International Journal of Medical Microbiology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.com/locate/ijmm

Gene cooption in Mycobacteria and search for virulence attributes: Comparative proteomic analyses of *Mycobacterium tuberculosis*, *Mycobacterium indicus pranii* and other mycobacteria

Yadvir Singh[a,1], Sakshi Kohli[a,1], Divya Tej Sowpati[b,1], Syed Asad Rahman[c], Anil K. Tyagi[d], Seyed E. Hasnain[a,b,*]

[a] *Kusuma School of Biological Sciences, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India*
[b] *Dr Reddy's Institute of Life Sciences, University of Hyderabad Campus, Gachibowli, Professor C.R. Rao Road, Hyderabad 500046, India*
[c] *European Molecular Biology Laboratory (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK*
[d] *Department of Biochemistry, University of Delhi South Campus, Benito Juarez Marg, New Delhi, India*

ARTICLE INFO

Article history:
Received 22 November 2013
Received in revised form 16 May 2014
Accepted 21 May 2014

Keywords:
Tuberculosis
M.tb
MIP
Virulence factors
Gene cooption
PE_PGRS

ABSTRACT

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is a leading infectious disease taking one human life every 15 s globally. Mycobacterium undergoes reductive evolution; the ancestors have bigger genome size and rich in metabolic pathways. *Mycobacterium indicus pranii* (*MIP*) is placed much above *Mycobacterium tuberculosis* (*M.tb*) in evolutionary scale and is a non-pathogenic, saprophytic mycobacterium. Our in silico comparative proteomic analyses of virulence factors of *M.tb* and their homologs in 12 different Mycobacterial species, including *MIP*, point toward gene cooption as an important mechanism in evolution of mycobacteria. We propose that adaptive changes in niche factors of non-pathogenic mycobacterium, together with novel gene acquisitions, are key players in the evolution of pathogenicity. Antigenic analyses between *M.tb* and *MIP* highlighted the importance of PE/PPE family in host immunomodulation, further supporting the likely potential of *MIP* as an effective vaccine against TB.

Introduction

The genus mycobacterium consists of more than 120 species, which can be easily categorized into strict pathogens like *Mycobacterium tuberculosis* (*M.tb*), opportunistic pathogens like *M. avium* and non-pathogens like *M. smegmatis*. The genome size of non-pathogens is larger than that of strict pathogens suggesting that loss of genes is an integral part of ongoing evolution of slow growing mycobacterial pathogens (Ahmed et al., 2008; Brosch et al., 2002). In addition to the known reductionist evolution in mycobacterium, few families like PE/PPE, MCE have expanded down the evolutionary course. Gene acquisition helps the bacteria to diversify and adapt in a varying environmental conditions (Juhas et al., 2009). *M.tb* has acquired genes specific to its survival in the host since its divergence from the common ancestor with the closest relative *M. kansasii* (Veyrier et al., 2011) or *M. canettii* (Supply et al., 2013) or *M. marinum* which can survive in a broader environmental niche (Stinear et al., 2008). Thus, mycobacterium evolution involves both losses of genes, not critical for survival in the host, from saprophytic predecessor, and gain of few to help the pathogens establish in the host.

Gene cooption involves gain of new function of a gene by duplication or without duplication. Without gene duplication, genes can be coopted for some other function by change in coding sequences and gain of novel domains. On the other hand gene duplication can assist in cooption by either sub-functionalization (conservation of function partially) or neo-functionalization (gain of a new function) of the paralogous genes (True and Carroll, 2002). Gene cooption has also been used to explain the presence of virulence factors in both pathogenic and environmental Rhodococci bacteria, which belong to the same taxonomical order as *M.tb* (Letek et al., 2010).

Mycobacterium indicus pranii (*MIP*), a saprophytic non-pathogenic mycobacterium, is placed in *M. avium* complex of mycobacteria and is evolutionary close to opportunistic pathogens of the MAC family (Ahmed et al., 2007; Saini et al., 2009). This

* Corresponding author at: Kusuma School of Biological Sciences, Indian Institute of Technology, HauzKhas, New Delhi 110016, India. Tel.: +91 1126597522; fax: +91 1126597530.
E-mail addresses: seyedhasnain@gmail.com, sehiitd@gmail.com (S.E. Hasnain).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ijmm.2014.05.006
1438-4221/

Table 1
List of thirteen mycobacterium species included in the analyses.

Mycobacterium species	Categorization based on virulence	NCBI Ref Seq accession number	Number of proteins
Mycobacterium tuberculosis H_{37}Rv	Strict pathogen	NC_000962	4018
Mycobacterium bovis subsp. bovis AF2122/97	Strict pathogen	NC_002945	3918
Mycobacterium leprae TN	Strict pathogen	NC_002677	1605
Mycobacterium ulcerans Agy99	Strict pathogen	NC_005916, NC_008611	4241
Mycobacterium marinum M	Strict pathogen	NC_010604, NC_010612	5452
Mycobacterium avium 104	Opportunistic pathogen	NC_008595	5120
Mycobacterium intracellulare ATCC 13950	Opportunistic pathogen	NC_016946	5144
Mycobacterium avium subsp. paratuberculosis K-10	Opportunistic pathogen	NC_002944	4350
Mycobacterium indicus pranii MTCC 9506	Non-pathogen	NC_018612	5254
Mycobacterium smegmatis MC2 155	Non-pathogen	NC_008596	6717
Mycobacterium gilvum PYR-GCK	Non-pathogen	NC_009338, NC_009339, NC_009340, NC_009341	5579
Mycobacterium vanbaalenii PYR-1	Non-pathogen	NC_008726	5979
Mycobacterium vaccae	Non-pathogen	ALQA00000000 (GenBank)	5949

unique placement of *MIP* on the mycobacterium evolutionary scale gives it the benefit of sharing more number of virulence factors of *M.tb* than other non-pathogenic members. The *MIP* genome has been recently annotated (Saini et al., 2012). This annotation has revealed the presence of homologs of *M.tb* virulence factors in *MIP* despite its non-virulent nature, which further supports its role as a predecessor of *Mycobacterium avium* complex (Ahmed et al., 2007).

MIP has a well-established role as an immunomodulator in various diseases (Ahmad et al., 2011; Katoch et al., 2008; Rakshit et al., 2012; Talwar, 1999). It has been used as a successful commercial immunotherapeutic vaccine 'Immuvac' against leprosy (Nath, 1998). Further, in animal trials *MIP* has been found to be protective against tuberculosis infection and has entered human clinical trials (Gupta et al., 2009, 2012; Saini et al., 2012). The broad-spectrum antigenic potential of *MIP* lies in the fact that it shares B and T cell antigens with other mycobacteria including *M.tb* and *M. leprae* (Saini et al., 2012; Singh et al., 1992; Yadava et al., 1991).

Comparative genomics can help shed light on the molecular basis of pathogenicity and non-pathogenicity of *M.tb* and *MIP*, respectively and the evolution and phenotypic differences between fast growing soil mycobacteria and slow growing pathogens. We compared the antigenic proteins of *MIP* with *M.tb*, based on the antigenicity index analyses by VaxiJen (Doytchinova and Flower, 2007). For a meaningful interpretation, we included another rapidly growing saprophytic mycobacterium '*Mycobacterium vaccae*', which has also been investigated as vaccine candidate against TB in various clinical trials (Yang et al., 2010, 2011). Our comparative proteomic analyses revealed the presence of homologs of some *M.tb* virulence factors in non-pathogenic and opportunistic mycobacteria as well. Conservation of virulence factors across mycobacteria point to gene cooption, besides gene acquisition and gene loss, as a driving force in mycobacterial evolution. Further, lateral gene transfer (LGT) analysis of *M.tb*, suggests that along with cooption of primary niche factors, *M.tb* has acquired other genes which helped it to adapt better in intracellular niche of host.

Materials and methods

Sequences and genomes

All proteome and chromosome files were obtained in FASTA format from NCBI (http://www.ncbi.nlm.nih.gov/). The different Mycobacterium species used in this study and their respective accession numbers are listed in Table 1. List of virulence factors was downloaded from Virulence Factors DataBase (VFDB, http://www.mgc.ac.cn/VFs/main.htm). MvirDB (http://mvirdb.llnl.gov/) and ARDB (http://ardb.cbcb.umd.edu/). A curated list of virulence factors from all the above databases was created based on the experimental evidence through text mining (Supplementary Table 1). Proteins with no experimental validation were not considered for further analyses in this study.

Sequence similarity and functional analysis

Mycobacterial proteins having similar amino acid sequences were procured using BLASTp program of NCBI BLAST suite (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/). Whole proteome BLAST database for each mycobacterial species was created using makeblastdb tool of BLAST suite, and BLASTp was run at an *E*-value cut-off of 10e−4. BLASTp results were parsed with 35% and 70% sequence similarity cut-off (Rost, 1999), for detecting similar proteins. Gene Ontology (GO) IDs for MIP proteins were obtained from UniProt (http://www.uniprot.org/uniprot/?query=taxonomy:35617) and were submitted in REViGO (http://REViGO.irb.hr/) to summarize and cluster them.

Globularity and hydrophobicity analysis

GlobPlot (http://globplot.embl.de/) was used to predict the globular and disordered regions of a protein, using the Russell/Linding propensities (Linding et al., 2003). The globular domains were predicted with a minimum domain length of 74 amino acids and a maximum linker region of 15 amino acids, while disordered regions were filtered with a minimum length of 5 amino acids and maximum linker gap of 4 amino acids.

The GRAVY (Grand Average of Hydropathicity) and Instability indices of proteins were calculated using the ProtParam tool of ExPASy (http://web.expasy.org/protparam/). ProtParam calculates GRAVY using Kyte-Dolittle's hydropathic indices of amino acids (Kyte and Doolittle, 1982). The instability index was measured by analyzing the dipeptide composition of a protein (Guruprasad et al., 1990).

Analysis of lateral gene transfer

Regions that are putatively acquired laterally are predicted using the Alien Hunter tool (http://www.sanger.ac.uk/resources/software/alien_hunter/). Alien Hunter uses an Interpolated Variable Order Motif distribution approach to calculate compositional bias in a genome, and defines an arbitrary threshold based on the sequence composition. The boundaries of regions thus predicted are further optimized by employing a 2-state 2nd order Hidden Markov Model (Vernikos and Parkhill, 2006). Gene

Please cite this article in press as: Singh, Y., et al., Gene cooption in Mycobacteria and search for virulence attributes: Comparative proteomic analyses of *Mycobacterium tuberculosis*, *Mycobacterium indicus pranii* and other mycobacteria. Int. J. Med. Microbiol. (2014), http://dx.doi.org/10.1016/j.ijmm.2014.05.006