1. INTRODUCTION:

The genus Mycobacterium comprises more than 120 species. Among them the most contagious and infectious bacteria in particular, *M. tuberculosis* (MTB) is the causal agent of tuberculosis (TB), which is an ancient microorganism infecting and killing humans for thousands of years. Several studies demonstrated that this bacterium is an intracellular microorganism restricted to mammals and its DNA is still detectable in the bones of Egyptian mummies [1]. It is noteworthy that the human TB could be also induced by *M. bovis*, which belongs to the MTB complex (MTBC) and principally infects cattle, but the zoonotic risk for human represents a serious problem predominantly, for those who are living at animal-human interface [2]. Furthermore, it has become clear that the members of MTBC were originated from a single ancestor resulted from an evolutionary bottleneck and a clonal expansion occurred 20,000 to 35,000 years ago [3]. In addition, the progenitor of MTBC offspring was restricted in a limited geographical region (East Africa) and called "*M. prototuberculosis"* [4].

Fig 1.1. : Mycobacterium tuberculosis

Importantly, most of the pathogenic or slow growing mycobacteria are sharing a high similarity and a strong phylogeny relationship [5] and interestingly several studies confirmed that the pathogenic mycobacteria were originated from a free living progeny and due to the genome reduction and the acquisition of new genes by horizontal gene transfer (HGT) [6] and gene rearrangement [7], their capacity of parasitism and infectiousness was developed for enabling them to cause severe and dangerous illnesses.

The most widely-used estimates of the national burden of tuberculosis in India are produced by the Revised National Tuberculosis Control Programme, WHO and the Global Burden of Disease Study. India accounts for an estimated 2.2 million of the 8.6 million new cases of tuberculosis that occur each year globally and harbors more than twice as many cases as any other country. Ideally, in any country, tuberculosis surveillance is based on a comprehensive monitoring system to which all new cases are reported and a vital registration system that collects accurate data on the causes of all deaths. Such case monitoring and vital registration systems allow evaluation of the incidence of new infections and the levels of tuberculosis-related mortality, respectively. In countries lacking such systems, estimates of the tuberculosis burden are typically based on national tuberculosis prevalence surveys and national mortality surveys that assess causes of death. Although many countries have either undertaken prevalence surveys since 2002 or are planning to undertake such surveys by 2017, India is not one of them. Currently, estimates of the tuberculosis burden in India are predominately based on the numbers of cases that are notified and expert opinion on the corresponding level of underreporting. Such estimates could be made more accurate and less subjective if expert opinions could be replaced with empirical estimates of the level of underreporting in the case notification system.[8]

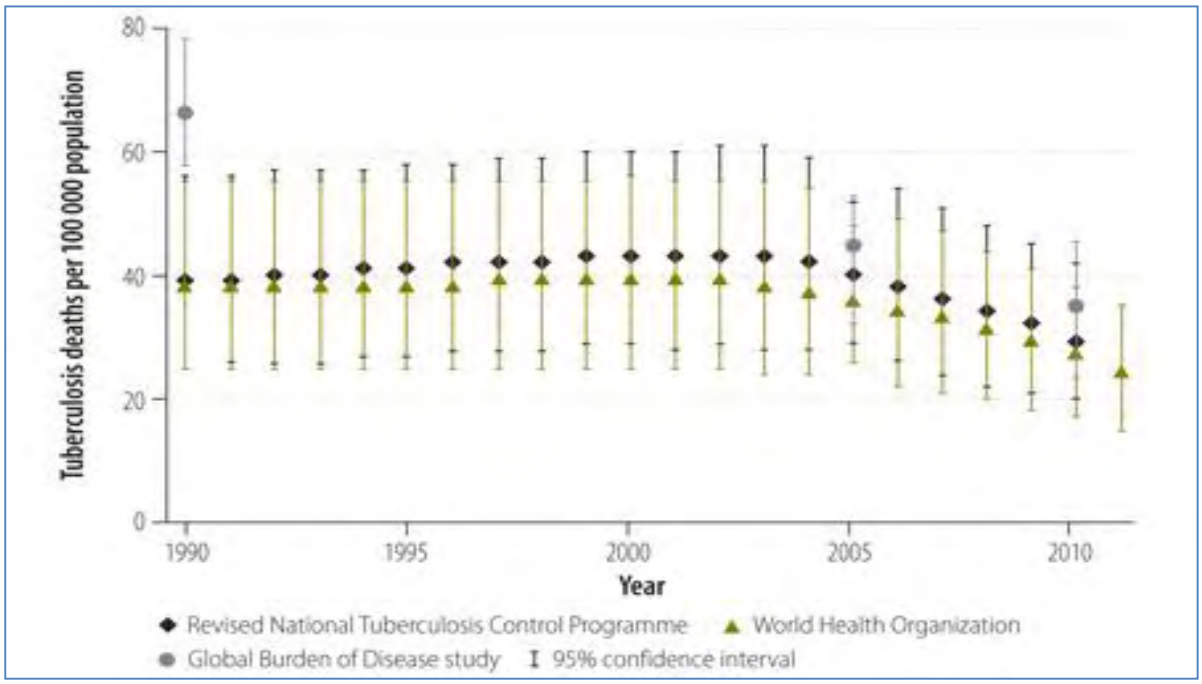

Fig.1.2 : Estimates of the mean level of tuberculosis-attributable mortality, India, 1990–2011

Note: Deaths caused by co-infection with the human immunodeficiency virus and tuberculosis have been excluded

Indeed, the efforts worldwide focus on the combat against TB, leprosy and other mycobacterial diseases and the medical care providers face a great challenge toward the achievement of this goal. As a result the first sequenced mycobacterial genome was that of the reference strain *M. tuberculosis* H37Rv and it was re annotated in 2002 by Camus *et al.* [9]. In 2002, the second MTB genome sequence of the clinical strain CDC1551 was completed and a whole comparative genome analysis was done with the reference strain H37RV based on Large Sequence Polymorphisms (LSPs) and Single Nucleotide Polymorphisms (SNPs) [10]. The SNPs were also used as comparative genome markers for studying the evolution, pathogenesis and molecular epidemiology of clinical MTB strains and a new phylogeny analysis based on SNPs arrangement was established by Alland *et al.*, [11]. More recently, Fillol *et al*. also described the same approach and they have identified six SNP cluster groups (SCGs) and five subgroups within the MTBC members [12].

Recently, different databases were established and provided the complete genome annotation of the reference strain and other TB strains, such as TubercuList and TB database (TBDB) and with the huge amount of visualization, simplification and comparative genomics of their data for better understanding their evolutionary events and consequently, the conception of their environmental niches, mechanisms of adaptation into human and animal being, pathogenicity, virulence determinants that paved the way for appropriate conditions of survival within their hosts and the development of new tools of diagnosis and drug targets for better controlling those threatening diseases [13].

Thus, a set of different approaches were used for studying the phylogeny of MTBC by fingerprinting the insertion sequence IS6110 or by the SNPs based analysis and the phylogeny of other mycobacterium species was performed based on the extracted 16S rRNA sequences [14].

Recently, the availability of Bioinformatics tools for genomic comparison facilitated handling, visualizing and analyzing of enormous amount of sequence information of multispecies bacterial genomes [15, 16]. Therefore, this present study aims to highlight a comparative genome, proteome and phylogeny analysis by the computational and bioinformatics tools (CLC Bio, ClustalW) and phylogeny analysis for the comparison between different *mycobacterium tuberculosis* strains.

2. OBJECTIVES:

1. To retrieve different pathogenic and free living Mycobacterium strains (Sequence).

2. To carry out genomic and proteomic analysis of retrieved Mycobacterium strains.

3. To carry out phylogenetic analysis based on the above analysis.

4. To identify mutagenic sites of the strains.