Commentary

The value of comparative genomics in understanding mycobacterial virulence: *Mycobacterium tuberculosis* H37Ra genome sequencing – a worthwhile endeavour

If the importance of a disease for mankind is measured by the number of fatalities it causes, then tuberculosis must be considered much more important than those most feared infectious diseases, plague, cholera and the like. If one only considers the productive middle-age groups, tuberculosis carries away one-third, and often more.

Robert Koch, 24 March 1882

It is indeed tragic that these words holds true to this day; nearly one-third of the world's population is latently infected with *Mycobacterium tuberculosis* and the annual TB death toll is ~2 million (Corbett *et al* 2003). Robert Koch established the etiology of TB in 1882 and the much debated but widely used BCG (bacille Calmette-Guérin) vaccine was developed in 1921. Several compounds with anti-TB activity were introduced as therapeutics during 1943 to 1963 and it was anticipated that this would mark the rapid exit of TB as a public health challenge (Ramakrishnan and Chandrashekhar 1999). However, the ability of *M. tuberculosis* to persist indefinitely in tissues in a metabolically quiescent state requires the implementation of long treatment regimens. This makes patient compliance difficult, leading to treatment failure and the emergence of drug resistant strains. Another attribute of persistent organisms, which poses a great obstacle to the world wide control and eradication of TB, is that the non-replicating bacilli may be in a state of 'drug indifference' wherein they are not killed by drugs (Wayne and Sohaskey 2001). Our understanding of the physiology of the persistor bacilli has improved in recent years thanks to the development of *in vitro* models, generation of mutant *M. tuberculosis* strains, global gene expression analysis and to the increasing availability of various mycobacterial whole genome sequences (Tyagi and Sharma 2004). Based on our improved understanding, novel approaches are being attempted to target the persistor bacterial population.

In this post-genomic era, now more than ever before, an understanding of strain-to-strain variation is expected to fast-forward the rational design and development of effective diagnostic reagents, vaccines and drugs in the case of the 'difficult-to-crack' TB pathogen. Therefore, not surprisingly the present era of mycobacterial genomics is rapidly evolving into a period of comparative genomics and transcriptomics. In comparative genomics, sequence data of related species are scrutinized from the perspective of correlating differences in the physiology, biochemistry and virulence with those found at the genetic level. Several genomes including two strains of *M. tuberculosis* (i.e. H37Rv and CDC1551), *M. bovis*, *M. leprae*, *M. smegmatis*, *M. ulcerans*, *Mycobacterium* sp. MCS, *Mycobacterium* sp. KMS, *M. vanbaalenii* PYR-1, *M. avium* and *M. avium* subsp. *paratuberculosis* have been sequenced and annotated completely (www.ncbi.nlm.nih.gov/genomes/lproks.cgi). Fifteen additional mycobacterial genome sequencing projects have either been completed and awaiting annotation or are underway.

The elucidation of the *M. tuberculosis* H37Rv genome sequence is unanimously acknowledged to be a landmark achievement in that it provided the first truly global insight into the tubercle bacillus (Cole *et al* 1998). However, given that *M. tuberculosis* H37Rv is a classical laboratory reference strain that has been passaged for many decades outside the human host, there were concerns relating to how accurately its sequence reflected the pathogenic potential of tubercle bacilli. This led to the sequencing of the *M. tuberculosis* CDC1551 genome (also known as the Oshkosh or CSU93 strain), a clinical isolate from a rural area in the United States (Fleischmann *et al* 2002). Although there has been some controversy over

Keywords. *Mycobacterium tuberculosis* H37Ra; genomics; virulence

whether or not the CDC1551 strain is more virulent than other strains in animals, it was demonstrated to show increased transmissibility and greater immunoreactivity (Bishai *et al* 1999). Comparison of the CDC1551 and H37Rv genome sequences revealed a number of notable differences (table 1). The high levels of polymorphism disproved the earlier notion that there is little sequence variation in *M. tuberculosis*. Four gene families including the Pro-Glu/Pro-Pro-Glu (PE/PPE) family showed higher substitution frequency than that observed for the genome as a whole, warranting further investigation into their role in virulence or immunogenicity.

Analysis of the genome sequence of the causative agent of leprosy, *M. leprae*, revealed an extreme case of reductive evolution (Cole *et al* 2001). Its genome contains only ~ 40% of the genes in comparison to *M. tuberculosis* H37Rv, leading to the loss of many important metabolic functions (table 1). While preserving genes required for its transmission, establishment and survival in the host, *M. leprae* has discarded genes that can be compensated for by a host-dependent parasitic lifestyle. Analysis of the *M. leprae* genome therefore provided an explanation for its obligate intracellularism and inability to grow *in vitro*.

A recent addition to the list of annotated complete mycobacterial genomes is that of *M. bovis* AF2122/97 (table 1), which causes TB in a range of animal species and also in man (Garnier *et al* 2003). Strikingly, its genome sequence is >99.95% identical to that of *M. tuberculosis* and comparison with *M. leprae* reveals a number of common gene losses, suggesting the removal of functional redundancy. Cell wall components and secreted proteins showed the maximum variation suggesting their potential relevance in host-pathogen interaction or immune evasion (Garnier *et al* 2003). Furthermore, there are no genes unique to *M. bovis*, implying that differential gene expression may be responsible for the host specificity of human and bovine bacilli. *M. bovis* was also the progenitor of the human BCG vaccine

Table 1. Insights from comparative genomics with *M. tuberculosis* H37Rv.

Organism	Strain (size)	Inference deduced
M. tuberculosis (Fleischmann *et al* 2002)	CDC1551 (4.4 Mb)	CDC1551 contains 49 insertions relative to H37Rv, 35 of which affect ORFs H37Rv contains 37 insertions relative to CDC1551, 26 of which affect ORFs Frequency of SNPs in *M. tuberculosis* is lower than in other bacteria 4 copies of IS*6110* are present in CDC1551 compared to 16 copies in H37Rv Polymorphic loci included a phospholipase C, a membrane lipoprotein, members of an adenylate cyclase family and members of PE/PPE gene family, some of which have been implicated in virulence or host immune response
M. leprae (Cole *et al* 2001)	Armadillo-derived Indian isolate (3.27 Mb)	Massive gene decay resulting in: – loss of many important regulatory and metabolic activities including siderophore production, part of the oxidative and most of the microaerophillic and anaerobic respiratory chains, and numerous catabolic and their regulatory circuits – restricted neural predilection
M. bovis (Garnier *et al* 2003)	AF2122/97 (4.35 Mb)	>99.95% identical to H37Rv; 11 regions of difference (RD), varying in size from ~1 to 12.7 Kb Only one region (TbD1) absent in majority of *M. tuberculosis* strains No unique genes, broad host range likely to be the result of differential gene expression Cell wall components and secreted proteins show greatest variation

strains, derived by serial passage of *M. bovis* in the laboratory over a 13 year period (Garnier *et al* 2003). An interesting comparative genome microarray analysis between *M. tuberculosis* and 13 different *M. bovis* BCG strains revealed that 16 regions (named RD1 to RD16) were deleted from different BCG genomes relative to the H37Rv genome (Behr *et al* 1999). Nine of the RDs (containing 61 ORFs) were absent from all the BCG strains as well as all virulent *M. bovis* strains tested. One key deletion, RD1, was missing from all the BCG strains but present in all other *M. tuberculosis* complex members including *M. bovis*. Such comparative studies between *M. tuberculosis* and *M. bovis* BCG have provided insight into the potential genetic basis of attenuation of BCG and also bestowed opportunities for experimentally examining the role of discrete genomic sequences in virulence and pathogenesis. Thus the RD1 deletion appears to have played a key role in the attenuation of BCG (Lewis *et al* 2003; Pym *et al* 2002). We expect to gain deeper insights into the mechanism of attenuation when the genome of *M. bovis* BCG Pasteur strain is analysed.

In a similar vein, sequencing of the *M. tuberculosis* H37Ra genome and its comparative genomics is also likely to expedite the identification of TB virulence factors. TB researchers have extensively worked with the highly attenuated H37Ra and virulent H37Rv strains of *M. tuberculosis* that were derived by serial passaging of the H37 clinical isolate through laboratory media (Brosch *et al* 1999). However, the genetic basis for the avirulent phenotype of *M. tuberculosis* H37Ra in the animal models is not known. Since the H37Ra strain does not revert to a virulent phenotype on passaging through animals, it is likely that the former has acquired multiple point mutations or small deletions/rearrangements in genes crucial for virulence. Despite extensive use of these strains only a few differences between H37Rv and H37Ra at the genetic and expression levels have been reported over the last several years. These include IS*6110* mediated polymorphisms (Brosch *et al* 1999; Bifani *et al* 2000). Among the five regions (RvD1-RvD5) deleted in the H37Rv genome relative to *M. bovis*, four (except for RvD2) were also absent in H37Ra (Brosch *et al* 1999). However, since RvD2 is present in H37Ra as well as in clinical isolates, this region is unlikely to be instrumental in the attenuation of H37Ra (Brosch *et al* 1999). Although (in contrast to BCG) the RD1 region is intact in H37Ra, marked down regulation was noted for genes coding for RD1 antigenic proteins CFP10 and ESAT6 as well as neighboring genes that likely encode a secretion apparatus devoted to their export (Mostowy *et al* 2004). These results provide a partial explanation for H37Ra's attenuation and reiterate the correlation between RD1 and attenuation. In our laboratory, we used subtractive RNA hybridization between H37Rv and H37Ra strains cultured *in vitro* to isolate genes that were differentially expressed in virulent strain [*dev* genes] (Kinger and Tyagi 1993). Several genes were identified which included the two-component system *devR-devS* [*Rv3133c*/*Rv3132c*; *dosR*/*dosS*] (Dasgupta *et al* 2000) and *acr* (Tyagi, unpublished data). *devR* was shown to modulate the virulence of *M. tuberculosis* (Malhotra *et al* 2004; Parish *et al* 2003) and to be a regulator of the hypoxia/NO dormancy response (Park *et al* 2003; Voskuil *et al* 2003). The underlying basis for the differential gene expression of *devR-devS* genes was not readily explained as the coding and promoter sequences were identical in H37Ra and H37Rv (Malhotra and Tyagi, unpublished data). This suggests the likelihood of a regulatory element(s) present elsewhere in the genome that may be altered in the avirulent H37Ra strain. The aggregation of mycobacteria into structures known as 'cords' has long been associated with virulence (Middlebrook *et al* 1947). Comparative gene expression analysis recently identified 22 genes that were consistently expressed at higher levels in H37Rv (a cording strain) than in H37Ra (a non-cording strain) under a variety of growth conditions (Gao *et al* 2004). More than half of these genes were either involved in lipid metabolism or were putative cell membrane proteins. The difference in cell wall methyl-branched lipids indeed formed the basis for neutral-red positive staining for H37Rv in contrast to negative staining for H37Ra (Cardona *et al* 2006). Rindi *et al* (1999) identified six genes down regulated in the H37Ra strain compared to the H37Rv strain, of which three code for proteins that can be predictively associated to immunological or pathogenetic aspects of *M. tuberculosis* infection. The absence of five ESAT-6 like proteins from the culture filtrates of H37Ra and their presence in H37Rv has also been related to the latter's virulence (He *et al* 2003). Another crucial virulence determining factor is the ability of *M. tuberculosis* to infect and grow in human macrophages. Infected macrophages respond by undergoing apoptosis, a cornerstone of effective antimycobacterial host defense. However, virulent mycobacteria override this reaction by inducing necrosis leading to dissemination of the pathogen and development of disease (Chen *et al* 2006).

Although both H37Rv and H37Ra were able to infect macrophages equally, only H37Rv and not H37Ra was able to grow intracellularly as well as cause significant mitochondrial damage leading to necrosis (Silver *et al* 1998; Chen *et al* 2006).

Elucidating the mechanism of TB pathogenesis is a central question for mycobacteriologists. Disease is an outcome of a delicate and dynamic balance between pathogen and host immune response. *M. tuberculosis* strains generated both by random transposon mutation and targeted gene deletion approaches have taught us that virulence is indeed a very complex phenotype. These mutant strains derived by deliberate engineering in the laboratory are each defective at a single genetic locus. One of the surprises of H37Rv genome sequencing was the large extent of functional redundancies present (Cole *et al* 1998). Hence, it is possible that a gene disruption may not display a phenotype due to functional compensation by a redundant 'backup' gene. Sequencing H37Ra and its comparison with that of H37Rv is likely to reveal the functional redundancies that possibly come into play *in vivo* (in H37Rv) in addition to presenting a kaleidoscopic view of the genetic spectrum necessary for the display of the complex phenotype of virulence.

In view of the (i) stark differences in the virulence phenotypes of H37Rv and H37Ra strains, (ii) failure to detect gross differences in their nucleotide sequences, and (iii) inability to restore virulence in H37Ra complemented with H37Rv sequences encompassing polymorphic regions (Brosch *et al* 1999), the H37Ra genome is a worthwhile candidate for whole genome sequencing. The genome sequence will help to answer pressing questions related to the virulence of *M. tuberculosis* by enabling bioinformatics-guided experimentation, and also provide valuable inputs to TB vaccine and drug development endeavours.

Acknowledgements

Deepak Sharma is thankful to Dr Debasisa Mohanty for his support, CSIR for a Senior Research Fellowship and DBT for a Postdoctoral Fellowship.

References

Behr M A, Wilson M A, Gill W P, Salamon H, Schoolnik G K, Rane S and Small P M 1999 Comparative genomics of BCG vaccines by whole-genome DNA microarray; *Science* **284** 1520–1523

Bifani P, Moghazeh S, Shopsin B, Driscoll J, Ravikovitch A and Kreiswirth B N 2000 Molecular characterization of *Mycobacterium tuberculosis* H37Rv/Ra variants: distinguishing the mycobacterial laboratory strain; *J. Clin. Microbiol.* **38** 3200–3204

Bishai W R, Dannenberg A M Jr, Parrish N, Ruiz R, Chen P, Zook B C, Johnson W, Boles J W and Pitt M L 1999 Virulence of *Mycobacterium tuberculosis* CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method; *Infect. Immun.* **67** 4931–4934

Brosch R, Philipp W J, Stavropoulos E, Colston M J, Cole S T and Gordon S V 1999 Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain; *Infect. Immun.* **67** 5768–5774

Cardona P J, Soto C Y, Martin C, Giquel B, Agusti G, Andreu N, Guirado E, Sirakova T, Kolattukudy P, Julian E and Luquin M 2006 Neutral-red reaction is related to virulence and cell wall methyl-branched lipids in *Mycobacterium tuberculosis*; *Microbes Infect.* **8** 183–190

Chen M, Gan H and Remold H G 2006 A mechanism of virulence: virulent *Mycobacterium tuberculosis* strain H37Rv, but not attenuated H37Ra, causes significant mitochondrial inner membrane disruption in macrophages leading to necrosis; *J. Immunol.* **176** 3707–3716

Cole S T, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V, Eiglmeier K *et al* 1998 Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence; *Nature (London)* **393** 537–544

Cole S T, Eiglmeier K, Parkhill J, James K D, Thomson N R, Wheeler P R, Honore N, Garnier T *et al* 2001 Massive gene decay in the leprosy bacillus; *Nature (London)* **409** 1007–1011

Corbett E L, Watt C J, Walker N, Maher D, Williams B G, Raviglione M C and Dye C 2003 The growing burden of tuberculosis: global trends and interactions with the HIV epidemic; *Arch. Int. Med.* **163** 1009–1021

Dasgupta N, Kapur V, Singh K K, Das T K, Sachdeva S, Jyothisri K and Tyagi J S 2000 Characterization of a two-component system, *devR-devS*, of *Mycobacterium tuberculosis*; *Tuber. Lung Dis.* **80** 141–159

Fleischmann R D, Alland D, Eisen J A, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, *et al* 2002 Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains; *J. Bacteriol.* **184** 5479–5490

Gao Q, Kripke K, Arinc Z, Voskuil M and Small P 2004 Comparative expression studies of a complex phenotype: cord formation in *Mycobacterium tuberculosis*; *Tuberculosis (Edinb)* **84** 188–196

Garnier T, Eiglmeier K, Camus J C, Medina N, Mansoor H, Pryor M, Duthoy S Grondin S *et al* 2003 The complete genome sequence of *Mycobacterium bovis*; *Proc. Natl. Acad. Sci. USA* **100** 7877–7882

He X Y, Zhuang Y H, Zhang X G and Li G L 2003 Comparative proteome analysis of culture supernatant proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra; *Microbes Infect.* **5** 851–856

Kinger A K and Tyagi J S 1993 Identification and cloning of genes differentially expressed in the virulent strain of *Mycobacterium tuberculosis*; *Gene* **131** 113–117

Lewis K N, Liao R, Guinn K M, Hickey M J, Smith S, Behr M A and Sherman D R 2003 Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation; *J. Infect. Dis.* **187** 117–123

Malhotra V, Sharma D, Ramanathan V D, Shakila H, Saini D K, Chakravorty S, Das T K, Li Q, Silver R F, Narayanan P R and Tyagi J S 2004 Disruption of response regulator gene, *devR*, leads to attenuation in virulence of *Mycobacterium tuberculosis*; *FEMS Microbiol. Lett.* **231** 237–245

Middlebrook G, Dubos R J and Pierce C 1947 Virulence and morphological characteristics of mammalian tubercle bacilli; *J. Exp. Med.* **86** 175–184

Mostowy S, Cleto C, Sherman D R and Behr M A 2004 The *Mycobacterium tuberculosis* complex transcriptome of attenuation; *Tuberculosis (Edinb)* **84** 197–204

Parish T, Smith D A, Kendall S, Casali N, Bancroft G J and Stoker N G 2003 Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*; *Infect. Immun.* **71** 1134–1140

Park H D, Guinn K M, Harrell M I, Liao R, Voskuil M I, Tompa M, Schoolnik G K and Sherman D R 2003 Rv313-3c/*dosR* is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*; *Mol. Microbiol.* **48** 833–843

Pym A S, Brodin P, Brosch R, Huerre M and Cole S T 2002 Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*; *Mol. Microbiol.* **46** 709–717

Ramakrishnan T and Chandrasekhar P 1999 The control of tuberculosis: a continuous game of snakes and ladders; *J. Biosci.* **24** 143–152

Rindi L, Lari N and Garzelli C 1999 Search for genes potentially involved in *Mycobacterium tuberculosis* virulence by mRNA differential display; *Biochem. Biophys. Res. Commun.* **258** 94–101

Silver R F, Li Q and Ellner J J 1998 Expression of virulence of *Mycobacterium tuberculosis* within human monocytes: virulence correlates with intracellular growth and induction of tumor necrosis factor alpha but not with evasion of lymphocyte-dependent monocyte effector functions; *Infect. Immun.* **66** 1190–1199

Tyagi J S and Sharma D 2004 Signal transduction systems of mycobacteria with special reference to *M. tuberculosis*; *Curr. Sci.* **86** 93–102

Voskuil M I, Schnappinger D, Visconti K C, Harrell M I, Dolganov G M, Sherman D R and Schoolnik G K 2003 Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program; *J. Exp. Med.* **198** 705–713

Wayne L G and Sohaskey C D 2001 Nonreplicating persistence of *Mycobacterium tuberculosis*; *Annu. Rev. Microbiol.* **55** 139–163

Deepak Sharma[#] and Jaya Sivaswami Tyagi*
Department of Biotechnology,
All India Institute of Medical Sciences,
Ansari Nagar, New Delhi 110 029, India
[#]*Present address: National Institute of Immunology,*
Aruna Asaf Ali Marg,
New Delhi 110 067, India
**(Email, jstyagi@gmail.com)*

*e*Publication: 9 February 2007