Indian Journal of Biotechnology
Vol. 5, January 2006, pp. 58-61

Phylogentetic analysis of *Mycobacterium leprae* genome for identification of novel drug targets

Aditya Saxena, Pritish Varadwaj, S Singh, Manoj Jaiswal[2] , K Misra[1] and T Lahiri[*]

Indian Institute of Information Technology, Allahabad 211 012, India
[1]Indo-Russian Centre for Biotechnology, Allahabad 211 011, India
[2]Centre of Physiology and Pathophysiology, University of Gottingen, Germany

Received 14 May 2004; revised 20 December 2004; accepted 14 January 2005

The reductive evolution provides a pathogen minimal gene set. Due to loss of genes, for drug target identification, biochemical techniques cannot be used since these pathogens are difficult to culture. Besides classical drug targets i.e. toxins, adhesions, some novel drug targets, using a molecular phylogeny approach, can be identified. Horizontally transferred genes in bacterial pathogen from their eukaryotic host has been classified as 'Novel drug target', and a reusable system for identifying horizontally transferred gene has been built using a molecular phylogeny approach. Using this system, two genes have been identified in *Mycobacterium leprae,* which have also been identified by other methods.

Keywords: horizontal gene transfer, novel drug targets, phylogenetic analysis
IPC Code: Int. Cl.[7] C12R1:32

Introduction

Leprosy, one of the oldest recorded diseases, remains a major public health problem till date. Although prevalence has been reduced extensively by WHO multidrug therapy and vaccination with BCG, the incidence of the disease remains worrying with more than 690,000 new cases reported annually. In 1873, in the first convincing association of a microorganism with a human disease, Armauer Hansen discovered the leprosy bacillus in skin biopsies but failed to culture *Mycobacterium leprae*. A century later, the nine-banded armadillo was used as a surrogate host, enabling large quantities of the bacillus to be isolated for biochemical and physiological studies. Subsequent efforts to demonstrate multiplication in synthetic media have been equally fruitless, although metabolic activity can be detected. The exceptionally slow growth of the bacillus, which has a doubling time of 14 days, may have contributed to these failures. The means of transmission of leprosy is uncertain but, like tuberculosis, the infection also thought to be spread by the respiratory route because lepromatous patients harbour bacilli in their nasal passages.

* Author for correspondence:
Tel: 91-9415618262; Fax: 91-532-2430006
E-mail- tlahiri@iiita.ac.in

The bacterium accumulates principally in the extremities of the body where it resides within macrophages and infects the Schwann cells of the peripheral nervous system. Lack of myelin production by infected Schwann cells, and their destruction by host-mediated immune reactions, leads to nerve damage, sensory loss and the disfiguration that, sadly are the hallmarks of leprosy.

The complete gnome sequence of *M. leprae* contains 3,268,203 base pairs (bp), and has an average G+C content of 57.8%. These values are much lower than those reported for the *M. tuberculosis* genome, which comprises 4,000 genes, 4,411,532 bp and 65.6% G+C from detailed pairwise comparisons of both genome and proteome sequences. Only 49.5% of the *M. leprae* genome contains protein-coding genes, whereas 27% contains recognizable pseudogenes (inactive reading frames with functional counterparts in the tubercle bacillus). The remaining 23.5% of the genome does not appear to be coding, and may correspond to regulatory sequences or even gene remnants mutated beyond recognition. The distribution of the 1,116 pseudogenes is essentially random and if these are excluded 1,604 potentially active genes remain, of which, 1,439 are common to both pathogens. Among the 165 genes with no orthologue in *M. tuberculosis* are 29 for which functions can be attributed[1]. Many of the 136 residual

coding sequences in *M. leprae*, which show no similarity to known genes, may also represent pseudogenes, as they are shorter than average sequence and occur in regions of low gene density. Pseudogenes are non-functional regions in the genome that have arisen as a consequence of accumulating mutations that either result in the premature termination of proteins during protein synthesis or the disruption of transcription[1].

The effort to completely sequence bacterial genomes is rapidly gaining momentum. Preliminary analysis of these genome sequences has provided valuable insight into the biology of the targeted organisms and numerous new 'classical' virulence factors e.g. toxins, adhesions and invasions have been identified, many novel virulence-associated genes still await discovery. The comparative analysis of a number of bacterial genome sequences has provided evidence that the horizontal transfer of genetic material between even evolutionarily disparate microbial lineages is a very common event. Evidence of horizontal gene transfer between prokaryotes[2-6], from bacteria to eukaryotes[7,8] and even from eukaryotes to bacteria has been presented previously[9-11]. The process of horizontal gene transfer allows instantaneous delivery of fully functional, complex metabolic capabilities to a bacterial lineage, and thus represents a powerful mechanism by which the outcome of a bacteria-host interaction can be permanently altered. A number of studies have discovered the presence of horizontally acquired genomic segments, known as pathogenicity islands, which play a major role in the virulence processes of many bacterial pathogens and can convert a benign bacterium into a pathogen upon incorporation[12].

Materials and Methods

South Africa National Bioinformatics Institute (SANBI) has engineered a reusable system for the detection of horizontal gene transfer, which can be explained as below:

1. Identification of Candidate Horizontally Transferred Genes by Comparative Similarity Searching

a) BLASTP organisms predicted proteins against subset of eukaryotes and bacterial subset of GENBANK.
b) Eukaryotes proteins that scored higher with at least 10 orders of magnitude difference in E-value from bacteria were selected as possible candidates.
c) All *Mycobacterial* protein sequences should be removed.

2. Confirmation of Absence of Bacterial Orthologous by More Sensitive Sequence Similarity Searches

Candidate proteins were compared against a complete *nr* protein database using the NCBI *PSI-BLAST* to ensure that those proteins matching eukaryotic proteins exclusively in the first step truly had no bacterial orthologous.

3. Phylogenetic Analysis

a) Bacterial sequences identified by *PSI-BLASTA* and representative protein sequences from the three kingdoms were aligned using CLUSTALW and subjected to phylogenetic analysis, using the neighbour joining and protein parsimony methods of the PHYLIP package.
b) Candidate that presented a non-congruent phylogeny with bootstrapping support => 70% was classified as horizontal gene transfers.

Results and Discussion

The authors have identified two predicted proteins, which are horizontally transferred from eukaryotic host into pathogen genome *Prolyle t-RNA Synthetase* and *Uridine phosphorylase.* Patho-biological relevance of these transfers for *leprae* requires some wet lab experiments, and due to unavailability of facility we did literature search and found few explanations for *Prolyle t-RNA Synthetase* as 1). Aminoacyl-tRNA synthetases have been divided into two evolutionary distinct families, each characterized by a class defining catalytic short conserved sequence motifs[12-14]. Within each class subclasses can be identified containing more closely homologues synthetases that often share a common anticodon-binding module[12]. In class II, the largest subclass is IIa, which contains the Ser-Thr-, Pro-, Gly- and His-tRNA synthetase (RSs)[15,16]. (ProRS) is a class IIa synthetase that, according to sequence analysis, occurs in different organisms with one of two quite distinct structural architectures: prokaryote-like and eukaryote/archaeon-like[12,17]. Differences between 'prokaryote-like' and 'eukaryote-like' IIeRSs accounts for the fact that pseudomonic acid is a potent inhibitor of the former but not the later[12]. So our inference is that the transfer makes pathogen resist to pseudomonic acid and enhance its pathogenicity. 2). Prolyl-tRNA synthetase, encoded by proS, is more

similar to the enzymes of *Borrelia burgdorferi* and eukaryotes such as drosophila, humans and yeast. It has been proposed that horizontal transfer of tRNA synthetase genes occurs frequently, and that the pathogen *B. burgdorferi* may have acquired proS from its host[17]. Comparing the genetic context provides further support for this hypothesis, as the *M. leprae* proS is both displaced and inverted with respect to the *M. tuberculosis* genome, consistent with a recent acquisition[1].

As the authors have shown that the eubacterium, *M. leprae* acquired, probably by horizontal transfer, so both these examples also illustrate how recognizing the species differences between human and pathogen synthetase is potentially of great importance in the design of new antibiotics that target these enzymes. For uridine phosphorylase, we could not find any paper, which could give relevance of transfer but one thing noticed in doing BLAST, was that we did not find any bacterial protein with significance scores. Even in PUBMED, no reference is there for prokaryotic uridine phosphorylase. This is the reason we could not collect prokaryotes orthologous for tree building so this cannot be claimed as a confirmed drug target by the current piece of work. However, it may be an initial guess for the identification of possible drug targets.

Conclusion

This system is very useful for identifying novel drug target especially like *M. leprae* type of organism which shows a drastic example of reductive evolution and have minimal gene set, due to this reason they are difficult to culture, so for drug target identification we will have to use comparative genomics. And molecular phylogeny has been proved to be very helpful in comparing microbe's genomes by the method of tree building. As a further work the authors have modelled both the protein using a combined approach of comparative modelling and threading and we are in a process of submitting in to PDB. Even in general protein analysis this type of work can be annexred in order to understand its function as weather it also have some role in pathology of that pathogen in study.

The following (Figs 1 & 2) trees have been generated using the neighbour joining method of PHYLIP with bootstrap value > 70, same type of trees are also generated by protein parsimony method. These show fairly well phylogenetic incongruence of leprae protein with bacterial orthologous.

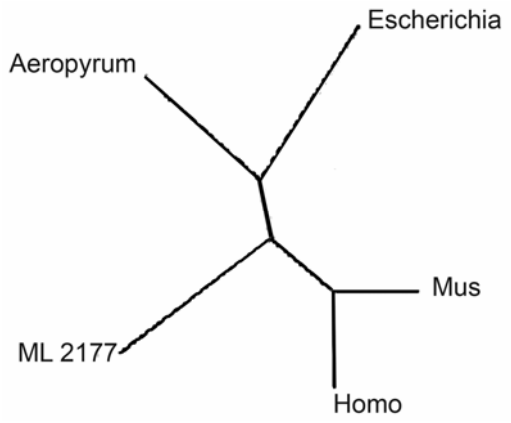

Fig 1—Tree generated by neighbour joining method for Prolyle t-RNA synthetase

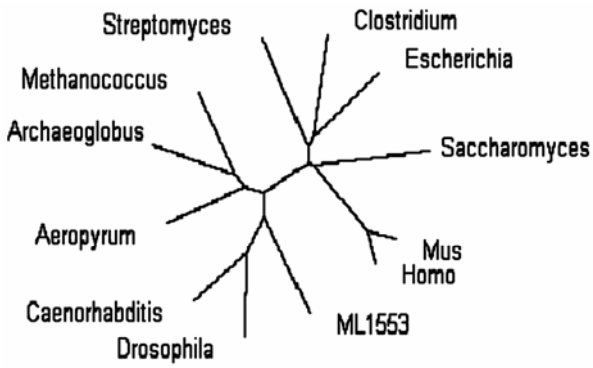

Fig 2—Tree generated by neighbour joining method for Uridine phosphorylase

References

1 Cole S T, Eiglmeier K, Parkhill J, James K D, Thomson N R *et al*, Massive gene decay in the leprosy bacillus, *Nature (Lond)*, 409 (2001) 1007-1011.

2 Kroll J S, Wilks K E, Farrant J I & Langford P R, Natural genetic exchange between *Haemophilus* and *Neisseria*: Intergeneric transfer of chromosomal genes between major human pathogens, *Proc Natl Acad Sci USA*, 95 (1998) 12381-12383.

3 Lawrence J G & Ochman H, Molecular archaeology of the *Escherichia coli* genome, *Proc Natl Acad Sci USA*, 95 (1998) 9413-9417.

4 Woese C, The Universal ancestor, *Proc Natl Acad Sci USA*, 95 (1998) 6854-6859.

5 Bork P, Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile molecules that cross-phyla horizontally, *Protein*, 17 (1993) 363-374.

6 Doolittle W F, You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genome, *Trends Genet*, 14 (1998) 307-311.

7 Stephens R S, Kalman S, Lammel C, Fan J, Marathe R *et al*, Genome sequence of an obligate intrncellular pathogen of humans: *Chlamydia trachomati, Science.* 282 (1998) 754-759.

8 Wolf Y I, Aravind L & Koonin E V, Rickettsiae and Chlamydiae, evidence of horizontal gene transfer and gene exchange, *Trends Genet*, 15 (1999) 173-175.

9 Wolf Y I, Kondrashov A S & Koonin E V, Interkingdom gene fusions, *Genome Biol*, 1(6) (2000) 131.1-131.13.

10 Groisman E A & Ochman H, Pathogenicity islands: Bacterial evolution in quantum leaps, *Cell*, 87 (1996) 791-794.

11 Yaremchuk A, Cusack S & TukNordeen M, Crystal structure of a eukaryote/archeon-like prolyle-tRNA synthetase and its complex with $tRNA^{Pro}$ (CGG), *EMBO J*, 19 (2000) 4745-4758.

12 Gamieldien J, Novel genomic approaches for the identification of virulence genes and drug targets, in pathogenic bacteria, Ph D Thesis, South African National Bioinformatics Institute, 1999.

13 Madan Babu M, Did the loss of sigma factors initiate pseudogene accumulation *in M leprae? Trends Microbiol*, 11 (2003) 59-61.

14 Vissa V D & Brnnnan P J, The genome of *Mycobacterium leprae:* A minimal mycobacterial gene set, *Lepr Rev*, 72 (2001) 387-398.

15 Tekaia F, Gordon S V, Garnier T, Brosch R, Barrell, Analysis of the proteome of *Mycobacterium tuberculosis in silico, Tubercle Lung Dis* 79 (1999) 329-342.

16 Quadri L E, Sello J, Keating T A, Weinreb P H & Walsh C T, Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly if the virulence-conferring siderophore mycobactin. *Chem BioI*, 5, (1998) 631-645.

17 Cole S T, Brosch R, Parkhill J, Garnier T, Churcher C *et al*, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature (Lond)*, 393 (1998) 537-544.