Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis complex* strains (MTBC). It is a highly contagious disease and primarily involves lungs; however, it can affect any part of the body except hairs and nails. It is the major cause of morbidity and mortality all over the world and accounts for nearly 1.3 million deaths every year. One third of the world’s total population is infected with this deadly pathogen and incidence of the TB is increasing at a rate of 1% per year (Dye *et al.*, 2006). According to revised national tuberculosis control programme (RNTCP) annual report (2010) or India every day about 5000 people develop the infection and around 1000 die because of this cruel disease.

The causative agent of TB is known since 1882 and there are effective medicines to treat the diseased patients but still the incidence of the disease is very high (Mathema *et al.*, 2006). In 2010, World Health Organization (WHO) reported around 9.4 million new TB cases. Out of this nearly 55% of cases were reported from Asia, whereas 30% new TB cases were recorded from Africa. The five high TB burden countries, showing new TB cases per year, are India (1.6-2.4 million), China (1.1-1.5 million), South Africa (0.40-0.59 million), Nigeria (0.37-0.55 million) and Indonesia (0.35-0.52 million) (WHO, 2010).

Prevalence of the disease varies in different parts of the world and is substantially low in the developed nations (WHO, 2010). The data from United States of America documented that between 1953 and 1985 there was a steady decline in the incidence rate of tuberculosis from 53 to 9.3 cases per 100,000 persons, respectively. In 1985 the incidence didn’t decrease and it slowly started to increase in subsequent years and peaked to 10.5/100,000 in 1992 (CDC, “Reported tuberculosis in the United States in 2009”, 2010). The main reason for the altered trend was increased TB among human immunodeficiency virus (HIV) infected persons. It was also observed that increase in the number of TB patients correlated with the emergence of acquired immunodeficiency syndrome (AIDS) (Mayans *et al.*, 1997). Reinstitution of good public health practices has once again brought back the incidence rate (1.7 / 100,000 persons) of TB in United States (CDC, “Reported tuberculosis in the United States in 2009”, 2010).
It is evident that a good TB control program and effective disease management rests mainly on two aspects:

i. Rapid diagnosis with effective treatment and

ii. Good public health practices including efficient contact tracing.

The reasons for high mortality in Indian TB populations are many and very diverse ranging from delayed diagnosis of the disease to poverty (inability of the patients to buy drugs), illiteracy (lack of awareness about the disease and its treatment), faulty prescription (an average patients visit three doctors before reaching a TB specialist) and spurious drugs (availability of cheap drugs that might not have the full bioavailability of the prescribed drug). The implementation of WHO sponsored directly observed therapy, short course (DOTS) program has really addressed all the listed issue to a large extent, except rapid diagnosis of the disease, and still made a perceptible different in decreasing the overall incidence of the disease in India (Vashishtha, 2009). Contact tracing remains a big hurdle keeping in view the large disease burden, social taboos associated with TB and poor record keeping of the patient data.

The unduly high incidence of TB in the western countries in early 1990’s, along with high disease burden in the developing world forced the WHO to declare TB as a global emergency. It was for the first time WHO declared global emergency due to an infectious disease. Consequent to this declaration, last decade has witnessed unprecedented advances in understanding the biology of the pathogen and pathophysiology of the disease. Some of the landmark advances include deciphering the complete sequence of *M. tuberculosis* (Cole et al, 1998); understanding the molecular mechanism of drug resistance in *M. tuberculosis* (Blanchard, 1996); development of novel methods for the rapid diagnosis of TB and renewed efforts are directed for drug and vaccine development against tuberculosis (Young, 2003; Young and Dye, 2006).

Remarkable progress has been made in developing new molecular tools for rapid detection of *M. tuberculosis* complex strains (Eisenach et al, 1991; El Amin et al, 2000; Dwivedi and Sehajpal, 2005, Helb et al, 2010) and implementation of anti tuberculosis therapy (ATT) under DOTS, but tracing contacts and monitoring disease transmission
in various regions of the world is posing a big hurdle. Data is available from developed nations where the incidence of the disease is very low but similar data is missing from resource poor countries. Given that vast majorities of TB patients reside in developing countries and disseminating the disease to other regions of the world, it is absolutely essential to understand the transmission dynamics of the disease in such areas. Since the route of transmission is from active patient to close or casual contact, therefore, contact tracing holds the key to effective management of the disease.

One of the fundamental requirements for such endeavors is to understand the genetic diversity present in the *M. tuberculosis* isolates prevalent in different geographical regions of the world. This would require molecular characterization of clinical isolates of *M. tuberculosis* to confirm the patient association with contacts. This has been attempted in the past employing methods that targeted physical characterization of the microbes or their colonies, phage typing and susceptibility to known antibiotics (Herold *et al*, 1996). These methods were of limited use due to their very low discriminatory power.

In mid 1980s, a molecular biology technique was employed to differentiate *M. tuberculosis* isolates. The genomic DNA of clinical isolates was digested with restriction enzymes and the resulting fragments were resolved using agarose gel electrophoresis and compared (Collins and De Lisle, 1984). The method yielded limited information due to the lack of adequate analytical tools and rather uncertain patterns obtained after restriction digestion. The technique was further refined using DNA repetitive elements present in variable numbers leading to high discrimination among the isolates (Zainuddin and Dale, 1989; Hermans *et al*, 1990).

Developments in the area of molecular epidemiology have helped immensely in generating data that could address the limitation of inadequate contact tracing and lack of information regarding disease transmission. The molecular epidemiology is based on the assumption that the *M. tuberculosis* strains for their successful survival adapted to the changing environment by making necessary changes in its genome which ultimately
contributed to the genetic diversity of the pathogen. Any two strains with the same genetic signatures would imply a common source of their origin.

A number of molecular approaches are reported to study the microbial diversity, but IS 6110- restriction fragment length polymorphism (RFLP) and variable number of tandem repeats (VNTRs) typing are the principle methods employed in such studies (Cowan et al, 2002; Hawkey et al, 2003). Besides studying molecular diversity, these methods are also useful in monitoring the prevalence and outbreak of the disease in a particular geographical region. In addition, these methods help in tracing transmission of the disease, based on the simple principle that the patients infected with identical strains are epidemiologically linked (Savine et al, 2002).

IS 6110 RFLP is currently the “gold standard method” for molecular typing and is widely used for strain differentiation of Mycobacterium tuberculosis clinical isolates (van Embden et al, 1993). A sizable number of M. tuberculosis clinical isolates from India are reported to be without or with low copies of IS 6110 elements (Radhakrishnan et al, 2001), thus suggesting the limited use of this method in sub speciation of M. tuberculosis isolates in South Asia (Cowan et al, 2002).

The H37Rv genome sequencing project revealed valuable data regarding the presence of polymorphic loci (Mathema et al, 2006). Majority of these loci occur in the form of tandem repeats and polymorphism in them mainly arise due to change in number of individual repeat units, hence creating allelic variants due to the variable number of tandem repeats unit (VNTRs). Supply et al, 1997 identified 41 mycobacterial interspersed repetitive units (MIRUs), which are 40-100 bp repetitive sequences distributed over the MTBC genome. Out of these, 12 MIRU loci show polymorphism among the epidemiologically unrelated clinical isolates of M. tuberculosis (Supply et al, 2000). Hence MIRU typing, 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 geographical regions with
discriminatory power close to or better than IS 6110 RFLP method (Mazar et al, 2001; Supply et al, 2001; Cowan et al, 2002).

Comparative genomics study of MTBC strains revealed presence of several insertions/deletions (InDels) sequences (Brosch et al, 1999; Gordon et al, 1999). Of reported sequences, a 2,153 bp long sequence was found to absent among 87% *M. tuberculosis* clinical isolates from different geographical region and therefore referred as tuberculosis specific deletion 1 (TbD1) (Brosch et al, 2000). Based on the presence or absence of this sequence, the clinical isolates were categorized as ancestral or modern, respectively. Recent reports from south and central Indian states showed majority of ancestral *M. tuberculosis* strains (Gutierrez et al, 2006; Ahmed et al, 2009). However, this phenotype was reported as less prevalent in other countries like Singapore and Bangladesh (23.2 and 27%, respectively). Therefore, it would be very interesting to explore the evolutionary history of clinical isolates from Punjab based on this particular sequence.

There are very few studies from India elucidating the molecular heterogeneity of *M. tuberculosis* isolates and virtually no data is available on the strains prevalent in Punjab (Chauhan et al, 2004; Gutierrez et al, 2006; Sharma et al, 2008). It is pertinent to mention that a sizable number of Punjabi residents travel abroad, as a large number of their relatives immigrated to foreign countries in pursuit of better living. Keeping in view the high rate of *M. tuberculosis* infection among adult Indian population, these visitors or their immigrated members could act as passive carriers of tubercle bacilli to other nations and a source of foreign strains in our country. It would therefore, be worth while studying the molecular heterogeneity of clinical isolates of Punjab to build a molecular data base to help in discriminating and tracing clinically relevant strains, and also to provide valuable insights into the phylogenetic relationship with other strains of the world.

Keeping the above in view, the objectives of this investigation are to study *M. tuberculosis* clinical isolates from Punjab for the:

- Distribution of IS 6110 copy number,
Introduction

- Molecular diversity based on the MIRU typing,
- IS 6110 and MIRU typing dependent phylogenetic relationship among various isolates,
- Distribution of modern/ancestor clinical isolates based on TbD1 sequence.
- Comparison of molecular heterogeneity with the strains from different geographical regions.