International Journal of Engineering & Scientific Research
Vol. 6 Issue 1, January 2018, (Special Issue ICON17-MCC)
ISSN: 2347-6532 Impact Factor: 6.660
Journal Homepage: http://www.esrjournal.com Email: esrjeditor@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Mapping of Unique Regions in the Genome of *Mycobacterium tuberculosis*

Suma Sarojini*

Abstract

Microbial genomes are much easier to study and understand due to the lack of introns and very less intergenic regions.*Mycobacterium tuberculosis* is one among the top pathogenic killers causing infectious diseases. The biggest challenge in the treatment of tuberculosis is the extremely slow growing nature of the pathogen, multiple drug resistance, association with HIV, very long treatment period with multiple drugs etc. In the past few decades, many studies have been conducted to understand the biology and pathogenesis of the microbe. The last two decades saw the emergence of a number of high throughput genome sequencing technologies which enabled easy sequencing of many different species and strains of *Mycobacterium*. India has a lot of cases of tuberculosis and hence a lot of variations can also be expected in the genomes of the clinical strains. Unique regions in the genomes of particular strains of pathogenic bacteria are responsible for traits like pathogenicity, virulence etc. The paper focuses on techniques like subtractive hybridization and chromosome walking which can be applied to find out unique regions in the genomes of *Mycobacterium tuberculosis*.

Keywords:

Mycobacterium tuberculosis;
clinical strains;
genome mapping.

Author correspondence:

Suma Sarojini,
Department of Life Sciences, Christ University
Bangalore, India

1. Introduction

Even at the beginning of the new millennium, tuberculosis remains one of the number one killers among infectious diseases. Out of the 8.7 million cases estimated worldwide in 2011 by World Health Organization (WHO) Global Tuberculosis Report for 2012, 1.4 million deaths were recorded as due to TB [1]. *Mycobacterium tuberculosis*, the bacterial pathogen causing this dreaded disease is one of the most successful and scientifically challenging human pathogens in the history of mankind. It is easier to tackle a pathogen by design of treatment strategy if the virulence factors are known. The greater virulence of some bacterial strains is often associated with pathogenicity islands in the genome. This menace has become all the more significant with the emergence of extremely drug resistant (XDR) and totally drug resistant (TDR) strains [2,3].

The Indian subcontinent has been a global hotspot for the growth and spread of the TB epidemic in recent times [4] and has served as the corridor of early world-wide dissemination of *M. tuberculosis* during the ancient era [5]. India with the highest prevalence of tuberculosis (TB) worldwide, might be expected to have a pool of clinical isolates with a lot of genetic heterogeneity. Extensive genomic level studies of the clinical isolates of this organism from India have not been carried out. Comparative genomics helps a long way in unraveling the molecular basis of pathogenesis, host range and evolution of this slow growing pathogen. The last two decades saw a lot of whole genome sequencing especially of prokaryotic organisms. This has generated a wealth of nucleotide sequence data. The number of microbial genomes that have been completely sequenced is increasing day by day.

Bacterial genome comparisons give valuable information on the evolutionary patterns and also about the physiological mechanisms of bacteria [6]. Many of the bacterial genomes sequenced are those of common lab strains of the pathogen which have been thoroughly studied by classical experiments. The sequence data generated can be used for comparing related clinical isolates of the pathogen using *in silico* techniques or by

*Associate Professor, Department of Life Sciences, Christ University, Bangalore, India

using techniques such as microarray and proteomics. Subtractive hybridization is an alternative method of comparing related pathogens wherein no prior sequence information is required and is relatively simple.

2. Research Method

i) Techniques to spot unique regions in the genome

One of the low cost and less technically demanding techniques to find out unique regions in a particular bacterial strain or species is genomic subtractive hybridization. In order for a subtractive hybridization to be successful, there should be about fifty to hundred-fold enrichment of the target sequences. This can be achieved only by sorting through a plethora of DNA sequences so that the unique region is identified [7]. One of the best method to achieve this goal is to combine subtractive hybridization with PCR. This can lead to exponential amplification of only the target DNA sequences [8].

For subtraction experiments the genomic DNA sample that contain the unique sequences of interest is called the "tester" and the reference sample is called "driver". This method entails digestion of DNA from the strain of interest and a reference strain (tester and driver, respectively) with a tetra cutting restriction endonuclease such as *Rsa*I to generate DNA fragment populations with average sizes of about 0.5 kb. Generally, tetra cutting restriction enzymes are preferred as the resultant products will have more of shorter DNA fragments. Two different PCR adaptors that can join only to 5' ends of target DNAs (because their own 5' ends lack phosphate groups) are ligated to different aliquots of tester DNA. These ligated DNAs are denatured, mixed with an excess of driver DNA (that has no adaptors) and allowed to anneal. The two DNA pools are then mixed together, and more denatured driver DNA is added to further bind tester sequences that are also present in the driver genome. The remaining complementary single strands of tester DNA are allowed to anneal, and the adaptor sequences are copied onto their 3' ends. PCR is then used to obtain exponential amplification of tester DNAs with different adaptors at each end. In contrast, amplification of DNAs with the same adaptor at each end is suppressed because self-annealing of inverted repeat adaptors inhibits binding of PCR primers. Tester DNAs with an adaptor at only one end undergo linear, but not exponential, amplification. For example, the clinical isolate of *Mycobacterium tuberculosis* can serve as the tester and a type strain of *Mycobacterium tuberculosis* as the driver. The quality of genomic DNA is very important for successful subtraction.

ii) Identification of subtracted regions

PCR using primers containing adapters at one end can serve to amplify the subtracted genomic regions. Care should be taken while designing the PCR primers. The forward and reverse primers should have the sequence designed in such a way that it amplifies only those DNA fragments which carry the adapters at both ends. Since a tetra cutting restriction enzyme is normally used to digest the genomic DNA, the amplicons expected will be ranging from about 0.1 to 0.5 kb generally. Once amplicons are obtained in the 2% agarose gel by running an aliquot of the PCR reaction mixture, one can be sure that some unique regions got selected by the subtraction process. The remaining part of the PCR mixture also can be loaded in a 2% agarose gel and the fragments can be separated. These amplicons can be eluted using Gel Extraction kits. This DNA can be used for colony hybridization using driver and tester as probes (in separate colony blot hybridization reactions). The unique regions will be those clones which show a positive signal only when probed with tester DNA and not with driver DNA. These clones can be used for further characterization. If it is a unique region in the tester which is not yet reported, then the sequencing primers cannot be obtained from the start of the sequence. In such a situation, the vector in which the unique region insert lies can be made use of. One of the easiest method is to sequence the inserts in these clones using the forward and reverse reaction primers of the vector, for eg, the T7 forward primer and SP6 reverse primer in pGEMT Easy vector. Once the sequence is obtained, it can be subjected to Bioinformatics analysis so as to check for nucleotide homology in the available bacterial DNA databases. One can also translate the new sequence and check for homology with protein databases. This will give an idea of the probable functional aspects of the unique region, by means of comparison with similar sequences reported in other strains or species or even genera.

3. Results and Analysis

Whole Genome sequence of *Mycobacterium tuberculosis*

The first species of the genus *Mycobacterium* for which whole genome sequencing (WGS) was done was *M. tuberculosis* H37Rv [9]. Until the whole genome sequencing was done, manipulation of the mycobacterial genome to study the mycobacterial metabolism was slow and difficult. After WGS of *M. tuberculosis* H37Rv, whole genome sequencing of many other members of the *Mycobacterium tuberculosis* Complex (MTC) such as *M. bovis*[10] and*M. leprae*[11] were achievable. Manipulations of these organisms at the genetic level and subsequently at the phenotypic level could be achieved as a result of these findings. Comparative genomics studies such as subtractive hybridization, microarrays, Whole Genome Sequencing

(WGS) etc., are used to study *Mycobacterium* at the genetic level. Every technique has its advantages as well as drawbacks, but every technique has been informative in adding up to the available genome information.

Subtraction Studies in different species of Mycobacterium

Previous mycobacterial genome subtractions have identified a wide variety of genes. In *M. avium* subsp. *paratuberculosis*, a pathogenicity island containing genes potentially involved in cellwall polysaccharide modification were identified [12]. More than twenty genes were found in *M. avium* that were absent in *M. intracellulare*. Some of these genes were hypothesized to play a role in the invasion of intestinal mucosa [13]. Genomic subtraction has been used to compare the genomes of *M. bovis* and *M. bovis* BCG Connaught with a view to find out the genes responsible for pathogenicity and virulence. Three regions – RD1, RD2 and RD3 - have been identified that were deleted in BCG during its attenuation [14]. Polyketides are important elements in the cell wall composition of Mycobacterium, contributing to their virulence. A type I polyketide synthase locus was identified in *M. ulcerans* when subtraction was done using *M.ulcerans* and *M. marinum* [15]. Thus genomic subtraction technique proves to be a prominent technique to identify and isolate unique genomic regions in various species of mycobacteria.The subtraction procedure was done as described in figure 1.

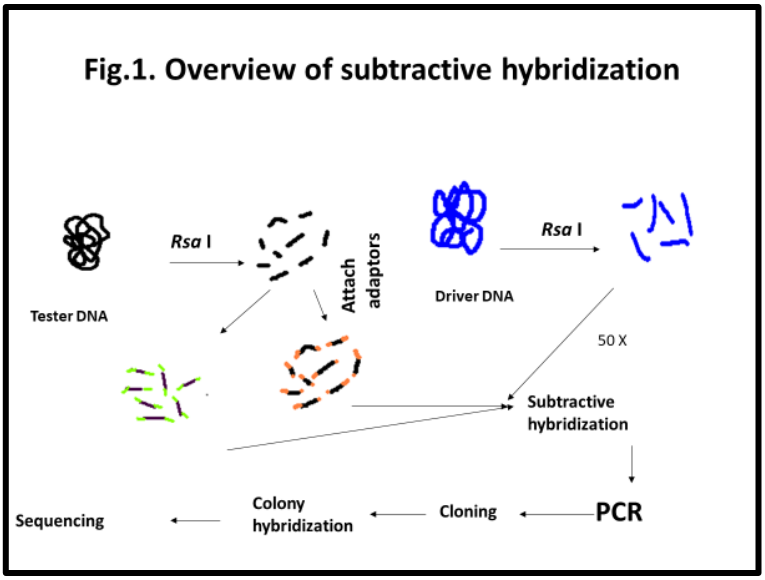

Figure 1: Scheme depicting genomic subtractive hybridization - The tester DNA is digested with *Rsa*I restriction enzyme and divided equllay into two sets. One set is ligated to adapter 1 and another set to adapter 2. Driver DNA is separately digested wuth *Rsa*I. Digested driver DNA is added 50 times in excess with tester DNA, sets 1 and 2 and hybridized, PCR amplified, cloned, screened and sequenced to get subtracted fragments.

4. Conclusion

Taking into consideration the results of these studies, one can conclude that subtractive hybridization is indeed a starting point in the detailed study of finding unique regions in the genomes of related species or strains of organisms. This is especially significant and easier in the case of prokaryotic organisms due to two reasons - i) they lack introns ii) they have very less of intergenic DNA. These days, a lot of new generation sequencing technologies have come up which can be used to find out the entire genome sequences of many species or strains in a faster manner. The significance of the subtraction and mapping studies lies in the fact that it is a lowcost technology and also less technical expertise is required.

References

[1] WHO. "Global tuberculosis report 2012," Geneva, Switzerland, 2012
[2] Sotgiu, G. Ferrara, A. Matteelli, M. D. Richardson, Centis, R., Ruesch-Gerdes, S., Toungoussova, O., Zellweger, J. P., Spanevello, A., Cirillo, D., Lange, C and Migliori, G. B. "Epidemiology and clinical management of XDR-TB: a systematic review by TBNET". Eur. Respir. J. 33 ,pp. 871–88, 2009.
[3] Velayati, A., Masjedi, R., Farnia, P., Tabarsi, P., Ghanavi, J., ZiaZarifi, A.H and Hoffner, S.E. "Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran", Chest..136, pp: 420–425, 2009.
[4] WHO Report. "Global tuberculosis control: surveillance, planning, financing", Infect Dis . 12, pp: 1367–1374, 2008.

[5] Gutierrez, M. C, Ahmed, N., Willery, E., Narayanan, S., Hasnain, S. E., Chauhan, D.S., Katoch, V. M., Vincent, C. Locht, C and Supply, P. "Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India", Emerg. Infect. Dis. 12, pp:1367–1374, 2006.
[6] Fraser, J. Eisen, FleischmanN, R.D., Ketchum, K.A and Peterson, S. "Comparative genomics and understanding of microbial biology", Emerg. Infect. Dis. 6, pp: 505–512, 2000.
[7] Myers, R, M. "The pluses of subtraction", Science 259, pp.942-943, 1993.
[8] Lisistsyn, N. A., Lisistsyn, N, M and Wigler, M. "Cloning the difference between two complex genomes", Science 259, pp.946–51, 1993.
[9] Cole, S.T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, A. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell *et al*.,. "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence", Nature 393, pp. 537-544, 1998.
[10] Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S. Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, L. R., Parkhill, J Barrell, B.G., Cole, S. T., Gordon, S. V and Hewinson R. G. "The complete genome sequence of *Mycobacterium bovis*, Proc Natl Acad Sci U S A 100, pp.7877-7882, 2003.
[11]. Cole, S.T., Supply, P and Honore, N. "Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity", Lepr Rev 72, pp. 449-461, 2001.
[12] Tizard, M., Bull, T., Millar, D., Doran, T., Martin, H., Sumar, N., Ford, J and Hermon-Taylor, J. "A low G+C content genetic island in *Mycobacterium avium* subsp. paratuberculosis and *M. avium* subsp. silvaticum with homologous genes in *Mycobacterium tuberculosis*", Microbiology 144 (12), pp.3413-3423, 1998.
[13] McGarvey, J. A., and Bermudez, L. E. "Phenotypic and genomic analyses of the *Mycobacterium avium* complex reveal differences in gastrointestinal invasion and genomic composition", Infect Immun 69, pp.7242-7249, 2001.
[14] Mahairas, G. G., Sabo, P. J. Hickey, M. J., Singh, D. C and Stover, C. K. "Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis"*, J Bacteriol 178, pp.1274-82, 1996.
[15]Jenkin, G.A., Stinear, T. P., Johnson P. D and Davies J. K. "Subtractive hybridization reveals a type I polyketide synthase locus specific to *Mycobacterium ulcerans."* J Bacteriol 185, pp.6870-6882, 2003.