A Novel Replication Clock for *Mycobacterium tuberculosis*

Wendy P. Gill, Nada S. Harik, Molly R. Whiddon, Reiling P. Liao, John E. Mittler and
David R. Sherman

Supplementary Materials.

Supplementary Methods.

Mathematical Model of *in vitro* and *in vivo* Plasmid Loss. To model *in vitro* growth
and plasmid loss, we made four assumptions: 1. Cell death is negligible during log phase
growth; 2. Growth rate in log phase is constant; 3. Plasmid free and plasmid bearing
strains have the same growth rates (experimentally verified in preliminary experiments,
data not shown); and 4. Segregation rate (i.e., plasmid loss per generation) is independent
of growth rate (**Fig. 1c** and **f**) and does not change with time. We obtain equations for
dynamics during log phase growth:

$$dN/dt = rN \tag{1}$$

$$dP/dt = r(1 - s)P \tag{2}$$

$$dF/dt = rF + rsP \tag{3}$$

where N = total number of bacteria, P = number of bacteria carrying the plasmid, F =
number of plasmid-free bacteria, r = population growth rate, s = segregation constant, δ =
population death rate (= 0 *in vitro*), and t = time in days. These equations were integrated
to yield:

$$N(t) = N(0)e^{rt} \tag{4}$$

$$P(t) = P(0)e^{r(1-s)t} \tag{5}$$

These equations predict that the frequency f of plasmid-bearing bacteria at time t is

$$f(t) = P(t)/N(t) = P(0)e^{r(1-s)t} / N(0)e^{rt} = f(0)e^{-rst} \tag{6}$$

From these equations, r and rs, respectively, were estimated by fitting regression lines through plots of $ln[N(t)]$ and $ln[f(t)]$ versus t (quantifies we refer to as *SlopeN* and *Slopef*). In our studies, f and N were estimated from CFU counts on selective and non-selective plates. Finally, s was obtained by dividing our estimate for rs by our estimate for r.

To assess *in vivo* growth, we altered the model to allow for changes in growth rate over time and nonzero time-dependent death (or removal) rates. As with *in vitro* model, we assumed that the plasmid-free and plasmid-bearing strains have the same growth rates and that segregation rate is independent of growth rate and does not change with time. We also assumed that no bacteria immigrate into the lung from other organs. With these assumptions, the equations become

$$dN/dt = r(t)N - \delta(t)N \tag{7}$$

$$dP/dt = r(t)(1-s)P - \delta(t)P \tag{8}$$

$$dF/dt = r(t)F + r(t)sP - \delta(t)F \tag{9}$$

where $r(t)$ and $\delta(t)$, respectively, are the time-dependent growth and death rates. Again, we use the CFU, the percentage of bacilli carrying plasmid and the segregation constant to calculate the number of bacterial replications and the total number of bacilli that have died. Using the segregation constant estimated from *in vitro* experiments, we estimated $r(t)$ and $\delta(t)$:

$$r(t) = [SlopeN - SlopeP] / s \tag{10}$$

$$\delta(t) = r(t) - SlopeN \tag{11}$$

where *SlopeN* and *SlopeP*, respectively, are estimated from plots of $ln[N(t)]$ and $ln[P(t)]$ versus t *in vivo*. We assumed that $r(t)$ and $\delta(t)$ were constant within time intervals, but could change between time intervals. We then calculated the number of dead cells, D, from

$$dD/dt \; = \; \delta(t)N(t) \tag{12}$$

where $N(t)$ is defined using equation (7) above. Integrating this equation, we find that the number of dead and removed cells at the end of each interval is

$$D(t_{\mathrm{f}}) = \; D(t_{\mathrm{s}}) \; + \; \delta N(t_{\mathrm{s}})[\; \mathrm{e}^{(r - \delta)(t_{\mathrm{f}} - t_{\mathrm{s}})} - 1 \,]\,/(\; r - \delta) \tag{13}$$

where δ and r are the death and growth rates during that interval, and t_{s} and t_{f} are the start and end times, respectively, of that time interval. We defined the cumulative bacterial burden at time t as the total number of bacilli (living, dead, and removed) within the host lung, $N(t) + D(t)$. $N(0)$ was set to the first measurement on day 1, while $D(0)$ was set to zero. Although setting $D(0)$ to zero underestimates the number of dead cells, this approximation has little effect on our final calculations given the large increases in bacterial density over the first four weeks of infection.

Bootstrap analysis. We constructed 1000 bootstrapped data sets by randomly sampling five values with replacement from the five mice sampled at each time point with R2.6.1 statistical computing software[1]. For each parameter, the 95% bootstrap confidence intervals were estimated from values lying at the 2.5 and 97.5 percentile.

References.

1.   http://www.r-project.org/
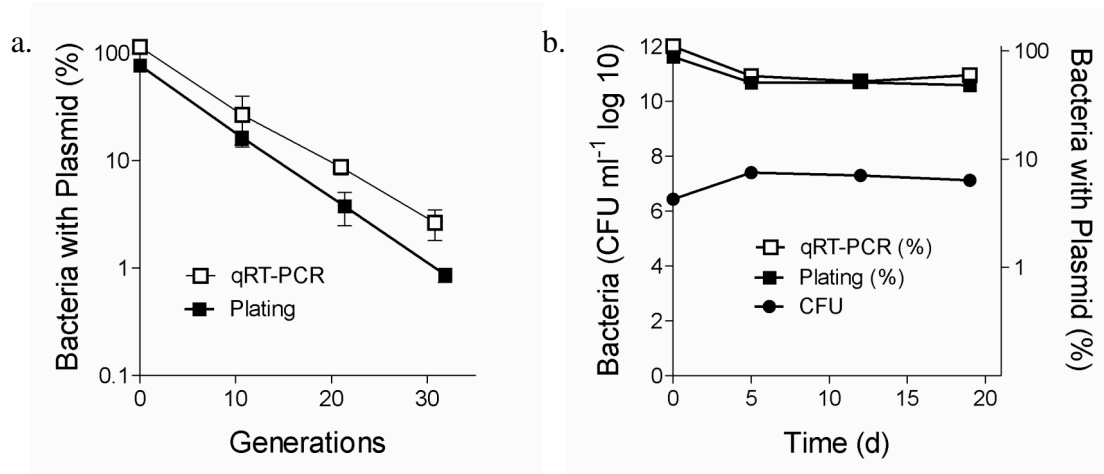
Supplementary Figure 1. *In vitro* stability of pBP10 in Mtb in the absence of antibiotic selection, as measured by qRT-PCR in log phase (**a**) and stationary phase (**b**). Data on bacterial CFU and plasmid stability as determined by plating are from Figures 1 and 2, and are presented here for comparison. PCR consistently measures slightly more plasmid per time point than CFU, probably because the plasmid copy number per cell is slightly >1. Both PCR and CFU measure identical rates of plasmid loss per generation.

Supplementary Equation 1: Measurable cell death rate *in vitro*.

Allowing for a non-zero death gives

$$dN/dt = rN \; - \; \delta N \tag{S1}$$

$$dP/dt = r(1 - s)P \; - \; \delta P \tag{S2}$$

$$dF/dt = rF + rsP \; - \; \delta F \tag{S3}$$

These equations can be integrated to yield:

$$N(t) = N(0)e^{(r - \delta)t} \tag{S4}$$

$$P(t) = P(0)e^{[r(1-s) - \delta]t} \tag{S5}$$

meaning that the frequency *f* of plasmid-bearing bacteria at time *t* is

$$f(t) = P(t)/N(t) = P(0)e^{[r(1-s) - \delta]t} / N(0)e^{(r - \delta)t} = f(0)e^{-rst} \tag{S6}$$

which is the same as equation (6) in the paper. Cell death does, however, affect estimates for r and s. In the absence of cell death, r and rs, respectively, can be estimated from *SlopeN* and *Slopef* [estimated from slopes of $ln[N(t)]$ and $ln[f(t)]$ versus t] With *in vitro* cell death, r and s should be estimated using the equations:

$$r = SlopeN + \delta \tag{S7}$$

and

$$s = - Slopef/r = - Slopef/(SlopeN + \delta). \tag{S8}$$

Failure to account for *in vitro* death, therefore, would lead to artificially low estimates for r and artificially high estimates for s. When fitting our *in vivo* model to data, we find that lower values for s lead to higher estimates for bacterial turnover *in vivo*. Therefore, assuming no bacterial death *in vitro* is conservative with respect to our conclusion that bacteria are turning over during the chronic phase in the mouse.

Supplementary Equation 2: Effect of assuming that the plasmid-carrying bacterium has a lower growth rate.

If we assume that the plasmid imposes a fitness cost, the equations become

$$dF/dt = rF + s(1-c)rP - \delta F \tag{S9}$$

$$dP/dt = (1-s)(1-c)rP - \delta P \tag{S10}$$

where r is the growth rate of the plasmid-free bacteria and c is the decrease in the growth rate caused by the plasmid.

Although these equations can be solved analytically, the solution does not yield a straight-forward expression for r. However, if we restrict ourselves to modelling the chronic

phase (the more interesting and relevant phase biologically), we can take advantage of the fact that plasmid-bearing bacteria have fallen to a low frequency. If $P << F$, segregation contributes little to F, leading to the following simplified equations:

$$dF/dt \; = \; [r - \delta]F \tag{S11}$$

$$dP/dt \; = \; [(1-s)(1-c)r \; - \; \delta]P \tag{S12}$$

Solving these equations, it is not hard to show that

$$r = \frac{SlopeF \; - \; SlopeP}{1 \; - \; (1-c)(1-s)} \tag{S13}$$

where $SlopeF = r - \delta$ and $SlopeP = (1-s)(1-c)r \; - \; \delta$, respectively, are estimated from the slopes of $ln[F(t)]$ and $ln[P(t)]$ versus t. From the mouse experiments, we know that $SlopeF = \sim 0.010/day$ and $SlopeP = \sim -0.022/day$ during the last interval. Using these estimates, we can use equation S13 to see how changing the fitness cost, c, (assumed to be zero in the text) would affect our estimates for r.

s	c	r
0.18	0.00	0.18
0.18	0.10	0.12
0.18	0.20	0.09
0.18	0.35	0.07
0.18	0.50	0.05

For example, assuming a 20% fitness cost during the chronic phase would reduce our estimate for the growth rate of bacteria during the last time interval from ~0.18 per day to ~0.09 per day. Numerical fits of the full model (equations S9 and S10) to this data gave

nearly identical results (with all five cost-dependent estimates for r during the last time interval being within 1.3% of the values above).