IS2-4

Evolutionary Genomics of Mycobacteria

Stewart T. Cole

Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Dr. Roux, 75724 Paris Cedex 15, France

There are over eighty known species of mycobacteria and many more awaiting discovery. They can be divided into two broad groups on the basis of their growth rate; the fast-growers have a doubling time of 3 - 5 hours whereas the slow-growers require ~24 hours. This division is also reflected in the number of rRNA operons present in the genome since the fast-growing mycobacteria generally contain two copies while the slow-growers only possess one. Mycobacteria are commonly found in soil and water where they adopt a saprophytic lifestyle. A small number of slow-growing species have become obligate human or veterinary pathogens although mycobacteria do not figure among the known components of the microbial flora of mammals. The reasons for this change in ecological niche and metabolic life style are not clear but comparative genomics of mycobacteria is helping us to understand some of the mechanisms involved.

There are three major human diseases that arise from infection with mycobacteria: tuberculosis, leprosy and Buruli ulcer. Tuberculosis is the leading cause of death from infectious disease in humans with roughly two million lives lost annually. Its etiologic agent, *Mycobacterium tuberculosis*, causes eight million new cases of disease every year and two billion people are believed to be latently infected with the tubercle bacillus (Dye *et al.*, 1999). These individuals are at risk of succumbing to tuberculosis later in life when their immune system wanes as a result of aging, infection with HIV or immuno-suppression. Tuberculosis is spread mainly by droplet infection and the lungs are the primary sites of disease with heavy bacterial loads found in macrophages and monocytes.

Leprosy, with its characteristic skin lesions, is not a fatal disease but 700,000 cases occur every year (Engers and Morel, 2003). It results from infection with *Mycobacterium leprae*, a pathogen that cannot be cultured in the laboratory, which displays an exceptionally long generation time of 14 days. The leprosy bacillus also resides within macrophages and monocytes but displays a predilection for the Schwann cells of the peripheral nervous system. Destruction of Schwann cells results in nerve damage that is often irreversible and sensory loss, which can lead to mutilations and disfigurement. The exact means of infection is unclear but may involve the respiratory route or penetration through the skin.

Unlike tuberculosis and leprosy, the source of infection of Buruli ulcer is environmental as its agent, *Mycobacterium ulcerans*, is found in biofilms on the surface of aquatic plants and in the bodies of insects and snails that browse thereon. Transmission to humans seems to occur following bites from water beetles such as *Naucoris cimicoides*, whose salivary glands are colonized by the bacilli (Marsollier *et al.*, 2002). *M. ulcerans* causes raised ulcers that are painless and devoid of inflammatory responses but undergo necrosis

leading to massive destruction of the skin. Pathogenesis is mediated exclusively by a macrolide toxin produced by the bacillus, called mycolactone. There are no preventive or curative treatments for Buruli ulcer; surgery and skin grafts are the sole medical interventions possible at present (Johnson *et al.*, 1999). *M. ulcerans* is closely related to *Mycobacterium marinum*, another water-borne pathogen that causes granulomatous disease in fish and amphibia (Stinear *et al.*, 2000).

The complete genome sequences are available for *M. tuberculosis, M. leprae* and *M. ulcerans,* and their comparison has provided a wealth of information about the genetics, biochemistry, physiology, and evolution of mycobacteria. Functions could be predicted for over 60% of the genes with the remainder encoding hypothetical, or conserved hypothetical, proteins many of which were restricted to mycobacteria. Prominent features of the genomes are an abundance of genes involved in lipid metabolism and several gene families that encode novel proteins of repetitive sequence that may correspond to variable surface antigens.

There are big differences in genome size and potential coding density (Table 1). *M. leprae* has the smallest genome (3.26 Mb) and has undergone reductive evolution as over 50% of its genes have decayed and lost their functions (Cole *et al.*, 2001). This explains in part its slow growth, our inability to culture the bacillus and its obligate human parasitism. The genome of *M. tuberculosis* (4.41 Mb) comprises a single, circular chromosome of which >91% codes for proteins (Cole *et al.*, 1998). In contrast, the genome of *M. ulcerans* is considerably larger and consists of a chromosome (5.63 Mb) and a large plasmid (0.17 Mb). *M. ulcerans* appears to be a descendant of *M. marinum* (6.78 Mb) that has embarked upon a course of reductive evolution and acquired by horizontal transfer a virulence plasmid that encodes giant polyketide synthases required for mycolactone production (Stinear *et al.*, 2004).

Table 1. General features of the genomes of the main mycobacterial pathogens.

Feature	*M. tuberculosis*	*M. tuberculosis*	*M. ulcerans* *
Chromosome size (bp)	3,268,203	4,411,532	5,631,607
Plasmid (bp)	-	-	174,155
GC content (%)	57.79	65.61	65.47
No. of protein genes	1605	3,993	~5,700
Gene density (bp/gene)	2037	1,106	ND
No. of pseudogenes	1116	6	ND
No. of stable RNA genes	50	50	50

* annotation still in progress

References

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., III, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, A., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.-A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K.,

Whitehead, S., and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Duthoy, S., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., McLean, A., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R., and Barrell, B.G. (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007-1011.

Dye, C., Sheele, S., Dolin, P., Pathania, V., and Raviglione, M.C. (1999) Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. *JAMA* **282**: 677-686.

Engers, H., and Morel, C.M. (2003) Leprosy. *Nat Rev Microbiol.* **1**: 94-95.

Johnson, P.D., Stinear, T.P., and Hayman, J.A. (1999) Mycobacterium ulcerans--a mini-review. *Med Microbiol.* **48**: 511-513.

Marsollier, L., Robert, R., Aubry, J., Saint Andre, J.P., Kouakou, H., Legras, P., Manceau, A.L., Mahaza, C., and Carbonnelle, B. (2002) Aquatic insects as a vector for *Mycobacterium ulcerans*. *Appl Environ Microbiol* **68**: 4623-4628.

Stinear, T.P., Jenkin, G.A., Johnson, P.D.R., and Davies, J.K. (2000) Comparative genetic analysis of *Mycobacterium ulcerans* and *Mycobacterium marinum* reveals evidence of recent divergence. *J. Bacteriol.* **182**: 6322-6330.

Stinear, T.P., Mve-Obiang, A., Small, P.L.C., Frigui, W., Pryor, M.J., Brosch, R., Jenkin, G.A., Johnson, P.D.R., Davies, J.K., Lee, R.E., Adusumilli, S., Garnier, T., Haydock, S.F., Leadlay, P.F., and Cole, S.T. (2004) Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *PNAS* **101**: 1345-1349.