 10.1111/j.1469-0691.2006.01637.x

Molecular evolution of *Mycobacterium tuberculosis*

C. Arnold

Applied and Functional Genomics, Centre for Infections, Health Protection Agency, London, UK

ABSTRACT

Tuberculosis continues to be the main cause of death from a single infectious agent in developing countries. The causative agent, *Mycobacterium tuberculosis*, is thought to have diverged from its common ancestor as recently as 15 000 years ago. Subsequently, various genetic elements have evolved over time at different rates and can be used to elucidate patterns of infection. When individual elements are studied within genetic families, very low rates of variation are observed for almost every marker. For example, when all *M. tuberculosis* genetic families are considered, the number of alleles observed at each mycobacterial interspersed repetitive unit (MIRU) locus usually drops when viewed within a single genetic family, indicating that the rate of repeat variation may be low, as each member of that family is a descendant of a single common ancestor. Also, the low level of silent nucleotide variation observed indicates that *M. tuberculosis* is, in evolutionary terms, very young. Mapping the variation of the different markers used in molecular epidemiology within a genetic framework enables the relative rates of variation of these markers to be determined and, together with a complete chronology, allows the identification of more informative panels of markers tailored to individual genetic families.

Keywords Evolution, genetic variation, molecular clock, molecular epidemiology, *Mycobacterium tuberculosis*, review

Accepted: 4 September 2006

Clin Microbiol Infect 2007; **13**: 120–128

INTRODUCTION

Although a cure for tuberculosis (TB) was developed >50 years ago, it still remains one of the world's deadliest infectious diseases. The WHO reports that TB kills 5000 people a day, and between two and three million people annually, 98% of whom live in the developing world (http://www.who.int/mediacentre/factsheets/fs104/en/). Approximately one-third of the world's population is infected with TB, and hundreds of thousands of children will become TB orphans this year. One in three patients infected with human immunodeficiency virus (HIV) or with AIDS has TB, and drug resistance is emerging at an alarming rate in some geographical areas. Effective treatment is available for the majority of cases, but the disease often goes undiagnosed. The standard test for *Mycobacterium tuberculosis* infection is culture; however, the organism can take several weeks, or even months, to grow on solid culture. The advent of rapid liquid culture methods has reduced the time required to confirm a diagnosis to just 1–2 weeks.

Currently, the fastest method for confirming *M. tuberculosis* infection is a sputum smear test, stained for acid-fast bacilli, which can be carried out in just a few hours. However, a bacterial load of 10 000 cells/mL in a sputum sample is required for detection using this method, and >50% of 'smear-negative' patients are found subsequently to be culture-positive. Different strain-typing methods have been employed to link epidemiologically-related strains in order to follow and better understand the evolution and spread of this disease, and to perform the more difficult task of differentiating epidemiologically-unrelated strains of this organism.

The molecular clocks of genetic markers exploited to study the spread of *M. tuberculosis* differ, and can be used to investigate molecular evolution over shorter or longer periods. Study of the patterns of

Corresponding author and reprint requests: C. Arnold, Applied and Functional Genomics, Centre for Infections, Health Protection Agency, 61 Colindale Avenue, London NW9 5EQ, UK
E-mail: catherine.arnold@hpa.org.uk

variation and evolution of different genetic markers gives insight into their usefulness for different applications, and analysis of the way in which this genetic variation occurs provides further information concerning their efficacy for epidemiological purposes. This review considers some of the genetic variation strategies employed by *M. tuberculosis*, and the effect of these strategies on the relative molecular clocks of some of the commonly used markers within an evolutionary framework.

ORIGIN OF *M. TUBERCULOSIS*

Mycobacteria can be divided into two groups: fast-growing and slow-growing. Sequence data from various genes, including 16S rRNA, *rpo*B and *hsp*65 genes, have been used to construct phylogenetic relationships [1,2]. One of the sequenced strains of *M. tuberculosis*, H37Rv, has been shown to contain 20 cytochrome P450-containing mono-oxygenases that catalyse mixed oxidation of hydrophobic compounds; this is an activity that is associated with free-living saprophytes in soil, which perhaps indicates that the ancestor of *M. tuberculosis* was a soil mycobacterium [3].

M. tuberculosis is part of the *M. tuberculosis* complex (MTBC), a group of closely-related slow-growing mycobacteria that includes *M. tuberculosis*, *Mycobacterium bovis*, *Mycobacterium africanum*, *Mycobacterium microti* and *Mycobacterium canettii*. Until recently, it was assumed that cattle transmitted the disease to man, as the host range of *M. bovis* was much broader than that of *M. tuberculosis*. However, genetic information has revealed that the reverse is the case [4]. Gutierrez *et al*. [5] used sequence analysis of seven genes to show that *M. tuberculosis* appears to be a composite assembly of a relatively diverse group of smooth tubercle bacilli, including *M. canettii* strains. However, the generally low level of genetic variation seen in *M. tuberculosis* indicates that the total population resulted from clonal expansion following an evolutionary bottleneck, estimated to have occurred between 15 000 and 35 000 years ago [4–7].

GENERAL PRINCIPLES OF BACTERIAL EVOLUTION

There is significant evolutionary pressure towards smaller bacterial genomes, as smaller chromosomes can replicate faster, resulting in the bacteria being able to grow faster and out-populate bacteria with larger chromosomes. If a bacterial population moves into an environment where, for example, an essential amino-acid is abundant, some members of the bacterial population, after many generations, will lose the ability to synthesise that amino-acid. Eventually, every member of the population will also lose by deletion the DNA sequence encoding the pathway necessary for producing that amino-acid. However, this process occurs extremely slowly, as the evolutionary pressure to decrease genome size needs to also protect the organism from becoming extinct by degeneration of its genome [8].

Within the general premise that bacterial species change in response to their environment, it is clear that there are different molecular mechanisms for growth-dependent mutation and adaptive or stationary-phase mutation [9,10]. When growing cells are confronted with a change in their environment, they still have metabolic capabilities for a specific compensatory response (e.g., increased levels of transcription and mutation). However, after 4–5 days with no exogenous source of energy, cells resort to non-specific increases in all mutation rates in a final effort to produce a mutant that will survive. Mutation rates will also be affected by other factors, e.g., oxidising agents, UV light, the activity of DNA repair enzymes, and variables that could be influenced by starvation conditions (e.g., nucleotide pool levels). In a study of 26 structural genes of 842 *M. tuberculosis* isolates, >95% of nucleotide substitutions caused amino-acid substitutions in genes linked to antibiotic resistance [7]. A particular mutation arises independently in response to challenge with isoniazid, the front-line drug for tuberculosis treatment, more often than is expected by random mutation. Oxygen limitation induces dramatic and specific changes in mycobacteria, including enhanced resistance to isoniazid [11]. Thus, it appears that the rate at which mutations occur is elevated under conditions in which an excess of variation is most needed, so the question arises as to whether mechanisms have evolved to regulate the mutation rate. Most molecular processes leading to spontaneous and induced mutagenesis depend on the action of particular enzymes [12]. An accumulation of DNA lesions in cells under stress may lead to the saturation of DNA repair systems, which would lead to an increased mutation rate, as