

## Genomics and Proteomics of Virulent, Avirulent and Drug Resistant Strains of Tuberculous Mycobacteria

Lingaraja Jena\*, Gauri Wankhade and Pranita Waghmare, Bhaskar C Harinath

Bioinformatics Centre, Biochemistry and JBTDR, MGIMS, Sevagram, Maharashtra, India

\*Corresponding author: Harinath BC, JB Tropical Disease Research Centre, Mahatma Gandhi Institute of Medical Sciences, Maharashtra, India, Tel: +917152-284341-284355; Fax: (07152) 284038, E-mail: bc\_harinath@yahoo.com

Received date: June 28, 2016; Accepted date: July 25, 2016; Published date: August 01, 2016

Copyright: © 2016 Harinath BC, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

*Mycobacterium tuberculosis* (MTB), the causal organism of the oldest infectious disease tuberculosis is the leading cause of morbidity and mortality worldwide. This pathogenic organism has been evolved into variety of strains with diverse genotype, phenotype and pathogenic properties such as MTB H37Rv and CDC1551 strains which are virulent, while MTB H37Ra is an avirulent strain and MTB KZN strain is resistant to different anti-tuberculosis drugs. Due to the advancement in genome sequencing and molecular biology, whole genomes of different MTB strains have been completely sequenced. Genomic as well as proteomic comparison among the sequenced strains will help in understanding the differences between virulent, avirulent and resistant organisms. This article reviews the information available on completely sequenced MTB strains and presents the studies reported by researchers on genomic and proteomic comparison of various MTB strains.

**Keywords:** Virulence; Drug resistant; Tuberculosis; Genomics; Proteomics

### Introduction

Tuberculosis (TB), is the oldest infectious disease caused by a bacterium called *Mycobacterium tuberculosis* (MTB), that most commonly infects the lungs. It is transmitted from person to person via droplets from the throat and lungs of people with the active respiratory disease. Even after decades of discovery of MTB, TB infection remains a main cause for extensive morbidity and mortality. One in three of the world's population is considered to be infected with MTB, with 9.6 million people are estimated to have active TB infection in 2014 amongst which 12% were HIV-positive. Further, of the 4,80,000 cases of multidrug-resistant TB (MDR-TB) estimated to have occurred in 2014, only about a quarter of these -1,23,000 were detected and reported [1]. Outbreaks of extensively drug-resistant (XDR) tuberculosis have also been an increasing threat in certain regions around the world [2]. Even if MTB is very virulent, there is no simple answer found so far, that explains for the virulence of this organism. In course of time, this pathogenic organism has been evolved into different resistant strains with diverse genotype, phenotype and pathogenic properties. However, recent studies about MTB propose very little diversity in its genomic sequence [3,4].

### General Mechanism of Pathogenesis and Virulence of MTB

All the strains of MTB have developed several mechanisms to survive inside the host environment. First, MTB enters into the macrophages via cell surface molecules, including those of the integrin family CR1 and CR3 complement receptors [5]. Subsequently, after engulfment by macrophages, most tubercle bacilli are directed to phagolysosomes [6]. Then MTB bacilli bud out from the fused phagolysosomes into vacuoles which further fail to fuse to the

secondary lysosomes and thus escape from lysosomal killing. Due to this mechanism MTB bacilli able to reside temporarily in the phagolysosome which stimulate a response to the intracellular environment in MTB that helps its long-term survival and reproduction [7]. The dormancy or latency of MTB allows the bacterium to escape the activated immune system of the host. Though the detailed mechanisms by which MTB enters into the host cell, circumvents host defenses and reaches to neighboring cell are not fully understood, it has developed different effective survival approaches for surviving effectively in the host environment, which includes (a) The inhibition of phagosome-lysosome fusion; (b) The inhibition of phagosome acidification; (c) The recruitment and retention of tryptophan aspartate containing coat protein on phagosomes to prevent their delivery to lysosomes; and (d) The expression of members of the host-induced repetitive glycine-rich protein family of proteins [8].

Further, there are many genes of MTB are reported to be associated with pathogenesis and virulence. The protein codes by mce gene of MTB has the ability to invade HeLa cells and survive within the host macrophages [9] which suggests that this gene is involved in the invasion of host tissues [8]. The MTB erp gene which codes a secretory protein helps in survival of microorganism in host macrophages [10]. The iron-regulated genes of gram negative bacteria are reported as important factors for virulence [11] and the MTB also synthesizes two distinct iron-regulated siderophores [12] which helps in the growth and survival during the course of infection. Further, the fad D33 gene, encoding an acyl-coenzyme a synthase, plays a vital role in MTB virulence [13].

### Genome Sequencing of Different Strains of MTB

The complete genome sequence information [14,15] of different MTB strains, have provided valuable understanding of its biology. The accessibility of the genomic and proteomic information of MTB combined with high-throughput technologies might open the new

landscape for the development of novel diagnostic techniques, better vaccine and drugs against TB [16] With the decreased expenses of genome sequencing technology [17] and advancement in functional genomics & molecular biology, whole genome sequence information of

different strains of MTB has been released and available in the public domain. As of June 2016, complete genome sequence of several clinical and laboratory strains of MTB are available at “National Center for Biotechnology Information (NCBI)” (Table 1).

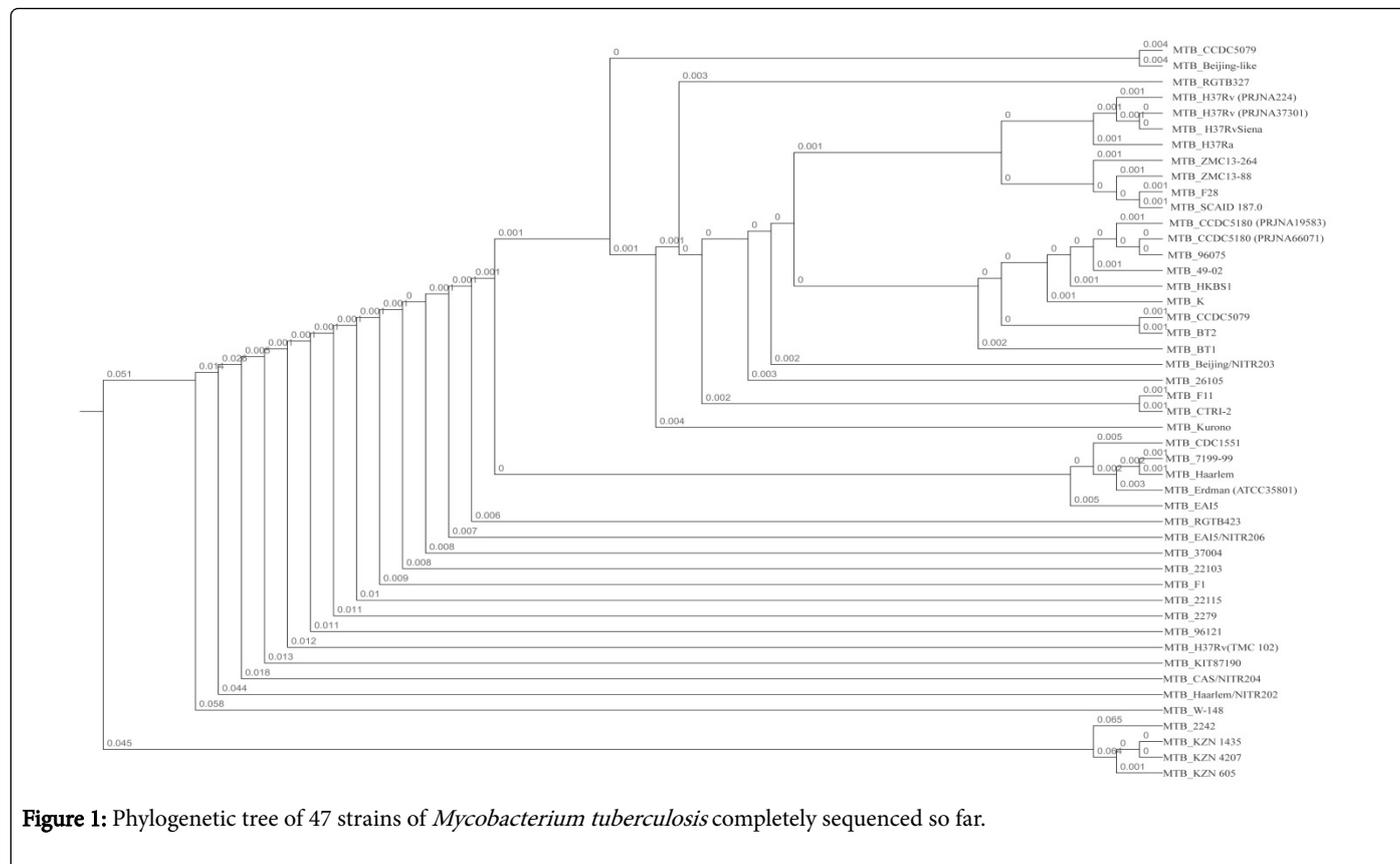
| Organism Name                                              | Strain             | Size (Mb) | GC%  | Genes | Proteins |
|------------------------------------------------------------|--------------------|-----------|------|-------|----------|
| <i>Mycobacterium tuberculosis</i> H37Rv                    | H37Rv              | 4.41153   | 65.6 | 4008  | 3906     |
| <i>Mycobacterium tuberculosis</i> CDC1551                  | CDC1551            | 4.40384   | 65.6 | 4113  | 3964     |
| <i>Mycobacterium tuberculosis</i> H37Ra                    | H37Ra; ATCC 25177  | 4.41998   | 65.6 | 4153  | 4069     |
| <i>Mycobacterium tuberculosis</i> F11                      | F11                | 4.42443   | 65.6 | 4139  | 4043     |
| <i>Mycobacterium tuberculosis</i> KZN 1435                 | KZN 1435           | 4.39825   | 65.6 | 4118  | 4014     |
| <i>Mycobacterium tuberculosis</i> str. Haarlem             | Haarlem            | 4.40822   | 65.6 | 4112  | 4015     |
| <i>Mycobacterium tuberculosis</i> KZN 4207                 | KZN 4207           | 4.39499   | 65.6 | 4115  | 4028     |
| <i>Mycobacterium tuberculosis</i> KZN 605                  | KZN 605            | 4.39912   | 65.6 | 4122  | 4016     |
| <i>Mycobacterium tuberculosis</i> W-148                    | W-148              | 4.41855   | 65.6 | 4133  | 4025     |
| <i>Mycobacterium tuberculosis</i> CTRI-2                   | CTRI-2             | 4.39853   | 65.6 | 4122  | 4028     |
| <i>Mycobacterium tuberculosis</i> CCDC5180                 | CCDC5180           | 4.40598   | 65.6 | 4131  | 3982     |
| <i>Mycobacterium tuberculosis</i> H37Rv                    | H37Rv              | 4.41171   | 65.6 | 4132  | 4058     |
| <i>Mycobacterium tuberculosis</i> 7199-99                  | 7199-99            | 4.4212    | 65.6 | 4121  | 4026     |
| <i>Mycobacterium tuberculosis</i> str. Erdman = ATCC 35801 | Erdman (ATCC35801) | 4.39235   | 65.6 | 4129  | 4010     |
| <i>Mycobacterium tuberculosis</i> str. Beijing/NITR203     | Beijing/NITR203    | 4.41113   | 65.6 | 4141  | 3937     |
| <i>Mycobacterium tuberculosis</i> EAI5/NITR206             | EAI5/NITR206       | 4.39031   | 65.6 | 4105  | 3869     |
| <i>Mycobacterium tuberculosis</i> CCDC5079                 | CCDC5079           | 4.41432   | 65.6 | 4136  | 4033     |
| <i>Mycobacterium tuberculosis</i> EAI5                     | EAI5               | 4.39117   | 65.6 | 4100  | 4002     |
| <i>Mycobacterium tuberculosis</i> HKBS1                    | HKBS1              | 4.40793   | 65.6 | 4126  | 4023     |
| <i>Mycobacterium tuberculosis</i> BT2                      | BT2                | 4.4019    | 65.6 | 4122  | 4019     |
| <i>Mycobacterium tuberculosis</i> BT1                      | BT1                | 4.3994    | 65.6 | 4125  | 4009     |
| <i>Mycobacterium tuberculosis</i> CCDC5180                 | CCDC5180           | 4.41435   | 65.6 | 4134  | 4032     |
| <i>Mycobacterium tuberculosis</i> K                        | K                  | 4.38552   | 65.6 | 4102  | 3991     |
| <i>Mycobacterium tuberculosis</i>                          | KIT87190           | 4.41079   | 65.6 | 4125  | 3976     |
| <i>Mycobacterium tuberculosis</i>                          | ZMC13-264          | 4.41151   | 65.6 | 4125  | 3997     |
| <i>Mycobacterium tuberculosis</i>                          | ZMC13-88           | 4.41151   | 65.6 | 4127  | 4005     |
| <i>Mycobacterium tuberculosis</i>                          | 96075              | 4.37938   | 65.6 | 4114  | 3994     |
| <i>Mycobacterium tuberculosis</i>                          | 96121              | 4.41094   | 65.6 | 4138  | 4007     |
| <i>Mycobacterium tuberculosis</i> 49-02                    | 49-02              | 4.41238   | 65.6 | 4126  | 4029     |
| <i>Mycobacterium tuberculosis</i> H37RvSiena               | H37RvSiena         | 4.41091   | 65.6 | 4133  | 4056     |
| <i>Mycobacterium tuberculosis</i> str. Kurono              | Kurono             | 4.41508   | 65.6 | 4139  | 4054     |
| <i>Mycobacterium tuberculosis</i> H37Rv                    | H37Rv; TMC 102     | 4.39612   | 65.6 | 4135  | 4023     |
| <i>Mycobacterium tuberculosis</i>                          | SCAID 187.0        | 4.37951   | 65.6 | 4099  | 3994     |

|                                                        |                  |         |      |      |      |
|--------------------------------------------------------|------------------|---------|------|------|------|
| <i>Mycobacterium tuberculosis</i>                      | F1               | 4.42862 | 65.6 | 4366 | 4318 |
| <i>Mycobacterium tuberculosis</i>                      | F28              | 4.4219  | 65.6 | 4153 | 4012 |
| <i>Mycobacterium tuberculosis</i>                      | 2242             | 4.41984 | 65.6 | 4512 | 4464 |
| <i>Mycobacterium tuberculosis</i>                      | 2279             | 4.40503 | 65.6 | 4647 | 4599 |
| <i>Mycobacterium tuberculosis</i>                      | 22115            | 4.40183 | 65.6 | 4136 | 3907 |
| <i>Mycobacterium tuberculosis</i>                      | 37004            | 4.41709 | 65.6 | 4151 | 3892 |
| <i>Mycobacterium tuberculosis</i>                      | 22103            | 4.39942 | 65.6 | 4115 | 3901 |
| <i>Mycobacterium tuberculosis</i>                      | 26105            | 4.42649 | 65.6 | 4150 | 3898 |
| <i>Mycobacterium tuberculosis</i> CDC5079              | CDC5079          | 4.39881 | 65.6 | 3696 | 3647 |
| <i>Mycobacterium tuberculosis</i> RGTB327              | RGTB327          | 4.38012 | 65.6 | 3739 | 3691 |
| <i>Mycobacterium tuberculosis</i> RGTB423              | RGTB423          | 4.40659 | 65.6 | 3670 | 3622 |
| <i>Mycobacterium tuberculosis</i> str. Haarlem/NITR202 | Haarlem /NITR202 | 4.40479 | 65.6 | 3730 | 3681 |
| <i>Mycobacterium tuberculosis</i> CAS/NITR204          | CAS/NITR204      | 4.39288 | 65.6 | 4008 | 3960 |
| <i>Mycobacterium tuberculosis</i>                      | Beijing-like     | 4.41122 | 65.6 | 4170 | 3829 |

**Table 1:** Different strains of *Mycobacterium tuberculosis* completely sequenced so far.

The phylogenetic tree (Figure 1) obtained by taking the whole genome of these strains using MAFFT multiple sequence alignment software version 7 [18] revealed the genetic divergence among them.

Although, these strains are phenotypically and genotypically different from each other and have different virulent power, they cause the same disease in humans.



**Figure 1:** Phylogenetic tree of 47 strains of *Mycobacterium tuberculosis* completely sequenced so far.

MTB H37Rv and CDC1551 strains are virulent, most causative agents of tuberculosis and susceptible to most of the anti-tuberculous drugs, while MTB H37Ra is an avirulent strain and MTB KZN (KwaZulu-Natal, South Africa) strain is resistant to different drugs like isoniazid, ofloxacin, rifampicin, kanamycin, pyrazinamide, ethambutol etc. [19]. Further, the F11 strain of MTB reported as predominant in South African epidemic while CCDC5180 was isolated from multidrug-resistant clinical isolate. There are three strains of MTB KZN completely sequenced so far out of which MTB KZN 1435 is multi drug resistance strain while KZN 605 is an extensively drug-resistant clinical isolate but KZN 4207 is a drug-susceptible clinical isolate. These dissimilarities in different strains may be due to mutation in gene resulting into generation of mutated proteins. So, there is a need for genomic as well as proteomic analysis among different strains of MTB to understand the mechanism of variation among them. Genomic and proteomic analysis of virulent, avirulent and drug resistance strains will help in developing more effective and safe drugs for TB control.

### Genome Biology of MTB H37Rv

In 1905, Edward R. Baldwin isolated H37 from a male nineteen years old pulmonary tuberculosis patient [20]. The MTB H37Rv strain obtained originally from the human-lung H37 isolate in 1934, since then it has been broadly used in biomedical research worldwide. MTB H37Rv preserves its complete virulence properties in animal model and is susceptible to anti tubercular drugs. The whole genome of this pathogenic strain was sequenced in 1998 [14]. The genome consists of 4411532 base pairs having 65.6% *guanine+cytosine* (GC) content. It contains more than 4000 protein coding genes and the gene density is at one gene per kilo bases. Genes in the genome are evenly dispersed on both forward and reverse strands. Almost one half of the coding sequences are due to gene duplication and domain shuffling [21].

### Genome Biology of MTB H37Ra

MTB H37Ra is an avirulent strain derived from the H37 which has several distinct properties as compared to MTB H37Rv. Those includes lack of cord formation [22], declined survival inside macrophages [23] or under anaerobic conditions [24], a "raised colony morphology", lack of neutral red dye binding [25] and decreased virulence in mice [26] and guinea pigs [27]. In spite of several genetic and biochemical studies in the past 70 years, the molecular mechanism responsible for the diminishment of virulence in MTB H37Ra is still under study [28]. The whole genome of the avirulent strain of MTB was sequenced by the Chinese National Human Genome Center at Shanghai. It has genome length of 4419977 base pairs with G+C content of 65.6%. Out of 4084 genes, 4034 are protein coding, 45 genes are responsible for coding tRNA whereas 3 for rRNA and 2 for others RNA.

### Genome Biology of MTB CDC1551

The MTB CDC1551 strain, also nicknamed "Oshkosh", is a recent clinical isolate Isolated from a large outbreak of tuberculosis from 1994 to 1996 around Tennessee and Kentucky in the US. The CDC1551 strain appears to be highly infectious in humans, is comparable in virulence to strain H37Rv in animal models [29]. However, this strain has not caused epidemics in man and is sensitive to a wide range of drugs. It is also highly virulent in a mouse lung model, producing several orders of magnitude resulting in more bacteria than the H37Rv strain when inoculated. The *Mycobacterium tuberculosis* CDC1551

genome was sequenced by TIGR and has a total length of 4403837 base pairs and 4113 genes [15].

### Genome Biology of MTB F11

MTB F11 strain is an aerobic, nonmotile, chemoorganotroph, rod-shaped, non-sporulating human pathogen. It was isolated in Tuberculosis patients during a TB epidemic in the Western Cape of South Africa in 1990s. The MTB F11 genome was sequenced by The Broad Institute and has a total length of 4424435 base pairs with 4139 genes. Isolates of F11 not only are a major contributor to the TB epidemic in South Africa but also are present in four different continents and at least 25 other countries in the world [30].

### Genome Biology of MTB KZN

Three strains of MTB isolated from patients in KwaZulu-Natal, South Africa have been sequenced using both Solexa and Sanger sequencing technology. These three strains were selected because they represent a range of important drug resistance phenotypes spanning fully drug-sensitive (DS) to multiply drug resistant (MDR), and to extensively drug resistant (XDR). The XDR (KZN 605), MDR (KZN 1435), and DS (KZN 4207) strains were selected for sequencing from among other strains in the KZN region. Genomic features of these strains are tabulated in Table 1.

### Genomic and Proteomic Comparison of Different Strains of MTB

Due to advances in genomics and associated novel technologies, vast amount of data sets are generating which provide new openings for indulgent and combating both genetic & infectious diseases in humans [31]. Comparative genomic analysis of different mycobacterial strains is also helpful in identifying the genetic basis of varying phenotypes which may further give new insights in the development of novel drugs and vaccines [32]. Comparative genomics is a powerful and novel tool for revealing microbial evolution and identifying genes which might be responsible for encoding novel drug targets [31]. The comparison study revealed that all members of MTB complex share 99.9% identity in their DNA sequence and having identical 16s rRNA [15,33].

A genomic approach was first carried out by Brosch et al., for identifying the variations between MTB H37Ra and MTB H37Rv at genetic level. Their study revealed dual polymorphisms in these two strains i.e., a fragment of 480 kilo bases in MTB H37Rv was found to be substituted by two segments of size 260 and 220 kilo bases in MTB H37Ra and presence of a *DraI* segment of size 7900 bases in MTB H37Ra which was absent in MTB H37Rv. The reported 7900 bases polymorphism was due to the removal of MTB H37Rv *RvD2* in MTB H37Ra. Three *IS6110* deletions (*RvD3* to *RvD5*) from the MTB H37Rv genome were also found in MTB H37Ra. Authors of this study also described the occurrence and mechanisms of genomic differences at genomic level between MTB H37Rv and MTB H37Ra but they were not clear about the role of variation in the MTB H37Ra attenuation [32].

Genomic comparison between MTB H37Rv and H37Ra also revealed that, the genome of MTB H37Rv is very similar to that of MTB H37Ra and is 8,445 base pair smaller than that of H37Ra. In H37Ra and H37Rv, only 98 "single nucleotide variations (SNVs)" were identified. Out of them, 119 were found identical between MTB

CDC1551 and MTB H37Ra and three were because of MTB H37Rv variation, leaving only 76 MTB H37Ra specific SNVs that affecting only 32 genes [28].

An *in silico* analyses of PE/PPE family of MTB H37Ra and MTB H37Rv revealed genetic variations in terms of numerous SNVs along with some deletions and insertions between these two strains. Due to these variations, changes are also observed in their physico-chemical properties, protein: protein interacting domains and phosphorylation, sites which can be correlated to differences in their virulence and pathogenesis [34].

A link between the avirulence of MTB H37Ra and a single amino acid substitution in the PhoP protein was observed by Gonzalo-Asensio et al. In this study, they focused on the *phoP* gene, which was found to have significant role in MTB virulence. This gene is completely conserved in all MTB complexes including MTB H37Rv except that of MTB H37Ra. There is point mutation in *phoP* gene resulting formation of mutilated protein with single amino acid variation i.e., replacement of the polar residue Ser219 by the nonpolar residue Leu [35].

Malen et al. compared membrane proteins of MTB H37Rv with its avirulent sister strain MTB H37Ra and identified more than seventeen hundred proteins. Among these proteins identified by them, majority were found to have comparable abundance in both the strains. There were 29 “membrane-associated proteins” reported with a five or more fold variation in their comparative abundance when compared one strain with the other. There were nineteen membrane and lipo proteins of MTB H37Rv and 10 other proteins of MTB H37Ra, observed with higher abundance in corresponding strains [36].

*In silico* comparative proteomic analysis between MTB H37Rv and MTB H37Ra revealed 3759 identical proteins in MTB H37Rv and H37Ra strains while 244 proteins of MTB H37Rv and 260 of MTB H37Ra were found to be non-identical. Among these non-identical proteins, 172 were identified with mutations (Insertions/Deletions/Substitutions) in MTB H37Ra. However, 53 proteins of MTB H37Rv and 85 proteins of MTB H37Ra were found to be distinct [37]. Non-identical proteins identified in MTB H37Ra and MTB H37Rv may have some important role for the variation in pathogenic property between these two strains. Further, out of 244 non-identical proteins of MTB H37Rv, 19 reported to have important biological function which showed mutation in MTB H37Ra. 40 proteins were identified with single amino acid variation in MTB H37Ra. Different mutation analysis systems and online Bioinformatics resources were employed to study the effect on protein variation and observed that five proteins of MTB H37Rv have lost their normal function in MTB H37Ra strains with the single amino acid mutation [38]. These proteins may have important contribution in pathogenicity of MTB H37Rv strain.

With the help of comparative genomics, two tandem duplications of 29 and 36 kb in the chromosome of *Mycobacterium bovis* BCG Pasteur strain have been revealed [32]. The entire genome comparison among different strains of MTB complex revealed the mutation (insertion/deletion/substitution), gene duplication and selection on the MTB strain evolution.

The CDC 1551 strain known to cause TB outbreak in the United States in 1990s [39] was observed to be comparatively less virulent than MTB H37Rv [40]. The genomic comparison of MTB H37Rv and MTB CDC1551 revealed 86 InDels and 1075 Single Nucleotide Polymorphisms (SNPs), of which 579 were observed to be

nonsynonymous, focusing on the association of genotypic changes with phenotypic variation [15].

The *Mycobacterium bovis* (affecting cattle) genome sequence was found to be 99.95% identical to the genomes of MTB CDC1551 and MTB H37Rv but with slightly smaller genome size. With the comparison of 2504 coding sequences (CDS) among these three genomes revealed 1600 CDS of *M. bovis* identical to MTB H37Rv and MTB CDC1551 respectively. There were 2400 SNPs identified between the two MTB strains and *M. bovis* [6,15]. The genome of *Mycobacterium leprae* (*M. leprae*) has undergone enormous gene loss, leaving only 1604 functional protein coding genes in the bacillus [41]. *M. leprae* is known to cause leprosy. Out of 1439 common genes of MTB and *M. leprae*, a set of 219 genes were found to be unique to mycobacteria through *in silico* comparative analysis [42]. Arnold et al., revealed the existence of short sequence repeats in MTB used for genotyping schemes through whole genome comparison [43]. Comparative genomics will also provide a proficient direction in making out the genetic based variation responsible for difference in phenotype, pathogenicity and host range among different mycobacterial species/strains. The current advances in comparative and functional genomics have also improved our understanding of genetic diversity among the MTB complex. Diaz et al., explored and identified genetic variability among different MTB strains through DNA microarrays technology [44].

The genome comparison of *M. bovis* BCG Pasteur 1173P2 (BCG Pasteur) with MTB H37Rv, MTB CDC1551, and *M. bovis* AF2122/97 discovered Large Sequence Polymorphisms (LSPs) which led to the loss of 133 genes in BCG Pasteur [45,46]. *In silico* analysis of MTB proteomes identified the existence of two novel protein families, PE and PPE [21]. Upon proteomic comparison of two *M. bovis* BCG non virulent strains (Chicago and Copenhagen) with two MTB virulent strains (Erdman and H37Rv), Jungblut et al. identified distinct proteins by mass spectrometry [47]. 27 diverse proteins specific to MTB were identified upon proteomic comparison of culture supernatant from MTB H37Rv and *M. bovis* BCG strain [48]. Miallau et al. identified “RelBE-like toxin-antitoxin complexes” associated with lethality of MTB [49].

Whole-genome sequencing of KwaZulu-Natal MDR and XDR outbreak strains revealed that the MDR strain (KZN-1435) has 150 short indels (<100 bp) and 46 large indels (>100 bp) relative to the wild-type strain, and the XDR sequence has 162 short indels and 37 large indels. The MDR and XDR strains contain typical mutations in *gyrA*, *rpoB*, *rrs*, *katG*, and the promoter of *inhA* that explain resistance to fluoroquinolones, rifampicin, kanamycin, and isoniazid [19]. Further, 22 novel mutations were identified which were unique to the XDR genome or shared only by the MDR and XDR genomes and not already known to be associated with drug resistance [50].

The computational genomics-proteomics analysis of 21 mycobacterial genomes, reported that 1250 *Mycobacterium* gene families are conserved across all species [51]. Further, a recent study by Viale et al. describes the contribution of genomics and functional genomics to studies of the evolution, virulence, epidemiology and diagnosis of *Mycobacterium bovis* and *Mycobacterium avium* subspecies paratuberculosis [52]. Genomic and transcriptomic analysis of the streptomycin-dependent MTB strain 18b provide insight into both the evolution of tubercle bacilli and the functioning of the ribosome [53]. Further another comparative genomics and proteomic analysis of four Non-tuberculous *Mycobacterium* spp. and MTB complex reported the occurrence ESX-1, ESX-3, and ESX-4 regions

(suppose to code for immunogenic proteins) in the genomes of most mycobacteria [54].

Comparative genomics revealed the genomic diversity among different MTB strains, specifically, the identification of particular genes that differ between virulent, avirulent or attenuated and drug resistance MTB strains. Though, most of the comparative genomics studies have been carried out on MTB H37Rv, MTB H37Ra, MTB Erdman, MTB CDC1551 and *Mycobacterium bovis* BCG, comparative genomics and proteomics studies are yet to be done for many strains of MTB sequenced recently. Therefore, the development of an *in silico* technology to study the gene and protein variations of different MTB strains will enhance understanding of virulence and drug resistance properties among them which may give insights the molecular mechanisms of pathogenicity, drug resistance and also give a new direction for the development of new drugs against TB.

## References

1. World Health Organization (2015) WHO Global Tuberculosis Report.
2. Shah NS, Wright A, Bai GH, Barrera L, Boulahbal F, et al. (2007) Worldwide emergence of extensively drug-resistant tuberculosis. *Emerg Infect Dis* 13: 380-387.
3. Musser JM, Amin A, Ramaswamy S (2000) Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. *Genetics* 155: 7-16.
4. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, et al. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci* 94: 9869-9874.
5. Ahearn JM, Fearon DT (1989) Structure and function of the complement receptors, CR1 (CD35) and CR2 (CD21). *Adv Immunol* 46: 183-219.
6. McDonough KA, Kress Y, Bloom BR (1993) Pathogenesis of tuberculosis: Interaction of *Mycobacterium tuberculosis* with macrophages. *Infect Immun* 61: 2763-2773.
7. Goren MB (1977) Phagocyte lysosomes: interactions with infectious agents, phagosomes, and experimental perturbations in function. *Annu Rev Microbiol* 31: 507-533.
8. Meena LS, Rajni (2010) Survival mechanisms of pathogenic *Mycobacterium tuberculosis* H37Rv. *FEBS J* 277: 2416-2427.
9. Arruda S, Bomfim G, Knights R, Huima-Byron T, Riley LW (1993) Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* 261: 1454-1457.
10. Berthet FX, Lagranderie M, Gounon P, Laurent-Winter C, Ensergueix D, et al. (1998) Attenuation of virulence by disruption of the *Mycobacterium tuberculosis* *erp* gene. *Science* 282: 759-762.
11. Litwin CM, Calderwood SB (1993) Role of iron in regulation of virulence genes. *Clin Microbiol Rev* 6: 137-149.
12. Wheeler PR, Ratledge C (1994) Metabolism of *Mycobacterium tuberculosis*. In *Tuberculosis: Pathogenesis, Protection and Control*, pp: 353-385.
13. Rindi L (2002) Involvement of the *fadD33* gene in the growth of *Mycobacterium tuberculosis* in the liver of BALB/c mice. *Microbiology* 148: 3873-3880.
14. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537-544.
15. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, et al. (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* 184: 5479-5490.
16. Ahmed N, Hasnain SE (2004) Genomics of *Mycobacterium tuberculosis*: old threats & new trends. *Indian J Med Res* 120: 207-212.
17. Ng PC, Kirkness EF (2010) Whole genome sequencing. *Methods Mol Biol* 628: 215-226.
18. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30: 772-780.
19. Ioerger TR, Koo S, No EG, Chen X, Larsen MH, et al. (2009) Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal 4: e7778.
20. Steenken WJr, Gardner LU (1946) History of H37 strain of tubercle bacillus. *Am Rev Tuberc* 54: 62-66.
21. Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG (1999) Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis* 79: 329-342.
22. Middlebrook G, Dubos RJ, Pierce C (1947) Virulence and morphological characteristics of mammalian tubercle bacilli. *J Exp Med* 86: 175-184.
23. Mackaness GB, Smith N, Wells AQ (1954) The growth of intracellular tubercle bacilli in relation to their virulence. *Am Rev Tuberc* 69: 479-494.
24. Hepler JQ, Clifton CE, Raffel S, Futrelle CM (1954) Virulence of the tubercle bacillus. I. Effect of oxygen tension upon respiration of virulent and avirulent bacilli. *J Infect Dis* 94: 90-98.
25. Dubos RJ, Middlebrook G (1948) Cytochemical reaction of virulent tubercle bacilli. *Am Rev Tuberc* 58: 698.
26. Larson C, Wicht WC (1964) Infection of mice with *Mycobacterium tuberculosis*, strain H37Ra. *Am Rev Respir Dis* 90: 742-748.
27. Alsaadi AI, Smith DW (1973) The fate of virulent and attenuated *Mycobacteria* in guinea pigs infected by the respiratory route. *Am Rev Respir Dis* 107: 1041-1046.
28. Zheng H, Lu L, Wang B, Pu S, Zhang X, et al. (2008) Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. *PLoS One* 3: 2375.
29. Bishai WR, Dannenberg AM Jr, Parrish N, Ruiz R, Chen P, et al. (1999) Virulence of *Mycobacterium tuberculosis* CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method. *Infect Immun* 67: 4931-4934.
30. Victor TC, de Haas PE, Jordaan AM, van der Spuy GD, Richardson M (2004) Molecular Characteristics and Global Spread of *Mycobacterium tuberculosis* with a Western Cape F11 Genotype. *Journal of Clinical Microbiology* 42: 769-772.
31. Cole ST (2002) Comparative mycobacterial genomics as a tool for drug target and antigen discovery. *Eur Respir J* 20: 78-86.
32. Brosch R, Gordon SV, Pym A, Eiglmeier K, Garnier T, et al. (2000) Comparative genomics of the mycobacteria. *Int J Med Microbiol* 290: 143-152.
33. Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, et al. (2002) Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. *Yeast* 17: 111-123.
34. Kohli S, Singh Y, Sharma K, Mittal A, Ehtesham NZ, et al. (2012) Comparative genomic and proteomic analyses of PE/PPE multigene family of *Mycobacterium tuberculosis* H37Rv and H37Ra reveal novel and interesting differences with implications in virulence. *Nucleic Acids Research* 40: 7113-7122.
35. Gonzalo-Asensio J, Soto CY, Arbués A, Sancho J, del Carmen Menéndez M, et al. (2008) The *Mycobacterium tuberculosis* *phoPR* operon is positively autoregulated in the virulent strain H37Rv. *J Bacteriol* 190: 7068-7078.
36. Malen H, De Souza GA, Pathak S, Söfteland T, Wiker HG (2011) Comparison of membrane proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra strains. *BMC Microbiol* 24: 11-18.
37. Jena L, Kashikar S, Kumar S, Harinath BC (2013) Comparative proteomic analysis of *Mycobacterium tuberculosis* strain H37Rv versus H37Ra. *Int J Mycobacteriol* 2: 220-226.
38. Jena L, Kashikar S, Kumar S, Harinath BC (2014) Effect of single amino acid mutations on function of *Mycobacterium tuberculosis* H37RV and H37RA by computational approaches. *Indian J Tuberc* 61: 200-206.

39. Valway SE, Sanchez MP, Shinnick TF, Orme I, Agerton T, et al. (1998) An outbreak involving extensive transmission of avirulent strain of *Mycobacterium tuberculosis*. *N Engl J Med* 338: 633-639.
40. Manca C, Tsenova L, Bergtold A, Freeman S, Tovey M, et al. (2001) Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN- $\alpha$ / $\beta$ . *Proc Natl Acad Sci U S A* 98: 5752-5757.
41. Cole ST, Supply P, Honore N (2001) Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity. *Lepr Rev* 72: 449-461.
42. Marmiesse M, Brodin P, Buchrieser C, Gutierrez C, Simoes N, et al. (2004) Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* 150: 483-496.
43. Arnold C, Thorne N, Underwood A, Baster K, Gharbia S (2006) Evolution of short sequence repeats in *Mycobacterium tuberculosis*. *FEMS Microbiol Lett* 256: 340-346.
44. Diaz R, Siddiqi N, Rubin EJ (2006) Detecting genetic variability among different *Mycobacterium tuberculosis* strains using DNA microarrays technology. *Tuberculosis (Edinb)* 86: 314-318.
45. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, et al. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284: 1520-1523.
46. Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, et al. (2007) Genome plasticity of BCG and impact on vaccine efficacy. *Proc Natl Acad Sci U S A* 104: 5596-5601.
47. Jungblut PR, Schaible UE, Mollenkopf HJ, Zimny-Arndt U, Raupach B, et al. (1999) Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol Microbiol* 33: 1103-1117.
48. Mattow J, Schaible UE, Schmidt F, Hagens K, Siejak F, et al. (2003) Comparative proteome analysis of culture supernatant proteins from virulent *Mycobacterium tuberculosis* H37Rv and attenuated *M. bovis* BCG Copenhagen. *Electrophoresis* 24: 3405-3420.
49. Miallau L, Jain P, Arbing MA, Cascio D, Phan T, et al. (2013) Comparative proteomics identifies the cell-associated lethality of *M. tuberculosis* RelBE-like toxin-antitoxin complexes. *Structure* 21: 627-637.
50. Motiwala AS, Dai Y, Jones-Lopez EC, Hwang SH, Lee JS, et al. (2010) Mutations in extensively drug-resistant *Mycobacterium tuberculosis* that do not code for known drug-resistance mechanisms. *J Infect Dis* 201: 881-888.
51. Zakhm F, Aouane O, Ussery D, Benjouad A, Ennaji MM (2012) Computational genomics-proteomics and Phylogeny analysis of twenty one mycobacterial genomes (Tuberculosis & non Tuberculosis strains). *Microb Inform Exp* 2: 7.
52. Viale MN, Zumárraga MJ, Araújo FR, Zarraga AM, Cataldi AA, et al. (2016) The genomics of mycobacteria. *Rev Sci Tech* 35: 215-240.
53. Benjak A, Uplekar S, Zhang M, Piton J, Cole ST, et al. (2016) Genomic and transcriptomic analysis of the streptomycin-dependent *Mycobacterium tuberculosis* strain 18b. *BMC Genomics* 17: 190.
54. Gcebe N, Michel A, Gey van Pittius NC, Rutten V (2016) Comparative Genomics and Proteomic Analysis of Four Non-tuberculous *Mycobacterium* Species and *Mycobacterium tuberculosis* Complex: Occurrence of Shared Immunogenic Proteins. *Front Microbiol* 7: 795.