Additional file 2

Contents

DATA ANALYSIS, INCLUDING BOOTSTRAPPING

For each combination of assay plate (KB, KB+rif), bacterial mutator strain (*WT* vs. *MutS*), resistance phenotype (*Rif-* vs. *Rif+*) and treatment (control, rif, phage, rif plus phage), we calculated the mean and variance of bacterial density over the 12 evolved replicate populations. These data gave us $2 \times 2 \times 2 \times 4 = 32$ pairs of values (mean, variance). In the same way, we calculated means and variances for the 12 replicate samples of the four ancestral populations *WT/Rif-; WT/Rif+; MutS/Rif-; MutS/Rif+* using each of the assay plates (KB, KB+rif). These data gave us an additional $4 \times 2 = 8$ (mean, variance) pairs.

The 40 pairs of means and variances of density (listed in Additional file 1) were base-10 log-transformed. In an Analysis of Covariance (ANCOVA), $\log_{10}$ variance was taken as response variable, $\log_{10}$ mean as covariate and assay plate environment, mutator strain, resistance phenotype and treatment as cofactors. Ancestral strain was analysed as a treatment in addition to the four long-term treatments. Table 1 lists all the cofactors.

A model with all two-way interactions between cofactors was statistically fitted to the data. By fitting two-way interactions between cofactors and log mean, we tested for differences in the slope of the log variance - log mean relationship. Backward model simplification was employed, starting with the removal of non-significant interactions with cofactors, to obtain a minimal adequate model (Table 2). Inverting roles by taking log mean as the response variable and log variance as covariate in the model produced qualitatively very similar results (not shown).

We also performed simple linear regressions of log variance on log mean to estimate intercept and slope separately for each combination of assay plate (presence or absence of rif), bacterial strain and treatment (Table 1).

In 7 populations (all *WT*) from the rif plus phage treatment, no phage was detected at the end of the experiment. However, neither mean nor variance of bacterial density differed from populations still harbouring phage ($p > 0.1$), and excluding these 7 populations from the main analysis produced nearly identical results (not shown). We therefore presented the results based on the complete data set, including these populations.

The validity of ANCOVA rests on assumptions that the residuals from each fitted function have a distribution that is independent of the abscissa and is normally distributed. To test the robustness of conclusions based on these assumptions, we used the bootstrap procedure to fit linear and quadratic functions to the 40 pooled data points. Because the minimal model (Table 2) identified only the interaction between log mean density and treatment as statistically significant, we also bootstrapped linear and quadratic functions separately to each of the 5 sets of 8 data points for each treatment. We performed 10,000 bootstrap samples for each analysis, and again independently 1,000 bootstrap samples, to see whether results were sensitive to bootstrap sample size in this range. As the results (not shown) differed little between 1,000 and 10,000 bootstrap samples, we reported only the results from 10,000 samples.

The bootstrap procedure samples the data points with replacement. All the data points sampled could be identical, producing a singular design matrix. In bootstrap samples of the 40 data points, this possibility (which has very low probability) never arose, but it did arise for several individual treatments, which had only 8 data points each. Such bootstrap samples were uninformative and were discarded.

ANCOVA analyses were performed using JMP v8.0 (SAS 2009). Bootstrap analyses were performed using MATLAB R2011b (Mathworks 2011).

RESULTS OF THE BOOTSTRAP ANALYSES

The bootstrap analyses confirmed the results from the above regression analyses, and hence are not tabulated here. Fitting quadratic functions of log mean gave no 95% confidence intervals for the coefficients of the quadratic terms that differed significantly from 0, for all 40 data points or all 5 treatments analysed separately. Hence log variance was approximately linear as a function of log mean. The median slopes (of log variance as a linear function of log mean) from the bootstrap samples did not differ to the nearest 0.01 or differed very slightly from the point estimates of the ordinary linear regressions in Table 1. The exception was the rif treatment, where the median bootstrap slope was 5.46, compared to the linear regression estimate of 5.43.

The bootstrapped confidence intervals of the slopes were always wider, and sometimes substantially wider, than the confidence intervals of the ordinary linear regressions. However, the confidence interval (3.68, 8.74) of the bootstrapped slope in the rifampicin treatment remained significantly above 2, like the confidence interval of the slope in the rifampicin treatment in Table 1.