type. Six substitutions were present in all isolates: one synonymous (Ala189Ala in *eccA2*) and five non-synonymous (Arg247Gly in *murD*, Thr365Ala in *ddlA*, Arg126Gln in *rpfE*, Ser127Pro in *Rv0192*, and Ser145Pro in *Rv0008c*).

Table 3. Principal synonymous and non-synonymous mutations identified in genes related to β -lactamase resistance.

Function	Functional Categories	Locus	Gene Name	AA Change
	Regulatory proteins	<i>Rv0015c</i>	<i>pknA</i>	Ser385Arg (agc/agG)
	Regulatory proteins	<i>Rv0324</i>		Thr168Ala (act/Gct)
	Lipid metabolism	<i>Rv0642c</i>	<i>mmaA4</i>	Asn165Ser (aac/aGc)
	Conserved hypotheticals	<i>Rv0791c</i>	-	Ser100Cys (tcc/tGc)
	Intermediary metabolism and respiration	<i>Rv0948c</i>	-	Lys59Thr (aag/aCg)
	Insertion seqs and phages	<i>Rv1128c</i>	-	Glu270Gly (gaa/gGa)
	Intermediary metabolism and respiration	<i>Rv1606</i>	<i>hisI</i>	Thr99Ile (acc/aTc)
	Cell wall and cell processes	<i>Rv1987</i>	-	Ser36Asn (agt/aAt)
	Conserved hypotheticals	<i>Rv2022c</i>	-	Val118Ala (gtg/gCg)
	Information pathways	<i>Rv2756c</i>	<i>hsdM</i>	Leu306Pro (ctg/cCg)
	Intermediary metabolism and respiration	<i>Rv3057c</i>	-	Asp112Ala (gat/gCt)/His111His (cac/caT)
	Conserved hypotheticals	<i>Rv3365c</i>	-	Ala266Thr (gcg/Acg)
	Cell wall and cell processes	<i>Rv3451</i>	<i>cut3</i>	Gly209Asp (ggc/gAc)/Leu259Arg (ctg/cGg)
	Lipid metabolism	<i>Rv3824c</i>	<i>papA1</i>	Leu35Phe (ctt/Ttt)
PG synthesis	Cell wall and cell processes	<i>Rv3884c</i>	<i>eccA2</i>	Ala189Ala (gcc/gcG)
PG synthesis	Cell wall and cell processes	<i>Rv2155c</i>	<i>murD</i>	Arg247Gly (cgg/Ggg)
PG hydrolysis	Cell wall and cell processes	<i>Rv2981c</i>	<i>ddlA</i>	Thr365Ala (aca/Gca)
PG hydrolysis	Cell wall and cell processes	<i>Rv1884c</i>	<i>rpfC</i>	His16Arg (cac/cGc)
PG assembly	Cell wall and cell processes	<i>Rv2450c</i>	<i>rpfE</i>	Arg126Gln (cgg/cAg)
PG assembly	Cell wall and cell processes	<i>Rv0050</i>	<i>ponA1</i>	Ala244Ala (gca/gcG)
Cell division	Conserved hypotheticals	<i>Rv0192</i>	-	Ser127Pro (tcg/Ccg)
Cell division	Cell wall and cell processes	<i>Rv0008c</i>	-	Ser145Pro (tcc/Ccc)
Cell division	Cell wall and cell processes	<i>Rv2748c</i>	<i>ftsK</i>	Met298Val (atg/Gtg)

The *ftsK* gene (*Rv2748c*) displayed the highest number of substitutions, with six non-synonymous and two synonymous substitutions. Among these, *Met298Val* substitution was the most frequent (47.0%, 24/51) and was only identified in the LAM isolates. In *murD* (*Rv2155c*), among the three nonsynonymous mutations, the most frequent was *Arg247Gly*, predominantly present in MDR isolates (52/88 isolates), with LAM and S-type isolates accounting for 67% of the cases. In contrast, *Phe76Leu* was exclusively present in S-type isolates. The *ponA1* (*Rv0050*) gene exhibited five substitutions, with *Pro631Ser* (43/95) and *Ala244Ala* (33/95) being the most common.

Additionally, with regard to substitutions distributed in all isolates, mutations were identified in different frequencies according to the lineages, as follows: In the LAM isolates more than 76% of them had mutations in genomic regions *pknA*, *mmaA4*, *hisI*, *rpfC*, *cut3*,

papA1, *ponA1*, *Rv0791c*, *Rv0948c*, *Rv1987*, *Rv2022c*, *Rv3057c*, and *Rv3365c*. For the Haarlem isolates, 100% presented mutations in the genomic regions *pknA*, *Rv1128c*, *pheT*, *hdsM*, *Rv3057c*, *cut3*, *eccA2*, *murD*, *Rv3915*, and *ponA1* and a deleterious effect mutation in the *eccA2* gene (Gln460*). Among S-type isolates, mutations in the genomic regions *pheT*, *murD*, *rpfE*, *ftsK*, *ftsH*, and *Rv0008c* were found in more than 90% of the isolates. Finally, in X-type isolates, mutations in *Rv0324*, *Rv1128*, *hdsM*, *cut3*, *eccA2*, *murD*, and *ponA1* were identified in 100% of the isolates, whereas *lpqK*, *glmU*, *rpfB*, and *ponA1* (Thr58Ala) were identified in 76% of the isolates.

4. Discussion

Several studies have been conducted to identify the causes of specific clinical behavior in different isolates of *Mtb* [48]. However, different results have been obtained depending on the geographical location [49–51]. The present study sought to characterize the genetic composition of Ecuadorian isolates and ascertain whether this pangenome displays variations associated with lineages or specific resistance patterns. This is the first comprehensive study analyzing the pangenome composition in Ecuadorian isolates. At the time of this study, a single investigation was undertaken involving 21 isolates [36], highlighting the necessity for an expanded collection of sequenced isolates from Ecuador.

Pangenome studies provide evidence of gene loss or gain events within a species, which depend on the environment in which they thrive [52]. In the case of *Mtb*, its environment is restricted, resulting in limited diversity [53]. Our analyses using Roary and BPGA reveal significant disparities in pangenome determination. Among the 88 *Mtb* isolates examined, the pangenome derived from the Roary algorithm encompassed 48% more accessory genes and unique genes compared to the results from the BPGA algorithm. Furthermore, the core genome constitutes 45% (Roary) or 70% (BPGA) of the total pangenome. These findings not only diverge from each other but also deviate from previous reports, which describe a core genome accounting for 25% [54], approximately 86% [55], or about 74% [35] of the pangenome. Roary assignment suggests that accessory genes will continue to increase. Different studies attribute variations in the number of annotated and assigned genes within pangenome partitions to the different thresholds used by the algorithms for gene annotation, clustering, and classification [29,34,56,57]. Therefore, it is important to standardize the processes used to define the pangenome and its partitions, particularly for prokaryotes.

Functional analysis showed that, among the lineages identified in our study, gene assignments were predominantly associated with metabolism (~45%) in the core genome, specifically lipid (~7.8%) and amino acid (~7.3%) transporters. In accessory and unique genes, assignments were related to secondary metabolites biosynthesis, catabolism, and transport (~15.9%), particularly in the LAM and S-type lineages. These findings are consistent with other studies where major genes are associated with lipid metabolism [35,55]. In fact, lipid metabolism holds significant relevance for *Mtb*, primarily because of its critical role as a component of the cellular membrane and mass, along with its involvement in vital processes such as cell invasion, evasion of the immune system, virulence, and growth retardation. Moreover, the identification of accessory and unique genes mostly associated with the biosynthesis, catabolism, and transport of secondary metabolites highlights the adaptive capacity of *Mtb* to the host or external environment [48,58,59]. These factors would be associated with differences in transmissibility and virulence characteristics observed in the members of the *Mtb* complex, playing a crucial role in disease recurrence [35,49].

Furthermore, the presence of cloud genes within these accessory and unique gene categories underscores the genetic diversity that contributes to the pathogen's adaptability. These cloud genes may encode functions that are not essential for survival under stan-

standard laboratory conditions but become critical under specific environmental pressures encountered during infection or treatment. For instance, cloud genes could facilitate enhanced metabolic pathways or provide mechanisms for resisting host immune responses or antibiotic action [60]. This genetic variability can lead to significant differences in strain behavior, including variations in virulence and transmission dynamics among different populations of *M. tuberculosis*. Understanding these cloud genes is essential for developing targeted interventions and improving our strategies for TB control by addressing the unique characteristics of various strains within this complex pathogen.

Studies have revealed that isolates from the LAM lineage (4.3.4.2) exhibit increased susceptibility to beta-lactams, whereas the Beijing and Haarlem lineages (4.1.2.1) display high resistance. Our results identified the highest number of mutations in isolates of LAM and S-type sublineages, suggesting that mutations in cell wall biosynthesis genes could be associated with higher resistance to beta-lactams, which should be further investigated if beta-lactams are considered for inclusion in TB treatment regimens. No mutations in the *blaC* gene were determined. Mutations in resuscitation-promoting factors (*Rpfs*), which are related to PG hydrolases, have been associated with increased outer membrane permeability and susceptibility to beta-lactams [61,62], and a similar pattern has been observed with substitutions in *EccA2*, an ESX-2 type VII secretion system component. Our findings suggest that Ecuadorian isolates related to the Euro-American and Haarlem sublineages may exhibit higher resistance to amoxicillin and meropenem because of the substitution Glu215Gly. Conversely, mutations in *lpqK*, a conserved lipoprotein similar to penicillin-binding proteins (PBPs), and *RpfC* were linked to increased susceptibility in the strains. In our study, we identified the substitutions Glu67Lys and His16Arg in the *lpqK* and *rpfC* genes, respectively, mostly in LAM isolates.

Despite the fact that we identified mutations associated with β -lactamase resistance by analyzing genomic sequences from clinical isolates, the absence of microbiological susceptibility results can be attributed to several limitations. First, most clinical microbiology laboratories do not routinely perform susceptibility testing for beta-lactams in *Mtb*, as these antibiotics are not part of standard TB treatment regimens. Traditional susceptibility testing methods, such as disk diffusion or broth microdilution, may not be effective in detecting beta-lactam resistance in *Mtb* due to its complex cell wall structure. Additionally, other resistance mechanisms, such as alterations in porins or efflux pumps, may confer resistance without being detected by conventional methods. Moreover, pharmacokinetic factors may limit the concentration of beta-lactams at the site of infection, hindering the ability to detect resistance in laboratory assays. Resistance may also not be fully expressed phenotypically under the conditions used for routine testing.

It is important to note that while we have identified lineage-associated signatures and mutations potentially linked to beta-lactam resistance, experimental beta-lactamase data were not included in this study due to the technical limitations mentioned above. As beta-lactams are not routinely tested in *Mtb* for clinical management, microbiological sensitivity testing was not conducted for these drugs in our isolates. Furthermore, existing methodologies in *Mtb* resistance testing often focus on first-line and second-line drugs rather than beta-lactams, which are not part of standard treatment protocols. Given the potential role of beta-lactams in TB treatment, we acknowledge the need for experimental validation of resistance mechanisms and suggest that further studies, including susceptibility testing, are warranted to assess the full susceptibility profile of these antibiotics in *Mtb*.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines13020313/s1>, Table S1. Genomic and clinical data from the Ecuadorian isolates are shown in Table S2. Distribution of major COG and KEGG pathways from Lineage and pangenome partitions; Table S3. Distribution of primary COG categories from

lineage and pangenome partitions; Table S4. Distribution of primary KEGG pathways from lineage and pangenome partitions; Table S5. Synonymous and non-synonymous mutations identified in genes related to β -lactamase resistance.

Author Contributions: G.M.-L. conceptualization, data curation, formal analysis, investigation and writing—original draft, D.A.-M. and J.C.F.-C. performed data analysis and writing—original draft, and L.B. performed conceptualization and writing—review and editing; the funding acquisition was performed by G.M.-L. and D.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The need to obtain informed consent was waived by The University Espírito Santo Review Board under code 2022-001A because the study included only isolates from a collection. The ethics committee ruled out informed consent because the data of the isolates were anonymized and no patient data were disclosed. This study was conducted under the ethical principles outlined in the Declaration of Helsinki.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw reads sequences analyzed in this study can be found in the SRA database under the following accession numbers: PRJNA827129.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization. *Global Tuberculosis Report*, 1st ed.; World Health Organization: Geneva, Switzerland, 2024; p. 68. Available online: <https://www.who.int/publications/i/item/9789240101531> (accessed on 15 January 2025).
2. De Jong, B.C.; Antonio, M.; Gagneux, S. *Mycobacterium africanum*—Review of an Important Cause of Human Tuberculosis in West Africa. *PLoS Negl. Trop. Dis.* **2010**, *4*, 4e744. [CrossRef]
3. Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry, C.E.; et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **1998**, *393*, 537–544. [CrossRef]
4. Garnier, T.; Eiglmeier, K.; Camus, J.C.; Medina, N.; Mansoor, H.; Pryor, M.; Duthoy, S.; Grondin, S.; Lacroix, C.; Monsempe, C.; et al. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 7877–7882. [CrossRef]
5. van Ingen, J.; Rahim, Z.; Mulder, A.; Boeree, M.J.; Simeone, R.; Brosch, R.; van Soolingen, D. Characterization of *Mycobacterium orygis* as *M. tuberculosis* Complex Subspecies. *Emerg. Infect. Dis.* **2012**, *18*, 653–655. [CrossRef]
6. Macedo, R.; Isidro, J.; Gomes, M.C.; Botelho, A.; Albuquerque, T.; Sogorb, A.; Bernardino, R.; Fernandes, T.L.; Mourato, T.; Durval, M.; et al. Animal-to-human transmission of *Mycobacterium pinnipedii*. *Eur. Respir. J.* **2020**, *56*, 2000371. [CrossRef]
7. Boritsch, E.C.; Brosch, R. Evolution of *Mycobacterium tuberculosis*: New Insights into Pathogenicity and Drug Resistance. *Microbiol. Spectr.* **2016**, *4*. [CrossRef]
8. Napier, G.; Campino, S.; Merid, Y.; Abebe, M.; Woldeamanuel, Y.; Aseffa, A.; Hibberd, M.L.; Phelan, J.; Clark, T.G. Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. *Genome Med.* **2020**, *12*, 114. [CrossRef]
9. Coll, F.; McNERney, R.; Guerra-Assunção, J.A.; Glynn, J.R.; Perdigão, J.; Viveiros, M.; Portugal, I.; Pain, A.; Martin, N.; Clark, T.G. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **2014**, *5*, 4812. [CrossRef]
10. Senghore, M.; Diarra, B.; Gehre, F.; Otu, J.; Worwui, A.; Muhammad, A.K.; Kwambana-Adams, B.; Kay, G.L.; Sanogo, M.; Baya, B.; et al. Evolution of *Mycobacterium tuberculosis* complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. *Sci. Rep.* **2020**, *10*, 327. [CrossRef]
11. Díaz Acosta, C.C.; Russomando, G.; Candia, N.; Ritacco, V.; Vasconcellos, S.E.G.; de Berrêdo Pinho Moreira, M.; de Romero, N.J.; Morcillo, N.; de Waard, J.H.; Gomes, H.M.; et al. Exploring the “Latin American Mediterranean” family and the RDRio lineage in *Mycobacterium tuberculosis* isolates from Paraguay, Argentina and Venezuela. *BMC Microbiol.* **2019**, *19*, 131. [CrossRef]
12. Ngabonziza, J.C.S.; Loiseau, C.; Marceau, M.; Jouet, A.; Menardo, F.; Tzfadia, O.; Antoine, R.; Niyigena, E.B.; Mulders, W.; Fissette, K.; et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat. Commun.* **2020**, *11*, 2917. [CrossRef] [PubMed]
13. Gonzalo-Asensio, J.; Pérez, I.; Aguiló, N.; Uranga, S.; Picó, A.; Lampreave, C.; Cebollada, A.; Otal, I.; Samper, S.; Martín, C. New insights into the transposition mechanisms of IS6110 and its dynamic distribution between *Mycobacterium tuberculosis* Complex lineages. *PLoS Genet.* **2018**, *14*, e1007282. [CrossRef]

14. David, S.; Duarte, E.L.; Leite, C.Q.F.; Ribeiro, J.N.; Maio, J.N.; Paixão, E.; Portugal, C.; Sancho, L.; Germano de Sousa, J. Implication of the RD(Rio) *Mycobacterium tuberculosis* sublineage in multidrug resistant tuberculosis in Portugal. *Infect. Genet. Evol.* **2012**, *12*, 1362–1367. [\[CrossRef\]](#)
15. Tatara, M.B.; Perdigão, J.; Viveiros, M.; Kritski, A.; da Silva, K.E.; Sacchi, F.P.C.; de Lima, C.C.; dos Santos, P.C.P.; Diniz, J.d.L.d.C.G.; Silva, P.E.A.; et al. Genetic Diversity and Molecular Epidemiology of *Mycobacterium tuberculosis* in Roraima State, Brazil. *Am. J. Trop. Med. Hyg.* **2019**, *101*, 774–779. [\[CrossRef\]](#)
16. Shanmugam, S.K.; Kumar, N.; Sembulingam, T.; Ramalingam, S.B.; Selvaraj, A.; Rajendhiran, U.; Solaiyappan, S.; Tripathy, S.P.; Natrajan, M.; Chandrasekaran, P.; et al. *Mycobacterium tuberculosis* Lineages Associated with Mutations and Drug Resistance in Isolates from India. *Microbiol. Spectr.* **2022**, *10*, e01594-21. [\[CrossRef\]](#)
17. Arias, A.P.J.; Lahiguera, M.J.; Borrás, R.; Cardona, C.G.; Silva, M.G.; López, M.J.V.; Serrano, M.d.R.G. Comparative study of the genetic diversity of *Mycobacterium tuberculosis* Complex by Simplified Amplified Fragment Length Polymorphism and Mycobacterial Interspersed Repetitive Unit Variable Number Tandem Repeat Analysis. *Rev. Ecuat. Med. Cienc. Biológicas* **2018**, *39*, 63–71.
18. Zurita, J.; Espinel, N.; Barba, P.; Ortega-Paredes, D.; Zurita-Salinas, C.; Rojas, Y.; Alcocer, I. Genetic diversity and drug resistance of *Mycobacterium tuberculosis* in Ecuador. *Int. J. Tuberc. Lung Dis.* **2019**, *23*, 166–173. [\[CrossRef\]](#)
19. Garzon-Chavez, D.; Zurita, J.; Mora-Pinargote, C.; Franco-Sotomayor, G.; Leon-Benitez, M.; Granda-Pardo, J.C.; Trueba, G.; Garcia-Bereguian, M.A.; de Waard, J.H. Prevalence, Drug Resistance, and Genotypic Diversity of the *Mycobacterium tuberculosis* Beijing Family in Ecuador. *Microb. Drug Resist.* **2019**, *25*, 931–937. [\[CrossRef\]](#)
20. Garzon-Chavez, D.; Garcia-Bereguian, M.A.; Mora-Pinargote, C.; Granda-Pardo, J.C.; Leon-Benitez, M.; Franco-Sotomayor, G.; Trueba, G.; de Waard, J.H. Population structure and genetic diversity of *Mycobacterium tuberculosis* in Ecuador. *Sci. Rep.* **2020**, *10*, 6237. [\[CrossRef\]](#)
21. Garcés, E.; Cifuentes, L.; Franco, G.; Romero-Sandoval, N.; Arias, P.J. Study of the distribution of lineages of *Mycobacterium tuberculosis* in a prison in Guayaquil, Ecuador. *Enfermedades Emergent* **2023**, *22*, 7–14.
22. Yang, T.; Gan, M.; Liu, Q.; Liang, W.; Tang, Q.; Luo, G.; Zuo, T.; Guo, Y.; Hong, C.; Li, Q.; et al. SAM-TB: A whole genome sequencing data analysis website for detection of *Mycobacterium tuberculosis* drug resistance and transmission. *Brief Bioinform.* **2022**, *23*, bbac030. [\[CrossRef\]](#)
23. Lam, C.; Martinez, E.; Crighton, T.; Furlong, C.; Donnan, E.; Marais, B.J.; Sintchenko, V. Value of routine whole genome sequencing for *Mycobacterium tuberculosis* drug resistance detection. *Int. J. Infect. Dis.* **2021**, *113*, S48–S54. [\[CrossRef\]](#)
24. Wang, L.; Yang, J.; Chen, L.; Wang, W.; Yu, F.; Xiong, H. Whole-genome sequencing of *Mycobacterium tuberculosis* for prediction of drug resistance. *Epidemiol. Infect.* **2022**, *150*, e22. [\[CrossRef\]](#)
25. Brown, A.C. Whole-Genome Sequencing of *Mycobacterium tuberculosis* Directly from Sputum Samples. In *Mycobacteria Protocols; Methods in Molecular Biology*; Springer: New York, NY, USA, 2021; pp. 459–480. [\[CrossRef\]](#)
26. Torres Ortiz, A.; Coronel, J.; Vidal, J.R.; Bonilla, C.; Moore, D.A.J.; Gilman, R.H.; Balloux, F.; Kon, O.M.; Didelot, X.; Grandjean, L. Genomic signatures of pre-resistance in *Mycobacterium tuberculosis*. *Nat. Commun.* **2021**, *12*, 7312. [\[CrossRef\]](#)
27. Freschi, L.; Vargas, R.; Husain, A.; Kamal, S.M.M.; Skrahina, A.; Tahseen, S.; Ismail, N.; Barbova, A.; Niemann, S.; Cirillo, D.M.; et al. Population structure, biogeography and transmissibility of *Mycobacterium tuberculosis*. *Nat. Commun.* **2021**, *12*, 6099. [\[CrossRef\]](#)
28. Iketleng, T.; Lessells, R.; Dlamini, M.T.; Mogashoa, T.; Mupfumi, L.; Moyo, S.; Gaseitsiwe, S.; de Oliveira, T. *Mycobacterium tuberculosis* Next-Generation Whole Genome Sequencing: Opportunities and Challenges. *Tuberc. Res. Treat.* **2018**, *2018*, 1298542. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Tonkin-Hill, G.; Corander, J.; Parkhill, J. Challenges in prokaryote pangenomics. *Microb. Genom.* **2023**, *9*, 001021. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Ferrés, I.; Iraola, G. Protocol for post-processing of bacterial pangenome data using Pagoo pipeline. *STAR Protoc.* **2021**, *2*, 100802. [\[CrossRef\]](#)
31. Brockhurst, M.A.; Harrison, E.; Hall, J.P.J.; Richards, T.; McNally, A.; MacLean, C. The Ecology and Evolution of Pangenomes. *Curr. Biol.* **2019**, *29*, R1094–R1103. [\[CrossRef\]](#)
32. Lapiere, P.; Gogarten, J.P. Estimating the size of the bacterial pan-genome. *Trends Genet.* **2009**, *25*, 107–110. [\[CrossRef\]](#)
33. Puyén, Z.M.; Santos-Lázaro, D.; Vigo, A.N.; Cotrina, V.V.; Ruiz-Nizama, N.; Alarcón, M.J.; Asto, B.; Huamán, T.; Moore, D.A.J. Whole Genome Sequencing of *Mycobacterium tuberculosis* under routine conditions in a high-burden area of multidrug-resistant tuberculosis in Peru. *PLoS ONE* **2024**, *19*, e0304130. [\[CrossRef\]](#)
34. Negrete-Paz, A.M.; Vázquez-Marrufo, G.; Gutiérrez-Moraga, A.; Vázquez-Garcidueñas, M.S. Pangenome Reconstruction of *Mycobacterium tuberculosis* as a Guide to Reveal Genomic Features Associated with Strain Clinical Phenotype. *Microorganisms* **2023**, *11*, 1495. [\[CrossRef\]](#)

35. Hurtado-Páez, U.; Álvarez Zuluaga, N.; Arango Isaza, R.E.; Contreras-Moreira, B.; Rouzaud, F.; Robledo, J. Pan-genome association study of *Mycobacterium tuberculosis* lineage-4 revealed specific genes related to the high and low prevalence of the disease in patients from the North-Eastern area of Medellín, Colombia. *Front. Microbiol.* **2023**, *13*, p1076797. [CrossRef]
36. Morey-León, G.; Andrade-Molina, D.; Fernández-Cadena, J.C.; Berná, L. Comparative genomics of drug-resistant strains of *Mycobacterium tuberculosis* in Ecuador. *BMC Genom.* **2022**, *23*, 844. [CrossRef]
37. Andrews, S. FASTQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 18 June 2022).
38. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. 2014. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
39. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **2014**, *15*, R46. [CrossRef]
40. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **2017**, *13*, e1005595. [CrossRef]
41. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]
42. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
43. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef]
44. Chaudhari, N.M.; Gupta, V.K.; Dutta, C. BPGA—An ultra-fast pan-genome analysis pipeline. *Sci. Rep.* **2016**, *6*, 24373. [CrossRef]
45. Kohl, T.A.; Utpatel, C.; Schleusener, V.; De Filippo, M.R.; Beckert, P.; Cirillo, D.M.; Niemann, S. MTBseq: A comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ* **2018**, *6*, e5895. [CrossRef]
46. Smith, I. *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. *Clin. Microbiol. Rev.* **2003**, *16*, 463–496. [CrossRef]
47. Olivença, F.; Nunes, A.; Macedo, R.; Pires, D.; Silveiro, C.; Anes, E.; Miragaia, M.; Gomes, J.P.; Catalão, M.J. Uncovering Beta-Lactam Susceptibility Patterns in Clinical Isolates of *Mycobacterium tuberculosis* through Whole-Genome Sequencing. *Microbiol. Spectr.* **2022**, *10*, e00674-22. [CrossRef]
48. Periwal, V.; Patowary, A.; Vellarikkal, S.K.; Gupta, A.; Singh, M.; Mittal, A.; Jeyapaul, S.; Chauhan, R.K.; Singh, A.V.; Singh, P.K.; et al. Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of *Mycobacterium tuberculosis* Pangenome. *PLoS ONE* **2015**, *10*, e0122979. [CrossRef]
49. Zakham, F.; Sironen, T.; Vapalahti, O.; Kant, R. Pan and Core Genome Analysis of 183 *Mycobacterium tuberculosis* Strains Revealed a High Inter-Species Diversity among the Human Adapted Strains. *Antibiotics* **2021**, *10*, 500. [CrossRef]
50. Warner, D.F.; Koch, A.; Mizrahi, V. Diversity and disease pathogenesis in *Mycobacterium tuberculosis*. *Trends Microbiol.* **2015**, *23*, 14–21. [CrossRef]
51. Gagneux, S. Genetic Diversity in *Mycobacterium tuberculosis*. In *Pathogenesis of Mycobacterium tuberculosis and Its Interaction with the Host Organism*; Pieters, J., McKinney, J.D., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 1–25. [CrossRef]
52. Rouli, L.; Merhej, V.; Fournier, P.E.; Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* **2015**, *7*, 72–85. [CrossRef]
53. Achtman, M. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. *Annu. Rev. Microbiol.* **2008**, *62*, 53–70. [CrossRef]
54. Dar, H.A.; Zaheer, T.; Ullah, N.; Bakhtiar, S.M.; Zhang, T.; Yasir, M.; Azhar, E.I.; Ali, A. Pangenome Analysis of *Mycobacterium tuberculosis* Reveals Core-Drug Targets and Screening of Promising Lead Compounds for Drug Discovery. *Antibiotics* **2020**, *9*, 819. [CrossRef]
55. Yang, T.; Zhong, J.; Zhang, J.; Li, C.; Yu, X.; Xiao, J.; Jia, X.; Ding, N.; Ma, G.; Wang, G.; et al. Pan-Genomic Study of *Mycobacterium tuberculosis* Reflecting the Primary/Secondary Genes, Generality/Individuality, and the Interconversion Through Copy Number Variations. *Front. Microbiol.* **2018**, *9*, 1886. [CrossRef]
56. Tonkin-Hill, G.; MacAlasdair, N.; Ruis, C.; Weimann, A.; Horesh, G.; Lees, J.A.; Gladstone, R.A.; Lo, S.; Beaudoin, C.; Floto, R.A.; et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **2020**, *21*, 180. [CrossRef]
57. Tettelin, H.; Riley, D.; Cattuto, C.; Medini, D. Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **2008**, *11*, 472–477. [CrossRef]
58. Sapriel, G.; Brosch, R. Shared Pathogenomic Patterns Characterize a New Phylotype, Revealing Transition toward Host-Adaptation Long before Speciation of *Mycobacterium tuberculosis*. *Genome Biol. Evol.* **2019**, *11*, 2420–2438. [CrossRef]

59. Kweon, O.; Kim, S.J.; Blom, J.; Kim, S.K.; Kim, B.S.; Baek, D.H.; Park, S.I.; Sutherland, J.B.; Cerniglia, C.E. Comparative functional pan-genome analyses to build connections between genomic dynamics and phenotypic evolution in polycyclic aromatic hydrocarbon metabolism in the genus *Mycobacterium*. *BMC Evol. Biol.* **2015**, *15*, 21. [\[CrossRef\]](#)
60. Behruznia, M.; Marin, M.; Farhat, M.; Thomas, J.C.; Domingo-Sananes, M.R.; Meehan, C.J. The *Mycobacterium tuberculosis* complex pangenome is small and driven by sub-lineage-specific regions of difference. *eLife* **2024**, *13*, RP97870.
61. Wivagg, C.N.; Hung, D.T. Resuscitation-Promoting Factors Are Required for β -Lactam Tolerance and the Permeability Barrier in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **2012**, *56*, 1591–1594. [\[CrossRef\]](#)
62. Hett, E.C.; Chao, M.C.; Deng, L.L.; Rubin, E.J. A Mycobacterial Enzyme Essential for Cell Division Synergizes with Resuscitation-Promoting Factor. *PLoS Pathog.* **2008**, *4*, e1000001. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.