Tuberculosis: evolution in millennia and minutes

S.H. Gillespie
Centre for Medical Microbiology, Hampstead Campus, Rowland Hill Street, London NW3 2PF, U.K.

Abstract
Tuberculosis remains a global public health threat: the causative organism, Mycobacterium tuberculosis, was once thought to show little genetic variation, but research in the last 10 years has demonstrated an ability to change in a series of different time frames. Related species of mycobacteria have undergone evolution by deletion of segments of DNA, allowing Mycobacterium bovis and other species to emerge from the M. tuberculosis complex, disproving the previously accepted theories. Deletions also affect the pathogenic potential of different lineages of M. tuberculosis. Over shorter time periods genetic variation is achieved by the movement of insertion sequences such as IS6110. Some lineages identified by this means are over-represented in patient populations, suggesting a genetic advantage, although the mechanism for this is not yet apparent. M. tuberculosis must also adapt to host and antibiotic selection pressure, and this is achieved by point mutations. Almost all antibiotic resistance emerges in this way, and data from clinical and in vitro studies indicate that M. tuberculosis exists with pre-existent mutants that remain as a small proportion of the population because of fitness deficits. Under certain physiological conditions, these rarer mutants may be favoured and, when antibiotic selection pressure is applied, will rise to dominate the bacterial population. M. tuberculosis is a highly effective pathogen that has caused disease in human populations for millennia. We are now starting to understand some of the genetic mechanisms behind this phenomenon.

Introduction
Tuberculosis has been part of the human experience for many thousands of years with the first recognizable description of the disease being found in early Chinese and Egyptian manuscripts. Hippocrates described it as phthisis or consumption, a name by which it was known until the early 20th Century [1]. Modern molecular methods have permitted the tuberculosis community to understand the way in which the genome of this pathogen has interacted with the human population since prehistory, allowing us to chart the evolution of the organism as it adapts to changes in the human population [2].

Until 1948 and the introduction of chemotherapy, tuberculosis was not curable and Hippocrates advised physicians not to visit such cases as their inevitable death might damage their reputation. International standard regimens will now produce a 95% cure. However, poorly prescribed regimens, substandard drugs and poor patient adherence permits resistance to emerge [3] and, in recent months, the scale of extensively resistant Mycobacterium tuberculosis has been recognized [4,5]. Resistance emerges as the bacterium adapts to the strong selective pressure applied by chemotherapy. This review describes the evolutionary processes acting on M. tuberculosis over these widely different time periods with the hope that the understanding gained will better equip us to create more effective treatment paradigms and control measures.

The evolution of the M. tuberculosis complex
M. tuberculosis and Mycobacterium leprae are of most interest to clinicians because of their disease-causing potential. Comparative analysis of their genomes indicate a core of 219 conserved genes, which are usually also found in the sequenced genomes of Mycobacterium marium, Mycobacterium avium and Mycobacterium smegmatis, suggesting a core of mycobacterial-specific genes. The M. tuberculosis complex of organisms includes members that are exclusively human pathogens (M. tuberculosis, Mycobacterium africanum and Mycobacterium canetti), a rodent pathogen, Mycobacterium microti, and another, Mycobacterium bovis, with a wide host range. There is a high degree of sequence conservation which argues for an evolutionary bottleneck some 10,000–20,000 years ago. Using analysis of sequences deleted from various species of the complex and the related species, it has been possible to show that the previous hypothesis that M. tuberculosis evolved from M. bovis...
Speculatively following the domestication of cattle [6] was incorrect. Rather, *M. bovis* has undergone many deletions of sequences that are present in *M. tuberculosis* [2], and the genome of *M. bovis* is smaller than that of *M. tuberculosis* [7]. It is now widely accepted that there was a previous species from which all current members of the complex evolved [8]. The modern tuberculosis strains share a common deletion, TD1, that other members of the complex have. The RD1 from which all current members of the complex evolved [8]. It is now widely accepted that there was a previous species that there are six phylogeographical lineages, each associated with a human sympatric population [13]. Thus there is a suggestion that the organism has adapted to specific human populations. One example of this is provided by the recent outbreak in Leicester by a strain that is part of the East African–Asian lineage and is defined by a deletion in Rv1519. The nature of the phenotype of this strain has been extensively studied. It grew less quickly, and was less acidic and H2O2-tolerant than reference strains, but its ability to grow inside macrophages was not impaired. Additionally the strain produced less protective IL-12p40 and more anti-inflammatory IL-6 and IL-10 gene transcription from macrophage-derived macrophages. Thus this deletion appears to produce an immune-subversive phenotype that contributes to persistence and outbreak potential in this lineage [14].

**Variation on shorter time scales**

The development in molecular epidemiological tools in the last 10 years has done much to unlock the secrets of the population genetics of tuberculosis described above by providing for the first time a reliable way to identify different strains of tuberculosis. The three techniques used for this purpose include typing using the insertion sequence IS6110, MIRUs (mycobacterial interspersed repetitive units) and spoligotyping [15–18]. Using these techniques, it has been possible to plot the transmission between hosts over shorter and longer time periods. This is illustrated by our study in a comprehensive survey of tuberculosis strains isolated in northern Tanzania. In this rural environment we had the opportunity of typing one-tenth of all the isolates made during 1 year. Using the IS6110 methodology, we were able to identify two lineages of isolates on the basis of 70% similarity [19]; these were named Kilimanjaro and Meru after the two largest mountains in the region. The centre of the lineage was made as the strain that was arbitrarily identified as the one most commonly isolated. By combining IS6110, MIRU and PGRS (polymorphic guanine-cytosine-rich repetitive sequence) typing methods, we were then able to track the putative mechanism by which IS elements were gained or lost, a process that is estimated to take between 3.5 years [20] and when the copy number of interspersed repeat units changed. None of the identical strains was epidemiologically related. Where a strain could be inferred, but for which we did not have an isolate, we put in a node. The Kilimanjaro lineage appeared to be the newest to have emerged, being less dispersed as well as the average putative time between the central and most diverse strain being approx. 20–30 years, while the Meru lineage was approx. 28–40 years. These data lead to the question of why are these two lineages over-represented in the population, constituting as they do 16.9% and 15.9% of all of the isolates studied. It may provide a way of identifying the pathogenic potential of these specific strains that give it such a significant biological advantage.

**Evolution in the face of antibiotic resistance**

From the perspective of *M. tuberculosis*, the appearance of antibiotics may have been a catastrophe, but it should be remembered that mycobacteria and related organisms are adapted to survival in the environment and as such have been
exposed to antibiotic agents produced by a wide range of antibiotic-producing organisms. Indeed, it was this insight that prompted Waksman and co-workers to search for agents that were active against *M. tuberculosis*, culminating in the discovery of streptomycin [21]. The importance of resistance was recognized in the first placebo-controlled trial of tuberculosis, where the initial improvement recognized in patients reversed within 5 years as a result of the emergence of streptomycin-resistant bacteria [22,23].

The molecular basis of rifampicin was discovered by Telenti and co-workers who identified mutations in a small segment of *rpoB* that encodes the β subunit of RNA polymerase in epidemiologically unrelated isolates that had developed resistance [24,25]. As *M. tuberculosis* does not possess plasmids and horizontal gene transfer is thought to be rare, all resistance emerges through mutations in chromosomal genes [3]. Mutations arise spontaneously at a low rate in all prokaryotes at a constant rate, but most do not survive [26]. It is only when a mutation provides an advantage that mutant numbers increase in the population, and this is the process at work in the evolution of drug resistance. The mechanism of resistance to most of the anti-tuberculosis agents has now been determined by sequencing of genes in drug-resistant strains [3]. For some agents, resistance is defined by mutations in a single gene, as in the case of rifampicin, pyrazinamide and ethambutol, [25,27,28] or in multiple genes, as in the case of isoniazid or streptomycin [29,30].

In clinical practice, the spectrum of rifampicin mutations so far identified have a characteristic pattern with approx. 70% of strains having one of three mutations [25]. We have measured the growth rate of each of the mutant strains and have shown that some grow at close to the same speed as the wild-type strains, whereas others have more significant fitness deficits [31]. These data were confirmed by other investigators [32]. We also showed that there was a significant correlation between the frequency of isolation in clinical practice and the degree of biological disadvantage. This is an important consideration as it provides us with the ability to predict the pathogenic and transmission-potential of different mutations among human populations. Bacteria that acquire an antibiotic-resistance determinant may pay an initial physiological price, but many studies have shown that adaptation takes place with further passages, a process that occurs by compensatory mutation or reversion [33]. We measured the fitness of strains of *M. tuberculosis* isolated from HIV-seropositive patients that were infected as part of a hospital outbreak. Each of the three strains had a significantly different fitness, suggesting that the strains had adapted to the host in which they were multiplying. It was notable that both of the patients infected with strains with a near normal fitness had progressed rapidly and had died shortly after infection [34].

The nature of the resistance mechanism and the particular mutation may have an effect on the fitness of the isolate. This has been shown to transfer through to the epidemiology of the disease, after measuring the degree of clustering associated with different resistant strains. Isoniazid resistance may arise through mutation in *katG* and *inhA* promoter. In populations from San Francisco and the Netherlands, it was shown that strains with a S315T mutation in *katG* or a mutation in the *inhA* promoter were more likely to be transmitted than strains resistant to isoniazid by other mechanisms [35,36]. The risk of particular mutations arising was related to the geographic lineage of the strain; for example, *katG* mutations other than S315T were more common in the East Asian lineage than in the Indo-Oceanic and Indo-European lineages [35]. Further work extending these observations indicates that both the position of rifampicin-resistant mutation and the genetic background of the strain affects the fitness measured [37].

**Risk of resistance**

All prokaryotes undergo spontaneous mutation at a low rate of 0.0033 per replication. Mutation rate per base pair is inversely proportional to genome size [26]. Previous studies have shown that the rate of mutation depends on the nature of the drug selection, but for most of the main anti-tuberculosis drugs, this occurs at a rate of $10^{-9}$ mutations per cell division [31,38]. This is an important reason for anti-tuberculosis drugs to be given in combinations as the risk of a mutant containing three resistance mutations is greater than $10^{-18}$. Thus it is only when through poor prescription or adherence that bacteria are exposed to monotherapy that resistance is likely to emerge [39].

Recently, fluoroquinolones have been suggested as a class of drugs that may result in shortened treatment time [40]. This is because *in vitro* animal and early-phase clinical studies indicate the efficacy and safety of these drugs [41–43]. Fluoroquinolones work by inhibiting the activity of DNA gyrase, producing lethal double-stranded DNA breaks [44]. Our group and others have shown that, when exposed to sub-inhibitory concentrations of fluoroquinolones, the mutation rate to resistance can be increased more than 100-fold [45]. These data have important implications for new quinolone-based tuberculosis treatments since it is possible that the widespread introduction of fluoroquinolone-based treatments could worsen the risk of drug resistance and thus need further investigation. The mechanism of hypermutability has always been thought to be due to the activation of the SOS response. When *M. tuberculosis* DNA is damaged by UV light, mitomycin C or hydrogen peroxide, DNA damage activates the *recA/lexA* system. Trans-lesional repair is completed by the DNA E2 polymerase, resulting in survival of the organism at the expense of an increase in the mutation rate. DNA E2 is crucial for *M. tuberculosis* as knockout strains are attenuated for mouse infection, confirming the importance of the ability to mutate to adapt to a new host noted above [34,46]. Using DNA arrays and reverse transcription–PCR, we have shown that, for doses above the MIC (minimum inhibitory concentration required to inhibit 90% of the pathogen) of ciprofloxacin, DNA E2 is as significantly up-regulated as it is in response to mitomycin C. In contrast, at sub-inhibitory concentrations, this gene is not up-regulated. Thus the polymerase responsible
for trans-leisional repair in response to sub-inhibitory concentrations of fluorquinolone still requires elucidation. It will be necessary for our group to identify the mechanism for hypermutability and, more importantly, to determine whether this process occurs in vivo through our clinical trial.

Summary

*Mycobacterium tuberculosis* has been interacting with the human population for many millennia. During that time the infection has waxed and waned, taking advantage, for example, of the industrialization that resulted in an enormous increase in cases worldwide. The appearance of HIV, which has produced a large population of immunocompromised subjects, has resulted in an explosive increase in cases in sub-Saharan Africa [47]. The genome of tuberculosis has adapted to different populations through deletions and mutations. Although limited in its response to change by the isolation imposed by its pathogenic process and lack of ability for horizontal gene exchange, *M. tuberculosis* has proved able to adapt rapidly to the evolutionary stress applied by antibiotics through mutation in chromosomal genes, and adaptation to the initial physiological deficit imposed by these mutations occurs rapidly. Modern molecular developments have opened up the prospect of understanding the complex relationship between *M. tuberculosis* and its human host. We know, however, that the adaptability of this pathogen is such that drug resistance is rising worldwide and that, if we are to reduce the burden of disease it imposes, we must develop new drugs and vaccines rapidly.

References