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, 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 spotlight and has led to efforts such as this Wellcome-Trost-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 coli6 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 and requires greater genetic content. Thus, the perception of M. tuberculosis as a plodding, monotonous microorganism capable only of inexorable aerobic growth and self-protecting ensonnement within a waxy coat requires some revision. The sequence reported by Cole et al. 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 synthases are apparent in the sequence. These large operons, best characterized in Streptomyces, encode multienzyme pathways required to generate lipids, secondary metabolites and antibiotics. 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...
(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 hyper-virulence 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 under way at TRENDS IN MICROBIOLOGY 465 VOL. 6 NO. 12 DECEMBER 1998

Table 1. Genome sizes of sequenced pathogens

<table>
<thead>
<tr>
<th>Pathogens</th>
<th>Genome size (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycobacterium tuberculosis</td>
<td>4.40</td>
</tr>
<tr>
<td>Nesseria meningitidis</td>
<td>2.20</td>
</tr>
<tr>
<td>Strepatooccus pneumoniae</td>
<td>2.20</td>
</tr>
<tr>
<td>Vibrio cholerae</td>
<td>2.50</td>
</tr>
<tr>
<td>Treponema pallidum</td>
<td>1.14</td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>1.66</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa</td>
<td>5.90</td>
</tr>
<tr>
<td>Mycobacterium avium</td>
<td>4.70</td>
</tr>
<tr>
<td>Legionella pneumophila</td>
<td>4.10</td>
</tr>
<tr>
<td>Enterococcus faecalis</td>
<td>3.00</td>
</tr>
<tr>
<td>Listeria monocytogenes</td>
<td>3.00</td>
</tr>
</tbody>
</table>

9 Smolen, W., Jr, Gurney, W.H., Jr and Prestoff, S.A. (1934) J. Exp. Med. 60, 533–540
10 Fraser, C.M. et al. (1998) Science 281, 1764–1770

Finally, another NIH-supported sequencing project to complete the genome sequence of Mycobacterium avium complex (MAC) microorganisms is widely distributed 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, prosers 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 cavitary destruction to covert latent infection. Although the genome sequence does not deliver an instant answer concerning the pathogenicity of TB, a full set of puzzle pieces 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 survival13,14.

References
9 Smolen, W., Jr, Gurney, W.H., Jr and Prestoff, S.A. (1934) J. Exp. Med. 60, 533–540
10 Fraser, C.M. et al. (1998) Science 281, 1764–1770

Finally, another NIH-supported sequencing project to complete the genome sequence of Mycobacterium avium complex (MAC) microorganisms is widely distributed 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, prosers 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 cavitary destruction to covert latent infection. Although the genome sequence does not deliver an instant answer concerning the pathogenicity of TB, a full set of puzzle pieces 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 survival13,14.

References
9 Smolen, W., Jr, Gurney, W.H., Jr and Prestoff, S.A. (1934) J. Exp. Med. 60, 533–540
10 Fraser, C.M. et al. (1998) Science 281, 1764–1770

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.