Chapter 4

# Early Diagnosis of *Mycobacterium tuberculosis* using Next Generation Sequencing

***Dileep Tiwari[1*]; Divakar Sharma[2]***

[1]*Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban-4000, South Africa.*

[2]*Department of Biochemistry, National JALMA Institute for Leprosy and Other Mycobacterial Diseases, Tajganj, Agra, 282004, India*

[*]***Correspondence to: Dileep Tiwari,*** *Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4000, South Africa.*

*Phone: +27 31 2608942; Email: Tiwarid@ukzn.ac.za*

## 1. Introduction

Tuberculosis is a chronic granulomatous infectious disease caused by tubercle bacteria known as *Mycobacterium tuberculosis* (*Mtb*). It is a facultative intracellular organism that is readily phagocytosed but is resistant to intracellular killing by the macrophage. It can remain dormant for years till hosts' immunity persistence and then reactivates itself. *Mycobacterium tuberculosis* was the first bacterial human pathogen described, which dates back to 1882 by Robert Koch but even today continues to produce devastating illness. Since then, a large number of Mycobacterial species responsible for causing pulmonary and extra-pulmonary infections have been identified in humans as well as in animals [1]. *Mycobacterium tuberculosis* creates global epidemic disease, which occurs in many parts of the body, but most commonly found (80%) in the lungs and is rampant throughout the world (Center for disease control and prevention, Morbidity Mortality weekly report, 1993a,b; [2-4]. Every third person on this earth is believed to be infected with *Mtb*, leading to eight million cases of active tuberculosis per year and approximately three million deaths annually. This situation has been more aggravated by the emergence of epidemic multi-drug resistant strains of *Mtb* in conjunction with HIV infection have rendered this problem all the more serious [5]. Furthermore, the accelerating and amplifying influence of HIV infection is contributing to the increasing incidence of disease caused by multi-drug resistant strains of *M. tuberculosis* [6]. Thus, Tuberculosis has become

a major concern worldwide with significant mortality and morbidity reported from various parts of the world [7-12]. We have suggested the theoretical basis of high TB burden among the healthcare workers of developing countries which can be used for further improvement in strategies for the prevention of TB infections in hospital settings [13,14]. Apart from that we have also suggested the potential diagnostics and therapeutics based strategies to combat the drug resistance TB [13,15,16]. Without increased of active commitment at National and International levels, tuberculosis will claim about 30 million more lives in the next decade and there will be about 90 million new cases of active tuberculosis. Such a vast epidemic creates challenges as it raises the demand for public health solutions. Recent advances in diagnosis done by deep sequencing namely Next-generation sequencing this technology decreases the time in diagnosis delay and improve treatment of MDR, XDR etc. outcome.

## 2. Next-Generation Sequencing

Like other medical fields, the role of next generation sequencing (NGS) techniques is about to revolutionize diagnostics of infectious diseases. Initially, the diagnosis of the microbial origin of diseases was based on the laboratory test by culture assay for bacteria, the lacunae were cultivability could only be detected rather than its hypotheses. In order to seek specific pathogens, specialized media rich or selective and culture conditions defined oxygen tension or temperature can be used but was laborious and their intrinsic inefficiency in the propagation of fastidious bacteria like *Mycobacterium leprae*. These difficulties have been progressively replaced by nucleic acid-based tests like PCR or NASBA. The advantages of PCR are numerous: speed, low cost, automation, sensitivity, and specificity main drawback of targeted, pathogen-specific PCR is that it is only able to identify predefined targets, which supposes that the physician has elaborated an etiological hypothesis. To broaden the range of detection methods like the direct hybridization of non-amplified or random amplified nucleic acids (NA) from samples on DNA arrays was used but this has not been proven satisfactory, due to relative lack of sensitivity for medical diagnosis. Bacterial typing achieved by sequencing the 16S gene or other regions of the genome that is sufficiently conserved to allows the definition of consensus primers yet sufficiently variable to allow for typing. Use of NGS has increased the depth of sequencing by several orders of magnitude and thereby the capacity to detect rare species. But, with 16S PCR, the taxonomic assignation remains often at the level of the genus, an intrinsic limit due to the conservation of the locus between species of the same genus. Multiplexed PCR assays for multiple loci have been developed to provide simultaneous detection of several agents. A meta-analysis study for 6012 patients with sepsis results in an overall sensitivity and specificity of 0.75 (95% CI: 0.65–0.83) and 0.92 (95%CI: 0.90–0.95), respectively, to detect bacteremia [17]. Multiplex PCR range can be improved by designing primers targeting numerous pathogens and varied loci within pathogens and resolving these amplicons using electrospray ionization-mass spectrometry [18] or NGS([19].

www.openaccessebooks.com

To this all setbacks an alternative strategy takes advantage of the increasing availability and speed and decreasing cost per base of NGS offered by deep sequencing machines which allows metagenomics, which is the study of the microbial genetic sequences recovered directly from a given human, animal, or environmental sample here the sequence of all the NA species of the sample are determined and compared with those in databases. At first, this technology was used to describe the dynamics of microbiomes from different origins, from the gut, other mucosal sites and the skin, as well as from various human-made (e.g., sewage) and natural (e.g., sea) environments. It has also been used to discover new infectious agents. At present such metagenomic study cover known but unexpected viruses, phages, bacteria, parasites or fungi [20], which paves the way to the application in the field of diagnosis of infectious diseases. In principle, a whole genome NGS (WG-NGS) would be advantageous in clinical diagnostics, as there is no need to design specific primers to pre-amplify target sequences. In the field of bacteriology, most studies have dealt with sequencing of clinical isolates cultured in vitro, but good results have been obtained by direct sequencing from clinical samples, for example for the diagnosis of tuberculosis lesions [17], fecal samples from diarrheic patients [21], or urinary samples from patients with suspected urinary tract infection [22]. Another advantage of the technique is its capacity to identify co-infections, which is of great help to adapt therapeutics. Old methods to identify strains of *M. tuberculosis* rely on the analysis of small level of the genome, and assumed that the DNA sequences in these level are of mostly variable so not enough to allow researchers to separate strains of *M. tuberculosis* that are evolutionarily close or distant and also the true complexity of disease dynamics cannot be resolved by tracking strains using a small section of the genome. The availability of next-generation sequencing platforms has allowed viewing the complete genetic information of the bacteria, which should improve the accuracy of efforts to monitor strains of *M. tuberculosis* over space and time [23]. It also has been proven where NGS platform are used to expose the genetic heterogeneity of Mycobacterium tuberculosis in extra pulmonary TB patients [24].

## 3. Epidemiological Sound

For epidemiological investigations, diagnosis, and for testing whether strains of bacteria are susceptible to particular drugs rapid whole genome sequencing would be a promising tool. It has been reported that long-term large-scale whole genome sequencing strategy has been used to decipher the tuberculosis epidemic in a high prevalence setting with multiple sources of infection [25]. The whole genome sequences of 1687 *M. tuberculosis* samples (isolates) from patients in the Karonga District of Malawi for 15 years, which represented 72% of the total number of confirmed tuberculosis cases during that time. Guerra-Assunc et al. found that the epidemic was largely driven by members of one lineage, which implies that either this lineage arrived in the area earlier than the others, or that the members of this lineage were more successful. The genome of *M. tuberculosis* consists of ~4.4 million bases and is generally

believed to be relatively stable [26]. Guerra-Assunc¸ et al. developed a clustering formula to group together directly related isolates. Using this formula in combination with network-analysis they found that strains from certain lineages were more likely to be transmitted between patients than others. This suggests that there are differences in the abilities of bacteria in the different lineages to cause disease. In this high-incidence setting, 66% of identified cases clustered together, of which 38% of the patients had evidence of recent infection, implying ongoing transmission of the bacteria. This indicates that reactivation of the previous infection was the primary driving force behind this epidemic. Glynn, also showed that the proportion of tuberculosis cases due to reactivation increased over the duration of the 15-year study, as demonstrated by a marked decrease in transmission between 1999–2001 (45%) and 2008–2010 (30%). Guerra-Assunc¸ et al. reported due to the implementation of the antiretroviral therapy and isoniazid preventative therapy in Karonga. Significantly, this study reported that the tuberculosis control program in Karonga has reduced transmission of the bacteria and also demonstrated that whole-genome sequencing can provide new insights into tuberculosis epidemics, which could be used to advice and fine tune control programs. Despite the advantages of whole genome sequencing, it is important to acknowledge the complexity of the technology and data analysis. This questions how useful it could be in high incidence settings where tens of thousands of cases are diagnosed annually. Furthermore, the current technology is restricted to clinical isolates that need to undergo a lengthy culturing and DNA extraction process, which prevents its use as a real-time monitoring tool. Additionally, whole genome sequencing is labor intensive and financially demanding, although costs have decreased significantly over the last decade. Regardless of these challenges, this technology has the potential to immediately revolutionize drug susceptibility testing by identifying the complete repertoire of mutations in target genes that confer drug resistance [27].

## 4. **Key Advantage of Next Generation Sequencing (NGS) for *Mycobacterium Tuberculosis:***

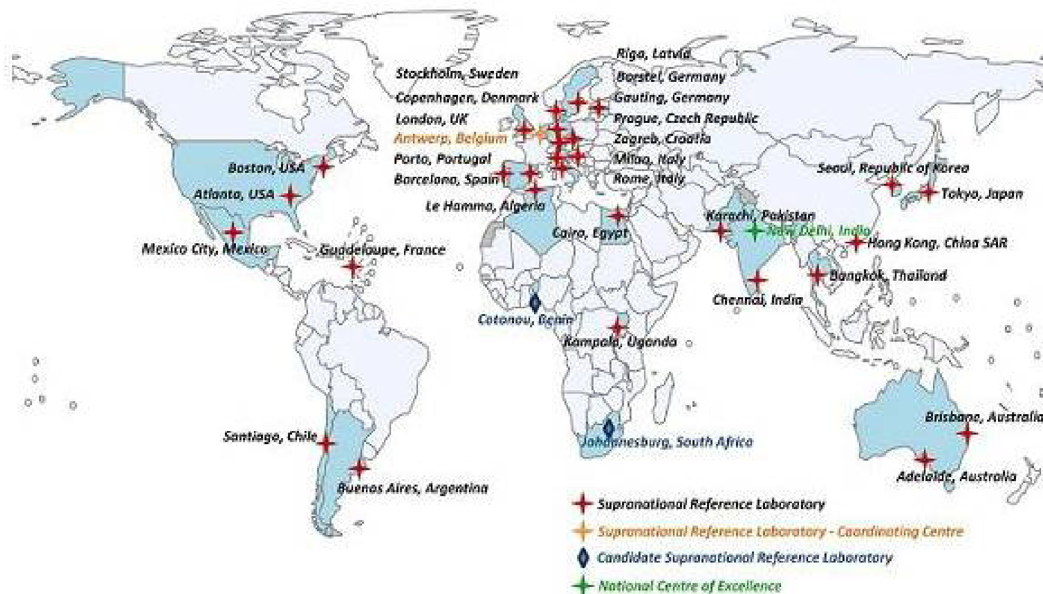
Next generation sequencing (NGS) approaches, where the complete genome is sequenced that not only assistances in pointing out minute variances between the various sequences but also saves time, cost, good Strengths – simpler, faster, Less expensive than WGS, Up to 200 gene targets and easy to handle.

**Following key advantages**

(i)-Sequence DNA direct from sputum, (ii)- Direct from sputum sample , no need BSL laboratory (iii) Detect primary 1st and 2nd line drug resistance mutations (iv) Rapidly sequence whole genomes (v)- 3.Utilize RNA sequencing (RNA-Seq) to discover novel RNA variants and splice sites, or quantify mRNAs for gene expression analysis (vi) Analyze epigenetic factors such as genome-wide DNA methylation and DNA-protein interactions (vii) These techniques enable

the identification of mycobacterial strains

Next Generation sequencing (NGS) is also found to be useful in identifying single nucleotide polymorphisms (SNPs), comparative genomics and also various aspects about transmission dynamics and also facilitate the study of their phylogenetic and evolutionary traits.




**Place of destination:** Culture-Free, NGS for Rapid DST in TB Reference Laboratories by 2020.

## 5. References

1. Wolinsky, E. and W.B. Schaefer, proposed numbering scheme for mycobacterial serotypes by agglutination. International Journal of Systematic Bacteriology, 1973. 23(2): p. 182-183.

2. Diagnosis of Tuberculosis BY Nucleic-acid amplification methods applied to clinical specimens (REPRINTED FROM MMWR, VOL 42, PG 686, 1993). Jama-Journal of the American Medical Association, 1993. 270(13): p. 1528-1528.

3. Estimates of future global tuberculosis morbidity and mortality. MMWR Morb Mortal Wkly Rep, 1993. 42(49): p. 961-4.

4. Gulland, A., More cases of tuberculosis than previously thought, WHO reports. Bmj-British Medical Journal, 2016. 355.

5. Kochi, A., The global tuberculosis situation and the new control strategy of the World Health Organization. Tubercle, 1991. 72(1): p. 1-6.

6. Webb, J.R., et al., Molecular cloning, expression, and immunogenicity of MTB12, a novel low-molecular-weight antigen secreted by Mycobacterium tuberculosis. Infection and Immunity, 1998. 66(9): p. 4208-4214.

7. Bisen, P.S., et al., Analysis of the shotgun expression library of the Mycobacterium tuberculosis genome for immunodominant polypeptides: potential use in serodiagnosis. Clin Diagn Lab Immunol, 2003. 10(6): p. 1051-8.

8. Garg, S.K., et al., Diagnosis of tuberculosis: available technologies, limitations, and possibilities. J Clin Lab Anal, 2003. 17(5): p. 155-63.

9. Tiwari, R.P., et al., Glycolipids of <em>Mycobacterium tuberculosis</em> Strain H37Rv Are Potential Serological Markers for Diagnosis of Active Tuberculosis. Clinical and Diagnostic Laboratory Immunology, 2005. 12(3): p. 465-473.

10. Tiwari, R.P., et al., Modern approaches to a rapid diagnosis of tuberculosis: promises and challenges ahead. Tuberculosis (Edinb), 2007. 87(3): p. 193-201.

11. Tiwari, D., et al., Efficient ELISA for diagnosis of active tuberculosis employing a cocktail of secretory proteins of Mycobacterium tuberculosis. Folia Biol (Praha), 2014. 60(1): p. 10-20.

12. Tiwari, D., et al., Fast and efficient detection of tuberculosis antigens using liposome encapsulated secretory proteins of Mycobacterium tuberculosis. J Microbiol Immunol Infect, 2017. 50(2): p. 189-198.

13. Sharma, D., et al., Prevalence and risk factors of tuberculosis in developing countries through health care workers. Microbial pathogenesis, 2018. 124: p. 279-283.

14. Sharma, D., et al., Interactome analysis of Rv0148 to predict potential targets and their pathways linked to aminoglycosides drug resistance: An insilico approach. Microb Pathog, 2018. 121: p. 179-183.

15. Sharma, D., et al., Repurposing and Revival of the Drugs: A New Approach to Combat the Drug Resistant Tuberculosis. Frontiers in Microbiology, 2017. 8(2452).

16. Sharma, D., D. Bisht, and A.U. Khan, Potential Alternative Strategy against Drug Resistant Tuberculosis: A Proteomics Prospect. Proteomes, 2018. 6(2).

17. Chang, S.S., et al., Multiplex PCR system for rapid detection of pathogens in patients with presumed sepsis - a systemic review and meta-analysis. PLoS One, 2013. 8(5): p. e62323.

18. Wolk, D.M., E.J. Kaleta, and V.H. Wysocki, PCR-Electrospray Ionization Mass Spectrometry The Potential to Change Infectious Disease Diagnostics in Clinical and Public Health Laboratories. Journal of Molecular Diagnostics, 2012. 14(4): p. 295-304.

19. Arena, F., et al., Rapid Resistome Fingerprinting and Clonal Lineage Profiling of Carbapenem-Resistant Klebsiella pneumoniae Isolates by Targeted Next-Generation Sequencing. Journal of Clinical Microbiology, 2014. 52(3): p. 987-990.

20. De Vlaminck, I., et al., Temporal Response of the Human Virome to Immunosuppression and Antiviral Therapy. Cell, 2013. 155(5): p. 1178-1187.

21. Loman, N.J., et al., A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104:H4. Jama, 2013. 309(14): p. 1502-10.

22. Hasman, H., et al., Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J Clin Microbiol, 2014. 52(1): p. 139-46.

23. Roetzer, A., et al., Whole Genome Sequencing versus Traditional Genotyping for Investigation of a Mycobacterium tuberculosis Outbreak: A Longitudinal Molecular Epidemiological Study. Plos Medicine, 2013. 10(2).

24. Das, S., et al., Genetic heterogeneity revealed by sequence analysis of Mycobacterium tuberculosis isolates from extra-pulmonary tuberculosis patients. Bmc Genomics, 2013. 14.

25. Guerra-Assuncao, J.A., et al., Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. Elife, 2015. 4.

26. Jagielski, T., J. van Ingen, and N. Rastogi, Current methods in the molecular typing of Mycobacterium tuberculosis and other mycobacteria. 2014. 2014: p. 645802.

27. Steiner, A., et al., KvarQ: tasrgeted and direct variant calling from fastq reads of bacterial genomes. BMC Genomics, 2014. 15(1): p. 881.