

The *Mycobacterium tuberculosis* genomic sequence: anatomy of a master adaptor

William Bishai

The elucidation of the complete genomic sequence of *Mycobacterium tuberculosis* H37Rv, reported by Stewart Cole, Bart Barrell and colleagues in *Nature*¹, is a technological triumph that elevates *M. tuberculosis* into the elite cadre of sequenced microorganisms. Poorly funded and under-researched for many years, the inexorable increase in tuberculosis (TB) incidence has vaulted this disease into the public spotlight² and has led to efforts such as this Wellcome-Trust-funded sequencing project. The sequence is undoubtedly a major milestone in the colorful history of TB research and will be recorded on a par with the discoveries of tuberculin, bacille Calmette-Guérin (BCG), streptomycin and multidrug therapy.

Weighing in at ~4.4 Mb, the *M. tuberculosis* genome, which is the fourteenth complete bacterial sequence reported since that of *Haemophilus influenzae* in 1995 (Ref. 3), is the second largest after that of *Escherichia coli*⁴ and one of the most technically demanding. Decoding DNA with a 65% GC content requires multiple sequencing chemistries and high coverage rates for reliable sequence ascertainment. From a sequencing standpoint, the unambiguous identification of the genome's 4 411 529 bp and annotation of its 3 924 open reading frames by the Sanger Centre (Hinxton, UK) team is a laudable accomplishment.

The sequence reveals the workings of a master adaptor and shows *M. tuberculosis* to be an unusually complex pathogen. Pathogens capable of causing disease in healthy, immunocompetent individuals are typically 'lean and

mean' with respect to genome size. They have evolved as highly adapted microorganisms whose niche is usually restricted to, and dependent on, the mammalian host⁵. Hence, pathogens, such as *Neisseria meningitidis*, *Streptococcus pneumoniae* or *Treponema pallidum*, typically have a limited adaptive repertoire. Colonization or full-blown disease are often the only two adaptive choices available to them. Living in such restricted environments, these pathogens operate with compact genomes that are usually no greater than 2.5 Mb (Table 1).

Although *M. tuberculosis* fits this definition of an obligate pathogen, its generous genome size is more akin to those of opportunistic pathogens, such as *Pseudomonas aeruginosa* (5.9 Mb) or *Legionella pneumophila* (4.1 Mb). These pathogens, which rely on extremes of age, underlying disease or immune dysfunction in their victims, survive readily in environmental reservoirs outside their mammalian hosts; infection and disease are minor facets of their rich adaptive repertoire. So, why does *M. tuberculosis*, which is not known to exist in environmental reservoirs, maintain such an ample genome? One answer could be that its lifestyle is more complex than expected and requires greater genetic capacity.

Thus, the perception of *M. tuberculosis* as a plodding, monotonous microorganism capable only of inexorable aerobic growth and self-protecting ensconcement within a waxy coat requires some revision. The sequence reported by Cole *et al.*¹ reveals a diverse genetic repertoire abounding in regulatory circuits; there are 13 RNA polymerase sigma factors, 30 two-component regulators, 14 protein kinases or phosphatases and >140 transcriptional regulators. Such regulatory complexity in the genome is consistent with sophisticated machinery for an intricate infection process.

M. tuberculosis invests much of its genomic inheritance in a complex lipid metabolic apparatus employing >250 enzymes. In addition to having elaborate lipid biosynthetic and degradative enzymes, a series of gene clusters for polyketide synthesis are apparent in the sequence. These large operons, best characterized in *Streptomyces*, encode multienzyme pathways required to generate lipids, secondary metabolites and antibiotics⁶. The discovery of four distinct polyketide synthase operons in *M. tuberculosis* raises the novel possibility that aspects of its pathogenicity might be associated with secondary metabolite secretion within the host.

Although *M. tuberculosis* is the fourteenth complete genome to be sequenced, it might gain the distinction of being the first microorganism to have two isolates completely sequenced and published for comparative purposes. An ongoing NIH-supported genome sequencing project at the Institute for Genomic Research

W. Bishai is in the Dept of International Health, Johns Hopkins School of Public Health, 615 N. Wolfe St, Baltimore, MD 21205, USA.
tel: +1 410 955 3507,
fax: +1 410 614 8173,
e-mail: wbishai@jhsph.edu

(TIGR) in Rockville, MD, USA is completing the genome sequence of a highly virulent *M. tuberculosis* strain known as CSU93 or Oshkosh. This strain has a hypervirulence phenotype in mice following aerosol infection and caused a major human TB outbreak in 1994–1996 in the USA (Ref. 7). In comparison, the recently sequenced H37Rv, which was first isolated at the turn of the century and has been passaged in the laboratory for over nine decades, is of modest virulence in animal models of TB (Refs 8,9). Hence, comparison of the two genomes holds the prospect of revealing virulence-associated differences between H37Rv and CSU93.

The genomic sequencing effort for *Mycobacterium leprae*, which is supported by the Wellcome Trust and the Heiser Foundation, is >90% complete. Preliminary observations on this uniquely human-adapted pathogen, which cannot be cultivated *in vitro*, reveal multiple apparently missing enzymes and mutated genes in an abbreviated mycobacterial genome of just 2.8 Mb. Analogous to *T. pallidum*, *M. leprae* might represent an extreme among human-adapted pathogens that has gradually shed all but a small core of genes needed for host survival^{10,11}.

Finally, another NIH-supported sequencing project to complete the genome sequence of *Mycobacterium avium* is under way at TIGR. *M. avium* complex (MAC) microorganisms are widely dis-

tributed in the environment but have a predilection for causing disseminated disease in late-stage AIDS patients. Why this particular pathogen, rather than others from the vast collection of mycobacteria, prospers so well in the specific milieu of HIV co-infection is poorly understood¹². Again, genomic comparisons among the *M. avium*, *M. tuberculosis* and *M. leprae* sequences will be informative in addressing this and a host of other questions.

With the completed database provided by the Sanger Centre, TB research now enters the post-genomic era. The sequence shows *M. tuberculosis* to be a master of adaptation, in spite of the fact that it cannot survive for long periods outside the human host. The large genome size and extensive regulatory apparatus reflect the capacity of this microorganism to cause a variety of disease syndromes, ranging from the overt cavitory destruction to covert latent infection. Although the genome sequence does not deliver an instant answer concerning the pathogenesis of TB, a full set of puzzle pieces has now been scattered on the table, and the task of interlocking them into a coherent image has been substantially advanced.

References

- 1 Cole, S.T. *et al.* (1998) *Nature* 393, 537–544
- 2 Raviglione, M.C., Snider, D.E., Jr and Kochi, A. (1995) *J. Am. Med. Assoc.* 273, 220–226
- 3 Fleischmann, R.D. *et al.* (1995) *Science* 269, 496–512
- 4 Blattner, F.R. *et al.* (1997) *Science* 277, 1453–1474
- 5 Finlay, B.B. and Falkow, S. (1997) *Microbiol. Mol. Biol. Rev.* 61, 136–169
- 6 Hopwood, D.A. and Khosla, C. (1992) *Ciba Found. Symp.* 171, 88–106
- 7 Valway, S.E. *et al.* (1998) *New Engl. J. Med.* 338, 633–644
- 8 North, R.J. and Izzo, A.A. (1993) *J. Exp. Med.* 177, 1723–1734
- 9 Steenken, W., Jr, Oatway, W.H., Jr and Petroff, S.A. (1934) *J. Exp. Med.* 60, 515–540
- 10 Fraser, C.M. *et al.* (1998) *Science* 281, 375–388
- 11 Young, D. (1998) *Trends Microbiol.* 6, 217–218
- 12 Shiratsuchi, H., Johnson, J.L. and Ellner, J.J. (1996) in *Mycobacterium avium Complex Infection* (Korvick, J.A. and Benson, C.A., eds), pp. 163–195, Marcel Dekker

Table 1. Genome sizes of sequenced pathogens

Pathogens	Genome size (Mb)
Obligate pathogens	
<i>Mycobacterium tuberculosis</i>	4.40
<i>Neisseria meningitidis</i>	2.30
<i>Streptococcus pneumoniae</i>	2.20
<i>Vibrio cholerae</i>	2.50
<i>Treponema pallidum</i>	1.14
<i>Helicobacter pylori</i>	1.66
Opportunistic pathogens	
<i>Pseudomonas aeruginosa</i>	5.90
<i>Mycobacterium avium</i>	4.70
<i>Legionella pneumophila</i>	4.10
<i>Enterococcus faecalis</i>	3.00
<i>Listeria monocytogenes</i>	3.00

Coming to *Trends in Microbiology* in 1999

- Pathogenic bacteria in protozoa and biofilms, by M.R.W. Brown and J. Barker
- Drug targets in trypanosomes, by M.P. Barrett, J.C. Mottram and G.H. Coombs
- *Escherichia coli* mutator genes, by J-P. Horst, T. Wu and M.G. Marinus
- HIV-1 attachment, by S. Ugolini, I. Mondor and Q.J. Sattentau
- Switching in bacterial chemotaxis, by R.E. Silversmith and R.B. Bourret

Don't miss these and many more articles of interest; subscribe to *Trends in Microbiology* using the form bound in this issue.

