Supplemental Material

Flow cytometry and Real-Time qPCR as tools for assessing plasmid persistence

Wesley Loftie-Eaton[1, 2], Allison Tucker[2, 3, 4], Ann Norton[2] and Eva M. Top[1, 2, 3, *]

[1]Department of Biological Sciences, [2]Institute for Bioinformatics and Evolutionary Studies (IBEST), [3] Bioinformatics and Computational Biology Program, [4] Departments of Mathematics and Statistics; University of Idaho,

Supplementary methods

Calculation of plasmid loss rat

The rate of decline of plasmid-bearing cells within a population, or 'plasmid loss rate', in each assay was calculated using a logistic regression model, adapted from Hughes *et al.* (1), as described below:

$$
\begin{aligned}
&Y_{ikl} \sim BIN\left(\pi_{ikl}, n_{ikl}\right), \\
&logit\left(\pi_{ikl}\right) = \frac{\pi_{ikl}}{1-\pi_{ikl}} = \alpha_i + \beta_i 10k + \gamma_{ikl}, \\
&i = 1,...,5, k = 0,...,K, l = 1,...,3 \\
&\alpha_i \sim N\left(0,1000\right), \\
&\beta_i \sim N\left(0,1000\right), \\
&\gamma_{ikl} \sim N\left(0,\sigma_\gamma\right)
\end{aligned}
$$

Here, i is an identifier of the sampling technique, k represents the day of observation, and l denotes three replicate stability assays. The observed counts of plasmid-containing bacteria, Y_{ikl}, were assumed to follow a binomial distribution (BIN) with the expected proportion of plasmid free cells equal to π_{ikl} and sample size equal to n_{ikl}. α_i and β_i represent the intercept and slope of the linear model, respectively, per sampling technique i and were assumed to be normally distributed. The parameter γ_{ikl} was added to capture variation in sampling not accounted for by the binomial model (overdispersion). The distribution of the standard deviation σ_γ was assumed to be uniform with an upper bound of 100.

Intercepts of the presented model provide an indicator of the initial proportion of cells in the population without a plasmid at time zero. The slopes provide an indicator of the rate of plasmid loss over time. Due to the nonlinear nature of this rate, to compare the

different analysis methods the maximum rate of change, occurring at the time where the plasmid-bearing fraction of the population was 50%, was used. This is equivalent to $\beta_i/4$. Stability (or instability) was determined by the magnitude of this value – the more negative the value, the less stable the plasmid.

Markov Chain Monte Carlo (MCMC) was used to implement the model, specifically using a Gibbs sampler (2), as implemented in JAGS (3). Three parallel chains were run for 1,000,000 iterations, discarding the first 900,000 as burn in. Samples were taken every 100^{th} iteration to overcome possible autocorrelation. From this target posterior distribution dataset 30 points were sampled at random without replacement to calculate and compare the mean maximum loss rate. Convergence was assessed using the Gelman-Rubin diagnostic (4). Computation was done using R (R Development Core Team, 2012. http://www.R-project.org/) utilizing the package rjags (M. Plummer. 2011. http://CRAN.R-project.org/package=rjags) for interface with JAGS and coda (5) for diagnostics.

Supplementary Figures

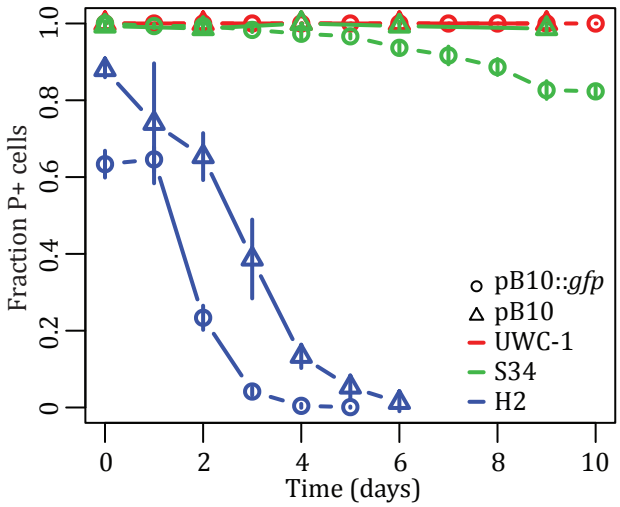

FIG. S1. Comparison of the persistence of pB10::*gfp* to wild type pB10 (6) in *P. putida* UWC-1, *P. veronii* S34 and *P. putida* H2 using the PC method. Note that in strain S34, the fraction of cells with wild type plasmid also eventually declined after 10 days (6). P+, plasmid-containing cells.

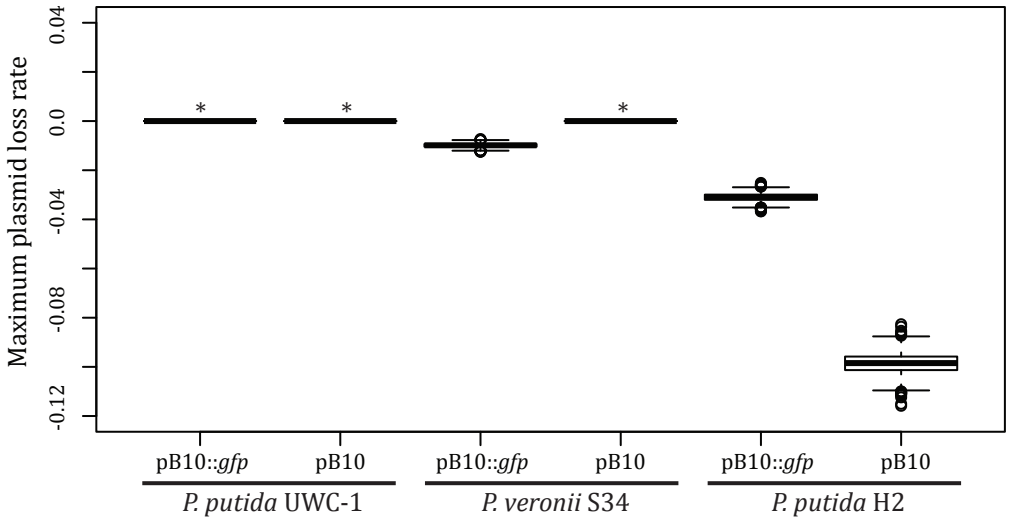

FIG. S2. Maximum plasmid loss rates (per generation) for pB10::gfp and wild type pB10 in the three hosts. * In *P. putida* UWC-1 no plasmid loss was observed, while in *P. veronii* S34 pB10 was reported to be only sporadically lost during the first 10 days of the persistence assay, with plasmid-containing cells varying at a fraction between 0.98 and 1.0 during this time (6). Therefore the maximum loss rate for these three strains was regarded as zero.

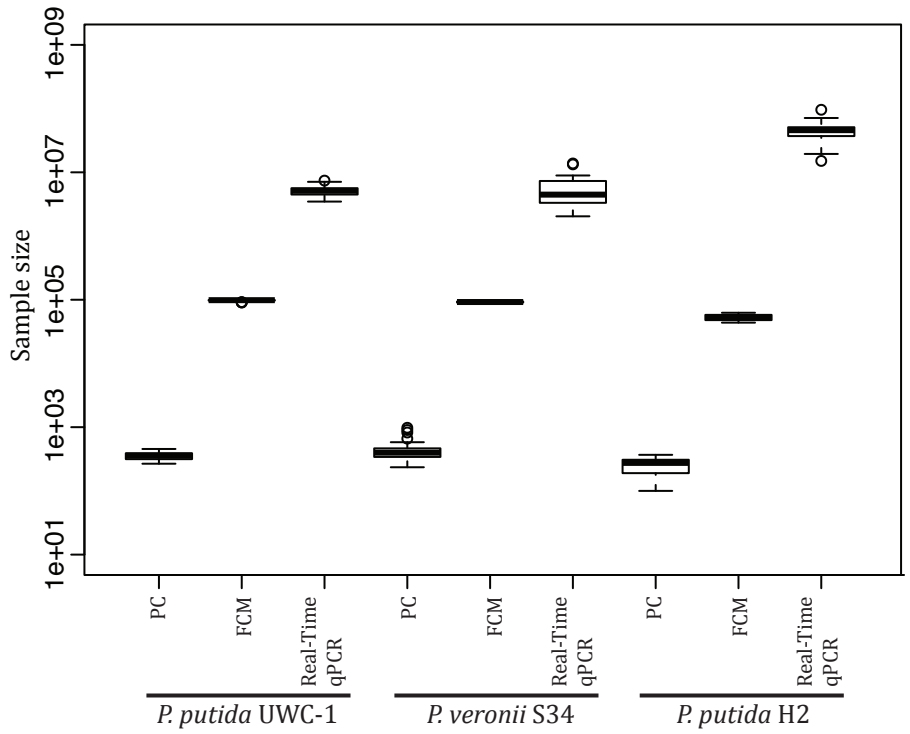

FIG. S3. Average sample size across all time points for each persistence assay. For PC sample size refers to the number of colonies screened, for FCM it refers to the number of events interrogated within the SSC-FSC-gated population and for Real-Time qPCR it is the number of chromosomes that were present in each Real-Time qPCR reaction.

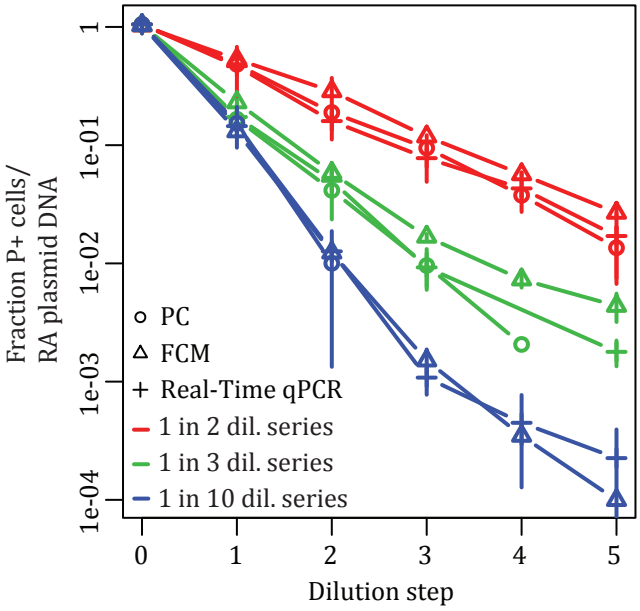

FIG. S4. Artificial plasmid stability assays consisting of known ratios of plasmid-containing and –free *P. putida* UWC-1 cells and measured by PC, FCM and real-Time qPCR. Different plasmid loss rates were achieved by diluting plasmid-containing cells into plasmid-free cells following a 1 in 2, 1 in 3 and 1 in 10 dilution series. Each dilution series was repeated in triplicate. Data are the same as in Fig. 5 but here fractions are plotted on a logarithmic scale. P+, plasmid-containing; RA, relative abundance.

Supplementary Tables

Table S1 ANOVA* comparisons of the maximum loss rates across the three different techniques for each of the persistence assays.

	df	sum sq	mean sq	f	p	Significant
P. putida UWC1 (pB10::*gfp*)	2	5.43E-05	2.71E-05	4.04E+02	<0.001	Yes
P. veronii S34 (pB10::*gfp*)	2	2.24E-03	1.12E-03	1.23E+03	<0.001	Yes
P. putida H2 (pB10::*gfp*)	2	6.70E-03	3.35E-03	5.26E+02	<0.001	Yes
1 in 2 dil. series	2	8.14E-04	4.07E-04	1.44E+02	<0.001	Yes
1 in 3 dil. series	2	1.72E-02	8.60E-03	3.07E+02	<0.001	Yes
1 in 10 dil. series	2	1.11E-02	5.55E-03	2.15E+01	<0.001	Yes

* Pairwise t-test; where df is the degrees of freedom, sum sq is the sum of the squares, mean sq is the mean of the squares, f is the f-statistic and p is the probability that the null hypothesis is true

Table S2 ANOVA* comparisons of the maximum loss rates across the three different dilution series for each of the three techniques.

	df	sum sq	mean sq	f	p	Significant
PC	2	0.2261	0.1131	676.2	<0.001	Yes
FCM	2	0.1604	0.0802	1828	<0.001	Yes
Real-Time qPCR	2	0.1846	0.0923	1177	<0.001	Yes

* Pairwise t-test; where df is the degrees of freedom, sum sq is the sum of the squares, mean sq is the mean of the squares, f is the f-statistic and p is the probability that the null hypothesis is true

Table S3 Comparison of variance* in the maximum loss rate calculated from the data gathered using each of the three techniques.

Dilution	Comparison	f	p	Result
1 in 2	PC vs FCM	1.1466	0.7151	Similar
	PC vs qPCR	0.9765	0.9495	Similar
	FCM vs qPCR	0.8517	0.6685	Similar
1 in 3	PC vs FCM	1.9203	0.08419	Similar
	PC vs qPCR	1.0142	0.9699	Similar
	FCM vs qPCR	0.5281	0.091	Similar
1 in 10	PC vs FCM	4.1659	<0.001	FCM smaller
	PC vs qPCR	2.3356	0.02561	qPCR smaller
	FCM vs qPCR	0.5606	0.125	Similar

* F test; f is the f-statistic and p is the probability that the null hypothesis is true

Supplementary References

1. **Hughes, J. M., B. K. Lohman, G. E. Deckert, E. P. Nichols, M. Settles, Z. Abdo, and E. M. Top**. 2012. The role of clonal interference in the evolutionary dynamics of plasmid-host adaptation. MBio **3(4)**:e00077-12.
2. **Geman, S., and D. Geman**. 1984. Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell **6(6)**:721-741.
3. **Plummer, M**. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing.
4. **Geweke, J**. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments, Bernardo, J. M., J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), Bayesian Statistics 4, Oxford University Press, Oxford UK.
5. **Plummer, M., Best, N., Cowles, K. and K. Vines.** 2006. R News 6: 7-11.
6. **De Gelder, L., J. M. Ponciano, P. Joyce, and E. M. Top.** 2007. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. Microbiol **153(2)**:452-463.