Supplemental Material for "Mutation rate variability as a driving force in adaptive evolution"

Dalit Engelhardt and Eugene I. Shakhnovich

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

S1 Evolution equations

We consider a population subdivided into wildtype with growth rate g_{wt} and baseline mutation rate μ_{bl}, a resistant phenotype with this baseline mutation rate and growth rate g_r, and corresponding phenotypes of equal growth rate but increased mutation rate $f\mu_{bl}$, $f > 1$. Population levels at time t are given by $x_{wt,1}(t)$, $x_{r,1}(t)$, $x_{wt,f}(t)$, and $x_{r,f}(t)$, respectively. Transitions (mutations) between the subpopulations are permitted in accordance with the schematic (Fig. 1 in the main text). We assume limited resources set by an environmental carrying capacity K, so that the subpopulation levels thus evolve in time according to the equations

$$\begin{cases} \dot{x}_{wt,1}(t) = \left(1 - \frac{x_{tot}(t)}{K}\right)\left[\left(1 - \mu_{bl}\left(p_{r,wt} + P_{del}\right) - r_\mu\right) g_{wt} x_{wt,1}(t) + \mu_{bl} p_{r,wt} g_r x_{dr,1}(t) + r_\mu g_{wt} x_{wt,f}(t)\right] \\ \dot{x}_{wt,f}(t) = \left(1 - \frac{x_{tot}(t)}{K}\right)\left[\left(1 - f \times \mu_{bl}\left(p_{r,wt} + P_{del}\right) - r_\mu\right) g_{wt} x_{wt,f}(t) + f \times \mu_{bl} p_{r,wt} g_r x_{r,f}(t) + r_\mu g_{wt} x_{wt,1}(t)\right] \\ \dot{x}_{r,1}(t) = \left(1 - \frac{x_{tot}(t)}{K}\right)\left[\left(1 - \mu_{bl}\left(p_{r,wt} + P_{del}\right) - r_\mu\right) g_r x_{r,1}(t) + \mu_{bl} p_{r,wt} g_{wt} x_{wt,1}(t) + r_\mu g_r x_{r,f}(t)\right] \\ \dot{x}_{r,f}(t) = \left(1 - \frac{x_{tot}(t)}{K}\right)\left[\left(1 - f \times \mu_{bl}\left(p_{r,wt} + P_{del}\right) - r_\mu\right) g_r x_{r,f}(t) + f \times \mu_{bl} p_{r,wt} g_{wt} x_{wt,f}(t) + r_\mu g_r x_{r,1}(t)\right] \end{cases} \tag{1}$$

where r_μ is the rate of mutation from a baseline-mutation rate ($f = 1$) phenotype to a $f > 1$ phenotype (rate at which mutations leading to elevated mutation rates $f\mu_{bl}$ occur) and its reverse (assumed to be equal), $p_{r,wt} \equiv p_{wt\to r} = p_{r\to wt}$ is the probability of mutation from wildtype to the resistant phenotype and backward (assumed to be equal as explained in the main text), P_{del} is the probability of mutation to deleterious phenotypes, $x_{tot} = x_{wt,1} + x_{wt,1f} + x_{r,1} + x_{r,f}$.

In order to compute the relative advantage or disadvantage conferred by hypermutation on the fixation of drug resistant subpopulations (Eq. (2) in the main text) we numerically compute the ratio of resistant mutants (combined non-hypermutant and hypermutant types) in the total population at $2 \leq f \leq 1000$ to the ratio that would result if no hypermutations were allowed in the system, i.e. if we set $r_\mu = 0$ and consider only phenotypes $x_{wt,1}$ and $x_{r,1}$. When computing these quantities for a system with an initial distribution of hypermutants of either phenotype, we assume that the $r_\mu = 0$ system has a corresponding distribution in which

$$x_{i,r_\mu=0}\left(t=0\right) = x_{i,1,r_\mu\neq 0}\left(t=0\right) + x_{i,f,r_\mu\neq 0}\left(t=0\right),$$

with i representing either wildtype or the resistant phenotype.

In the figures shown in the main text and here μ_{bl} was set at $2 \times 10^{-10} \times N_{\text{genome}}$ per generation per cell [1] where $N_{\text{genome}} = 4.64 \times 10^6$ is the number of basepairs in the E. coli genome. In the results shown in the main text, when the population is not purely wildtype at $t = 0$, the proportion of hypermutants chosen is assumed to be distributed proportionally amongst the wildtype and resistant populations.

S2 Rate of acquisition r_μ of increased mutation rate and initial proportion of hypermutants

To estimate a biologically reasonable r_μ we consider a simple system consisting of a wildtype $f = 1$ phenotype and a wildtype $f > 1$, both with fitness g_{wt}, which can mutate into each other with rate r_μ:

$$\begin{cases} \dot{x}_{wt,1}(t) = \left(1 - \frac{x_{wt,1}(t)+x_{wt,f}(t)}{K}\right)\left[(1-r_\mu)\, g_{wt} x_{wt,1}(t) + r_\mu g_{wt} x_{wt,f}(t)\right] \\ \dot{x}_{wt,f}(t) = \left(1 - \frac{x_{wt,1}(t)+x_{wt,f}(t)}{K}\right)\left[(1-r_\mu)\, g_{wt} x_{wt,f}(t) + r_\mu g_{wt} x_{wt,1}(t)\right] \end{cases} \tag{2}$$

The steady-state (stationary distribution) proportion of hypermutants in the total population will be given by

$$R = \frac{x_{wt,f}(\tau)}{x_{wt,1}(\tau) + x_{wt,f}(\tau)} \tag{3}$$

at time τ after resources have been saturated. Hypermutation can be caused by various mechanisms; studies focused on pathogenic E. coli have found comparatively high ($> 1\%$) proportions of mutators in bacterial isolates (3.6% in [2] and 1.9% in [3]); a separate study that looked specifically for MMR deactivation in E. coli found a much lower proportion (0.24%) when both commensal and pathogenic E. coli were included [4]. A later study [5] found, however, that when other sources of hypermutation were included besides MMR, E. coli cells exhibiting increased mutation rates – of up to two orders of magnitude from the baseline mutation rate – constituted as much as 14% of the total population, most being mild mutators, with both commensal and pathogenic strains included in the study. The highest mutation rates were found to correspond to MMR deficiencies, with lower increases due to other mechanisms. Note that since r_μ is a neutral-selection rate, studies of mutator proportions that were conducted under conditions of adaptive evolution will likely overestimate this parameter and we therefore restrict our data to studies of natural isolates, noting that even in those cases adaptive evolution in the recent past may have taken place.

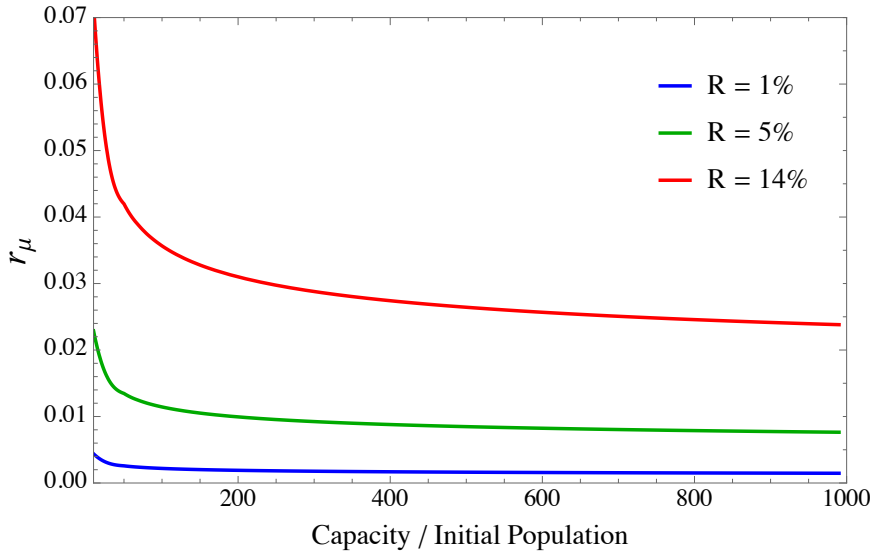

Figure S1: Computed values of r_μ from the system (2) for different expected steady-state proportions of resistant cells as the availability of resources is varied. Across a wide range of carrying capacities r_μ only varies from $\sim 0.5\%$ to $\sim 1.5\%$. Plots in the main body of the paper employ $r_\mu = 0.25\%$, corresponding to an initial hypermutant population of 1% of the total population.

We compute which r_μ values yield the stationary distribution ratio (3) for different carrying capacities by taking g_{wt} as in the main text to be 0.34 h^{-1} under no inhibition. The results for different values of R are shown in Fig. S1. Since we consider a uniform distribution[1] of mutation rate increase factors f and the

[1] A non-uniform distribution can be incorporated by multiplying r_μ by a probability distribution that depends on f - as this adds additional degrees of freedom to the model we avoid doing so here.

14% figure is heavily tipped towards mild mutators, using this figure will likely overestimate the mutation rate r_μ in our model for higher values of f. For the purpose of the plots in the main body of the paper we set on the lower end, at $r_\mu = 0.25\%$ and an initial proportion of 1% hypermutating cells in the population.

S3 Results for a spectrum of r_μ and P_{del}

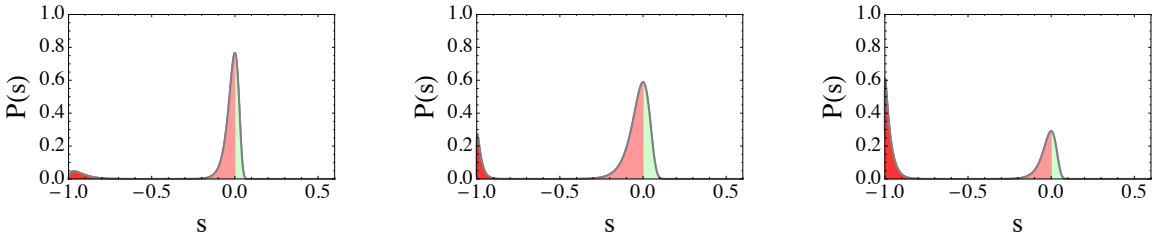

Figure S2: Sample bimodal distributions of fitness effects shown in terms of the probability of mutation as a function of the selection coefficient, $s \equiv g_{\text{mut}}/g_{wt} - 1$, with the green and light red regions indicating mutations with beneficial and mildly deleterious effects, respectively. The darker red region indicates strongly deleterious and lethal mutations.

Experimental findings on distributions of fitness effects (DFEs) across a variety of organisms and conditions point to a probability peak around neutral (wildtype) fitness that is skewed toward mildly to somewhat deleterious mutations (Fig S2). The proportion of lethal or near-lethal mutations varies greatly in the literature and exhibits a dependence on both the organism studied and the experimental conditions: earlier studies (e.g. [6]) found a bimodal distribution to be typical, whereas a recent study has observed far fewer lethal mutations in E. coli, consituting only about 1% of all mutations recorded [7]. In view of these differing data and the possibility that ambient conditions may affect the DFE, it is important to establish the extent to which our conclusions in this study are robust against variations in the balance between the rate of deleterious, neutral, and beneficial mutations. The probability of beneficial mutations was one of the main indicators of selective advantage that we explored in the main text as a component of advantage-conferring conditions $\vec{\rho}$; here we consider changes in P_{del} – the probability assigned to strongly deleterious or lethal mutations in our model.

Figs. S3 and S4 show plots corresponding to Fig. 2(c)-(d) and Fig. 3 in the main text for different combinations of P_{del} and r_μ with corresponding choices of initial hypermutant proportion, $x_{h,0}/x_{tot,0}$ based on the analysis in the previous section. We see from Fig. S3 that the effects of increasing P_{del}, i.e. altering the balance of beneficial to deleterious mutations to result in a higher rate of deleterious mutations, are (1) to shift the optimal mutation rate (i.e. to more strongly constrain the range of mutation rate increases f that benefit resistance fixation) to lower values for a given level of baseline-rate advantage ($R_r(f = 1)$); and (2) to increase the overall impact of hypermutation ($R_r(f > 1)/R_r(f = 1)$) for a given level of baseline-rate advantage. Importantly, however, the level of advantage at which hypermutation no longer plays a role does not significantly change over a large range of variations in P_{del}, indicating that our main results about the role of hypermutation as a function of evolutionary advantage are qualitatively robust against changes in the distribution of fitness effects. We see that changes in r_μ and $x_{h,0}/x_{tot,0}$ carry a significant quantitative impact on $R_r(f > 1)/R_r(f = 1)$ and that higher values of r_μ and $x_{h,0}/x_{tot,0}$ result in little to no loss in $R_r(f > 1)/R_r(f = 1)$ even at very high mutation rates. The effects due to changes in P_{del} and $(r_\mu, x_{h,0}/x_{tot,0})$ carry over to the behavior of $R_r(f > 1)/R_r(f = 1)$ and $R_h([I] > 0)/R_h([I] = 0)$ as shown in Fig. S4.

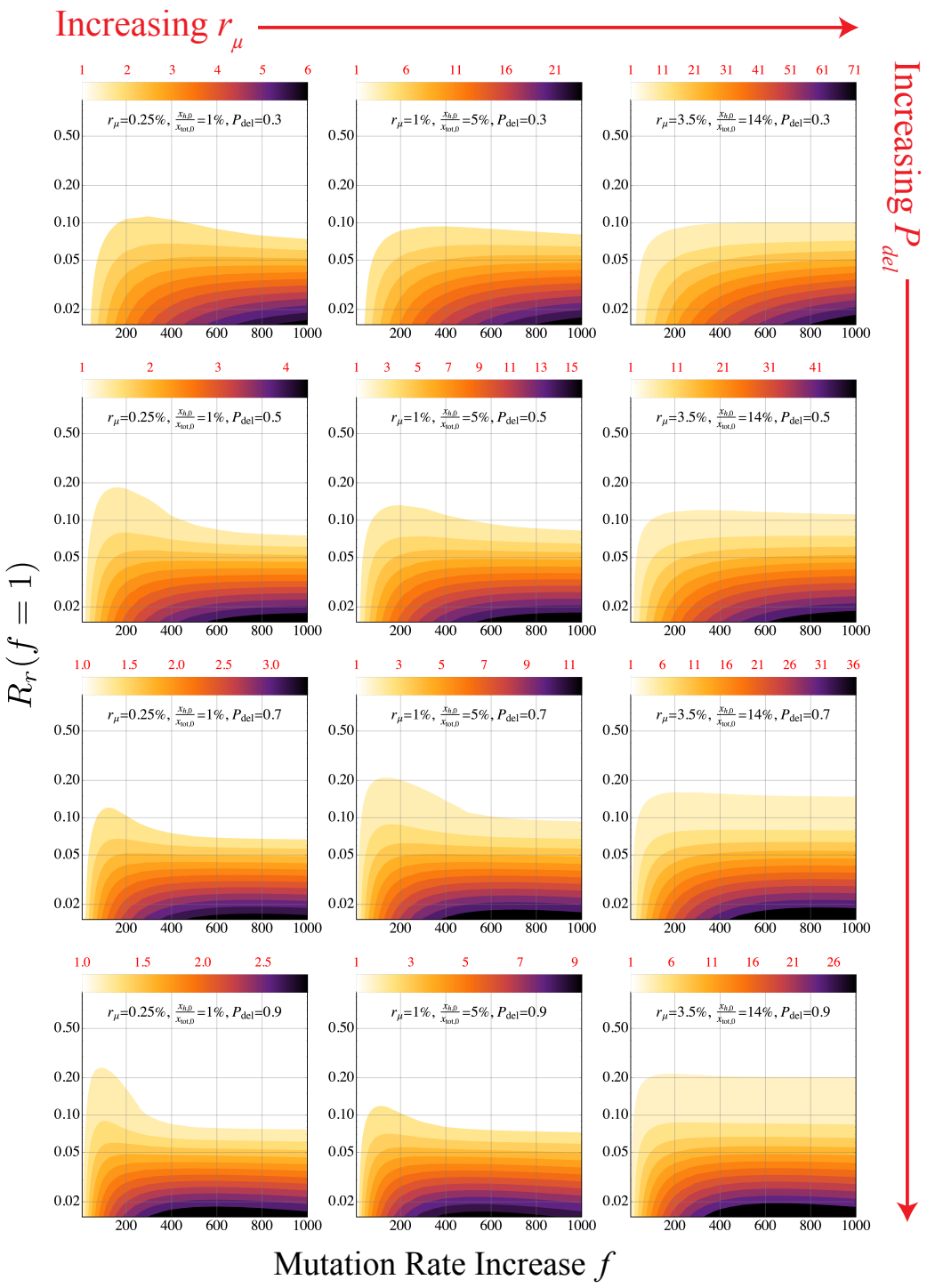

Figure S3: Contour plots of $R_r(f > 1)/R_r(f = 1)$ (Eqn. (2) in the main text) in f-$R_r(f = 1)$ space corresponding to those shown in Fig. 2(c)-(d) in the main text (with parameters noted there) at different combinations of (r_μ, P_{del}) and the approximate[2] proportion of initial-state hypermutants $(x_{h,0}/x_{tot,0})$ corresponding to this choice of r_μ based on the analysis plotted in Fig. S1 above.

[2] Since we consider a range of K values in calculating $R_r(f > 1)/R_r(f = 1)$, the choice of $x_{h,0}/x_{tot,0}$ corresponding to r_μ is approximate.

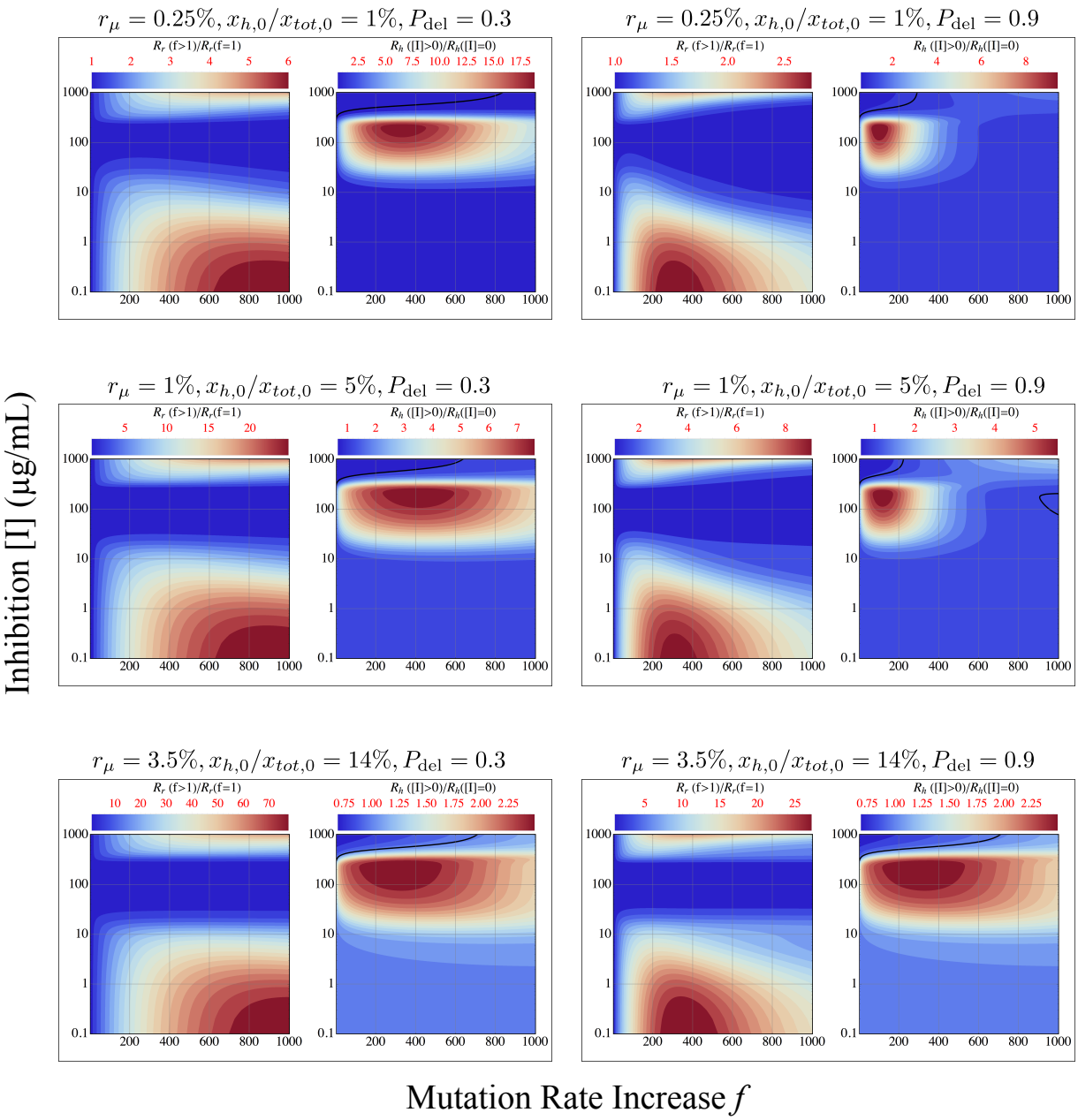

Figure S4: Side-by-side contour plots for the effects of inhibition on resistance fixation ($R_r(f > 1)/R_r(f = 1)$) and hypermutation fixation ($R_r([I] > 0)/R_r([I] = 0)$) in f-$[I]$ space corresponding to those shown in Fig. 3(a)-(b) in the main text (with parameters noted there) at different combinations of (r_μ, P_{del}) and the approximate proportion of initial-state hypermutants ($x_{h,0}/x_{tot,0}$) corresponding to this choice of r_μ based on the analysis plotted in Fig. S1 above. The location of $R_r([I] > 0) = R_r([I] = 0)$ is shown in the black curve, indicates the point of no hitchhking of hypermutants on resistant mutations.

S4 Subpopulation fixation dynamics

Shown below are the frequencies of each of the four subpopulations in the model as a function of time at different mutation rates. In the main text, proportions were calculated from the stationary population distribution.

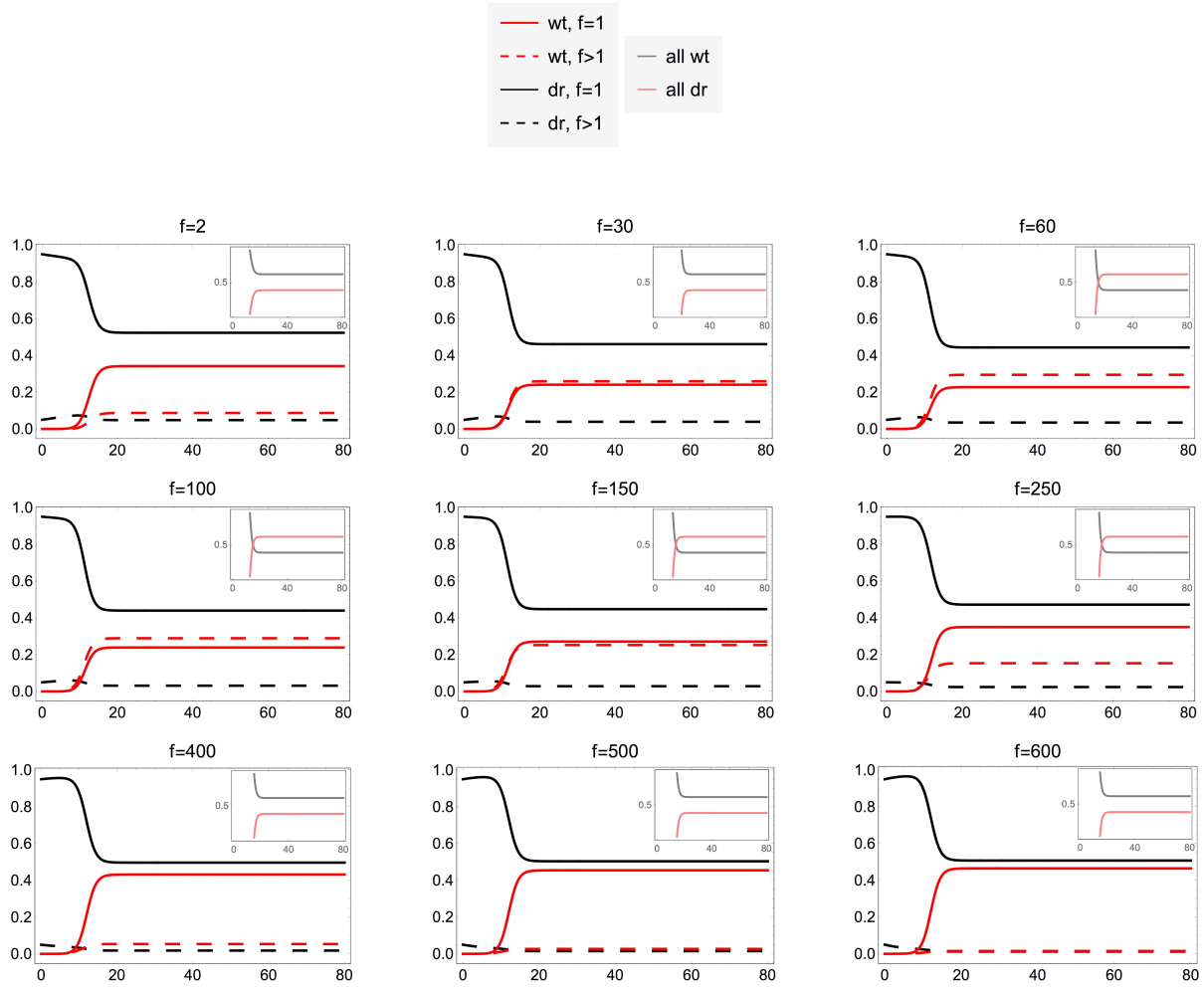

Figure S5: Frequencies (proportions in total population) over time of each of the four subpopulations in the model: baseline-mutation rate wildtype (solid red), baseline-mutation rate resistant (solid black), hypermutant wildtype (dashed red), and hypermutant resistant (dashed black). The inset plots show the frequencies of all wildtype (hypermutant and baseline combined) cells (gray) and all resistant cells (light red). Intermediate mutation rate increases ($f = 60, 100, 150, 250$) in this case result in the resistant mutant becoming dominant by the time that resource saturation and hence a stationary-state distribution are reached, whereas lower and higher mutation rate increases have a wildtype-dominated stationary-state distribution. Model parameters were set at $g_r = 3.5 \times g_{wt}$ with $g_{wt} = 0.34\ \mathrm{h}^{-1}$, $x_{tot}(t = 0) = 10^6$, $K = 10^8$, $P_{wt \to r} = 0.01$, $x_r(t = 0) = 0$, $x_h(t = 0) = 5\%$ (all wildtype hypermutants), $r_\mu = 1\%$, and $P_{del} = 0.9$.

References

[1] H. Lee, E. Popodi, H. Tang, and P. L. Foster, "Rate and molecular spectrum of spontaneous mutations in the bacterium escherichia coli as determined by whole-genome sequencing," *Proceedings of the National Academy of Sciences* **109** no. 41, (2012) E2774–E2783.

[2] K. Jyssum, "Observations on two types of genetic instability in escherichia coli," *APMIS* **48** no. 2, (1960) 113–120.

[3] J. E. LeClerc, B. Li, W. L. Payne, and T. A. Cebula, "High mutation frequencies among escherichia coli and salmonella pathogens," *Science* **274** no. 5290, (1996) 1208–1211.

[4] M. D. Gross and E. C. Siegel, "Incidence of mutator strains in escherichia coli and coliforms in nature," *Mutation Research Letters* **91** no. 2, (1981) 107–110.

[5] I. Matic, M. Radman, F. Taddei, B. Picard, C. Doit, E. Bingen, E. Denamur, and J. Elion, "Highly variable mutation rates in commensal and pathogenic escherichia coli," *Science* **277** no. 5333, (1997) 1833–1834.

[6] A. Eyre-Walker and P. D. Keightley, "The distribution of fitness effects of new mutations," *Nature Reviews Genetics* **8** no. 8, (2007) 610.

[7] L. Robert, J. Ollion, J. Robert, X. Song, I. Matic, and M. Elez, "Mutation dynamics and fitness effects followed in single cells," *Science* **359** no. 6381, (2018) 1283–1286.