Article

Spatial Distribution and Clustering of Drug-Resistant *Mycobacterium Tuberculosis* Infections in Rural Eastern Cape Province of South Africa

Lindiwe M. Faye [1], Mojisola C. Hosu [1], Sandeep Vasaikar [1], Anzaan Dippenaar [2], Selien Oostvogels [2] and Teke Apalata [1,]

[1] Department of Laboratory Medicine and Pathology, Faculty of Health Sciences and National Health Laboratory Services, Walter Sisulu University, Private Bag X5117, Mthatha 5099, Eastern Cape, South Africa; fayelindiwe@yahoo.com, mojisolahosu@gmail.com, sandeepvasaikar@yahoo.com,

[2] Family Medicine and Population Health (FAMPOP), Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; selienoostvogels@hotmail.com, anzaan.dippenaar@uantwerpen.be

* Correspondence: ruffinapalata@gmail.com ;+27-47-502-1995

Abstract: Tuberculosis (TB), an infectious airborne disease caused by *Mycobacterium tuberculosis* (Mtb), is a serious public health threat reported as the leading cause of morbidity and mortality worldwide. South Africa is a high TB burden country with TB being the highest infectious disease killer. The study investigated the distribution and clustering of Mtb mutations and spoligotypes in rural Eastern Cape Province. The Mtb isolates included were 1,157 from DR-TB patients and analysed by LPA followed by spoligotyping of 441 isolates, of these, 36 were whole genome sequenced. Distribution of mutations and spoligotypes was done by spatial analysis and clustering analysis was done by Bayesian model-based clustering of allele frequencies at heterozygous sites, using Mclust package in R. The *rpoB* gene had highest number of mutations. The distribution of *rpoB* and *katG* mutations was more prevalent in four health care facilities, *inhA* mutations were more prevalent in three healthcare facilities and heteroresistant isolates were more prevalent in five healthcare facilities. The Mtb was genetically diverse with Beijing more prevalent and largely distributed. Spatial analysis and mapping of gene mutations and spoligotypes revealed better picture of distribution. Clustering of isolates indicates that there is transmission of mixed infection in this area.

Keywords: tuberculosis; spatial analysis; mutations; spoligotypes; heteroresistance; cluster

1. Introduction

Tuberculosis (TB), a chronic inflammatory infectious disease caused by *Mycobacterium tuberculosis* (Mtb), is a serious public health threat and is reported as the leading cause of morbidity and mortality worldwide. It is easily transmitted from one person to another by airborne droplet nuclei [1,2]. In 2021, the most current global TB data reported the incidence of TB to be 10.6 million new cases with 1.6 million attributable fatalities globally [3]. The distribution of TB differs geographically both within counties and on different continents of the world. Africa accounts for 29% and 34% of all TB cases and deaths respectively worldwide with the highest recorded incidence rate of 275 cases per 100,000 people [4,5].

South Africa is a high TB burden country with TB being the highest infectious disease killer. According to the global burden of disease study, TB is the fifth leading cause of years of life lost (YLL) and disability-adjusted life-years (DALY) in the country [6]. With an estimated population of 60.6 million by the end of June 2022 [7], South Africa (SA) shares borders with six other nations, including Botswana, Lesotho, Namibia, Mozambique, Swaziland, and Zimbabwe where TB is also endemic. In 2019, the TB incidence in South Africa was estimated to be 615 cases per 100,000 population, ranging from 427 to 835 cases per 100,000 and with estimated 360,000 people who developed TB [8]. SA had

the second-highest absolute number of notified rifampicin (RIF)-resistant (RR) and multi-drug-resistant (MDR) cases globally with 18,734 cases reported in 2015 [9]. Eastern Cape is one of the three provinces in South Africa that have the highest TB incidence rates [1].

A better understanding of local geographic heterogeneity in routinely identified TB cases and the correlation of that heterogeneity with the location of undiagnosed prevalent cases may therefore be useful in directing active case-finding interventions to high-risk areas [10]. Furthermore, there is a need for accurate and early detection of drug-resistant TB (DR-TB) for minimizing the development of drug resistance, effective patient care and preventing the spread of DR strains [11].

Our knowledge of the epidemiology of TB on a local and global level has been substantially improved by the development and use of genotyping methods for Mtb; likewise, with the help of geospatial analytical technologies, the understanding of public health issues can be enhanced [12,13]. Spoligotyping combined with geospatial analytical methods such as geographic information systems (GIS) and DOTS strategy can help assess the transmission of Mtb strains and are promising essential tools for helping to understand the distribution of drug-resistant strains as well as the drivers of drug resistance [14]. Previous spatial studies have used GIS, whole genome sequencing (WGS) and spatial statistics to identify transmission hotspot areas with elevated risk for prioritisation of control and intervention measures [15-17]. These spatial studies have also shown that MDR-TB is clustered in specific geographical areas associated with location, socio-economic status and population density [4]. Even though there is an increase in the number of studies that use geospatial analytical methods to understand TB and other public health problems [18-20], however, the transmission dynamics of prevalent Mtb strains in rural Eastern Cape are not well understood.

Understanding such spatial variations in TB prevalence is crucial for improved targeting of interventions and resources for the prevention and management of TB in a particular area. The geographical distribution of MDR-TB in high TB burden settings is very important for effective control of TB epidemics. This can inform a basis for understanding DR gene migrations within populations since the frequency of mutations varies geographically. In this study, we sought to investigate the spatial patterns of Mtb and conduct a genomic clustering analysis in order to determine the transmission patterns and mixed-strains infections. We present the first spatial analysis of DR-TB and mutations associated with RIF and INH causing heteroresistance and spoligotype distributions and formation of clusters of Mtb in healthcare facilities (HCFs) of Mthatha and surrounding areas.

2. Methods

2.1. Study Design and setting

This was an ecological study design conducted in four districts (OR Tambo, Alfred Nzo, Amatole, Chris Hani) and one metropolitan municipality (Buffalo City) with 101 healthcare facilities in rural Eastern Cape Province (ECP), South Africa; the distribution of the healthcare facilities is as shown (Figure 1).

 doi:10.20944/preprints202212.0359.v1

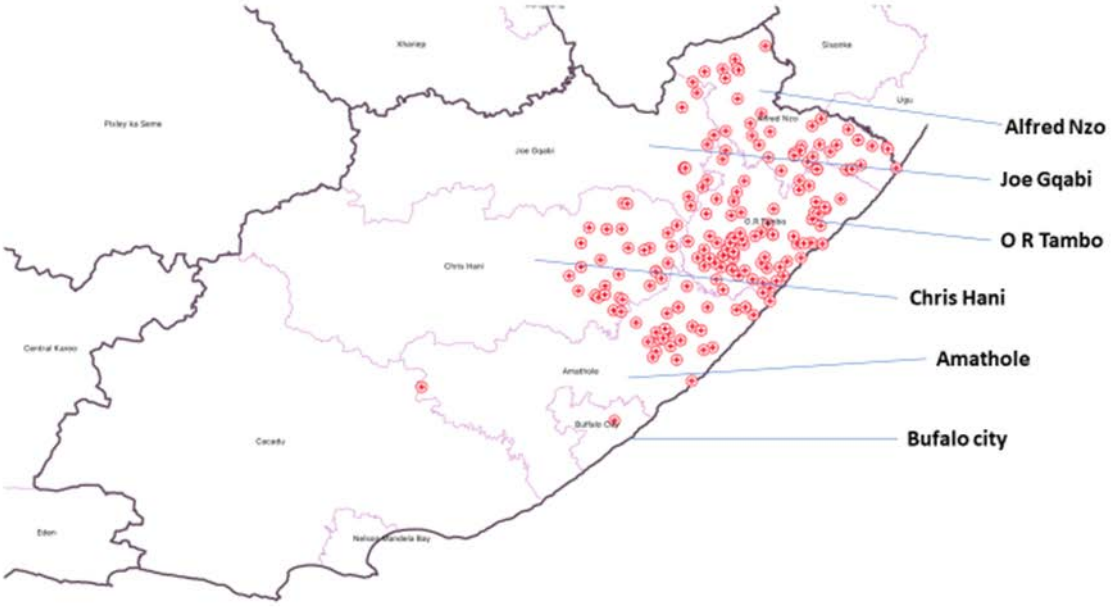

Figure 1. Distribution of healthcare facilities in the study area

This province is the second largest province in the country and serves a population of 7,130,480. O.R Tambo district is one of the 7 districts of the ECP, located on the coastline. The seat of O.R Tambo is in Mthatha. The population is about 1 676 463. Five local municipalities namely King Sabata Dalindyebo, Nyandeni, Mhlontlo, Port St Johns and Ingquza form O R Tambo district municipality.

2.2. Study population and sampling strategy

The study population included all DR-TB cases registered and living in the catchment area of the five municipalities' TB treatment centers and Nelson Mandela Academic Hospital NHLS diagnostic services catchment areas, registered between the years 2018 and 2020.

2.3. Data collection and isolates identification

Sputum samples were collected from consecutive clinically diagnosed pulmonary TB patients reporting to 101 selected health facilities within the district municipalities named above. The samples were submitted to National Laboratory Services (NHLS) TB Laboratory for diagnostic testing. The information of patients on the laboratory requisition form was recorded for spatial analysis in order to determine the distribution of mutations in *rpoB*, *katG* and *inhA* genes and their spoligotypes and formation of clusters among isolates. Isolates were collected over a 3-year period and investigated using first line probe assay (LPA) and WGS.

2.4. Isolates analysis

The GenoType MTBDRplus VER 2.0 is a DNA-strip based in-vitro assay for identifying the Mtb complex (MTBC) and its resistance to RIF and INH from smear positive pulmonary sputum samples and positive culture samples. DNA was extracted using Genolyse® kit (Hain Life Science GmbH, Nehren, Germany). The extracted DNA was pro-

cessed by the LPA using DRplus [21] to detect MTBC and RIF and/or INH resistance according to the manufacturer's instructions. Out of 1,157 isolates, 441 and 36 were randomly selected for spoligotyping and WGS to determine genotypes and clustering of Mtb isolates circulating in Mthatha and surrounding areas. Spoligotyping was performed using microbeads from TB-SPOL Kit (Beamedex®, Orsay, France), and the fluorescence intensity was measured using Luminex 200® (Austin, TX). Generated patterns were assigned to families using the SITVIT2 international database of the Pasteur Institute of Guadeloupe and compared [22].

WGS was performed on 36 DR-TB isolates. The analysis of pairwise genetic differences was performed using a cut-off ≤ 10 SNPs. Mixed infections were detected by estimating the number of Mtb strains present in a sample through Bayesian model-based clustering of allele frequencies at heterozygous sites, using the mclust package in R. For each sample, the major and minor allele frequencies of mapped reads at each heterozygous base call was calculated (Phred P error > 0.05) and used as a univariate input for clustering. The allele frequencies of heterozygous sites in mixed infection samples will cluster at similar frequencies in a set number of groups depending on the number and proportion of strains present. On the other hand, the allele frequencies of heterozygous sites in pure samples, though there may be a high number of heterozygous sites in samples with high clonal heterogeneity, will be more randomly distributed without clustering.

The LPA and spoligotype results of Mtb isolates were analysed for distribution of the mutations and spoligotypes using the QGIS 3.14 software. Clinics within hospitals with the same coordinates were merged in the analysis. We assessed the distribution of mutations in *rpoB, katG* and *inhA* as well as the distribution of heteroresistant strains and spoligotypes.

The WGS data was then analysed with the TB-profiler software (https://github.com/jodyphelan/TBProfiler). The software uses H37Rv as the default reference to determine the sequence variants and resistance profiles. Strain-specific single nucleotide variants were used for lineage identification and a curated library was used to infer drug resistance using TB-Profiler [23]. Raw WGS data of 36 Mtb clinical isolates were also analyzed with USAP, a pipeline developed by the TB-Genomics group. Briefly, reads were trimmed with Trimmomatic [24] using a sliding window approach and an average phred quality score of 20 and aligned to the Mtb H37Rv reference genome (GenBank NC000962.3) with BWA [25] SMALT [26] and Novoalign (Novocraft). Genomic variants (single nucleotide variants and 1-10 base pair insertions and deletions) identified in all three alignments with SAMTools [27] and the Genome Analysis Toolkit [28] were considered with high confidence.

Concatenated sequences of high confidence variable sites (single nucleotide polymorphisms [SNPs] only) were used to determine clustering among the study isolates. Genomic transmission clusters were evaluated under a 5 SNP threshold. Cluster analysis was done using the R-programming language, implementing the APE (Analyses of Phylogenetics and Evolution) and adegenet (Exploratory Analysis of Genetic and Genomic Data) packages.

3. Results

3.1. Distribution of mutations, lineages of isolates

A total of 1157 DR-TB, MDR-TB, extensively drug-resistant (XDR) TB and heteroresistant clinical isolates were isolated from the different healthcare facilities. All three (rpoB, katG and inhA) genes under investigation had one or more mutations from each isolate. The total number of mutations that occurred in rpoB gene were 761, representing the highest number of mutations, followed by katG gene with 683 mutations while inhA gene accounted for 286 mutations.

 doi:10.20944/preprints202212.0359.v1

The distribution of these rpoB mutations are shown in figure 2. Four health care facilities had more gene mutations namely HCF 3 (n = 83), HCF 2 (n = 77), HCF 4 (n = 60) and HCF 1 (n = 53).

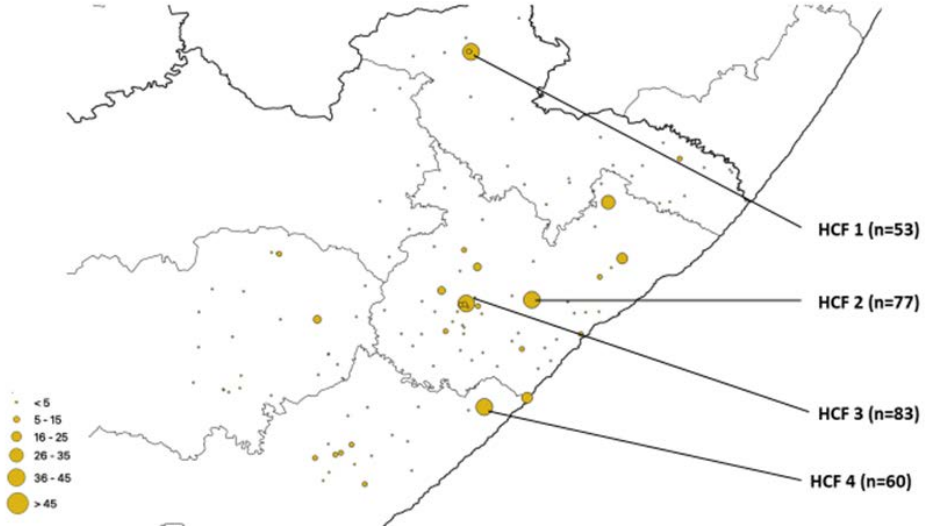

Figure 2. rpoB mutation distribution.

katG gene mutations is shown below (Figure 3). Four health care facilities had more mutations namely HCF 4, n =36, HCF 2, n = 42, HCF 1, n = 43 and HCF 3, n = 56.

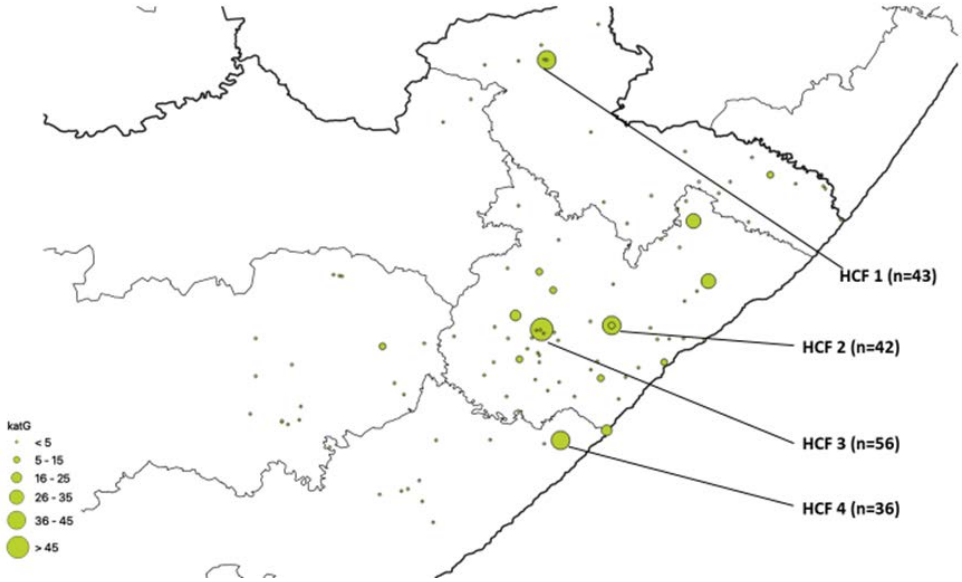

Figure 3. katG mutation distribution.

There were three healthcare facilities that had the highest number of inhA gene mutations, HCF 5, HCF 4 and HCF 2 with 28, 32 and 37 mutations respectively indicated in Figure 4.

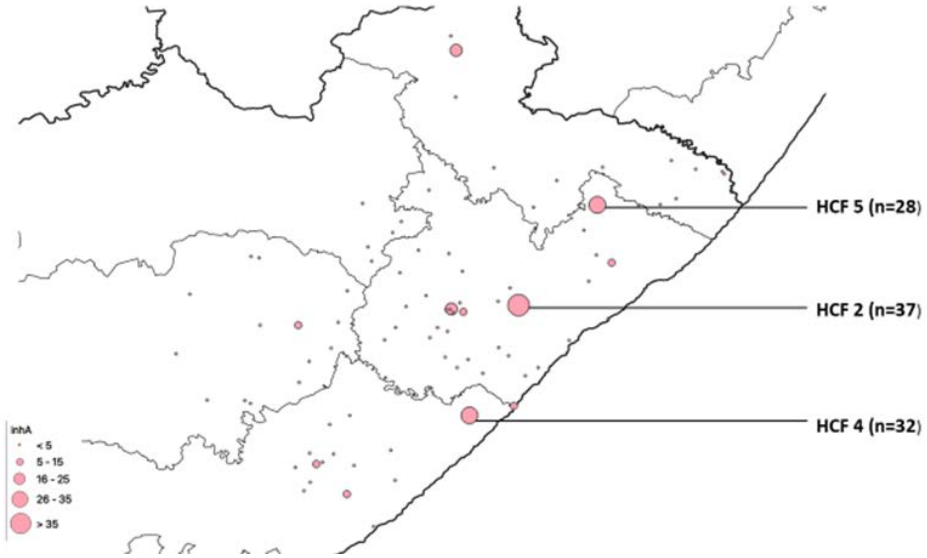

Figure 4. inhA mutation distribution.

The total number of heteroresistant isolates was 207, which was 17.9 % of the total number of isolates. These isolates had one and or more mutations in different genes. The following clinics had the most mutations in heteroresistant isolates, HCF 1, HCF 4, HCF 2, HCF 5 and HCF 3 with 78, 83, 93, 94 and 114 mutations respectively (Figure 5).

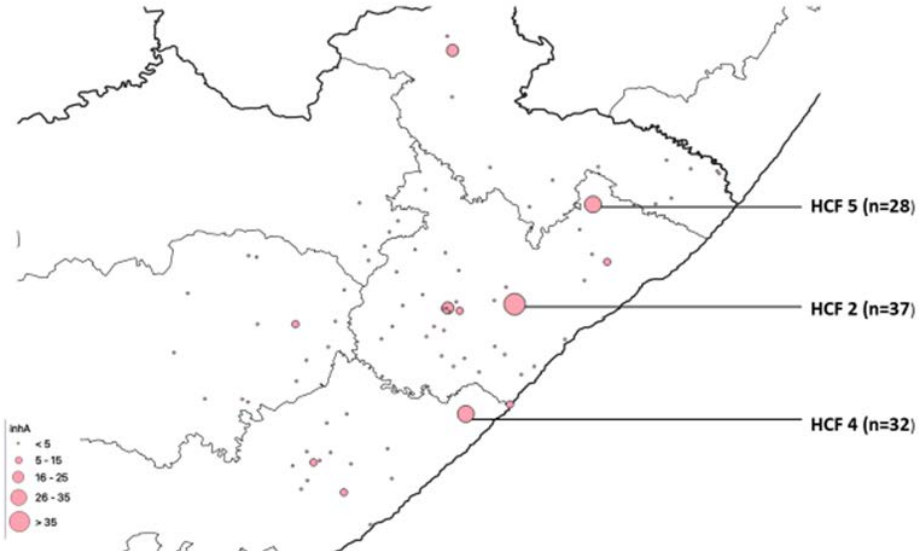

Figure 5. Geospatial distribution of heteroresistance in rural ECP.

The prevalent gene mutations in HCF within three municipalities were captured in the table below (Table 1)

Table 1. Distribution of prevalent gene mutations in HCF in rural ECP.

HCF	Municipality	*rpoB*	*katG*	*inhA*	Heteroresistant genes
1	Alfred Nzo	53	43	nil	78
2	O R Tambo	77	42	**37**	93
3	O R Tambo	83	56	0	114
4	Amathole	60	36	32	83

 doi:10.20944/preprints202212.0359.v1

5	O R Tambo	nil	nil	28	94

Based on the 441 Spoligotyping results, eight families were identified. Of 441 isolates queried for the lineage assignment, 59 (81.9%) were classified into the previously known lineages and 13 (18.1%) were not assigned to any known lineages. The Beijing family was the predominant group, representing 42.0% (n = 185) of all isolates, followed by the LAM family, 18.8% (n=83), X family,10.9% (n=48), T family, 7.7% (n=34), S family, 7.0% (n=31), EAI family, 3.6% (n=16), H, 1.4% (n=6) and CAS family, 1.1% (n=5) (Figure 6). Four (0.9%) isolates showed unknown patterns that were not assigned to any known major lineages in the SITVIT2 database, and 27 (6.1%) had no results.

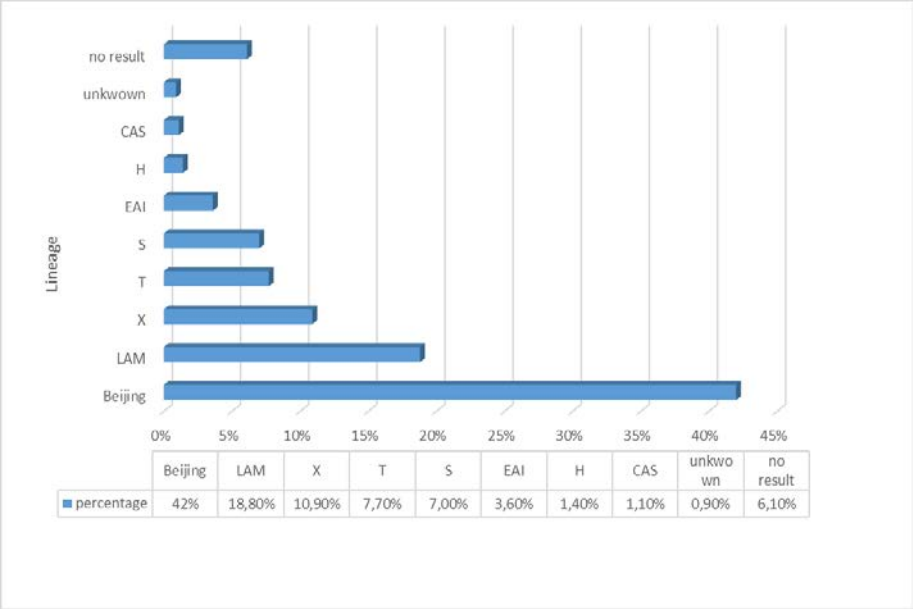

Figure 6. Distribution of Mtb sublineages/families (LAM: Latin American; EAI: East-African Indian; Delhi/CAS: Delhi/Central Asian, H: Haarlem).

Although spoligotyped isolates received from many HCFs of study setting, HCFs with greater than n=30 isolates were from two municipalities, namely OR Tambo (HCF 2, HCF 3 and HCF 6) and Buffalo city (HCF 4) (Figure 7).

 doi:10.20944/preprints202212.0359.v1

Figure 7. Geospatial distribution of spoligotyped isolates in healthcare facilities.

Our results showed that from the 441 clinical isolates that were spoligotyped the Beijing family strains accounted for 42.0% (185/441) of all the strains circulating in Mthatha and surrounding areas. The healthcare facilities with most Beijing family strains were HCF 2 (n=23), HCF 3 (n=18), HCF 5 (n=15) and HCF 6 (n=13) (Figure 8).

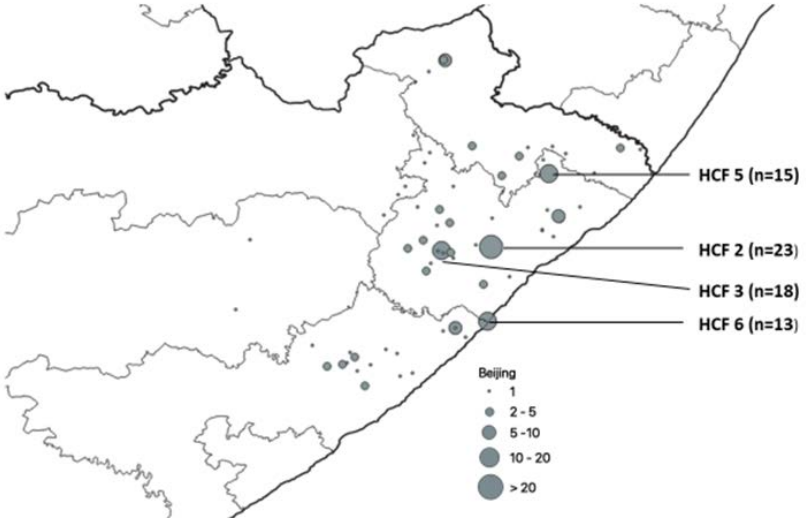

Figure 8. Geospatial distribution of Beijing family.

The LAM family being the second most prevalent was mostly identified in two HCFs (HCF 4 and HCF 5) (Figure 9).

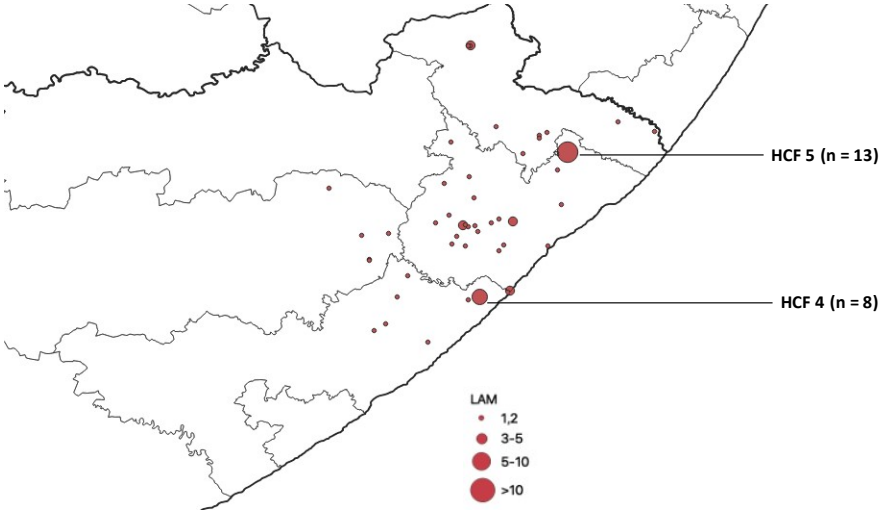

Figure 9. Geospatial distribution of LAM family in healthcare facilities in rural ECP.

3.2. Clustering analysis

Whole genome sequencing results analysis of pairwise genetic differences showed a high number of strains that differed by large amounts of single nucleotide polymorphisms (SNPs). Using a cut-off ≤ 10 SNPs, the cluster analysis of 28 Mtb strains (8 others were excluded because they failed quality control). Most strains were not related at genetic level

 doi:10.20944/preprints202212.0359.v1

with only 4 strains (Figures 10, 11, 13 and 14) forming part of 2 transmission clusters (clustering rate of 14.3%), each one comprising of two strains. These clusters consisted mainly of strains from Lineage 4 (Euro-American).

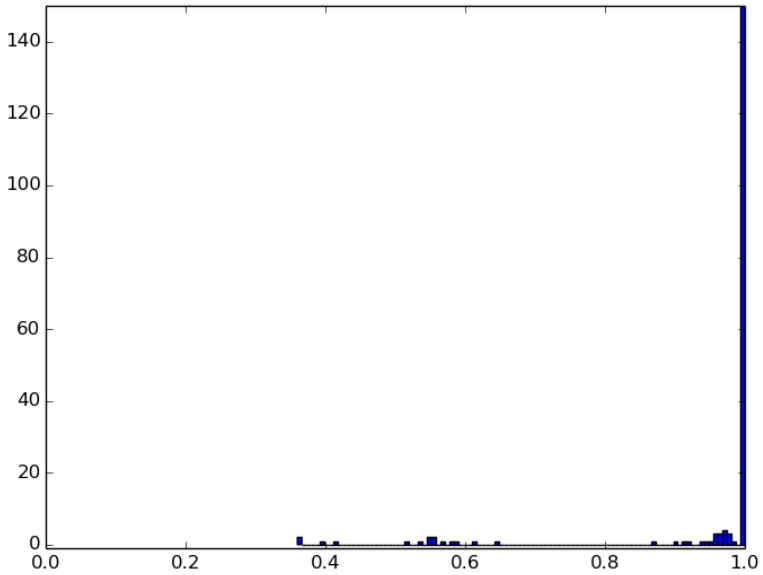

Figure 10. LF-8 hetero-graph.

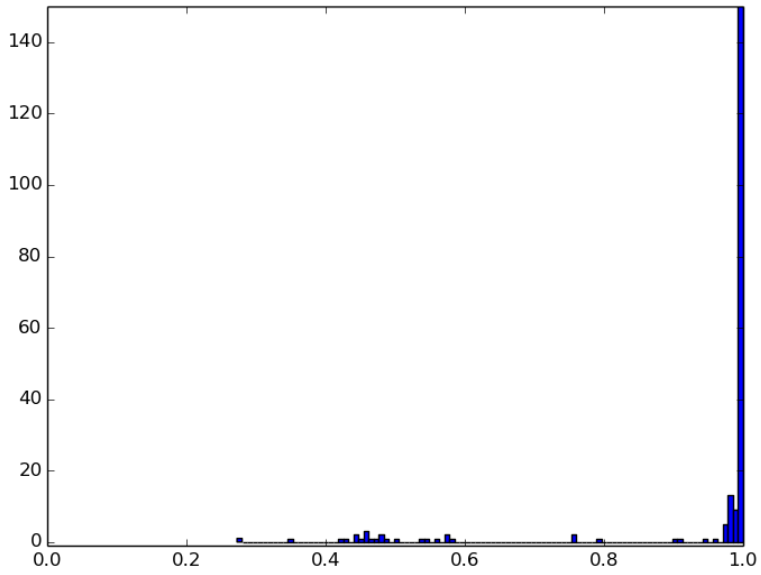

Figure 11. LF-25 hetero-graph.

Filtered variants were used to determine the variant distance between the isolates, using a variant distance of 12 for clustering. When applying a cut-off of 5, no clusters were observed. Clustering analysis suggests that isolates LF-8 and LF-25 (Figure 12) are closely related (distance of 6 single nucleotide variants). The analysis also shows LF-3 and LF-4 (Figure 12) to be related, both consist of mixed populations (Figures 13 and 14).

 doi:10.20944/preprints202212.0359.v1

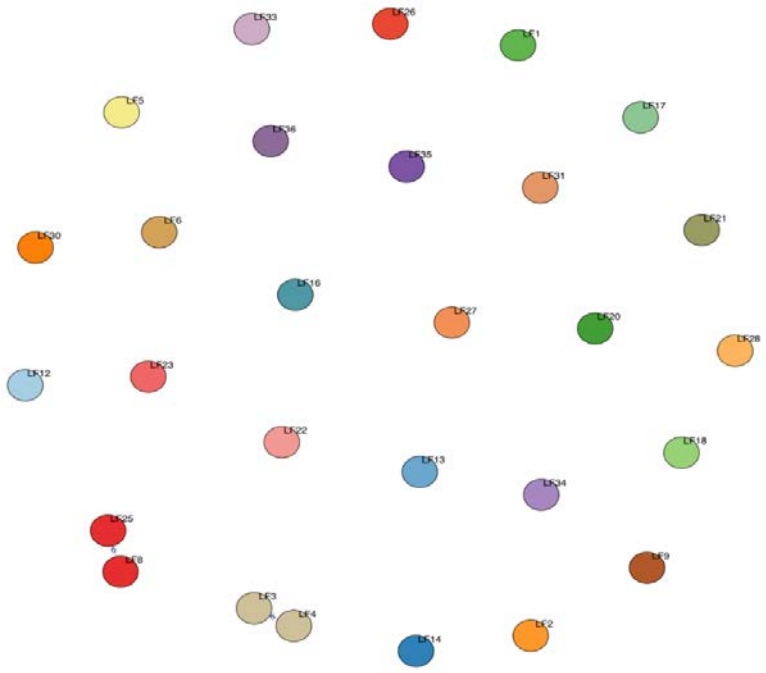

Figure 12. Clustering analysis of isolates.

In total 3/28 samples were classified as containing two different strains each with one sample that failed the quality control. The hetero graphs were generated using the frequency (a value between 0 and 1) of the detected variants for each isolate. A bimodal distribution suggested the presence of mixed infections (presence of two populations at distinct proportions). If the data points are centered on 0.5, it is likely that the two populations exist at more or less equal frequencies. LF-3 (Figure 13) isolate has two strain populations, one at a frequency of ±0.4 and another at ±0.6 whilst LF-4 (Figure 14) isolate has also two strain populations, but one at a frequency of ±0.3 and another ±0.7.

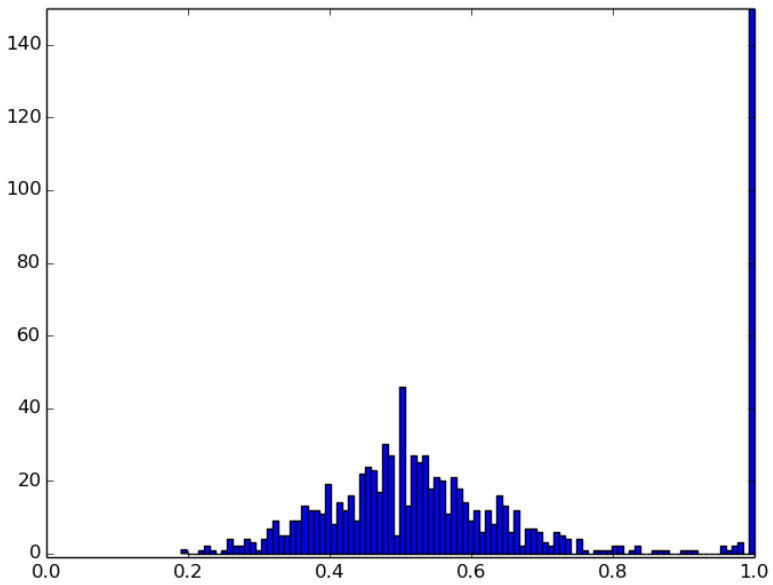

Figure 13. LF-3 hetero-graph.

 doi:10.20944/preprints202212.0359.v1

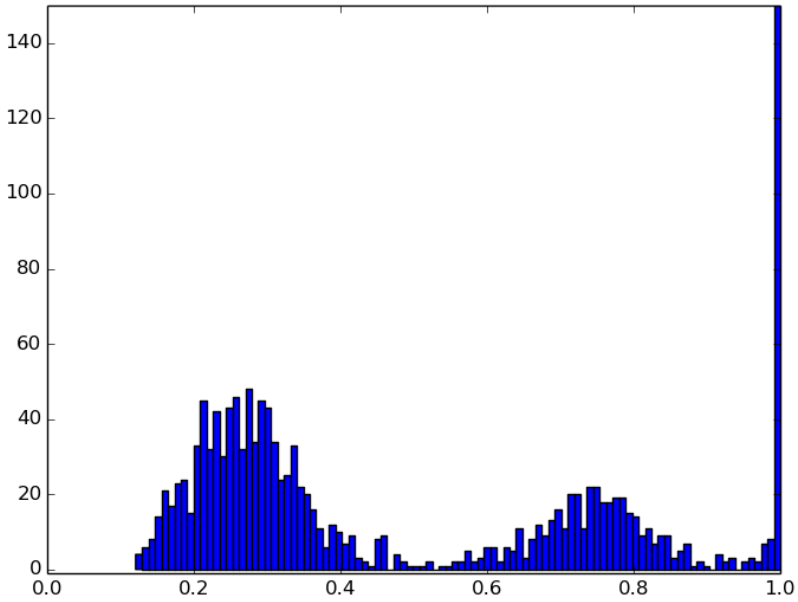

Figure 14. LF-4 hetero-graph.

4. Discussion

The substantial gap in the detection and treatment of MDR/RR-TB means that most patients are missed. Hence, identifying geographical areas with a high incidence of disease and adopting active case finding in such areas could help to reduce detection gap among DR-TB patients. By highlighting high-burden regions with poor public health initiatives and low case detection rates, knowledge of the spatial distribution of tuberculosis could help with control and prevention efforts. This identification enables decision-makers to undertake targeted measures for prevention and management of MDR-TB hotspots in order to stop the spread of the disease [4,29,30]. This may be crucial in areas with limited resources and nations with high MDR-TB burden [4]. This is the first epidemiological study in Eastern Cape combining GIS analysis with molecular-based methods to describe the distribution of DR-TB by mapping the distribution of DR gene mutations and genotypes.

Strategic measures in combating the spread of DR-TB include active surveillance, screening for DR-TB, early isolation, and management of confirmed cases [31]. The first step in easing the burden of DR-TB is to comprehend the geographical distribution of the disease and identify regions with the greatest prevalence of notified cases [29]. Drug resistance in Mtb is singularly mediated by chromosomal mutations [31]. Our study detected mutations in *rpoB, katG* and *inhA* from the four municipalities. Most of the mutations are concentrated in OR Tambo district municipality. This may be because it is more populous than the other municipalities. HCF 5, HCF 2 and HCF 4 had the highest number of mutations in *inhA* gene (Figure 4), this means INH can still be used but in high dose for treatment of TB in patients from the catchment areas of the clinic, however due to development of mutations in genes during treatment, treatment management is needed. Published evidence reviewing the frequency and distribution of gene mutations in Africa reported that out of the 22 gene mutations reported from 25 countries in Africa, *rpoB* ranked the highest. It further reported that mutations of *rpoB*, *katG*, *gyrA* and *inhA* are documented in almost all regions of Africa, which is an indication of the widespread of rifampicin- and isoniazid-resistant TB throughout the entire continent. Hence, TB treatment in Africa using rifampicin, isoniazid, fluoroquinolone and bedaquiline should be handled with caution [32].

The coexistence of both drug-susceptible and drug resistant bacteria within the same patient was observed at 17.9% of the total isolates of the study with 5 HCF that had most genes; the *rpoB* gene (RIF resistance) the *katG* gene (high-level INH resistance) and the promoter region of the *inhA* gene (low-level INH resistance) with heteroresistance. This may contribute to the difficulty in treating tuberculosis, as it is the precursor to full resistance [33-35]. Heteroresistance is a result of mixed infection, with two or more Mtb distinct strains in the same patient or the presence of different subpopulations caused by the microevolution of the single strain within the host [36]. This may endanger the effective treatment of patients with both RIF and INH thereby leading to the development of anti-TB drug resistance in the study area, which underscores surveillance of heteroresistance from patients prior and after treatment. The changing drug resistance patterns detected in patients with tuberculosis also confirmed the possibility that heteroresistance can persist over a long period [37]. Studies on heteroresistance have reported a prevalence ranging from 20% to 57% [38-40], which was higher than our study but observed to be increasing each year. Most of the studies focused on samples taken before treatment was initiated, presumably to show that the presence of heteroresistance should be considered in formulating treatment regimens. In our study, the samples were collected during the treatment period, which means that heteroresistance can be present even if the patient is on treatment and agrees with van Rie et al. [37] who reported persistence of heteroresistance over a long period. The detection of heteroresistance is vital to preventing the selection of drug resistance during antibiotics treatment [41].

Genomic analysis of Mtb clinical isolates worldwide has revealed differences in the geographical distributions and apparent host preference of distinct phylogenetic lineages [22]. According to [43-45] genomic analysis helps to identify clinical features of predominant or emerging genotypes and are important in public health perspective because they describe epidemiological associations with outbreaks and transmission routes. Studies that are investigating relationship among Mtb across different geographical areas are impacting positively the programs set to end TB because they help to understand transmission of TB [46].

Of 441 isolates spoligotyped, 437 revealed distinct spoligotype patterns. Patterns of 410 (93.1%) isolates matched a pre-existing SIT in the SITVIT2 database, while 27 unique patterns (6.1%) were not in the database. The Mtb population in this study area was genetically diverse with the Beijing lineage and its members, which is regarded as a successful clone of Mtb that is associated with drug resistance in some parts of the World [47]. A comparison of lineages among all clinics/ hospitals in the study shows that the Beijing family was the only genotype found in all the hospitals/ clinics, with HCF 2, HCF 3, HCF 5 and HCF 6 having the highest number of Beijing isolates. The Beijing family is known to be more transmissible than other families and more prevalent in Eastern Cape and Western Cape Provinces which are neighbors [48,49] this confirms its high prevalence in this study (42%). This family has been detected in studies reported from other parts of South Africa including Limpopo, Western Cape and Mpumalanga [49,50]. Nelson et al. [51] reported that human movement between rural and urban areas in search of employment in South Africa is common and serves as a bridge for transporting pathogen across long distances. LAM family was the second prevalent. This strain is widely distributed in KwaZulu Natal which is a neighboring province to Eastern Cape [52].

Table 2 compares the distribution of lineage of Mtb from our study with that in other studies including other provinces in South Africa and Sub-Saharan Africa. The Beijing family belonging to Lineage 2 is the most prevalent in our study, which is the same with Western Cape, Gauteng and North-West but a different outlook is portrayed in countries like Zambia and Botswana with LAM family predominating. Major lineages, L1 to L7, have been identified from analyses of Mtb strains worldwide [53-55], but recently this has been updated to include Lineages 8 and 9 [55]. Lineage 2 includes strains, the majority of which are members of the so-called Beijing family. More than a quarter of the world's

tuberculosis epidemic is attributable to the Beijing family, which is the most prolific genotype of Mtb. The widespread proliferation of this strain family in recent decades, its propensity to spread disease, association with drug resistance, treatment failure, early relapse, recurrence, fever during early therapy and increased risk of transmission chains globally have all garnered considerable attention according to reports from several clinical trials. Evidence from both experimental and clinical data points to Beijing strains' hypervirulent phenotype and increased mutation rate when compared to other strains [53,55,56].

Table 2. Distribution of Mtb lineages in this study in comparison with other studies

		Our study n (%)	**Western Cape n (%) [49]**	**Gauteng n (%) [49]**	**KZN n (%) [49]**	**Free State n (%) [57]**	**Limpopo n (%) [50]**	**North-West n (%) [49]**	**Zambia n (%) [58]**	**Botswana n (%) [59]**
Lineage	Family	441	897	142	230	86	226	358	274	458
2	Beijing	185 (42)	599 (66.8)	44 (31.0)	57 (24.8)	5 (5.8)	34 (15.0)	88 (24.6)	1 (0.4)	41 (9.0)
4	LAM	83 (18.8)	53 (5.9)	29 (20.4)	42 (18.3)	18 (20.9)	60 (26.5)	54 (15.1)	149 (54.4)	150 (32.8)
4	X	48 (10.9)	88 (9.8)	9 (6.3)	14 (6.1)	5 (5.8)	12 (5.3)	27 (7.5)	19 (6.9)	75 (16.4)
4	T	34 (7.7)	61(6.8)	18 (12.7)	29 (12.6)	14 (16.3)	43 (19.0)	60 (16.8)	39 (14.2)	73 (15.9)
4	S	31 (7.0)	23 (2.6)	9 (6.3)	49 (21.3)	6 (7.0)	21 (9.1)	37 (10.3)	4 (1.5)	62 (13.5)
1	EAI MANU	16 (3.6) nil	6 (0.7) Nil	12 (8.5) 3 (2.1)	6 (2.6) 2 (0.9)	Nil nil	11 (4.9) 3 (1.3)	24 (6.7) 6 (1.7)	6 (2.2)	31 (6.8) 2 (0.4)
4	H	6 (1.4)	10(1.1)	6 (4.8)	8 (3.5)	1 (1.2)	31 (13.7)	26 (7.3)	nil	21 (4.6)
3	CAS	5 (1.1)	8 (0.9)	2 (1.4)	5 (2.2)	nil	10 (4.4)	6 (1.7)	44 (16.1)	2 (0.4)
3	U	nil	7 (0.8)	nil	1 (0.4)	nil	1 (0.4)	1 (0.3)	nil	3 (0.7)

Mtb genotypes differ amongst populations and are strongly influenced by geography. In terms of host immune response modulation, transmissibility, and disease severity, different Mtb lineages frequently exhibit distinct traits and virulence profiles. A better understanding of phenotypic variations caused by the genetic diversity of Mtb strains is important when attempting to improve TB control measures [55]. Previous research has revealed that immune responses are significantly variable amongst genetically diverse Mtb strains. Lineage 2 Mtb strains are the most virulent and were shown to only elicit a weak immune response in mice. Evidence revealed that patients who were infected with Lineage 2 strains were more likely to die of TB compared to patients infected with other strains. Investigating the pathogenicity of distinct lineages of Mtb is therefore crucial [56]

Clustering of four cases in our study is evidence of the presence of continuous emergence and transmission of MDR-TB in our study setting which underscores the investigation of effectiveness of systems currently in place to identify if MDR-TB strains are shared or clustered among TB patients. The two clusters that were observed from four cases were characterized by strains from Euro-American family. Our study setting is a high TB setting; this agrees that there is high transmission happening due to person-to-person transmission.

5. Conclusions

The key control for DR-TB is to interrupt its transmission, this was done by identifying hotspots of gene mutations, lineages especially those that are drivers of DR-TB transmission, cluster formations and mixed infections. This study highlights the relevance and usefulness of performing whole genome sequencing as a high-resolution method to carry out genetic analysis of DR-TB strains circulating in rural ECP as it will facilitate understanding the transmission dynamics of DR-TB and guide the implementation of effective

control measures. The identification of areas where DR-TB is concentrated could assist policy makers to implement targeted interventions aimed at prevention and management of TB transmission. This is particularly important in resource-limited settings and in high DR-TB burden areas like rural areas of Eastern Cape. Targeted interventions to rural community may be necessary as these areas find it impossible to provide DR-TB services across the communities as the diagnosis and treatment of DR-TB is challenged by factors such as poverty and co-infection with HIV.

Authors' Contributions: Conceptualization, L.M.F. and T.A.; methodology, L.M.F.; formal analysis, L.M.F.; investigation, L.M.F.; resources, L.M.F.; data curation, A.D. and S.O.; writing—original draft preparation, L.M.F. and M.C.H.; writing—review and editing, T.A., L.M.F. and M.C.H.; visualization, L.M.F. and M.C.H.; supervision, S.V. and T.A.; project administration, L.M.F.; funding acquisition, L.M.F. All authors have read and agreed to the published version of the manuscript.

Funding: Financial support for the study was obtained from the South African Medical Research Council (SAMRC) Pilot grant.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics and Biosafety Committee of the Faculty of Health Sciences of Walter Sisulu University (Ref. No. 026/2019) and Eastern Cape Department of Health (Ref No EC_201904_011).

Informed Consent Statement: Not applicable. The study used routine samples received in National Health Laboratory Services (NHLS) TB laboratory.

Data Availability Statement: Data can be requested from the corresponding author.

Acknowledgments: The authors are grateful to the NHLS TB laboratory staff and participating clinics for their support during sample analysis and data collection.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. World Health Organization. Global Tuberculosis Report **2019**. https://apps.who.int/iris/bitstream/handle/10665/329368/9789241565714 [Assessed 21 January 2022].
2. Olmo-Fontánez, A.M.; Turner, J. Tuberculosis in an Aging World. *Pathogens* **2022,** 11(10), 1101
3. WHO Global tuberculosis report. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO. [accessed 28 September 2022].
4. Alene, K.A.; Viney, K.; McBryde, E.S.; Clements, A.C.A. Spatial patterns of multidrug resistant tuberculosis and relationships to socio-economic, demographic and household factors in northwest Ethiopia. *PLoS ONE* **2017**, 12, e0171800.
5. Maja, T.F.; Maposa, D. An Investigation of Risk Factors Associated with Tuberculosis Transmission in South Africa Using Logistic Regression Model. *Infect. Dis. Rep*. **2022**, 14, 609-620.
6. Murray, C.J.; Vos, T.; Lozano, R.; Naghavi, M.; Flaxman, A.D.; Michaud, C.; Ezzati, M.; Shibuya, K.; Salomon, J.A.; Abdalla, S.; Aboyans, V. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, 380, 2197-2223.
7. Statistics South Africa. Mortality and causes of death in South Africa. Findings from death notification. Pretoria, South Africa. Stats SA's June 2021 report [Assessed 21 August 2022].
8. World Health Organisation. The Global tuberculosis report 2020. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO. Available: https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf [accessed 28 September 2022].
9. World Health Organization Global Tuberculosis Report 2016. Geneva World Health Organization [accessed 15 June 2022].
10. Robsky, K.O.; Kitonsa, P.J.; Mukiibi, J.; Nakasolya, O.; Isooba, D.; Nalutaaya, A.; Salvatore, P.P.; Kendall, E.A.; Katamba, A.; Dowdy, D. Spatial distribution of people diagnosed with tuberculosis through routine and active case finding: a community-based study in Kampala, Uganda. *Infect. Dis. Poverty* **2020**, 9, 73.
11. Nguyen, T.N.; Berre, A.L.; Bañuls, A.L.; Nguyen, T.V. Molecular diagnosis of drug-resistant tuberculosis; a literature review. *Front Microbiol*. **2019**, 10, 794.
12. Chihota, V.N.; Niehaus, A.; Streicher, E.M.; Wang, X.; Sampson, S.L.; Mason, P.; Källenius, G.; Mfinanga, S.G.; Pillay, M.; Klopper, M.; Kasongo, W. Geospatial distribution of Mycobacterium tuberculosis genotypes in Africa. *PLoS One* **2018**, 13(8), e0200632.

13. Cui, Z.; Lin, D.; Chongsuvivatwong, V.; Zhao, J.; Lin, M.; Ou, J.; Zhao, J. Spatiotemporal patterns and ecological factors of tuberculosis notification: A spatial panel data analysis in Guangxi, China. *PLoS One* **2019**, 14(5), e0212051.
14. Porter, J.D.H.; Ogden, J.A.; Pronyk, P. The way forward: an integrated approach to tuberculosis control. In *Tuberculosis – an Interdisciplinary Perspective*, Porter J.D.H., Grange, J.M., Eds.; Imperial College Press, London, 1999; pp. 359-378.
15. Asare, P.; Otchere, I.D.; Bedeley, E.; Brites, D.; Loiseau, C.; Baddoo, N.A.; Asante-Poku, A.; Osei-Wusu, S.; Prah, D.A.; Borrell, S.; Reinhard, M. Whole genome sequencing and spatial analysis identifies recent tuberculosis transmission hotspots in Ghana. *Front. Med*. **2020**, 7, 161.
16. Gwitira, I.; Karumazondo, N.; Shekede, M.D.;, Sandy, C.; Siziba, N.; Chirenda, J. Spatial patterns of pulmonary tuberculosis (TB) cases in Zimbabwe from 2015 to 2018. *PLoS One* **2021**, 16(4), e0249523.
17. Brooks, M.B.; Millones, A.K.; Puma, D.; Contreras, C.; Jimenez, J.; Tzelios, C.; Jenkins, H.E.; Yuen, C.M.; Keshavjee, S.; Lecca, L.; Becerra, M.C. Mapping local hot spots with routine tuberculosis data: A pragmatic approach to identify spatial variability. *PLoS One* **2022**, 7(3), e0265826.
18. Caprarelli, G.; Fletcher, S. A brief review of spatial analysis concepts and tools used for mapping, containment and risk modelling of infectious diseases and other illnesses. *Parasitology* **2014**, 141(5), 581-601.
19. Chan-Yeung, M.; Yeh, A.G.; Tam, C.M.; Kam, K.M.; Leung, C.C.; Yew, W.W.; Lam, C.W. Socio-demographic and geographic indicators and distribution of tuberculosis in Hong Kong: a spatial analysis. *Int. J. Tuberc. Lung Dis.* **2005**, 9(12), 1320-1326.
20. Zulu, L.C.; Kalipeni, E.; Johannes, E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010. *BMC Infect. Dis*. **2014**, 14, 285.
21. Hain Lifescience. Company history and product releases. http://www.hain-lifescience.de/en/company/history.html. 2016 (Accessed 11 February 2021).
22. Couvin, D.; David, A.;, Zozio, T.; Rastogi, N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the Mycobacterium tuberculosis genotyping database. *Infect. Genet. Evol.* **2019**, 72, 31-43.
23. 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.
24. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, 30, 2114-2120.
25. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform *Bioinformatics*, **2009**, 25(14), 1754-1760
26. Ponstingl, H.; Ning, Z. SMALT - A new mapper for DNA sequencing reads. *F1000Posters* **2010**, L313.
27. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, 25(16), 2078-2079.
28. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M,.; DePristo, M.A. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **2010**, 20(9), 1297-1303.
29. Alene, K.A.; Xu, Z.; Bai, L.; Yi, H.; Tan, Y.; Gray, D.; Viney, K.; Archie, C.A.; Clements, A.C.A. Spatial clustering of drug-resistant tuberculosis in Hunan province, China: an ecological study. *BMJ Open* **2021**, 11, e043685.
30. Tafess, K.; Beyen, T.K.; Girma, S.; Girma, A.; Siu, G. Spatial clustering and genetic diversity of Mycobacterium tuberculosis isolate among pulmonary tuberculosis suspected patients, Arsi Zone, Ethiopia. *BMC Pulm. Med.* **2021**, 21, 206.
31. Kumwenda, G.; Chipungu, G.; Sloan, D.; Kaimila, Y.; Chiumya, K.; Pangani, H. The occurrence and frequency of genomic mutations that mediate Isoniazid and Rifampicin resistance in Mycobacterium tuberculosis isolates from untreated pulmonary Tuberculosis cases in urban Blantyre, Malawi. *Malawi Med. J.* **2018**, 30, 1-5.
32. Nam. B.H.; Nishizato, K.; Sugimoto, K.; Ueki, M.; Karibe, R.; Yashio, T.; Nishi, Y.; Munkhtuya, T. Frequency and distribution of drug-resistant tuberculosis gene mutations in Africa between 2010 and 2020. *South Afr. J Pub. Health* **2022**, 5, 60-64.
33. Chen, Y.; Jiang, Q.; Zou, J.; Yang, T.; Liu, Q.; Luo, G.; Gan, M.; Jiang, Y.; Takiff, H.E.; Lu, L.; Gao, Q. Deep whole-genome sequencing reveals no evidence for heteroresistance influencing treatment outcomes among drug-susceptible tuberculosis patients. *Tuberculosis* **2021**, 130, 102120.
34. Tessema, B.; Beer, J.; Emmrich, F., Sack, U.; Rodloff, A.C. Analysis of gene mutations associated with isoniazid, rifampicin and ethambutol resistance among Mycobacterium tuberculosis isolates from Ethiopia. *BMC Infect. Dis.* **2012**, 12, 37.
35. Brhane, M.; Kebede, A.; Petros, Y. Molecular detection of multidrug-resistant tuberculosis among smear-positive pulmonary tuberculosis patients in Jigjiga town, Ethiopia. *Infect. Drug Resist*. **2017**, 10, 75-83.
36. Liu, Q.; Via, L.E.; Luo, T.; Liang, L.; Liu, X.; Wu, S.; Shen, Q.; Wei, W.; Ruan, X.; Yuan, X.; Zhang, G. Within patient microevolution of Mycobacterium tuberculosis correlates with heterogeneous responses to treatment. *Sci. Rep*. **2015**, 5, 17507.
37. van Rie, A.; Victor, T.C.; Richardson, M.; Johnson, R.; van der Spuy, G.D.; Murray, E.J.; Beyers, N.; van Pittius, N.C.G.; van Helden, P.D.; Warren, R.M. Reinfection and mixed infection cause changing Mycobacterium tuberculosis drug-resistance patterns. *Am. J. Respir. Crit. Care Med*. **2005**, 172, 636 – 642.
38. Shin, S.S.; Modongo, C., Baik, Y., Allender, C.; Engelthaler, D.M.; Warren, R.M.; Klausner, J.; Zetola, N. Deep sequencing to estimate the prevalence of heteroresistance among mixed-strain and single-strain Mycobacterium tuberculosis infections. *Am. J, Respr. Crit. Care Med.* **2017**, 195, A1177.

39. Operario, D.J., Koeppel, A.F.; Turner, S.D.; Bao, Y.; Pholwat, S.; Banu, S.; Foongladda, S.; Mpagama, S.; Gratz, J.; Ogarkov, O.; Zhadova, S.; Heysell, S.K.; Houpt, E.R. Prevalence and extent of heteroresistance by next generation sequencing of multidrug-resistant tuberculosis. *PLoS One*, **2017**, 12, e0176522.
40. Hofmann-Thiel, S.; van Ingen, J.; Feldmann, K.; Turaev, L.; Uzakova, G.T.; Murmusaeva, G., van Soolingen, D.; Hoffmann, H. Mechanisms of heteroresistance to isoniazid and rifampin of Mycobacterium tuberculosis in Tashkent, Uzbekistan. *Eur. Respir. J.* **2009**, 33, 368–374.
41. Rigouts, L.; Miotto, P.; Schats, M.; Lempens, P.; Cabibbe, A.M.; Galbiati, S.; Lampasona, V.; de Rijk, P., Cirillo, D.M.; de Jong, B.C. Fluoroquinolone heteroresistance in Mycobacterium tuberculosis: detection by genotypic and phenotypic assays in experimentally mixed populations. *Sci. Rep.* **2019**, 9, 11760.
42. Couvin, D.; David, A.; Zozio, T.; Rastogi, N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the Mycobacterium tuberculosis genotyping database. *Infect. Genet. Evol.* **2019**, 72, 31-43.
43. Caws, M.; Thwaites, G.; Dunstan, S.; Hawn, T.R.; Lan, N.T.; Thuong, N.T.; Stepniewska, K.; Huyen, M.N., Bang, N.D.; Loc, T.H.; Gagneux, S.; van Soolingen, D.; Kremer, K.; van der Sande, M.; Small, P.; Anh, P.T.; Chinh, N.T.; Quy, H.T.; Duyen, N.T.; Tho, D.Q.; Hieu, N.T.; Torok, E.; Hien, T.T.; Dung, N.H.; Nhu, N.T.; Duy, P.M.; van Vinh Chau, N.; Farrar, J. The Influence of Host and Bacterial Genotype on the Development of Disseminated Disease with Mycobacterium tuberculosis. *PLoS Pathog*. **2008**, 4, e1000034.
44. Thwaites, G.; Caws, M., Chau, T.T.; D'Sa, A., Lan, N.T.; Huyen, M.N.; Gagneux, S.; Anh, P.T.; Tho, D.Q.; Torok, E.; Nhu, N.T.; Duyen, N.T.; Duy, P.M.; Richenberg, J.; Simmons, C.; Hien, T.T.; Farrar, J. Relationship between Mycobacterium tuberculosis Genotype and the Clinical Phenotype of Pulmonary and Meningeal Tuberculosis. *J Clin. Microbiol*, **2008**, 46, 1363-1368.
45. Von Groll, A., Martin, A.; Felix, C.; Prata, P.F.S.; Honscha, G., Portaels, F., Vandame, P.; da Silva, P.E.; Palomino, J.C. Fitness study of the RD^{Rio} lineage and Latin American- Mediterranean family of Mycobacterium tuberculosis in the city of Rio Grande, Brazil. *FEMS Immunol Med Microbiol*. **2010**, 58, 119-127.
46. Maung, H.M.; Palittapongarnpim, P.; Aung, H.L.; Surachat, K.; Nyunt, W.W., Chongsuvivatwong, V. Geno-Spatial Distribution of Mycobacterium Tuberculosis and Drug Resistance Profiles in Myanmar–Thai Border Area. *Trop. Med. Infect. Dis.* **2020**, 5(4), 153.
47. Hu, Y.; Mathema, B.; Zhao, Q.; Zheng, X.; Li, D.; Jiang, W.; Wang, W.; Xu, B. Comparison of the socio-demographic and clinical features of pulmonary TB patients infected with sub-lineages within the W-Beijing and non-Beijing Mycobacterium tuberculosis. *Tuberculosis* **2016**, 97, 18-25.
48. Karmakar, M.; Trauer, J.M.; Ascher, D.B.; Denholm, J.T. Hyper transmission of Beijing lineage Mycobacterium tuberculosis: systematic review and meta-analysis. *J. Infect*. **2019**, 79(6), 572-581.
49. Said, H.; Ratabane, J.; Erasmus, L.; Gardee, Y.; Omar, S.; Dreyer, A.; Ismail, F.; Bhyat, Z.; Lebaka, T.; van der Meulen, M.; Gwala, T. Distribution and Clonality of drug-resistant tuberculosis in South Africa. *BMC Microbiol*. **2021**, 21, 157.
50. Maguga-Phasha, N.T.; Munyai, N.S.; Mashinya, F., Makgatho, M.E., Mbajiorgu, E.F. Genetic diversity and distribution of Mycobacterium tuberculosis genotypes in Limpopo, South Africa. *BMC Infect. Dis.* **2017**, 17, 764.
51. Nelson, K.N.; Shah, N.S.; Mathema, B.; Ismail, N.; Brust, J.C.; Brown, T.S.; Auld, S.C.; Omar, S.V.; Morris, N.; Campbell, A.; Allana, S. Spatial patterns of extensively drug-resistant tuberculosis transmission in KwaZulu-Natal, South Africa. *J. Infect. Dis*. **2018**, 218(12), 1964-1973.
52. Pillay, M.; Sturm, A.W. Evolution of the extensively drug-resistant F15/LAM4/KZN strain of Mycobacterium tuberculosis in KwaZulu-Natal, South Africa. *Clin. Infect. Dis.* **2007**, 45(11), 1409-1414.
53. Luo, T.; Comas, I.; Luo, D.; Lu, B.; Wu, J.; Wei, L.; Yang, C.; Liu, Q., Gan, M., Sun, G.; Shen, X.; Liu, F.; Gagneux, S.; Mei, J.; Lan, R.; Wan, K.; Gao, Q. Southern East Asian origin and coexpansion of Mycobacterium tuberculosis Beijing family with Han Chinese. *PNAS*, **2015**, 112, 8136-8141
54. Kone, B.; Somboro, A.M.; Holl, J.L., Baya, B.; Togo, A.A., Sarro, Y.D.; Diarra, B.; Kodio, O.; Murphy, R.L.; Bishai, W.; Maiga, M. Exploring the usefulness of molecular epidemiology of tuberculosis in Africa: a systematic review. *Int. J Mol. Epidemiol. Genet.* **2020**, 11, 1-15..
55. Le Hang, N.T.; Hijikata, M.; Maeda, S., Miyabayashi, A.; Wakabayashi, K.; Seto, S.; Diem, N.T.; Yen, N.T.; Van Duc, L.; Thuong, P.H.; Van Huan, H.; Hoang, N.P.; Mitarai, S.; Keicho, N.; Kato, S. Phenotypic and genotypic features of the Mycobacterium tuberculosis lineage 1 subgroup in central Vietnam. *Sci. Rep*. **2021**, 11, 13609.
56. Chen, H.; He, L.; Huang, H.; Shi, C.; Ni, X.; Dai, G.; Ma, L.; Li, W. Mycobacterium tuberculosis lineage distribution in Xinjiang and Gansu Provinces, China. *Sci. Rep*. **2017**, 7, 1068.
57. Solo, E.S.; Suzuki, Y.; Kaile, T.; Bwalya, P.; Lungu, P.; Chizimu, J.Y.; Shah, Y.; Nakajima, C. Characterization of Mycobacterium tuberculosis genotypes and their correlation to multidrug resistance in Lusaka, Zambia. *Int. J. Infect. Dis.* **2021**, 102, 489-496.
58. Mogashoa, T.; Melamu, P.; Ley, S.D.; Streicher, E.M.; Iketleng, T.; Kelentse, N.; Mupfumi, L.; Mokomane, M.; Kgwaadira, B.; Novitsky, V.; Kasvosve, I. Genetic diversity of Mycobacterium tuberculosis strains circulating in Botswana. *PLoS One*, **2019**, 14(5), e0216306.