BioMed Central

Research article

Evidence for positive selection on Mycobacterium tuberculosis within patients

Mark M Tanaka*

Address: School of Biotechnology and Biomolecular Sciences, University of New South Wales, NSW 2052, Australia

Email: Mark M Tanaka* - m.tanaka@unsw.edu.au

* Corresponding author

Abstract

Background: While the pathogenesis and epidemiology of tuberculosis are well studied, relatively little is known about the evolution of the infectious agent *Mycobacterium tuberculosis*, especially at the within-host level. The insertion sequence IS*6110* is a genetic marker that is widely used to track the transmission of tuberculosis between individuals. This and other markers may also facilitate our understanding of the disease within patients.

Results: This article presents three lines of evidence supporting the action of positive selection on *M. tuberculosis* within patients. The arguments are based on a comparison between empirical findings from molecular epidemiology, and population genetic models of evolution. Under the hypothesis of neutrality of genotypes, 1) the mutation rate of the marker IS*6110* is unusually high, 2) the time it takes for substitutions to occur within patients is too short, and 3) the amount of polymorphism within patients is too low.

Conclusions: Empirical observations are explained by the action of positive selection during infection, or alternatively by very low effective population sizes. I discuss the possible roles of antibiotic treatment, the host immune system and extrapulmonary dissemination in creating opportunities for positive selection.

Background

How actively do populations of *Mycobacterium tuberculosis* cells undergo adaptive evolution on the spatial and temporal scales of individual infections? On the one hand, the long generation time and limited sequence diversity of this organism might suggest a slow pace of adaptive evolution. On the other hand, the rapidity and ease with which antibiotic resistance is generated during infection suggests otherwise. The physiology and immunology of tuberculosis pathogenesis have been well studied. The infectious agent *M. tuberculosis* is known to invade and replicate within alveolar macrophages. There is a spectrum of responses by the immune system, corresponding

to the relative involvement of *Th1* and *Th2* immune cells, which respectively stimulate the cytotoxic response (more effective against infected cells), and the humoral/antibody response (more effective against extracellular pathogens) [1,2]. Some progress has been made in describing the population dynamics of mycobacterial infection quantitatively [3-5]. At the wider spatial and temporal scales of populations, the molecular epidemiology and the evolution of *M. tuberculosis* have been carefully studied. Genotypic data are rapidly accumulating in the molecular epidemiology of infectious diseases. These are usually compiled and summarised to make inferences about the state of an epidemic in a given geographic

location, or at the global level. For example, epidemiologists seek to identify risk factors for infection, and to locate particular strains that are especially transmissible or pathogenic [6-8]. The evolutionary history of *M. tuberculosis* has also been characterised. For example, it has been argued that the limited variation at the nucleotide is due to a recent population bottleneck [9], and that the common ancestor of *Mycobacterium bovis* and *M. tuberculosis* may well have been a human rather than bovine pathogen [10].

Less understood is the evolution of *M. tuberculosis* at the cellular level *inside* bodies. There has been little integration of the genetic information from markers with the population genetics of the bacterial population within hosts. In this article, I examine data collected for the purposes of molecular epidemiology to present three lines of evidence supporting the action of positive selection on *M. tuberculosis*. The data come from the marker IS*6110*, which is currently the standard method of typing tuberculosis isolates. These genotypic data will be considered under assumptions of neutrality, and then under the assumption that positive selection is acting. The case for the action of selection is based on the following three arguments.

• Under the assumption of neutrality, the observed mutation (or transposition) rate of the genetic marker IS*6110* is unusually high; the estimated mutation rate is lower if selection is acting.

• The observed times associated with change are too low to be explained by neutrality; positive selection lowers the expected substitution time.

• The observed level of polymorphism is too low to be explained by neutrality.

Results: Models and observations
In each of the following sections a comparison is made between the strictly neutral model and a generalised model including selection through a single parameter *s* (described in the Appendix). Although the analyses start with strict neutrality (*s* = 0) in each argument, alleles for which *s* < 1/*N*, where *N* is the effective population size, can be considered *nearly neutral*, in that the effects of drift outweigh the force of selection [11]. In each case, explaining observations in this range of selective coefficients requires very low effective population sizes.

Transposition rates of IS6110
When genetic mutations are selectively neutral, the substitution rate is equal to the mutation rate [11]. In the present case, the *within-host* substitution process is of interest. Rosenberg *et al.* [12] determined the within-host substitution rate of the *IS6110* marker to be around

0.00184 to 0.0390 events per copy per year, with the maximum likelihood estimate at 0.0287. Under neutrality, therefore, this rate corresponds to a *per insertion* mutation rate of $\mu_i \sim 7.9 \times 10^{-5}$ events per site per generation, assuming a generation time of 1 day in active infections. This figure comes from a measured doubling time of close to 24 hours, based on clinical isolates grown in human monocyte cultures and in culture media [13-15]. Rates of point mutation (events *per nucleotide* per generation) are usually in the vicinity of 10^{-9}. In mutator strains, that is, genomes in which the DNA repair machinery is damaged, leading to elevated mutation rates, the mutation rate rises orders of magnitude, up to $\sim 10^{-7} - 10^{-6}$ [16]. The mutation rate of IS*6110* under neutrality therefore seems suspiciously high, although this is only "circumstantial evidence", since it is not inherently problematic. Indeed, mutation rates as high as 10^{-4} per element per generation have been measured for IS*10 in vitro* [17]. Nevertheless, if positive selection is allowed the estimated mutation rate decreases. Leaving aside the complicating influence of clonal interference [18], the rate of substitution is

$$K = uN\,\mu_i \quad (1)$$

where *u* is the probability of fixation of a mutant, μ_i is the mutation rate and *N* is the population size [11]. An estimate of the mutation rate when mutants have advantage *s* is $\hat{\mu}_i = K/(uN)$. The diffusion model of drift provides an expression for *u* as a function of the population size *N* and selective coefficient *s* (see the Appendix). Figure 1 plots $\hat{\mu}_i$ over *s* for a few different values of *N*. In each curve the estimated mutation rate decreases as the selective coefficient rises. According to this analysis, lower mutation rates are possible when there is some selection and a large population size, or when selection is strong and the population size is small. Note that the estimated mutation rate remains high if mutations are nearly neutral.

Fixation times
Various studies have measured the stability of IS*6110* as a genetic marker by examining genotypes of serial isolates from patients with persistent infection. A small number of changes in the genotypes between serial isolates indicates a stable marker. Differences in genotypes due to exogenous reinfection by unrelated strains are excluded from consideration. In the data of Niemann *et al.* [19] and Rosenberg *et al.* [12], the median time interval associated with changes in IS*6110* genotypes from serial samples of *M. tuberculosis* is 212 days, and the maximum is 683 days. Because the second sample is taken some time after fixation of the mutant, the actual substitution times are unknown, but they were clearly all under 683 days. I will now show that the expected substitution times under strict neutrality are well in excess of this value.
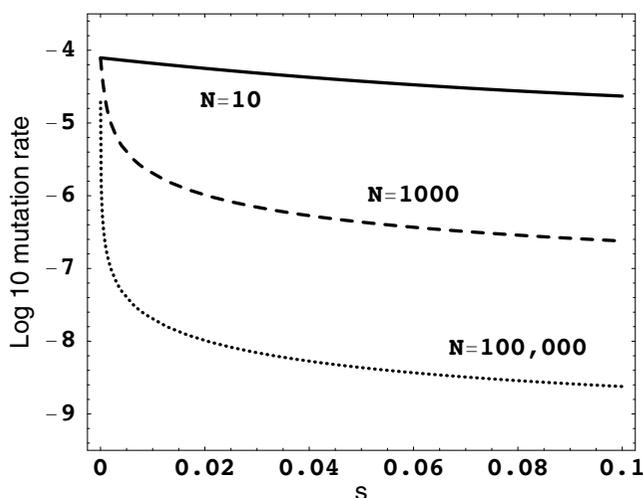
Figure 1
Estimate of mutation rate when positive selection is acting. The estimate $\hat{\mu}_i$ is plotted on a logarithmic scale in base 10. Solid curve: $N = 10$; Dashed; $N = 1000$; Dotted: $N = 10^5$.

Let us start with the assumption that the expected time for substitution to occur is the average time taken for the successful mutant to appear plus the time taken for that mutant to reach fixation conditional on its eventual fixation. (I will later drop the assumption about waiting for the mutant to appear). The average appearance time is $1/(\mu N u) = 1/\mu$ since $u = 1/N$ under strict neutrality. The average time for a successful neutral mutant to reach fixation is $4N$ generations. The mutation rate of interest in this context is the rate *per genome* per generation, since what is of concern is whether any of the elements in a given genome produce change. For simplicity, assume that the genomic mutation rate scales linearly with copy number. (At the resolution of this analysis, this is a reasonable approximation.) Considering a typical strain has 10 copies of the IS element, the relevant mutation rate here is $\mu = \mu_i \times 10 = 7.9 \times 10^{-4}$. Therefore, for $N = 10$, 10^3, 10^5, the expected substitution times are roughly 1300, 5300, 4×10^5 generations, respectively. With the generation time set to one day, the upper bound of observed substitution times was 683 generations, which is well below theoretical expectations.

Now consider the possibility of positive selection under two alternative conservative assumptions. The earlier assumption that there are no successful mutants at the time of the first sample is favourable to the parental strain. A more conservative approach (favouring mutants) would be to say that the mutant destined to reach fixation appears exactly at the time of the first sample. We can then

ask how long it takes on average for this mutant to reach fixation if it is positively selected. An even more conservative model would be that not only is the successor strain present at the time of the first sample, but is present at a frequency of 30%. Furthermore, let us say the subdominant strain only needs to be at 70% at the time of the second sample to be considered to have replaced the parental strain.

A model of the sojourn times of alleles in populations conditional on fixation must now be specified. Again using the diffusion model of drift (see Appendix), the mean time spent by a mutant in the range of frequencies (a, b) (provided a is greater than the initial frequency), conditional on fixation, was found by Ewens [20] and by Maruyama [21] to be

$$T(a,b) = \int_a^b \frac{(1 - e^{-4Nsx})(1 - e^{-4Ns(1-x)})}{sx(1-x)(1 - e^{-4Ns})} dx. \qquad (2)$$

Figure 2 shows the two conservative models, corresponding to two different boundary values for (a, b). Even in the extremely conservative model shown in the right-hand plot, the effective population size must be below 400 in order to explain the observed substitution times under strict neutrality. The data are difficult to account for even in terms of nearly neutral mutations ($s < 1/N$) and an effective population size of $N = 1000$. The alternative explanation is that the effective population size is larger, but positive selection is acting to make changes sweep through the population faster.

Polymorphism
Many analyses of pathogen genotypes assume isolated strains to be clonal, that is, to be monomorphic. This assumption has been scrutinised by De Boer *et al.* [22], who showed that, in fact, a large proportion (93%) of *M. tuberculosis* isolates are monomorphic using IS*6110* as the marker. They also show that the limits of detection of a second strain are around frequencies of 0.1 to 0.3. More sensitive instruments and refined genotyping procedures are likely to reveal greater polymorphism. The current information can be used, however, to study the population of the organism in hosts by using ranges of *detectable polymorphism*. In this section, two ranges will be considered in examining predictions from models: first, 0.1 to 0.9, and second, 0.3 to 0.7.

The polymorphism argument rests on the assumption that the isolates reported in [22] can be viewed as a random sample from a set of populations in mutation-drift equilibrium. It should be noted that because the isolate represents a sample of cells from the patient, it presumably does not always reflect the diversity of cells in the
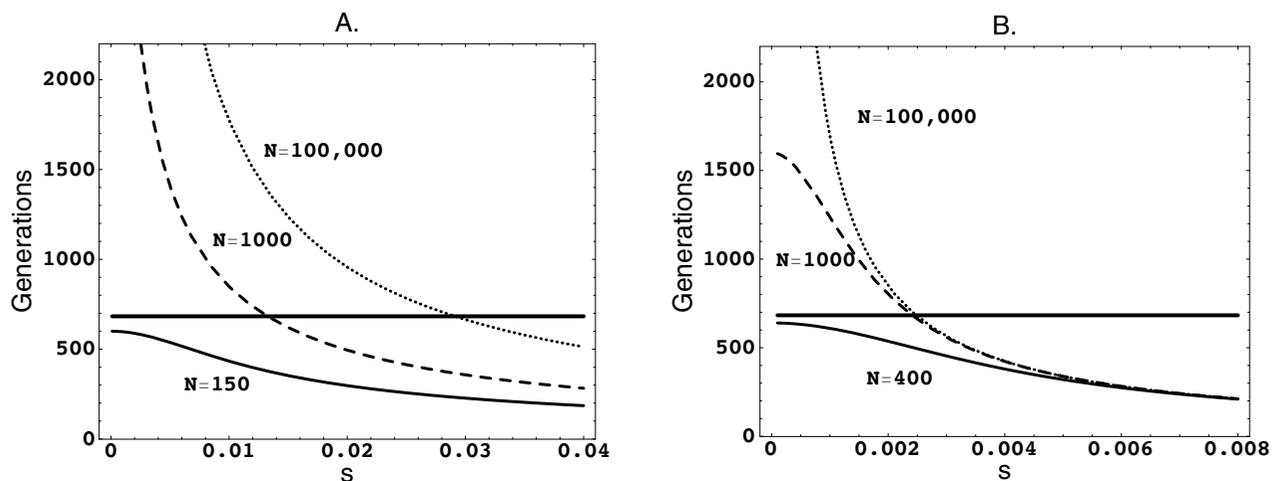
Figure 2
Mean sojourn times as functions of selective coefficient *s*, for different values of *N*. Left: from *a* = 1/*N* to *b* = 1 - 1/*N*; Right: from *a* = 0.3 to *b* = 0.7.

greater within-host population. Thus the polymorphism or heterogeneity observed from isolates is an underestimate of the actual levels.

Wright [23] found the stationary probability distribution of allele frequencies under the diffusion model with mutation and two alleles. Let *f(x)* be the probability density function of this distribution and *F(x)* be the cumulative probability function *F(x)* (see Appendix). The probability that a given population (patient) is between frequencies *a* and *b* (where *a* <*b*) is

$$P_{ab} = F(b) - F(a) = \int_a^b f(\gamma)d\gamma. \qquad (3)$$

This quantity can be alternatively interpreted as the proportion of populations observed to be polymorphic according to the detection limits set by (*a*, *b*).

First consider the neutral case. When there is no selection (*s* = 0), the distribution described by *f(x)* is a Beta distribution. Figure 3 shows the probability of an isolate being scored as a polymorphic population, using two alternative detectable polymorphism ranges (*a*, *b*) = (0.1, 0.9) and (0.3, 0.7), and a mutation rate of *μ* = 7.9 × 10⁻⁴ per cell per generation.

Next, consider the model that includes selection. For the two detectable polymorphism ranges, Figure 4 shows how selective coefficient *s* and effective population size *N* are related to the probability of observing polymorphism. As *s* increases, the predicted polymorphism decreases dra-
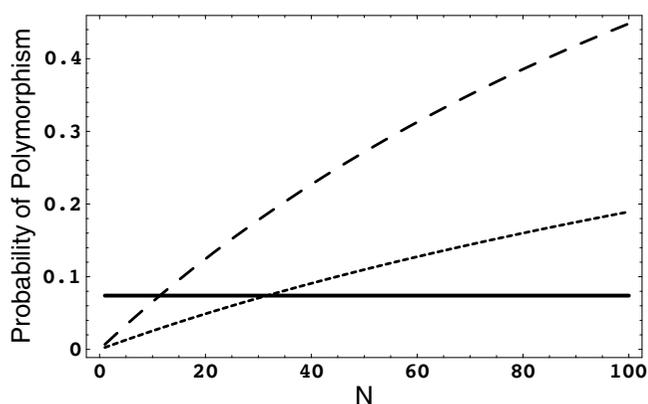

Figure 3
Probability of detecting polymorphism in the absence of selection, as a function of *N*. Two different ranges of detectable polymorphism were used. Dashed curve: (0.1, 0.9); dotted: (0.3, 0.7). We use *μ* = 7.9 × 10⁻⁴. The horizontal bar indicates the observed fraction of polymorphic populations (0.074) from de Boer *et al.* [22].

matically, particularly for large *N*. Again, an explanation of the observed level of polymorphism is only possibly by setting *N* to be extremely low.

Discussion
The three lines of evidence presented in this article suggest positive selection on *M. tuberculosis* within hosts. There are, however, limitations to these analyses. In the first argument there is no inherent problem with finding transposition rates that are high. In the second argument, it is
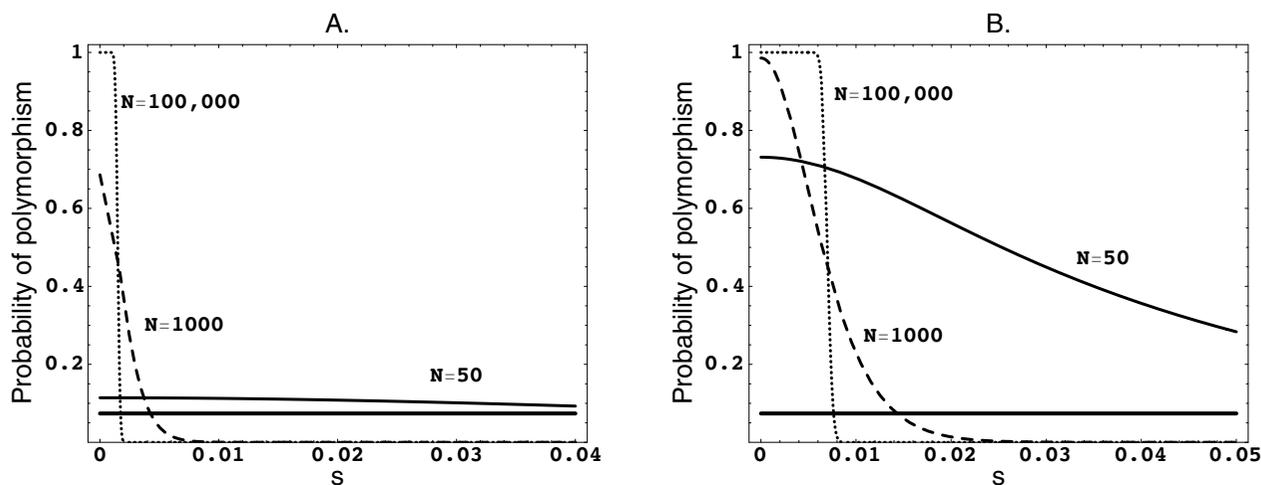
Figure 4
Probability of polymorphism as a function of *s*. Left: the detection threshold is set at 0.3; Right: the detection threshold is set at 0.1. The mutation rate is set to $\mu = 7.9 \times 10^{-4}$

possible to lower the effective population size far enough to explain the speed of substitution. In the third argument, 1) the bacterial populations sampled in [22] might not be close to mutation-drift equilibrium, 2) the sampled cells might not reflect the true diversity of the bacterial population in a patient, and 3) the levels of polymorphism again may be explained by very low effective population sizes. Consistency with observations nevertheless requires *N* values of around 100 or lower, which seems grossly at odds with the usually large census population sizes of bacteria. In mouse models of TB infection, for instance, bacterial loads reach around $10^5 - 10^7$ colony forming units per lung [2,24]. It has been noted, however, that effective population sizes of bacteria can be much lower than actual sizes [11,25].

I will also comment on why I have not attempted to statistically fit the model to data to estimate *N* and *s*. First, from the plots shown here, it is clear that different combinations of the two parameters can explain the observations. This would make it difficult to locate the best fit. Second, although the model can be used to assess the possibility of neutrality in the current context, it cannot adequately serve as a framework for estimation given the intricacies of host-pathogen interactions. Further, adding more parameters to the model would increase the complexity of the analysis beyond what can be sustained by the resolution of the currently available data.

Taken together, the results suggest positive selection, although the evidence is not conclusive. A possible alternative is that the effective population sizes of *M. tuberculosis* within patients are very low due to population

structure, background selection, or other factors. If there is indeed detectable adaptive evolution of tuberculosis within patients, what are the sources of selection? Two important candidates are antibiotic treatment and the host immune system. Studies using serial isolates have found no correlation between IS*6110* genotype instability and (a) drug resistance/susceptibility of the isolate [26,27], (b) *change* in drug resistance status [19] or (c) drug adherence by the patient [26]. It is still possible, however, that the collection of observed changes involve a variety of different genetic loci, with at least some conferring drug resistance, although such events may not be statistically detectable. Further, mutation in drug resistance loci will not necessarily be revealed by a marker. Genetic analysis of isolates of *M. tuberculosis* from the lung lesions of six patients has shown heterogeneity in resistance-associated alleles, but not with respect to IS*6110* [28].

Alternatively, fingerprint changes may reflect (evolutionary) escape from the immune system. The analysis here hints at low effective population sizes – perhaps the immune system induces a heavy decline in population sizes of *M. tuberculosis* within patients, i.e., bottlenecks – which is overcome by survivors with new genotypes. If the observed patterns are to be explained by severe bottlenecks, the surviving cells are not necessarily better adapted to residing in the host than the parental cells that were eliminated by the immune response. It is noteworthy that Yeh *et al.* [26] found no relationship between HIV status of the patient and genotype instability. This suggests that genetic changes in *M. tuberculosis* are not primarily driven by the immune system. However, the extraordinary ability

of *M. tuberculosis* to manipulate the T cell response [2] suggests the role of adaptation to the immune system in the deeper evolutionary history of the organism.

Interestingly, de Boer *et al.* [27] found an association between IS*6110* change and extrapulmonary disease or pulmonary+extrapulmonary disease and extrapulmonary origin of isolates. Dissemination is a major factor in the pathogenesis of tuberculosis. Since the lungs are the preferred environment of the organism, the new environments outside lungs may create opportunities for adaptive evolution. Adaptive evolution leading to specialisation to tissue types is to be expected. A recent article [29], for example, has found the occurrence of tissue-specific adaptations in *Streptococcus pyogenes* by examining ratios of non-synonymous to synonymous substitution rates (d_n/d_s).

Is IS*6110* directly responsible for adaptive mutations? On one hand, the apparently strict asexuality of *M. tuberculosis* implies that all genes good, bad or neutral are tightly linked to each other. It is likely then that IS-induced changes hitchhike to fixation with other mutations that confer advantage to the genome. On the other hand, it has recently been demonstrated that IS*6110* carries a promoter that can modify the the expression of neighbouring genes, raising the possibility of a direct role for the element in adaptive evolution [30]. Note that changes caused by IS*6110* can be not only beneficial, but also neutral or deleterious [31].

At the within-patient level, the best studied pathogen is perhaps HIV. While *M. tuberculosis* shares with viruses the characteristic of replicating within cells, a major difference is that mutation rates in viruses are much higher, particularly in retroviruses, which depend on reverse transcriptase (a low-fidelity enzyme) to copy their genomes. Hence, the extent of nucleotide variation of *M. tuberculosis* is not expected to be the same as is commonly observed for example in HIV [32]. There is ongoing controversy among HIV researchers about the role of stochasticity due to low effective population sizes in the evolution of the virus [33-35]. In any case, investigating the ratio of non-synonymous to synonymous substitutions (d_n/d_s) has established the action of positive selection on HIV within patients [32,36].

In *M. tuberculosis*, the level of polymorphism at synonymous sites has been noted to be extremely low [9]. It would be of interest to measure the ratio of non-synonymous to synonymous polymorphisms in key genes, such as loci conferring resistance to drugs, or those implicated in interactions with the immune system. These d_n/d_s ratios may provide further insight into the nature of positive selection in *M. tuberculosis*.

Appendix: Bacteria and the Wright-Fisher process

The analyses here rely on the commonly used diffusion model of drift and selection in a population, based on the Wright-Fisher process [37,38]. It is also possible to use the Moran model, in which at each time step an individual is chosen randomly to reproduce, and then another individual is chosen to die. The individual to die may be the same as the individual that reproduced, but not the offspring. Selection can be incorporated by including differential probability of birth or death for different genotypes. As noted by Ewens [38], the Moran model closely resembles the Wright-Fisher model; the critical difference between the two models arises from differences in the distribution of offspring number. The theory is usually discussed in relation to a diploid population of size N_e in which there are $2N_e$ copies of the (autosomal) gene in question. The diffusion model is used here with minor adjustments to describe bacterial populations. Let the number of bacterial cells in a population be N. Each mutant appears in the population at frequency $1/N$. Realistically, population sizes fluctuate and only a subset of cells actively divide. The number N should therefore be considered to be the effective rather than the actual population size, which may be much larger than N.

Mean and variance of change

It can be shown that the deterministic dynamics of selection in a haploid model are well approximated by a logistic model. The mean change in frequency x of an allele per generation is $m(x) = sx(1-x)$, which is identical to the diploid model with additive fitnesses (no dominance) if each copy of the advantageous allele adds s to the fitness (see [[37], p. 192]). That is, heterozygotes enjoy a fitness advantage s and homozygotes have advantage $2s$.

The variance component $v(x)$, the variance in change of allele frequency per generation, can also be taken from diploid theory. Replacing the diploid model of the random union of gametes with choosing cells randomly from each generation to the next, the effective population size is adjusted according to the distribution of offspring number under a given model of cellular division.

Binary fission

There are numerous ways to model drift in populations of organisms that reproduce by dividing to produce two daughters [39]. Here, cells are assumed to undergo fission synchronously and daughter cells are chosen randomly at each generation. In the absence of selective effects, the offspring distribution is $p_0 = 1/4$, $p_1 = 1/2$, $p_2 = 1/4$, where p_i is the probability of producing i offspring. The variance in offspring number here is $1/2$ and the variance-effective

population size equals $N/\left(\dfrac{1}{2}\right) = 2N$. Thus in this case,

the diploid theory can be directly used as far as $v(x)$ is concerned (replacing $2N_e$ with $2N$). Johnson and Gerrish [39] consider alternative models. These alternatives are associated with different rates at which drift proceeds in a population, and would not affect the qualitative conclusions drawn here.

Fixation probability

Diffusion models of genetic drift have shown [11] that the probability of fixation of an allele at frequency p in a randomly mating diploid population of size N_e (with $2N_e$ copies of the gene in question) is

$$u = \frac{1 - e^{-4N_e sp}}{1 - e^{-4N_e s}}. \qquad (4)$$

Therefore, using $p = 1/N$ rather than the usual $p = 1/(2N_e)$ the probability of fixation of a mutant bacterial cell with selective advantage s is

$$u = \frac{1 - e^{-4s}}{1 - e^{-4Ns}}. \qquad (5)$$

Note that when $4Ns \gg 1$, $u \sim 4s$. This agrees with a result of Gerrish and Lenski [18], using a branching process model (rather than the diffusion model) to find the fixation probability under this same model of binary fission. See [39] for discussion of u for alternative models.

Steady state distribution

Let the mutation rates from any genotype to any other be equal (μ). As stated above, selection is additive. As shown by Wright [23], the steady state distribution of allele frequency is then given by the density function

$$f(x) = \frac{x^{4N\mu-1}(1-x)^{4N\mu-1}e^{4Nsx}}{\int_0^1 y^{4N\mu-1}(1-y)^{4N\mu-1}e^{4Nsy}dy}. \qquad (6)$$

See also [37,38] for further details.

References
1. Bloom BR: *Tuberculosis: pathogenesis, protection and control* Washington: ASM Press; 1994.
2. Flynn JL, Chan J: **Tuberculosis: Latency and reactivation.** *Infect Immun* 2001, **69(7):**4195-4201.
3. Antia R, Koella JC, Perrot V: **Models of the within-host dynamics of persistent mycobacterial infections.** *Proc R Soc Lond B Biol Sci* 1996, **263(1368):**257-263.
4. Wigginton JE, Kirschner D: **A model to predict cell-mediated immune regulatory mechanisms during human infection with *Mycobacterium tuberculosis*.** *J Immunol* 2001, **166(3):**1951-1967.
5. Gammack D, Doering CR, Kirschner DE: **Macrophage response to *Mycobacterium tuberculosis* infection.** *J Math Biol* 2004, **48(2):**218-242.
6. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, Schecter GF, Daley CL, Schoolnik GK: **The epidemiology of tuberculosis in San Francisco: A population-based study using conventional and molecular methods.** *N Engl J Med* 1994, **330:**1703-1709.
7. Foxman B, Riley L: **Molecular epidemiology: focus on infection.** *Am J Epidemiol* 2001, **153(12):**1135-1141.
8. Seidler A, Nienhaus A, Diel R: **The transmission of tuberculosis in the light of new molecular biological approaches.** *Occup Environ Med* 2004, **61(2):**96-102.
9. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM: **Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination.** *Proc Natl Acad Sci U S A* 1997, **94:**9869-9874.
10. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Gamier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST: **A new evolutionary scenario for the *Mycobacterium tuberculosis* complex.** *Proc Natl Acad Sci U S A* 2002, **99(6):**3684-3689.
11. Kimura M: *The Neutral Theory of molecular evolution* Cambridge: Cambridge University Press; 1983.
12. Rosenberg NA, Tsolaki AG, Tanaka MM: **Estimating change rates of genetic markers using serial samples: applications to the transposon IS 6110 in *Mycobacterium tuberculosis*.** *Theor Popul Biol* 2003, **63(4):**347-363.
13. James BW, Williams A, Marsh PD: **The physiology and pathogenicity of *Mycobacterium tuberculosis* grown under controlled conditions in a defined medium.** *J Appl Microbiol* 2000, **88(4):**669-677.
14. Laochumroonvorapong P, Paul S, Manca C, Freedman VH, Kaplan G: **Mycobacterial growth and sensitivity to H2O2 killing in human monocytes in vitro.** *Infect Immun* 1997, **65(11):**4850-4857.
15. Zhang M, Gong J, Lin Y, Barnes PF: **Growth of virulent and avirulent *Mycobacterium tuberculosis* strains in human macrophages.** *Infect Immun* 1998, **66(2):**794-799.
16. Sniegowski PD, Gerrish PJ, Johnson T, Shaver A: **The evolution of mutation rates: separating causes from consequences.** *Bioessays* 2000, **22(12):**1057-1066.
17. Shen MM, Raleigh EA, Kleckner N: **Physical analysis of Tn10 and IS10 promoted transpositions and rearrangements.** *Genetics* 1987, **116:**359-369.
18. Gerrish PJ, Lenski RE: **The fate of competing beneficial mutations in an asexual population.** *Genetica* 1998, **102-103:**127-144.
19. Niemann S, Richter E, Rusch-Gerdes S: **Stability of *Mycobacterium tuberculosis* IS 6110 restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis.** *J Clin Microbiol* 1999, **37:**409-412.
20. Ewens WJ: **Conditional diffusion processes in population genetics.** *Theor Popul Biol* 1973, **4:**21-30.
21. Maruyama T: **The average number and the variance of generations at particular gene frequency in the course of fixation of a mutant gene in a finite population.** *Genetics* 1972, **19:**109-113.
22. de Boer AS, Kremer K, Borgdorff MW, de Haas PE, Heersma HF, van Soolingen D: **Genetic heterogeneity in *Mycobacterium tuberculosis* isolates reflected in IS 6110 restriction fragment length polymorphism patterns as low-intensity bands.** *J Clin Microbiol* 2000, **38(12):**4478-4484.
23. Wright S: **Evolution in Mendelian populations.** *Genetics* 1931, **16:**97-159.
24. Moreira AL, Tsenova L, Aman MH, Bekker LG, Freeman S, Mangaliso B, Schroder U, Jagirdar J, Rom WN, Tovey MG, Freedman VH, Kaplan G: **Mycobacterial antigens exacerbate disease manifestations in *Mycobacterium tuberculosis*-infected mice.** *Infect Immun* 2002, **70(4):**2100-2107.
25. Selander R, Levin B: **Genetic diversity and structure in *Escherichia coli* populations.** *Science* 1980, **210(4469):**545-7.

26. Yeh RW, Ponce De Leon A, Agasino CB, Hahn JA, Daley CL, Hopewell PC, Small PM: **Stability of *Mycobacterium tuberculosis* DNA genotypes.** *J Infect Dis* 1998, **177(4):**1107-1111.
27. de Boer AS, Borgdorff MW, de Haas PEW, Nagelkerke NJD, van Embden JDA, van Soolingen D: **Analysis of rate of change of IS *6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates.** *J Infect Dis* 1999, **180:**1238-1244.
28. Kaplan G, Post FA, Moreira AL, Wainwright H, Kreiswirth BN, Tanverdi M, Mathema B, Ramaswamy SV, Walther G, Steyn LM, Barry CE 3rd, Bekker LG: ***Mycobacterium tuberculosis* growth at the cavity surface: a microenvironment with failed immunity.** *Infect Immun* 2003, **71(12):**7099-7108.
29. Kalia A, Bessen DE: **Natural selection and evolution of streptococcal virulence genes involved in tissue-specific adaptations.** *J Bacteriol* 2004, **186:**110-121.
30. Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, Howard ST: **IS *6110* functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*.** *Mol Microbiol* 2004, **52(4):**999-1012.
31. Tanaka MM, Rosenberg NA, Small PM: **The control of copy number of IS *6110* in *Mycobacterium tuberculosis*.** *Mol Biol Evol* 2004 in press.
32. Bonhoeffer S, Holmes EC, Nowak MA: **Causes of HIV diversity.** *Nature* 1995, **376(6536):**125.
33. Brown AJ: **Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population.** *Proc Natl Acad Sci U S A* 1997, **94(5):**1862-1865.
34. Rouzine IM, Coffin JM: **Linkage disequilibrium test implies a large effective population number for HIV *in vivo*.** *Proc Natl Acad Sci U S A* 1999, **96(19):**10758-10763.
35. Rodrigo AG: **HIV evolutionary genetics.** *Proc Natl Acad Sci U S A* 1999, **96(19):**10559-10561.
36. Nielsen R: **Changes in ds /dn in the HIV-1 env gene.** *Mol Biol Evol* 1999, **16(5):**711-714.
37. Crow JF, Kimura M: *An introduction to population genetics theory* New York: Harper and Row; 1970.
38. Ewens WJ: *Mathematical population genetics* Berlin: Springer-Verlag; 1979.
39. Johnson T, Gerrish P: **The fixation probability of a beneficial allele in a population dividing by binary fission.** *Genetica* 2002, **115(3):**283-287.