Review Article
Indian J Med Res 120, October 2004, pp 207-212

Genomics of *Mycobacterium tuberculosis*: Old threats & new trends

Niyaz Ahmed & Seyed E. Hasnain

Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India

Received July 22, 2003

Tuberculosis (TB) has been declared as a global health emergency by the World Health Organization (WHO). This has been mainly due to the emergence of multiple drug resistant strains and the synergy between tubercle bacilli and the human immunodeficiency virus (HIV). Genomic analysis of strains for outbreak investigations is in vogue for about a decade now. However, information available from whole genome sequencing efforts and comparative genomics of laboratory and field strains is likely to revolutionize efforts towards understanding molecular pathogenesis and dissemination dynamics of this dreaded disease. Genomic information is also going to fuel discovery projects where new targets will be identified and explored towards a new drug for TB. Besides this, efforts of information technologists, chemists, population biologists, freelance workers, media persons, non-governmental organizations and administrators to needed to handle the problem of tuberculosis to prevent it from becoming a pandemic.

Key words Genomics - *Mycobacterium tuberculosis* - multidrug resistant strains

Mycobacterium tuberculosis is reputed to have the highest annual global mortality among all of the pathogens[1]. The rise in tuberculosis (TB) incidence over the last two decades is partly due to TB deaths in HIV-infected patients and partly due to the emergence of multidrug resistant strains of the bacteria. This rapid increase in the disease has led to potential funding arrangements aimed at a large effort towards stopping this disease before it becomes a global epidemic. Due to its slow growth and high virulence it is extremely difficult to work with the TB bacterium. However, rapidly evolving mycobacterial genomics with complete genome sequences known along with powerful bioinformatics approaches, one can realize better therapeutics and prophylactics in the near future[2]. A comparative genomic analysis of these species has a potential to reveal the genetic basis of disease phenotypes, which will be invaluable for the development of much needed drugs and newer vaccines.

Problem of multi drug resistant tuberculosis (MDR-TB)

One of the classical threats of the tuberculosis epidemic has been the MDR-TB. Use, and often abuse or misuse, of antimicrobial agents has encouraged the evolution of bacteria toward resistance, resulting often in therapeutic failure. There are evidences that bacteria have the ability to adapt to this deficit and recover fitness on serial passage[1]. Resistance to antituberculosis drugs has been a problem since the era of chemotherapy began. After dramatic outbreaks of MDR-TB in the early 1990s, resistance became recognized as a global problem. MDR-TB now threatens the inhabitants of countries in Europe, Asia, Africa, and the Americas[2]. An understanding of the molecular basis of drug resistance may contribute to the development of new drugs[3]. Management of MDR-TB relies on prompt recognition and treatment with at least 3 drugs to which an isolate is susceptible.

The roles of drug containing environments, and the immunological status of the host and bacterial molecular mechanisms of development of drug resistance to *M. tuberculosis* have been examined and results are helpful in implementation of modified drug regimens in tuberculosis control programmes. Multidrug resistant strains of *M. tuberculosis* seriously threaten tuberculosis control and prevention efforts. Molecular studies of the mechanism of action of antitubercular drugs have elucidated the genetic basis of drug resistance. Drug resistance in *M. tuberculosis* has been primarily attributed to the mutations in the drug target genes, however, the presence of efflux pumps in clinical MDR isolates cannot be ruled out[4]. These mutations lead either to an altered target or to a change in titration of the drug. A diverse array of strategies is already available to assist in rapid detection of drug resistance-associated gene mutations. In spite of remarkable advances in this area, much remains to be learned about the molecular genetic basis of drug resistance in *M. tuberculosis*. During the last decade, there has been a marked increase in the number and gravity of tuberculosis cases both in developing countries and in industrialized nations. One of the more insidious consequences of this resurgence has been the recent emergence of nosocomial transmission of multi drug resistant strains of *M. tuberculosis*, thus creating untreatable forms of the disease and these strains may become widespread. That the various clinical isolates of *M. tuberculosis* are geographically partitioned at the global level. Ahmed *et al*[5,6] has provided evidence to the concept of geographic genomics[7].

The TB- HIV symphony: Supporting the idea of low virulent clones

The emergence of the HIV/acquired immunodeficiency syndrome (AIDS) pandemic has led to major shift in our approaches towards epidemiological studies of TB both in resource poor and developed countries, with HIV infection becoming a risk factor for the development of active TB infection. This might be a consequence of increased re-activation of previously acquired, dormant bacilli and an increase in susceptibility to both re-infection and primary infection. Studies involving DNA fingerprinting of *M. tuberculosis* isolates obtained from AIDS patients by the technique of restriction fragment length polymorphism (RFLP) analysis showed that re-infection and new infection both occur in AIDS patients[5]. It has been speculated that the HIV/AIDS patients constituting an ecological niche for *M. tuberculosis*, where less virulent strains multiply freely without the selection pressure provided by an immunocompetent patient. Consequently, it should be possible to differentially identify pathogenic bacterial clones and differentiate them on the basis of epidemiological parameters related to co-infection, relapse versus recent infection and multiple drug resistance. Recently, one study[6] has been conducted under the auspices of the *M. tuberculosis* evolutionary genomics interest group, where significantly different genotypes were observed for HIV-associated tubercle bacilli as compared to bacilli recovered from non-HIV patients. This has been the only study of its kind in the post genomic scenario where high resolution genomic pattern analysis was systematically applied to a well characterized strain collection from a clinically tested patient population. While statistically significant association of a clone to the disease outbreak in a community may reflect increased fitness of the genome, it can be suggested that this phenomenon should be observed in multiple settings before they can be linked with certainty to a particular host population. Frequent travel and social links among the people of economically productive and sexually active age groups as seen in some African and South American countries, often provide for fast multiplication of low virulent strains and therefore high clustering is seen in such settings.

The post genomic approaches

New developments in molecular biology and functional genomics have fostered major advances in our understanding of genetic variability among *M. tuberculosis*. Complete genome sequence information[8] has been released for this organism opening new vistas in diagnosis, epidemiology and vaccine development. Automated DNA sequencing allows for the direct comparison of specific genes among large populations of isolates, and determination of complete genomic sequence of two strains together with new technologies like DNA

microarrays and computational biology, has provided a whole genome perspective on genomic content, gene regulation, and metabolism.

The face of infection epidemiology has changed in the post genomic scenario. Bacterial genome signatures permit whole genome comparisons between different bacterial strains or isolates by using differential amplification techniques and automated genotyping for the determination of genetic signatures or barcodes. These signatures could be a group of markers or DNA fragments of unique sizes. In this way, gene deletions may be readily identified that correlate with alteration in virulence or other phenotypes.

The use of differential expression to monitor whole genome expression is now a powerful approach, which will result in definition of differentially expressed genes important in pathogenesis, and will provide useful targets for rational design of new drugs and vaccine candidates for *M. tuberculosis*. These studies are likely to confirm the reality of "transcriptomics" as a technical deliverable for hypothesis driven research into mycobacterial pathogenesis, vaccine and drug research.

The emergence of protein-based analysis or proteomics is likely to facilitate determination of changes in the expression profile that leads to pathogenicity, drug resistance and immune responses. This approach has a very high potentiality and accuracy to identify proteins involved in pathogenicity and drug resistance leading to development of medical intervention. Similar approaches can be adopted to understand *M. tuberculosis* pathogenicity and multidrug resistance among different strains.

The recent accumulation of copious data about the extent and nature of genetic and phenotypic variation in tubercle bacilli provides an unprecedented opportunity to determine the phenotypic consequences of genetic polymorphism in this organism. However, a coherent synthesis of this knowledge is lacking, reflecting the formidable challenges of this endeavour. Because environmental and host factors clearly contribute to the clinical and epidemiologic behaviour of strains, these factors must be carefully integrated into the investigative process.

Post genomic databases for epidemiologists

A large number of molecular markers have been made available for epidemiological and evolutionary studies as a result of comparative genomics studies (Fig.). These markers include Mycobacterial interspersed repeat units-variable number tandem repeats (MIRU-VNTRs)[9], fluorescent amplified fragment length polymorphism (FAFLPs)[6] and large genomic deletions[10]. The present day challenge is to compile standardized fingerprint patterns originating from highly networked, multi-centric, genotypic analyses in the databases for interlaboratory use and future references[7]. This requires an accurate measurement of fragment lengths. Analysis via molecular weight markers in adjacent lanes is straightforward and can be done automatically on digitated images. A correction for the variation in migration rates and gel distortion is achieved by co-electrophoresis in each lane of both sample and marker fragment. One option is the use of invariant fragments with known lengths as internal markers. Digitated images in a standard graphical file may be obtained as scanned autoradiographs. These data can be imported in gel analysis software, which carries out the normalization and fragment sizing for analysis by the dedicated software. Many databases have been developed recently for rapid genotypic searches and comparisons on the world-wide web: SpolDB: Spoligotyping resource database at Institut Pasteur of Guadilope (http://www.cdc.gov/ncidod/EID/vol7no3/sola_data.htm) Tbase: RFLP resource database at RIVM, The Netherlands (https://hypocrates.rivm.nl/bnwww/IS6110-RFLP-bands.htm) AmpliBASE: AFLP resource database at CDFD Hyderabad (http:www.cdfd.org.in/amplibase)

MiruDB:VNTR resource database at IBL Lille, France (http://www.ibl.fr/mirus/mirus.html)

Post genomic discovery efforts: TB structural genomics

It is quite unfortunate to realize that while the world is faced with the horrors of TB epidemic, no novel compound has been introduced for widespread

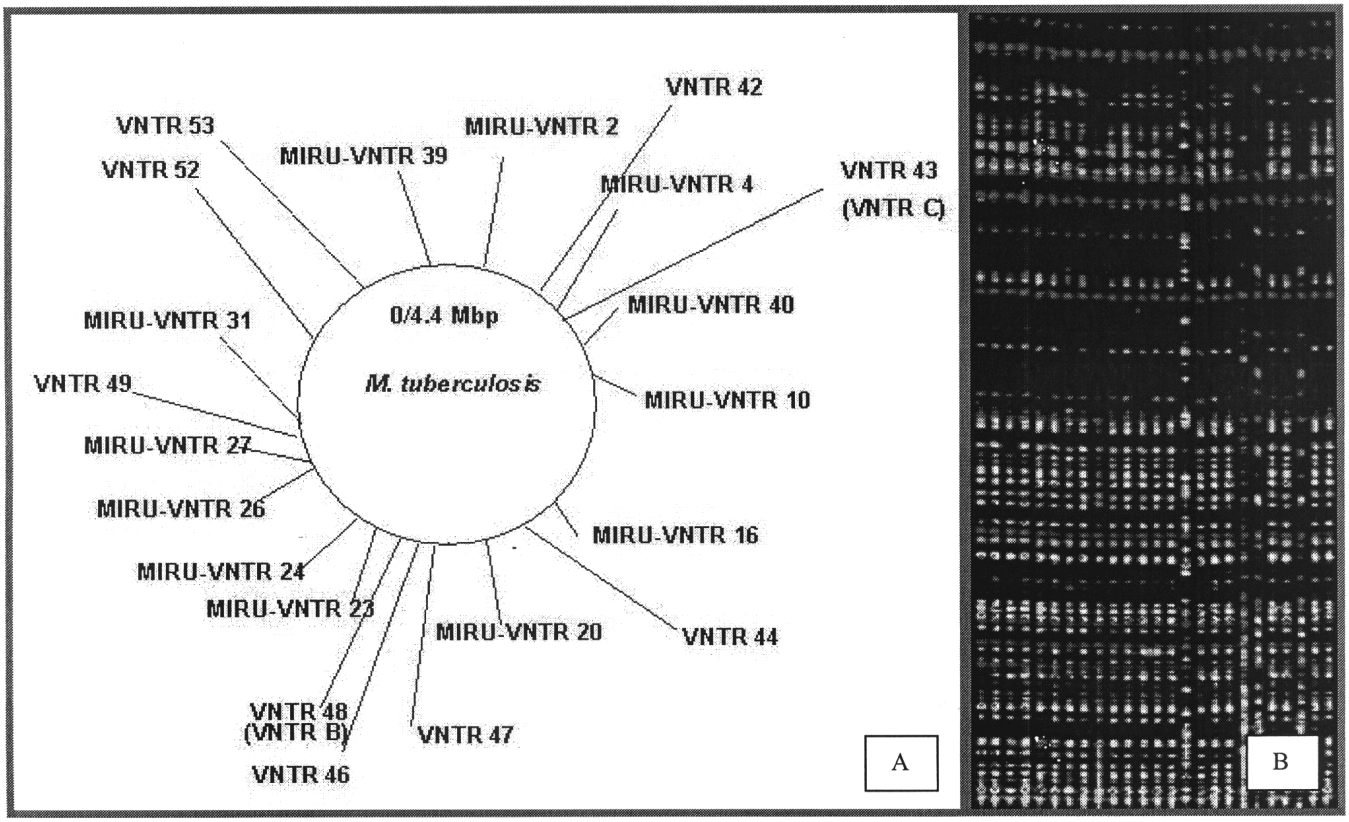

Fig. Screen dump of AmpliBASE MT®, a *M. tuberculosis* diversity knowledgebase, that is based on MIRU(A) and FAFLP(B) markers.

treatment of this disease after rifampicin was developed in 1963. At this point, however, it is believed that new chemotherapies for treatment of TB should have the following major objectives: (*i*) development of faster acting drugs to shorten the duration of treatment; (*ii*) development of novel antimicrobials to counter the emergence of bacteria resistant to current therapies; and (*iii*) development of chemotherapeutics that specifically target dormant bacilli and hence provide more effective treatment of latent TB infections.

In the post genomic scenario, a strong emphasis has been paid on structural genomics based discovery of new powerful drugs. Structural genomics is the large scale determination and analysis of protein structures. It is a new field that has been boosted by major technological advances in structure determination and by the availability of as many as 95 to 200 completed genome sequences related to many microbial and protozoan species. Several computational methods can predict the biological roles of proteins by analyzing functional relationships among proteins rather than sequence similarities. Three of these methods are phylogenetic profiling, domain fusion and gene clustering. These methods can help in identifying potential drug targets such as proteins functionally linked to known targets of anti-TB drugs and to proteins known to be essential for bacterial survival. Inhibiting these functionally related proteins should have a similar effect on the organism as inhibiting the present drug targets, since the same processes or pathways would be disrupted. An international effort in the form of the TB Structural Genomics Consortium[11] has been a most successful partnership of investigators working on different targets throughout the world. Objectives of this partnership include determining the structures of many of the mycobacterial proteins facilitating drug discovery efforts using high throughput screening, genetic identifications of key TB genes, and structure-based approaches. Other proteins that might be good targets for anti-TB therapy include extracellular proteins that are involved in virulence, and persistence determinants, since the *M. tuberculosis* cell envelope is impermeable to many

antibacterial agents. Potential therapeutic targets also include secreted antigens and proteins involved in iron acquisition, which is a critical process for *M. tuberculosis* survival. These proteins include iron regulatory proteins and enzymes involved in the production of secreted iron siderophores such as exocholins and mycobactins. Proteins specific to tubercle bacilli might constitute possible targets, since these may provide adaptations exclusive to the virulence and pathogenicity of mycobacteria. In particular, 10 per cent of the *M. tuberculosis* genome consists of genes that encode *M. tuberculosis*-specific PE, PPE and PE-PGRS proteins. The PE family is named after the presence of the motif proline-gluamic acid (PE) at positions 8 and 9 in a highly conserved N-terminal domain of approximately 110 amino acids, the C-terminal region is however, variable within the family. The PPE family resembles the PE family with a highly conserved N-terminal region (about 180 amino acids), containing the motif proline-prolineglutamic acid at positions 7-9. Lastly, the (PE-PGRS) (polymorphic GC-rich sequence) family represents an extension of the PE protein family with multiple repeats of glycine-glycine-alanine (asparagine) motifs. The function or functions of these proteins are unknown though they have been implicated in virulence and immune surveillance[12,13].

Conclusion and expert opinion

In the post genomic scenario, future of TB elimination holds a lot of promise but it is yet to be realized. With the unravelling of the *M. tuberculosis* genome, newer paradigms in the areas of epidemiology, evolution and diagnostics are almost certain. On the discovery front, scientists can now selectively target enzymes of *M. tuberculosis* which display strong antigenic attributes[14]. Genetic products of mycobacteria can be modified to cause their own death. The international initiative under the auspices of the Structural Genomics Consortium is all set to harvest the fruits of such a new biology. In the future, the distinction between drugs, immunomodulators and vaccines may disappear leading to amalgamation of all these as a super antimycobacterial to rid the world from the deadly shadow of TB. With the unprecedented growth of bioinformatics and comparative genomics, state of the art technologies are essential to utilize the vast gene pools sitting in the databases. Genomic discovery alone will not be sufficient to tackle the problem. Apart from basic scientists, the role of health care workers, clinicians and epidemiologists, public-private partnerships and the non- government organizations will be highly needed on all the frontiers.

Acknowledgment

Authors thank all the collaborators of *M. tuberculosis* evolutionary genomics interest group and Centre for DNA Fingerprinting and Diagnostics, Hyderabad for supporting this initiative. The generous support from Department of Biotechnology (DBT) and Council for Scientific & Industrial Research (CSIR), New Delhi is gratefully acknowledged.

References

1. Sharma SK, Mohan V. Multidrug-resistant tuberculosis. *Indian J Med Res* 2004; *120* : 354-76.

2. Chakhaiyar P, Hasnain SE. Defining the mandate of tuberculosis research in a post genomic era. *Med Princ Prac* 2004; *13* : 177-84.

3. Siddiqi N, Shamim M, Hussain S, Choudhary RK, Ahmed N, Prachee, *et al*. Molecular characterization of multidrug-resistant isolates of *Mycobacterium tuberculosis* from patients in north India. *Antimicrob Agents Chemother* 2002; *46* : 443-50.

4. Siddiqi N, Das R, Pathak N, Banerjee S, Ahmed N, Katoch VM, *et al*. *Mycobacterium tuberculosis* isolate with a distinct genomic identity overexpresses a TAP like efflux pump. *Infection* 2004; *32* : 109-11.

5. Ahmed N, Alam, M, Rao KR, Kauser F, Kumar NA, Qazi, NN, *et al*. Molecular genotyping of a large, multi-centric collection of tubercle bacilli indicates geographical partitioning of strain variation: Implications in global epidemiology of *M. tuberculosis. J Clin Microbiol* 2004; *42*: 3240-7.

6. Ahmed N, Caviedes L, Alam M, Rao KR, Sangal V, Sheen P, *et al*. Distinctiveness of *Mycobacterium tuberculosis* genotypes from human immunodeficiency virus type 1-seropositive and -seronegative patients in Lima, Peru. *J Clin Microbiol* 2003; *41*: 1712-6.

7. Majid AA, Ahmed N, Rao KR, Ghousunnissa S, Kausar F, Bose B, *et al*. AmpliBASE MT: A *Mycobacterium tuberculosis* diversity knowledgebase. *Bioinformatics* 2002; *20*: 989-92.

8. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, *et al*. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998; *393* : 537-44.

9. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughout genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial inter-spersed repetitive units. *J Clin Microbiol* 2001; *39* : 3563-71.

10. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, *et al*. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci USA* 2002; *99* : 3684-9.

11. Goulding CW, Apostol M, Anderson DH, Gill HS, Smith CV, Kuo MR, *et al*. The TB Structural Genomics Consortium: providing a structural foundation for drug discovery. *Curr Drug Targets Infect Disord* 2002; *2* : 121- 41.

12. Choudhary RK, Mukhopadhyay S, Chakhaiyar P, Sharma N, Murthy KJ, Katoch VM, *et al*. PPE antigen Rv2430c of *Mycobacterium tuberculosis* induces a strong B-cell response. *Infect Immun* 2003; *71* : 6338-43.

13. Chakhaiyar P, Nagalakshmi Y, Aruna B, Murthy KJ, Katoch VM, Hasnain SE. Regions of high antigenicity within the hypothetical PPE_MPTR open reading frame, Rv2608 show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. *J Infect Dis* 2004; *190* : 1237-44.

14. Banerjee S, Nandyala A, Podili R, Katoch VM, Murthy KJ, Hasnain SE. *Mycobacterium tuberculosis* isocitrate dehydrogenases show strong B-cell response and distinguish vaccinated controls from TB patients. *Proc Natl Acad Sci USA* 2004; *101*: 12652-7.

Reprint requests : Dr S.E. Hasnain, Director, Centre for DNA Fingerprinting & Diagnostics, ECIL Road, Nacharam, Hyderabad 500076, India
e-mail: hasnain@cdfd.org.in