Tuberculosis (TB) is a curable infectious disease caused by *Mycobacterium tuberculosis* complex strains. It is a major cause of morbidity and mortality all over the world and account for nearly 2 million deaths every year. One third of the world’s total population is infected with this deadly pathogen. According to a recent report, India contributes one fifth of the global tuberculosis burden resulting in 0.5 million deaths annually.

A proficient TB control program and effective disease management rests mainly on two aspects:

i) rapid diagnosis and early institution of effective treatment and

ii) ascertaining epidemiology of the pathogen.

Remarkable advances have been made to develop novel technologies to assist rapid diagnosis of TB. Beside this, the timely institution of anti-tuberculosis treatment (ATT) and its monitoring under the direct observed therapy short course (DOTS) program are some very welcome steps in fight against this deadly disease. However, little effort has been made to understand the epidemiology of the disease in various regions of our very diverse country. This aspect is practically neglected in India due to the high burden of the disease, poor record keeping, inability to trace contacts efficiently and variable transmission dynamics of the disease.

The pathogenicity of *M. tuberculosis* isolates appear to vary from one region to another. The emerging data suggest that the present genetic diversity of *M. tuberculosis* strains reflect the adaptation of these highly infectious microorganisms to its hosts inhabiting different ecological habitats. Therefore, knowledge of these genetic variations may help in understanding the transmission dynamics and pathogenesis of tuberculosis.

One of the molecular tool for strain differentiation that has found wide acceptability among public health workers and TB management groups is the restriction fragment length polymorphism (RFLP) based on the insertion element (IS) 6110. A standardized methodology for this approach was made available in 1993 and widely used for differentiation of *M. tuberculosis* strains in the western world. However, its applicability in Indian scenario was much reduced as substantial numbers of *M.
Summary and Conclusion

tuberculosis clinical isolates were reported to be without or with low copies of IS 6110 elements. Additionally, IS 6110 based studies has number of other limitations, like the methodology is time consuming, cumbersome and requires substantial amount of DNA. Comparison of data from different laboratories is not straight forward.

The *M. tuberculosis* genome sequencing project followed by comparative genomic studies released valuable data for studying genetic diversity of this deadly bacillus. It contains repetitive sequences, like 41 mycobacterial interspersed repetitive unit (MIRU) and 11 exact tandem repeat (ETR) loci. These are 40-100 bp repetitive sequences that are distributed over the *M. tuberculosis* complex genome. Out of these, 12 MIRU and 6 ETR loci were initially observed to show polymorphism among the epidemiologically unrelated clinical isolates of *M. tuberculosis*. Polymorphism in these loci mostly exist due to change in the number of individual repeat units, hence creating allelic variants known as variable number of tandem repeats (VNTRs). So VNTR typing, which is a polymerase chain reaction (PCR) based method, proved to be a rapid and convenient method for studying molecular diversity of *M. tuberculosis* clinical isolates. It requires little culture growth and generates easily comparable numerical data among clinical isolates from different geographic regions with discriminatory power close to or better than IS-6110 RFLP method.

There are genomic sequences which are lost in some of the isolates of *M. tuberculosis* complex genome during the course of evolution. Tuberculosis specific deletion 1 (TbD1) is known to be absent in H37Rv genome but present in *M. bovis* and some *M. tuberculosis* complex isolates. It is a 2.1 Kb long sequence and present between mmpS6 and mmpL6 genes. This sequence is considered primitive, therefore, based upon the presence or absence of TbD1 region, *M. tuberculosis* isolates are classified as ancestral or modern, respectively.

In addition to IS 6110, ETR and MIRU-VNTR typing, are also useful molecular tools in monitoring the prevalence and outbreak of the disease in a particular geographical region. These methods also help in tracing transmission of the disease, based on the simple principle that the patients infected with identical strains are epidemiologically linked.
There are few reports documenting the genetic diversity of \textit{M. tuberculosis} isolates from various regions of India. However, there is no report documenting the molecular heterogeneity of \textit{M. tuberculosis} clinical isolates from Punjab, hence the present investigation was designed to fill this lacuna. The objectives of this investigation were, therefore, to analyze \textit{M. tuberculosis} clinical isolates from Punjab for the distribution of IS 6110 copy number; to study molecular diversity based on MIRU and ETR typing and to look for a phylogenetic relationship among isolates and to compare the obtained molecular heterogeneity with strains from different geographical regions.

To achieve the proposed objectives, 263 sputum samples were collected randomly from suspected tuberculosis patients visiting different TB hospitals and DOTS centers located in different regions of the Punjab state. All the sputum specimens were decontaminated following modified Petroff’s method. A smear of processed sputum was examined microscopically with Ziehl-Neelson staining method using conventional guidelines. Loop full of decontaminated specimen was plated onto Lowenstein Jenson (L-J) slants, in duplicate, and the growth of the culture was monitored weekly, for at least 8 weeks. DNA was then isolated from a loop full of biomass following CTAB-NaCl method and quantified subsequent to U-V spectrophotometric analysis.

DNA from all the clinical isolates was digested using PvuII restriction enzyme. Southern hybridization was performed on digested DNA with IS 6110 specific probe using non radioactive hybridization and detection kit. The polymerase chain reaction (PCR) for typing 12 MIRU’s, 6 ETR’s and tuberculosis specific deletion 1 was performed using published primer sequences with minor modifications. The amplicons were analyzed in 3 % agarose gel along with standard 20 and 100 bp molecular weight markers and visualized with ethidium bromide staining. Product size was estimated and corresponding repeat number was calculated using MIRU-VNTR allele table. Similarly, copy number for each ETR locus was also calculated. The copy number was based on the number of repeats which could theoretically be present in a PCR product of a given size. Genotypes were hence expressed as a 12-digit code for MIRU and 6 digits code for ETR. Hunter-Gaston discriminatory index (HGDI) was used as a numerical index for calculating discriminatory power based on ETR, MIRU-VNTR and both. Based upon
the MIRU and ETR digital code each isolate was then compared with the online database at www.miru-vntrplus.org to categorize each isolate into the lineage.

SALIENT RESULTS

- The study is based on 81 acid fast bacilli (AFB) and culture positive *M. tuberculosis* clinical isolates collected from different regions of Panjab.
- Low IS6110 copy number (<5 copies) was found among 19.7% of strains whereas no isolate was found with zero copy number from Punjab.
- Only 21% of *M. tuberculosis* clinical isolates of ancestral type were found from Punjab.
- Hunter Gaston discriminatory index (HGDI) for ETR was found to be 0.924 for the clinical isolates from Punjab. ETR locus A and E, found to be highly discriminatory whereas loci B, C, D and F were designated as moderately discriminant and poorly discriminant respectively.
- MIRU 10, 16, 26 and 31 were highly discriminant with HGDI value of >0.6; MIRU 4, 23, 24, 39, 40 were moderately discriminant having HGDI value between 0.6 and 0.3 and MIRU 2, 20 and 27 were poorly discriminant with HGDI value <0.3.
- The combined HGDI value for MIRU typing was calculated to 0.994.
- Out of total ancestral strains (TbD1+), 87.5% were found to contain less copies of IS 6110 (≤5) whereas 95.4% of modern strains (TbD1-) from Panjab were found to contain higher copies (>5) of this insertion sequence.
- Similarly, 88.2% ancestral and 98.4% modern *M. tuberculosis* isolates from Panjab carried ≥ 2 and < 2 copies of MIRU 24, respectively.
- All the ancestral strains from Panjab showed strong relationship (100%) with higher copy (>2) of MIRU 4.
- Furthermore, 95.4% of modern strains have more than two copies of MIRU 26 whereas 94.1% of ancestral strains have ≤ 2 copies for this MIRU locus.
MIRU locus 24 showed association with IS 6110 copy number as 93.8 % of *M. tuberculosis* strains with high copies of IS 6110 (≥ 5) carried < 2 repeat of MIRU 24, whereas 75% strains with < 5 copies of IS 6110 contained ≥ 2 copies of MIRU 24 locus.

Additionally, 87.5 % of strains from Punjab with less copies of IS6110 (< 5) carried ≤ 2 and >2 repeats of MIRU26 and 4 respectively; whereas 95.4 % of strains with high copies of IS6110 found to contain ≤ 2 repeat units of MIRU 4. Similarly, 92.3 % of strains with high copies of IS6110 found to contain > 2 repeat units of MIRU 26.

Comparison of ETR, MIRU decimal code and IS6110 RFLP pattern depict that all the isolates collected from Punjab were unique.

Based on the 12 MIRU and 4 ETR loci, majority of *M. tuberculosis* clinical isolates from Punjab could be classified in to four main lineages {Central-Asian (CAS)/ Delhi (23.5 %), Beijing (19.8 %), East African-Indian (EAI) (9.9 %) and Latin American and Mediterranean (LAM) (5 %)}. In addition 3.7 % of strains belonged to Uganda I lineage whereas 2.5 % of both Cameroon and NEW 1 lineage and 1.2 % of Turkey, Haarlem, and URAL clades were identified among *M. tuberculosis* isolates from Punjab. Besides, 29.6 % of strains were not classified in any of known lineage using online database (www.miru-vntr.org) and therefore require an alternative method for further confirmation and identification.

**CONCLUSIONS**

- Ancestral strains (TbD1⁺) of *M. tuberculosis* are less prevalent in Punjab than other parts of India.
- Modern (TbD1⁻) and ancestral strains strongly correlated with MIRU locus 4, 24 and 26. These loci also helped in predicting the low or high IS6110 copy numbers among clinical isolates.
Strains with low IS6110 copy number (less than 5) were less prevalent among the studied clinical isolates from Punjab as compare to those represented from India. No clinical isolate lacking this region was observed in this investigation.

IS6110, ETR and MIRU typing revealed that *M. tuberculosis* clinical isolates from Panjab were highly heterogeneous.

Comparison of MIRU and ETR digital code with online data base predicted CAS/Delhi strain to be the most prevalent lineage in Panjab and 29.6% strains could not be assigned any lineage.