Population Heterogeneity in Mutation Rate Increases the Frequency of Higher-Order Mutants and Reduces Long-Term Mutational Load

Helen K. Alexander, *[,†,‡] Stephanie I. Mayer,[†] and Sebastian Bonhoeffer
Institute of Integrative Biology, Department of Environmental Systems Science, ETH Zürich, Switzerland
[†]These authors contributed equally to this work.
[‡]Present address: Department of Zoology, University of Oxford, Oxford, United Kingdom
***Corresponding author**: E-mail: helen.alexander@zoo.ox.ac.uk
Associate editor: Yuseob Kim

Abstract

Mutation rate is a crucial evolutionary parameter that has typically been treated as a constant in population genetic analyses. However, the propensity to mutate is likely to vary among co-existing individuals within a population, due to genetic polymorphisms, heterogeneous environmental influences, and random physiological fluctuations. We review the evidence for mutation rate heterogeneity and explore its consequences by extending classic population genetic models to allow an arbitrary distribution of mutation rate among individuals, either with or without inheritance. With this general new framework, we rigorously establish the effects of heterogeneity at various evolutionary timescales. In a single generation, variation of mutation rate about the mean increases the probability of producing zero or many simultaneous mutations on a genome. Over multiple generations of mutation and selection, heterogeneity accelerates the appearance of both deleterious and beneficial multi-point mutants. At mutation-selection balance, higher-order mutant frequencies are likewise boosted, while lower-order mutants exhibit subtler effects; nonetheless, population mean fitness is always enhanced. We quantify the dependencies on moments of the mutation rate distribution and selection coefficients, and clarify the role of mutation rate inheritance. While typical methods of estimating mutation rate will recover only the population mean, analyses assuming mutation rate is fixed to this mean could underestimate the potential for multi-locus adaptation, including medically relevant evolution in pathogenic and cancerous populations. We discuss the potential to empirically parameterize mutation rate distributions, which have to date hardly been quantified.

Key words: **transient mutagenesis, population genetics, cancer, bacteria, hypermutator, adaptation.**

Introduction

Mutation rate is a key evolutionary parameter that affects the level of genetic diversity in a population. Genetic diversity in turn affects both the population's current mean fitness and its capacity to adapt to changes in the environment. Most theoretical work to date has assumed that mutation rate takes on a fixed value in all members of the population. Nonetheless, mutation rate, like any other trait, can be expected to vary among individuals, due to genetic, environmental, and stochastic effects. The recognition that mutation rate can vary within a population is implicit in the long-standing study of mutation rate evolution, and more recently in considerations of stress-induced mutagenesis, especially in bacteria. However, a comprehensive conceptual understanding of how mutation rate heterogeneity within a population affects the *de novo* appearance of mutations and long-term availability of standing genetic variation is lacking.

The existence of rare individuals with high mutation rate could be particularly important when a combination of several mutations is relevant for adaptation (Ninio 1991; Boe 1992; Drake et al. 2005; Drake 2007). Given that the mutation rate is typically low, higher order mutants are generally rare, yet they can be crucial for adaptation to complex new environments. For example, when multiple drugs are applied in combination—a common treatment approach for cancer (Al-Lazikani et al. 2012) and several major infectious diseases (Goldberg et al. 2012)—resistance in the targeted pathogens/cells generally requires multiple mutations. The prevalence of such multi-point mutants in the standing genetic variation before drug treatment starts, when they are generally expected to carry fitness costs, is predicted to be crucial to the emergence of resistance during treatment (Ribeiro and Bonhoeffer 2000; Komarova and Wodarz 2005). Multiple mutations are also involved in the initiation and progression of many cancers (Knudson 2001).

There is clearly a genetic contribution to mutation rate via genes involved in replication, proofreading, and repair of the genetic material. This can result in variation of mutation rate even among closely related individuals. Laboratory investigations have identified "mutators" and "antimutators" (having higher, respectively lower, mutation rate than the wild type), attributable to one or few specific genetic changes, in a variety of organisms. Effect sizes range up to hundreds- to thousands-fold variation in eukaryotic cells, bacteria and DNA viruses,

Article

Open Access

and up to around 5-fold in retro- and RNA viruses (supplementary text I.1, Supplementary Material online). The abundance of such variants in natural populations is less clear. Mutators are expected to arise frequently *de novo* due to the large target size for mutations causing defects in replication or repair genes (Drake 1993; Denamur and Matic 2006). Theoretically, under constant conditions, alleles that alter mutation rate can be expected at mutation-selection balance in the long term (Johnson 1999; Denamur and Matic 2006; Lynch and Abegg 2010; Desai and Fisher 2011). Moreover, by hitchhiking with beneficial alleles they generate during phases of adaptation, mutators may rise to higher frequency in the short term (Taddei et al. 1997; Loeb and Loeb 2000; Desai and Fisher 2011). In experimental populations of bacteria, mutators have indeed been observed to spontaneously arise and persist (Sniegowski et al. 1997) and be enriched through selective sweeps (Mao et al. 1997). Surveys of clinical and other natural isolates in several bacterial species indicate that strains exhibiting a range of mutation rates also exist outside the laboratory (LeClerc et al. 1996; Matic et al. 1997; Oliver et al. 2000; Björkholm et al. 2001; Denamur et al. 2002; Richardson et al. 2002; Morosini et al. 2003; Prunier et al. 2003; Baquero et al. 2004; Watson et al. 2004). In RNA virus populations, mutators appear rapidly in laboratory settings (Suárez et al. 1992; Combe et al. 2015) and are expected to be present in heterogeneous natural populations (Suárez et al. 1992; Mansky and Cunningham 2000), but we are not aware of any surveys of natural isolates. Cancerous tumors, which are characteristically genetically unstable and highly heterogeneous (Lengauer et al. 1998; Gillies et al. 2012; Barber et al. 2015), are also anticipated to be polymorphic in genes affecting mutation rate. It has been hypothesized that a mutator phenotype arises early in carcinogenesis, and moreover increases the chances of successive mutations affecting genomic stability, leading to further non-uniform increases in mutation rate (Loeb et al. 1974; Loeb 1991, 2001; Loeb et al. 2003). However, there does not appear to be any study quantifying mutation rate in a representative sample of co-existing individuals from a single natural population (e.g., pathogens within one infected patient or cells within one tumor).

Many environmental factors—including temperature, pH, oxygenation, UV radiation, and chemicals—have also been implicated in modulating mutagenesis in bacteria, viruses, and cancerous cells (supplementary text I.2, Supplementary Material online). In addition to these abiotic factors, there is some evidence that bacterial mutation rate is modified by intercellular interactions, which depend on population density (Krašovec et al. 2014). Viral mutation rate could also be affected by its host cell's type, physiological state, and antiviral defenses (supplementary text I.2.3, Supplementary Material online). However, few quantitative estimates relating environmental variables to mutation rate are available. Some antibiotics appear to increase bacterial mutation rates by 2- to around 100-fold (Gillespie et al. 2005; Kohanski et al. 2010; Long et al. 2016), whereas certain antiretrovirals increase HIV-1 mutation rate by roughly 5-fold (Mansky et al. 2003). The mutation rate of *E. coli* varied over a 3-fold range with population density (Krašovec et al. 2014). While it is clear that the relevant environmental factors may be heterogeneously distributed in a population's habitat, inducing different mutation rates in co-existing individuals, the precise distribution will be highly context-dependent.

Finally, mutation rate may vary randomly and nonsystematically in a population, due to stochastic effects on individuals' physiological states (Boe et al. 2000; Drake 2007). For example, the SOS response, which is associated with production of error-prone polymerases in bacteria (Tenaillon et al. 2004), exhibited a distribution of induction levels in *E. coli*, with 0.3% of the population at least 20-fold above the average level (McCool et al. 2004). Even constitutively expressed replication and repair genes are subject to random errors in transcription and translation that affect the protein's fidelity (Ninio 1991; Boe 1992; Miller 1996). Rough calculations suggested that bacterial populations contain resulting "transient mutators" at a total frequency of around 5×10^{-4}, with mutation rates expected to be enhanced to similar degrees as in genetic mutators (Ninio 1991). Fluctuations in low copy number proteins, particularly upon cell division, could also yield temporary reduction in repair capacity (Drake 2007), and imbalanced concentrations of protein subunits could produce polymerases missing the proofreading subunit (Aoki and Furusawa 2001). Thus, even isogenic populations in uniform macroenvironments seem likely to contain individuals with differing propensities to generate mutations, although the few direct tests to date have yielded mixed results (Elez et al. 2010; Kennedy et al. 2015; Uphoff et al. 2016).

Taken together, this evidence suggests that mutation rate variation within populations is probably common, though few quantitative estimates are available. DNA-based organisms appear to have the capacity to vary mutation rate over a few orders of magnitude, whereas RNA-based viruses appear to tolerate only modest (up to around 5-fold) changes in their already high baseline mutation rates (Drake and Holland 1999; Mansky and Cunningham 2000). The frequency of mutators in a population could vary widely depending on the source of mutation rate variation and the selective conditions. Furthermore, a broad spectrum arises in the extent to which mutation rate is correlated between parent and offspring. At one extreme, if mutation rate is entirely genetically controlled, the offspring will inherit its parent's mutation rate. At the other extreme, erroneously translated polymerases or other intracellular components will have limited if any intergenerational effects before degrading and/or being diluted by new production. If mutation rate is primarily determined by the external environment, parent-offspring correlation could vary widely, depending on the extent to which they share a common environment. If spatial variation in the environment is fine-grained relative to the typical offspring dispersal distance, correlation will be low, while if variation is coarse-grained, parent and offspring are likely to experience the same environment and thus mutation rate.

The general evolutionary consequences of mutation rate heterogeneity cannot readily be determined experimentally, due to the diversity of its forms and the difficulty of establishing a comparable population that differs only in the extent of heterogeneity. This calls for a theoretical approach to

elucidate its effects. A large body of work on the evolution of mutation rate [reviewed by Sniegowski et al. (2000)] takes into account the existence of heritable variants. Though the majority of this literature assumes the mutation rate is constitutive, evolution of stress-induced mutagenesis has also been considered (Bjedov et al., 2003; Ram and Hadany, 2012, 2014). A key factor considered to drive evolution of mutation rate is indirect selection through linkage to fitness-determining loci, but the focus of these studies is on the dynamics of a mutator allele itself. Far fewer studies have considered the converse: how the existence of mutation rate variability in the population, regardless of its source, affects mutational dynamics at other loci (Ninio 1991; Furusawa and Doi 1992; Cairns 1998; Aoki and Furusawa 2001; Gonzalez et al. 2008; Lynch and Abegg 2010). These few existing models allow only two possible values of mutation rate, and are mostly designed for particular organisms and mechanisms of variation.

In the present study, we develop a more general theoretical framework to understand the effects of population heterogeneity—that is, variability about the mean mutation rate among co-existing individuals. We address both the probability of mutation at single or multiple loci and the temporal dynamics of mutants under ongoing production and selection. To keep our results as general as possible, we do not make any assumption as to the biological mechanism underlying this heterogeneity, in particular whether it is an adaptive/regulated response or an unavoidable byproduct of random processes or external environmental factors. Instead we determine how mutant frequency and population mean fitness depend on the moments of an arbitrary mutation rate distribution and the degree to which mutation rate is "inherited" (correlated) from parent to offspring. We consider haploid, asexually reproducing individuals, which is a reasonable first approach for bacteria (neglecting horizontal gene transfer in some species), viruses (neglecting complementation and in some cases recombination), and cancerous cells (neglecting dominance effects).

We find that variability of mutation rate about the population mean does not affect dynamics at a single focal locus, but has increasingly large relative effects as more loci are taken into consideration. Heterogeneity increases the probability of zero or many simultaneous mutations occurring on a genome, at the expense of intermediate (usually single-point) mutations. Over multiple generations, this altered *de novo* production in conjunction with selection turns out both to accelerate the first appearance of multi-point mutants and, in the long term, to enhance the frequency of costly higher-order mutants at mutation-selection balance. Inheritance of mutation rate strengthens these effects, especially when mutations tend to be accumulated sequentially rather than simultaneously. The clustering of mutations among population members also reduces the population's long-term mutational load. Our analytical approximations allow quantification of these effects in terms of moments of the mutation rate distribution and selection coefficients. Finally, we discuss the evolutionary implications of these findings, as well as empirical approaches to quantify mutation rate heterogeneity.

Results

We consider the effect of mutation rate heterogeneity on the frequency of mutants at various timescales. First, in a single generation before selection operates, we examine the probability of generating multiple mutations on a genome. Next, we extend to modeling temporal dynamics of mutants over multiple generations of mutation and selection. We consider the initial emergence of double mutants, where mutations are either beneficial or deleterious, as well as the mutation-selection balance achieved in the long term by deleterious mutations.

Since we are interested in the effects of heterogeneity itself—that is, variation about the population mean mutation rate—we focus on comparing a "heterogeneous" population, in which mutation rate has a given distribution, to a baseline "homogeneous" population with mutation rate fixed to the mean of this distribution. This approach contrasts with many previous studies that compared the presence versus absence of a "mutator" type, thus changing the mean as well as variance of mutation rate. In order to focus on the effects of inter-individual heterogeneity, we ignore other forms of heterogeneity, such as differences in mutation rate across genomic sites or systematic changes in the distribution of mutation rate over time (see "Discussion" section). To reflect the distribution of mutation rate across co-existing individuals, we treat mutation rate per locus (U) or per genome (Λ) as a random variable, with a fixed or realized value drawn from this distribution denoted by u or λ, respectively. We use the notation $\langle\cdot\rangle$ to denote the expectation over the distribution of mutation rate.

Probability of Simultaneous Mutations on a Genome

We model the occurrence of mutations in a single generation ("simultaneously") across n loci on a genome with a binomial distribution, i.e. each locus mutates independently with a probability given by the individual's mutation rate. (In the limit as $n \to \infty$ and the per-locus mutation rate $\to 0$, the number of mutations that occur is thus Poisson-distributed, and our results extend naturally to this case.) An individual's mutation rate in turn is drawn from the population distribution (fig. 1, left). The overall probability $p_{n,j} = \langle p_{n,j}(U)\rangle$ of j simultaneous mutations on a genome is then obtained by averaging the binomial probabilities $p_{n,j}(U)$ ["Methods" section, Equation (8)] over the distribution of U. These probabilities can also be interpreted as the expected frequencies of j-point mutants produced (before selection) by a purely wild type starting population.

As may be intuitively expected, variability of the mutation rate about a fixed mean does not change the expected number of mutations that occur, but does increase the variance when $n \geq 2$ (supplementary text II.1.1, Supplementary Material online). More precisely, we can ask how variability affects the probability of exactly j mutations.

For $n = 1$, the functions $p_{1,j}(U)$ are linear. Thus, the overall probability of mutation at a single locus is fully determined by the mean mutation rate and independent of the extent of variability: specifically $p_{1,0} = 1 - \langle U\rangle$ and $p_{1,1} = \langle U\rangle$. On

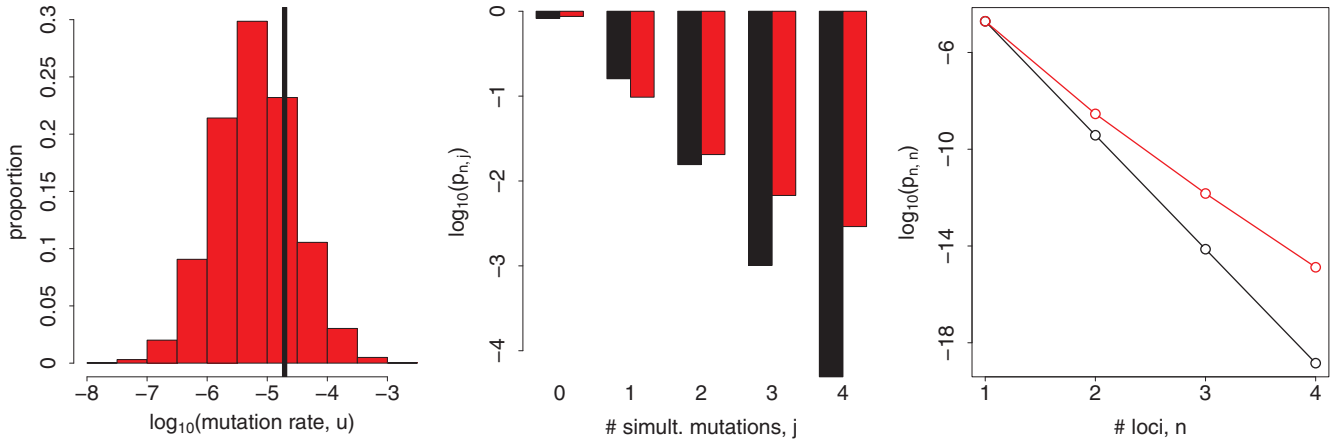

Fig. 1. Effect of mutation rate heterogeneity on the probability of simultaneous mutations. *Left*: the distribution of per-locus mutation rate (U) obtained by sampling 10000 times from a log-normal distribution, with the thick black vertical line indicating the sample mean. Sample mean 1.9×10^{-5} (close to the base substitution rate estimated for HIV-1; Mansky and Temin 1995), sample variance 2.5×10^{-9}, range 2.3×10^{-8}–1.4×10^{-3}. *Centre*: probability of j simultaneous mutations among $n = 10\,000$ loci (approximately the number of base pairs in the HIV-1 genome), i.e., $p_{n,j}$ as a function of j, if U is heterogeneous following the chosen distribution (red) versus fixed to the sample mean (black). The directional effects agree with analytical predictions (keeping in mind that higher probabilities will be less negative on the log scale). *Right*: probability of all n particular loci under consideration mutating simultaneously, i.e., $p_{n,n}$, as a function of n, again for the heterogeneous (red) and homogeneous (black) cases. The relative effect size increases with n.

the other hand, for $n \geq 2$, the functions $p_{n,j}(U)$ are nonlinear, and in general $\langle p_{n,j}(U)\rangle \neq p_{n,j}(\langle U\rangle)$: that is, the overall probability of simultaneous mutations at multiple loci is not fully determined by the mean. Applying Jensen's Inequality (Cover and Thomas 2006, p. 27), we can determine whether variability in mutation rate increases or decreases this probability, for each n and j, by examining the second derivative of $p_{n,j}(U)$ (supplementary text II.1.2, Supplementary Material online).

Since $p_{n,0}(U) = (1-U)^n$ and $p_{n,n}(U) = U^n$ are convex for all $U \in (0,1)$ and $n \geq 2$, the probabilities of either all or none of the loci mutating are clearly always increased by variability in U. Logically, the probability of at least some intermediate numbers of mutations must be reduced. We find that $p_{n,j}$ will generally be increased by heterogeneity for the smallest and largest values of j, and decreased in some intermediate range of j, with the exact switching points of the directional effect depending on n and on the particular distribution of U. For realistic ranges of mutation rate in most organisms, heterogeneity will increase the chance of zero or of two or more simultaneous mutations and decrease the chance of a single mutation occurring, even with many loci under consideration (fig. 1, centre). For certain RNA viruses with high mutation rates, and possibly cellular populations containing strong mutators, the switching points may be shifted upward.

We further analyze the magnitude of the effect of heterogeneity in the case of all n loci mutating simultaneously. Rewriting this probability (supplementary text II.1.3, Supplementary Material online):

$$p_{n,n} = \langle U^n\rangle = \langle U\rangle^n + \sum_{i=0}^{n-1} \binom{n}{i} c_{n-i}\langle U\rangle^i \qquad (1)$$

where $c_i = \langle (U - \langle U\rangle)^i\rangle$ is the i^{th} central moment of the mutation rate distribution. (In particular, $c_0 = 1$, $c_1 = 0$, and c_2 is the variance.) Thus the probability of n simultaneous mutations depends on the first n central moments. For instance, the "boost" in triple mutations increases with the variance, and is larger when the distribution is right-skewed than when it is left-skewed. Note that even-numbered central moments must be positive, whereas odd-numbered central moments may be positive or negative; however, according to Jensen's Inequality, any negative terms must be outweighed by the positive terms. It can also be shown (supplementary text II.1.3, Supplementary Material online) that the relative effect of heterogeneity on the probability of simultaneous mutation at all n loci under consideration ($p_{n,n}$) increases with n (fig. 1, right).

Initial Emergence of Mutants

We now turn to modeling dynamics over multiple generations. Our starting point is a standard deterministic model consisting of recursive equations that describe how the frequency x_i of each genotype i, defined by the non-mutant (0) or mutant (1) allele at each locus, changes from generation to generation under the forces of mutation and selection ("Methods—Deterministic Model of Genotype Frequency Dynamics" section). Equivalently, the vector of genotype frequencies $x(t)$ is multiplied by a "mutation-selection matrix", M, which incorporates mutation probabilities and selection coefficients s_i (relative fitness $1 - s_i$). We modify the standard equations to account for mutation rate heterogeneity within the population (supplementary text II.2.1, Supplementary Material online). Note that this model incorporates acquisition of multiple mutations both simultaneously (in a single generation) and stepwise (over multiple generations).

When considering dynamics over more than one generation, the extent to which mutation rate is inherited or correlated from parent to offspring becomes relevant. As described

in the "Introduction" section, this correlation could vary over a broad spectrum, but for mathematical analysis, we model the two extremes in which mutation rate is inherited either perfectly or not at all. When mutation rate is non-inherited, in each generation every individual independently draws anew from the mutation rate distribution. Each individual thus has the same overall mutation probabilities, obtained by taking the expectation over the distribution of mutation rate ($p_{n,j}$ as described in the previous section). These probabilities are incorporated into the matrix M representing the entire population. On the other hand, when mutation rate is perfectly inherited, a fixed mutation rate is maintained in each lineage. Then the population can be divided into disconnected subpopulations defined by distinct mutation rates, which combined with the subpopulation frequencies yield the population-level mutation rate distribution. The dynamical equations are then iterated separately in each subpopulation, with the mutant frequencies in the total population obtained at any given time point by taking a weighted average over subpopulations (equivalently, taking the expectation over the mutation rate distribution, now as the final step). By treating each subpopulation independently and holding subpopulation frequencies fixed, we prevent indirect selection on mutation rate itself, thus ensuring that the mutation rate distribution does not change over time and remains directly comparable to the non-inherited case. If subpopulations are actually competing together, this approach can be considered an approximation, whose accuracy we discuss in individual scenarios to follow and in greater detail in supplementary text II.2.1 and II.2.5, Supplementary Material online.

For the one- and two-locus models, we derive approximate solutions valid for the short-term temporal dynamics starting from a purely wild type population, for arbitrary mutational fitness effects (supplementary text II.2.2 and II.2.3, Supplementary Material online). At a single locus, dynamics are driven by the mean mutation rate and heterogeneity has a negligible effect, in the sense that the leading-order term in the mutant frequency is proportional to $\langle U \rangle$, and terms involving higher moments are of smaller orders of magnitude. We therefore focus on the two-locus case, in which dynamics are driven by both mean and variance of the mutation rate distribution. More generally, with n loci, M contains nonlinear terms up to U^n, and n-point mutants will have frequency on the order of $\langle U^n \rangle$. Thus, we expect the first n moments of the mutation rate distribution generally to play a non-negligible role in genotype frequency dynamics.

Figure 2 illustrates the temporal dynamics of double mutant frequency, $x_{11}(t)$, in the two-locus model, in an example where heterogeneous populations have two distinct mutation rates. During the initial rise in frequency, our analytical approximations ["Methods" section Equations (13)–(15)] show excellent agreement to results obtained by numerical iteration of the recursions, for both deleterious (upper panels) and beneficial (lower panels) mutants. At later times, the approximations break down for a beneficial mutant as it approaches fixation. This deviation occurs smoothly for fixed or non-inherited mutation rate, in which the population is well-mixed. In the perfectly inherited case, a shoulder appears in the frequency trajectory as the fittest mutant approaches fixation first in the subpopulation with higher mutation rate, followed by that with lower mutation rate.

During the early phase for which the approximations hold, double mutant frequency is higher when mutation rate is heterogeneous, for all choices of selection coefficients. Indeed, our analytical approximations indicate that the *absolute* increase in double mutant frequency compared with the homogeneous case is proportional to the variance in mutation rate, V, whereas the *relative* increase is proportional to the squared coefficient of variation, $c^2 = V/\langle U \rangle^2$ [compare Equations (14) and (15) to Equation (13)]. The magnitude of this increase is larger (for $t > 1$) when mutation rate is perfectly inherited than when it is non-inherited. However, while the non-inherited case falls substantially below the perfectly inherited case when single mutants are beneficial or mildly deleterious, the two cases become similar when single mutants are strongly deleterious. These effects will be clarified later ("Mutation-Selection Balance in the Two-Locus Model" section). Note that after a brief transient reflecting initial *de novo* production of the double mutant, its exponential rate of increase (slope in the log-linear plot) when beneficial is approximately the same across mutation models (fig. 2, lower panels). This reflects the fact that mutation rate determines the initial appearance of a mutant, but selection plays the predominant role in driving the rise of a beneficial mutant once it is present.

Except in very large populations, there will be a substantial and variable waiting time for the first appearance of a higher-order mutant. The initial dynamics can thus better be described by a stochastic model. For this purpose we develop a multi-type branching process model that captures mutation rate heterogeneity in an analogous way to the deterministic model (supplementary text II.3, Supplementary Material online). Numerical calculations yield the probability that a given mutant has appeared in an exponentially growing but finite population by a given time. For illustration we take a simple two-locus model in which individuals have a fixed total number of offspring, which mutate at either of two rates ("Methods—Branching Process Model" section). Fixing the mean, $\langle U \rangle$, fully determines the homogeneous case, while varying q_h (the probability or fraction of individuals with high mutation rate) and ρ (the ratio of the high to the low mutation rate) yields heterogeneous cases with a range of variances V.

Figure 3 illustrates the difference in waiting time for the first double mutant between the heterogeneous and homogeneous cases (specifically, the difference $\Delta T_{0.5}$ in time until the probability of appearance reaches 50%), as a function of q_h and ρ. In all cases, we observe that $\Delta T_{0.5} \leq 0$, indicating that heterogeneity accelerates the first appearance. The isoclines of $\Delta T_{0.5}$ closely match those of V, indicating that for fixed mean mutation rate, variance appears to determine the extent of the acceleration. The role of mutation rate inheritance is consistent with the deterministic model: when single mutants are neutral (upper panels), perfect inheritance yields a significant acceleration in the appearance of double mutants, whereas non-inheritance yields hardly any difference

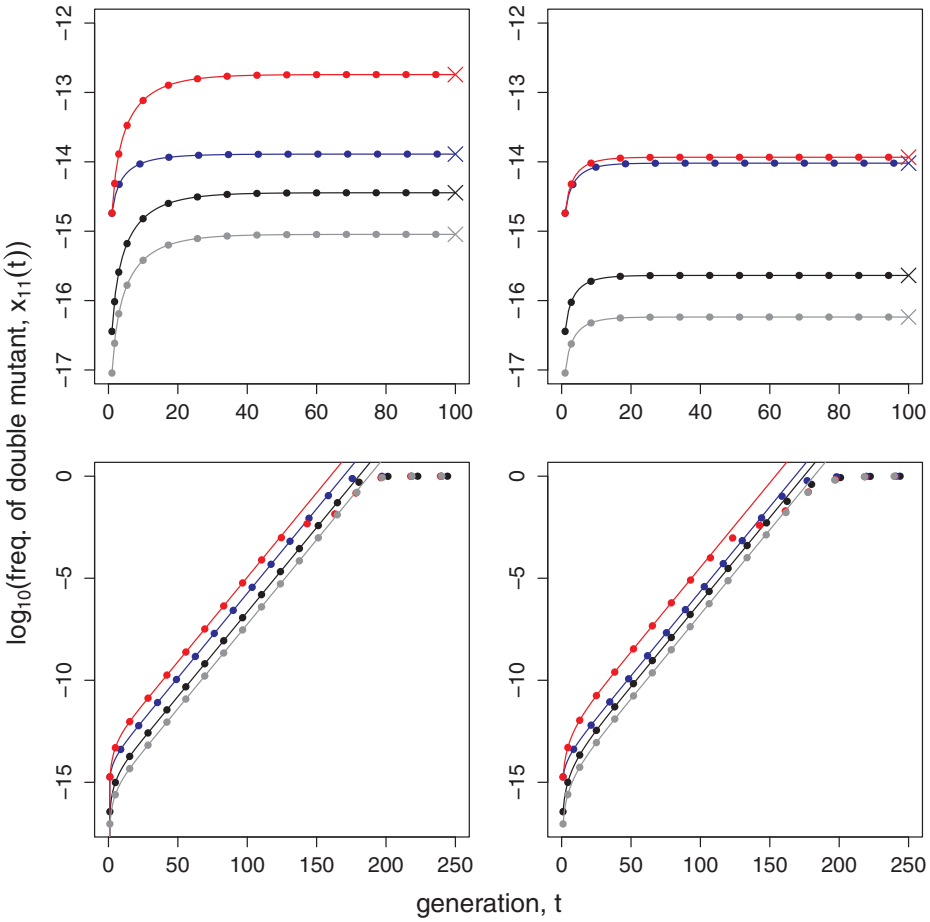

Fig. 2. Temporal dynamics of the double mutant. The population is initially composed entirely of the wild type. Mutation rate takes on either of two values: $U = u_\ell = 3 \times 10^{-9}$ with frequency 0.995 or $U = u_h = 6 \times 10^{-7}$ with frequency 0.005, thus $\langle U \rangle = 6.0 \times 10^{-9}$ and $V = 1.8 \times 10^{-15}$. The rates u_ℓ and u_h are parameterized approximately from *E. coli* normal and hypermutator (mismatch repair defective MutL$^-$ strain) mutation rates to rifampicin resistance (Lee et al. 2012), while the chosen hypermutator frequency maximizes c^2 ("Methods—Quantifying Effects with Two Mutation Rates" section). The selection coefficients vary across panels—*top left*: $s_{01} = s_{10} = 0.1$ and $s_{11} = 0.19$ (all mutants deleterious relative to the wild type; low-cost single mutants); *top right*: $s_{01} = s_{10} = 0.9$ and $s_{11} = 0.19$ (all mutants deleterious; high-cost single mutants); *bottom left*: $s_{01} = s_{10} = 0.1$ and $s_{11} = -0.21$ (deleterious single mutants, beneficial double mutant); *bottom right*: $s_{01} = s_{10} = -0.1$ and $s_{11} = -0.21$ (all mutants beneficial; double mutant fittest). Black shows the homogeneous case with mutation rate fixed to $\langle U \rangle$; blue shows the heterogeneous case with no inheritance; red shows the heterogeneous case with perfect inheritance; and grey shows the case when $U \equiv u_\ell$. Thus comparing black to grey indicates the effect of changing the mean mutation rate by adding a hypermutator, while comparing blue/red to black indicates the effect of increasing variance with fixed mean. The dots in each case show the result of numerically iterating the recursions, while the solid lines show the analytical approximation for the short-term temporal dynamics. When all mutants are deleterious, the "×" at the end of each curve shows the analytical approximation of the equilibrium.

from the homogeneous case. On the other hand, when single mutants produce no surviving offspring (lower panels), the acceleration due to heterogeneity is virtually identical in both cases of inheritance. Although we are limited to illustrating selected cases of the stochastic model numerically, we expect the reduction in waiting time due to heterogeneity to arise more generally, in correspondence with the deterministic model.

Recall that we neglected mutation rate evolution (when heritable) in order to clarify the effect of a fixed level of variation. Since indirect selection on mutation rate can only operate once mutations are present, it cannot affect the waiting time for the first mutation in a purely wild type population, but could conceivably affect the waiting time for multi-point mutants via selection on single-point mutants. However, when mutations are deleterious with selection coefficients that are large compared with the absolute differences in mutation rate among subpopulations, mutation rate evolution occurs on a much slower timescale than dynamics at the focal loci and can reasonably be neglected in the short term (supplementary text II.2.1 and II.2.5 and fig. S4, Supplementary Material online). If mutations are beneficial, indirect selection favoring higher mutation rates, given initial heritable variation, would accelerate the appearance of subsequent mutations beyond our prediction when holding the mutation rate distribution fixed. Of course, these issues do not affect the results when mutation rate is non-inherited.

Mutation-Selection Balance and Mutational Load

Finally, returning to our deterministic model, we analyze the mutation-selection balance achieved in a population in the long term when all mutants are less fit than the wild type.

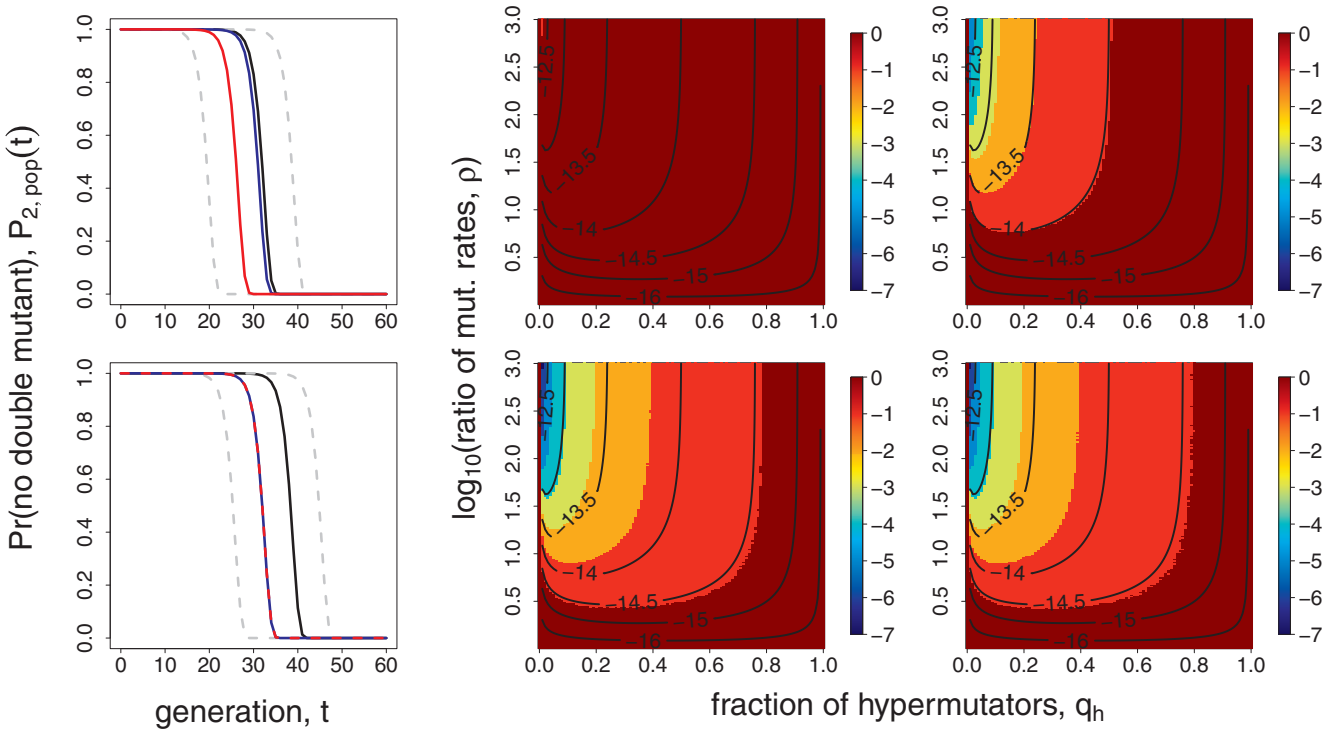

Fig. 3. Waiting time for first appearance of a double mutant in a branching process model. The population initially consists of $N = 100$ wild type individuals. Wild types have exactly two offspring; single mutants are neutral (top row) or lethal (bottom row). The mean mutation rate $\langle U \rangle$ is fixed to 10^{-7}, fully determining the homogeneous case, while the fraction/probability of high mutation rate (q_h) and the ratio of high to low mutation rate (ρ) define a range of heterogeneous populations. *Left*: the probability that a double mutant has not appeared in the population by generation t, $P_{2;pop}(t)$, when $q_h = 1/100$ and $\rho = 1000$ (yielding the maximal variance illustrated in the color plots). Black curve—homogeneous; blue—heterogeneous, no inheritance; red—heterogeneous, perfect inheritance. For comparison, the grey dashed curves show the results for a population consisting entirely of high- or low-mutators. The color plots (centre: no inheritance; right: perfect inheritance) indicate the difference in waiting time to reach 50% chance of appearance of a double mutant, $\Delta T_{0.5}$, for a heterogeneous versus homogeneous population, as a function of q_h and ρ. The resulting variance V in mutation rate is indicated by the black contour lines [labeled with $\log_{10}(V)$ values].

This amounts to finding the equilibrium solutions, x^*, to the genotype frequency recursions (or equivalently, the eigenvectors of the mutation-selection matrix M). For sufficiently large selection coefficients relative to mutation rates, as also required for the analytical approximations we make in the one- and two-locus models to hold, we again expect dynamics at the focal loci to equilibrate on a faster timescale than mutation rate evolution would occur in the inherited case (supplementary text II.2.1 and II.2.5, Supplementary Material online).

At a single focal locus, regardless of the inheritance assumption, the classic mutant frequency of u/s is simply replaced by $\langle U \rangle / s$ where $\langle U \rangle$ is the mean mutation rate in the population and s is the cost of the mutation. Population mean fitness at equilibrium is correspondingly given by $\bar{w}^* = 1 - \langle U \rangle$.

When multiple loci are involved, again higher moments come into play. We conduct an in-depth analysis of the two-locus case, while a brief consideration of an infinite-locus model confirms our key qualitative conclusions. Detailed mathematical results are given in supplementary text II.2, Supplementary Material online. As a short-hand to distinguish models under comparison, "het" will indicate a heterogeneous mutation rate characterized by a distribution, and "hom" a homogeneous mutation rate fixed to the mean of this distribution. Further, $H = 0$ will indicate that mutation rate is non-inherited and $H = 1$ will indicate perfect inheritance.

Mutation-Selection Balance in the Two-Locus Model

At equilibrium, double mutant frequency is increasing with variance and single mutant frequency is decreasing with variance, regardless of mutational costs. Interestingly, however, the effect on wild type frequency can go in either direction. Generally, mutation rate heterogeneity boosts the wild type when the double mutant is sufficiently costly relative to the single mutants, with the precise condition depending on the inheritance case (supplementary eq. S22 and S25, Supplementary Material online). This result can be explained intuitively by the competition exerted by mutants on the wild type. Mutation rate heterogeneity produces more double mutants at the expense of single mutants, so to yield a higher frequency of wild types, the double mutants must exert sufficiently less competition than the single mutants. Similarly, the effect of variance on the frequency of the mutant allele at any given locus and on the mean number of mutations per genome depends on the extent of epistasis (supplementary text II.2.3.4, Eq. S28, Supplementary Material online). Epistasis is positive when the double mutant is fitter than the

Table 1. Approximate Equilibrium Frequency of Double Mutants (x_{11}^*) in the Two-Locus Model.

	With Simultaneous Mutation	Without Simultaneous Mutation
Homog. mut. rate[a]	$\left(\frac{1}{s_{01}}+\frac{1}{s_{10}}-1\right)\frac{\langle U\rangle^2}{s_{11}}$	$\left(\frac{1}{s_{01}}+\frac{1}{s_{10}}-2\right)\frac{\langle U\rangle^2}{s_{11}}$
Heterog. mut. rate, no inheritance[b,c]	$\left(\frac{1}{s_{01}}+\frac{1}{s_{10}}-1\right)\frac{\langle U\rangle^2}{s_{11}}+\frac{V}{s_{11}}$	$\left(\frac{1}{s_{01}}+\frac{1}{s_{10}}-2\right)\frac{\langle U\rangle^2}{s_{11}}$
Heterog. mut. rate, perfect inheritance [b,d]	$\left(\frac{1}{s_{01}}+\frac{1}{s_{10}}-1\right)\frac{\langle U\rangle^2+V}{s_{11}}$	$\left(\frac{1}{s_{01}}+\frac{1}{s_{10}}-2\right)\frac{\langle U\rangle^2+V}{s_{11}}$

Selection coefficients $s_{01}, s_{10} > 0$ for single mutants, $s_{11} > 0$ for double mutants; cf. figure 4.
[a]Mutation rate fixed to $\langle U\rangle$; error $\mathcal{O}(\langle U\rangle^3)$, cf. supplementary Eq. S17, Supplementary Material online.
[b]Mutation rate distribution with mean $\langle U\rangle$ and variance V.
[c]Error $\mathcal{O}(\max(\langle U\rangle^3, \langle U\rangle V, V^2))$, cf. supplementary Eq. S21, Supplementary Material online.
[d]Error $\mathcal{O}(\max_k q_k u_k^3)$, cf. supplementary Equation S24, Supplementary Material online.

multiplicative expectation from the two single mutants' fitness values, and negative when the double mutant is less fit than this expectation. When epistasis is sufficiently positive (exceeding zero in the perfectly inherited case or exceeding a positive threshold in the non-inherited case, given by Equations S31 and S33, Supplementary Material online), the gain in double mutants outweighs the loss in single mutants, such that the mutant allele frequency and mean number of mutations show a net increase with variance. The magnitude of the variance effect on the frequencies of the wild type (order 1), single mutants (order $\langle U\rangle$), and mutant allele (order $\langle U\rangle$) is relatively small (supplementary fig. S1, Supplementary Material online), since $V < \langle U^2\rangle \ll \langle U\rangle$ for $U \ll 1$. Nonetheless, our reasoning points to effects that should also play out, potentially with larger magnitude, when more loci are under consideration.

Double mutant frequency (x_{11}^*) is of order $\max(\langle U\rangle^2, V)$ and may thus be substantially affected by variance. Analytical approximations of x_{11}^* under each model case are summarized in table 1. Although these approximations can break down for extreme mutation rate distributions, particularly when at least one selection coefficient is small (not illustrated), we find a good match to numerical results over a wide parameter range (e.g., top panels of fig. 2) and thus we will base our following analysis on these approximate solutions.

The absolute increase in the equilibrium double mutant frequency is proportional to the variance of the mutation rate distribution (V) and the relative increase is proportional to the squared coefficient of variation (c^2), as also found for the short-term frequency. The form of the proportionality constants, in particular the roles of mutation rate inheritance and selection coefficients, can be understood by considering how the pathways generating the double mutant are affected by heterogeneity in the mutation rate. The intuitive explanations also apply to the short-term dynamics, but can most readily be seen from the equilibrium expressions.

In the absence of inheritance, the mutation rate experienced by multiple loci on a genome only shows an association over one generation. Any boost in multi-point mutant frequency due to heterogeneity must thus be achieved through a boost in simultaneous mutations. Indeed, if we modify our model to disallow simultaneous mutations (supplementary text II.2.3.5, Supplementary Material online), we find that the frequency of double mutants is no different to the homogeneous case (right-hand column of table 1 and dashed curves in fig. 4). Conversely, in the full model, the absolute increase in double mutant frequency

$$\Delta^*_{\text{abs}}(H=0) = V/s_{11} \tag{2}$$

is due entirely to the increased influx of simultaneous double mutations, which are filtered with selection coefficient s_{11}, and independent of the selection coefficients of the single mutants (s_{01} and s_{10}). On the other hand, the relative increase in double mutant frequency

$$\Delta^*_{\text{rel}}(H=0) = c^2\left(\frac{1}{s_{01}}+\frac{1}{s_{10}}-1\right)^{-1} \tag{3}$$

increases with s_{01} and s_{10}, because the less fit single mutants are, the more relatively important simultaneous double mutation becomes. In figure 4, we see that $x_{11}^*(\text{het}, H=0)$ approaches $x_{11}^*(\text{hom})$ as the single mutant cost approaches zero, but shows a growing gap above the homogeneous case as the single mutant cost increases.

When mutation rate is inherited, correlations arise across generations and the accumulation of multiple mutations can be boosted not only by simultaneous acquisition, but also by stepwise acquisition over several generations. Perfect inheritance manifests itself as distinct subpopulations with differing fixed mutation rates, but the same action of selection. Thus heterogeneity affects all mutational pathways equally, and double mutant frequency in the total population is simply scaled up by a constant factor, independent of selection coefficients:

$$\Delta^*_{\text{rel}}(H=1) = c^2 \tag{4}$$

Plotting x_{11}^* on a log scale as a function of single mutant cost, this effect manifests itself as parallel curves in the homogeneous and heterogeneous cases (fig. 4). The absolute difference, on the other hand, is decreasing with all selection coefficients:

$$\Delta^*_{\text{abs}}(H=1) = \left(\frac{1}{s_{01}}+\frac{1}{s_{10}}-1\right)\frac{V}{s_{11}} \tag{5}$$

When single mutants are sufficiently fit, double mutants are mainly generated by stepwise accumulation of mutations, and blocking simultaneous mutations has little effect; however, when single mutants are very costly, blocking simultaneous mutations has a drastic effect. As $s_{01} = s_{10} \to 1$, the non-inherited and perfectly-inherited cases converge, since all double mutants must be generated directly by simultaneous mutation from the wild type.

The precise ranking of x_{11}^* across model cases is fully determined by the selection coefficients of the single mutants and c^2 of the mutation rate distribution (supplementary text

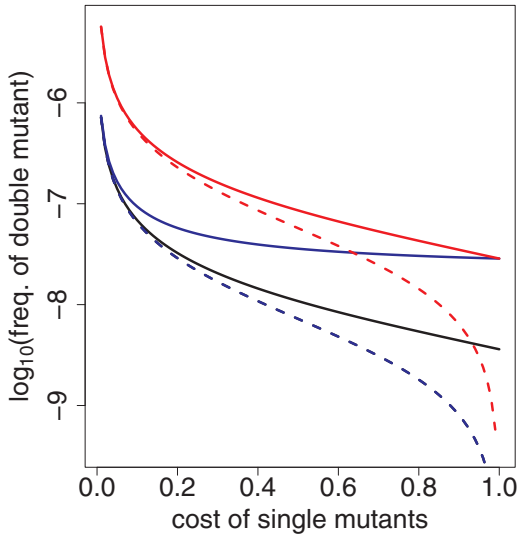

Fig. 4. Frequency of the double mutant at mutation-selection balance. The analytical approximations for the equilibrium double mutant frequency in the two-locus model, x_{11}^* (table 1), are plotted as a function of single mutant cost for each mutation model: black—homogeneous; blue—heterogeneous, no inheritance; red—heterogeneous, perfect inheritance. Simultaneous mutations are allowed (solid lines) or blocked (dashed lines; blue and black overlap). $\langle U \rangle = 1.9 \times 10^{-5}$ and $V = 2.5 \times 10^{-9}$ as in figure 1. The double mutant cost is fixed to $s_{11} = 0.1$, but does not affect relative differences among mutation models, since $x_{11}^* \propto 1/s_{11}$ in all cases.

II.2.3.6, Supplementary Material online). Importantly, the above reasoning implies that double mutant frequency is expected to increase with the degree of inheritance, which more generally falls on a continuum. Thus, the $(1 + c^2)$-fold increase over the homogeneous case obtained with perfect inheritance in fact provides an upper bound on the effect of mutation rate heterogeneity across all choices of selection coefficients and inheritance assumptions. This upper bound holds not only at equilibrium, but also during the short-term dynamics in the deterministic model.

Mutation-Selection Balance in an Infinite-Locus Model

Although an analysis of increasingly many loci with arbitrary selection coefficients becomes infeasible, we can gain some insights into the general behaviour at many loci by considering an infinite-locus model in which fitness is simply determined by number of mutations.

In the particularly simple case where each mutation has cost s with multiplicative fitness effects, under a fixed genomic mutation rate λ, the equilibrium mutant frequencies are Poisson-distributed with mean λ/s (Haigh 1978). In this case, we find similar effects of heterogeneity on the equilibrium distribution as on the mutational production in a single generation: the frequency of zero- or possibly few-point mutants, as well as many-point mutants, increases, while that of intermediate mutants decreases (supplementary text II.2.4, Supplementary Material online and fig. 5). The switching points in the directional effect of heterogeneity shift to higher mutant classes as s decreases. Among the higher-order mutants, the relative magnitude of this boost in frequency is again increasing with the number of mutations. The mean number of mutations per genome is unchanged by heterogeneity with perfect inheritance, but reduced in the case of no inheritance. Since there is no epistasis in this model, these results are consistent with the two-locus results.

Clearly, increasing the variance of the mutation rate distribution generally increases the effects of heterogeneity (supplementary fig. S2, Supplementary Material online). However, the differences in mutant frequencies are not directly proportional to variance, since all higher moments of the mutation rate distribution now also play a role. Consistent with the intuition developed for the two-locus results, heterogeneity has a greater impact on the mutant frequency distribution when the mutation rate is inherited: more precisely, there is a larger increase in the frequency of the lowest- and highest-order mutants, and correspondingly larger decrease in the frequency of intermediates. Again, the relative importance of inheritance varies with the strength of selection: the non-inherited case is similar to the homogeneous case when s is small, and becomes more similar to the perfectly inherited case as s increases (supplementary fig. S2, Supplementary Material online).

If we allow epistatic fitness effects, it is possible to find more complex patterns in the mutant frequency distribution, though we are generally limited to numerical investigations. In particular, it is possible to find examples where the wild type frequency is instead decreased by heterogeneity; where the directional effect of heterogeneity switches more than twice over the mutant classes; and where the mean number of mutations per genome is either increased or decreased (supplementary fig. S3, Supplementary Material online). Nonetheless, the frequency of sufficiently high-order mutants always appears to be boosted by heterogeneity.

These results should be taken with the caveat that for realistic mean genomic mutation rates and the levels of variance considered here, evolution of the mutation rate in the perfectly inherited case, if subpopulations were to compete with one another, could occur on the same timescale as evolution at the focal loci (supplementary text II.2.5 and fig. S5, Supplementary Material online). Thus, heritable variation in mutation rate would be lost and the population would approach an equilibrium determined by the lowest available mutation rate. For consistency in our modeling approach, such changes in subpopulation frequencies are nonetheless neglected in our main results.

Mutational Load

The reduction in mean fitness compared with a purely wild type population due to the production of deleterious mutants, i.e., $1 - \bar{w}$, is known as the "mutational load" (Bürger 2000, p. 105). We find that heterogeneity in mutation rate always reduces the mutational load at equilibrium.

Specifically, in the two-locus model, the equilibrium population mean fitness is given by:

$$\bar{w}^* = \langle (1-U)^2 \rangle = (1 - \langle U \rangle)^2 + V \qquad (6)$$

regardless of whether mutation rate is inherited (supplementary text II.2.3, Supplementary Material online). Thus,

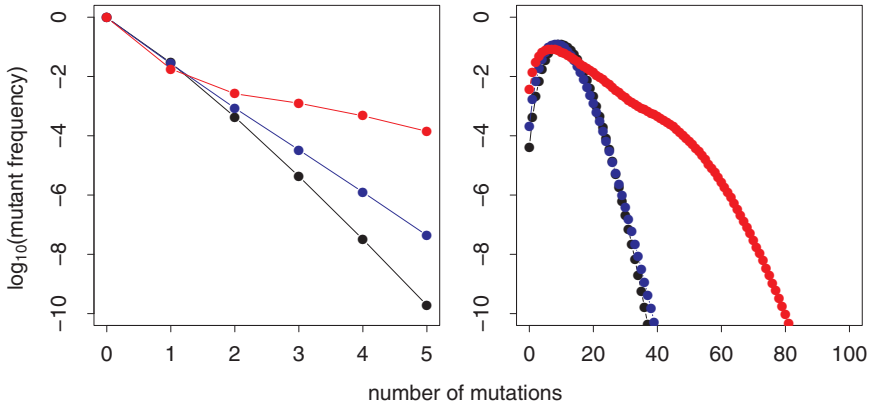

Fig. 5. Equilibrium mutant frequencies in an infinite-locus model with no epistasis. The genotype frequency at the numerically determined equilibrium is plotted as a function of number of mutations carried, in each model case (black—homogeneous; blue—heterogeneous, no inheritance; red—heterogeneous, perfect inheritance). The cost per mutation is $s = 0.1$. *Left*: Per-genome mutation rate Λ takes on two values, 0.0015 with frequency 0.99 or 0.15 with frequency 0.01. Mean = 0.0030 (bacteria-like; Drake et al. 1998), variance = 2.2×10^{-4}. The mean number of mutations per genome at equilibrium is $m^*(\text{hom}) = m^*(\text{het}, H = 1) = 0.0298$ and $m^*(\text{het}, H = 0) = 0.0289$. *Right*: Λ is given by 1000 draws from a log-normal distribution. Sample mean = 1.01 (RNA-/retrovirus-like; Drake et al. 1998), sample variance = 0.29, range = 0.19–4.2. Here, $m^*(\text{hom}) = m^*(\text{het}, H = 1) = 10.1$ and $m^*(\text{het}, H = 0) = 9.0$.

heterogeneity decreases mutational load by an amount equal to the variance of mutation rate in the population. This effect is relatively small in magnitude ($\bar{w}^*$ is of order $\langle U \rangle \gg V$), but in a consistent direction.

In the infinite-locus model in which fitness is determined by number of mutations (with arbitrary costs relative to the wild type), the equilibrium population mean fitness is given by

$$\bar{w}^* = \langle e^{-\Lambda} \rangle \tag{7}$$

again regardless of inheritance (supplementary text II.2.4, Supplementary Material online). An application of Jensen's Inequality demonstrates that $\bar{w}^*$ is enhanced by mutation rate heterogeneity. Since Λ is not necessarily $\ll 1$ (Drake et al. 1998), this effect may be non-negligible: for example, with the log-normal mutation rate distribution used in figure 5 (right), $\bar{w}^*$ is increased from 0.364 to 0.408. (In the perfectly inherited case, this points to the strong selection on mutation rate that could be anticipated, as mentioned above, when deleterious mutations are occurring at many loci.)

It is not immediately obvious that heterogeneity should consistently increase mean fitness. Transiently, mean fitness can indeed be either increased or reduced by heterogeneity, depending on the initial genotype frequencies, the relationships among selection coefficients, and the inheritance assumption. Furthermore, as seen in the previous section, heterogeneity does not necessarily increase the equilibrium frequency of mutation-free genomes, nor does it necessarily reduce the total number of mutations carried by the population. The result also cannot be explained simply by heterogeneity providing more material on which selection can act, because higher variance in the number of mutations does not necessarily lead to higher variance in fitness. (As a simple counterexample, suppose that single mutants are very costly while double mutants are nearly as fit as the wild type. Then mutation rate heterogeneity that produces more wild types and double mutants will reduce the population's variance in fitness.) In fact, at equilibrium, population mean fitness is independent of selection coefficients and only depends on mutational production. Heterogeneity then seems to have the advantage of clustering deleterious mutations and producing more mutation-free genomes in any given generation.

Discussion

The critical role of mutations in producing the raw material for evolution has long been recognized by biologists and mathematically analyzed by population geneticists. In the vast majority of analyses, mutation rate is assumed to be constant. However, given the available evidence for genetic, environmental, and random physiological influences on mutation rate, there is a strong case to suggest that a constant mutation rate is the exception rather than the norm. We therefore investigated the population genetic consequences of heterogeneity in mutation rate, and found that it will generally increase the frequency of higher order mutants and reduce long-term mutational load relative to a homogeneous population with the same mean mutation rate. However, the magnitude of mutation rate heterogeneity under natural conditions and hence its typical significance for evolution is largely unknown. We thus proceed to discuss empirical approaches to tackle this question.

The Effect of Heterogeneity on Mutant Frequencies

Mutation rate heterogeneity promotes the production of multi-point mutants. This effect of variability itself is to be distinguished from the effect of simply raising mutation rate in all or a subset of individuals: while the latter increases the frequency of all mutants, the former disproportionately increases the frequency of multi-point mutants relative to single-point mutants. This effect is fairly intuitive and has occasionally been pointed out in the literature (Ninio 1991; Drake et al. 2005; Drake 2007; Elez et al. 2010). Clearly, a heterogeneous mutation rate will also increase the chance

of no mutation. With a more general model and rigorous analysis, we could give precise conditions for the threshold numbers of mutations at which heterogeneity has increasing versus decreasing effects, depending on the population's range of mutation rates ("Probability of Simultaneous Mutations on a Genome" section).

Previous calculations of the contribution of mutators to single- and multi-point mutations have usually considered only a single generation (Ninio 1991; Cairns 1998; Drake et al. 2005; Gonzalez et al. 2008). It is less obvious how heterogeneity will affect mutant frequencies over multiple generations, in which mutations can accumulate both simultaneously and sequentially, and their frequencies are modified by selection. In a two-locus model, we found that heterogeneity accelerates the first appearance of a double mutant, regardless of whether mutations are beneficial or deleterious ("Initial Emergence of Mutants" section). At mutation-selection balance, when all mutants are less fit than the wild type, heterogeneity boosts the frequency of the double mutant in the two-locus model, and generally increases the frequency of higher order mutants, with increasingly large relative effects, in an infinite-locus model ("Mutation-Selection Balance and Mutational Load" section). However, the subtle interplay of altered mutational production and selection can yield more complex effects of heterogeneity on the wild type and lower order mutants. Roughly speaking, when the over-produced (higher order) mutants are sufficiently fit relative to the under-produced (lower order) mutants, the increased competition faced by the wild type can outweigh the increased chance of mutation-free reproduction, such that the net effect of heterogeneity is to reduce wild-type frequency. Likewise, depending on the extent of epistasis, the total number of mutations carried by the population at equilibrium may be increased or decreased by heterogeneity, even though the mean number of *de novo* mutations produced in each generation is unchanged. Though the magnitude of these effects is small when considering only two loci, the intuition developed here should carry over to more loci, where effect sizes could be larger.

Our analytical approximations clarify how these effects depend on moments of the mutation rate distribution and selection coefficients. We could thus generalize conclusions to an arbitrary distribution of mutation rate (not limited to two distinct values) and separate the effect of changing mean mutation rate from variability itself. The dynamics of mutants at n focal loci are driven by the first n moments of the mutation rate distribution. Thus if examining only one locus, the population mean mutation rate is sufficient to predict mutant dynamics, but to predict the joint dynamics at two loci, variance must be considered, and so on. We analyzed the two-locus case in detail and found that the frequency of double mutants in a deterministic model is boosted both in the short term and, for deleterious mutations, at equilibrium. The absolute increase in frequency is proportional to the variance, while the relative increase is proportional to, and at most equal to, the squared coefficient of variation. In a stochastic model, the reduction in waiting time for the first double mutant was likewise determined by variance. Since variance can mathematically be of comparable or even larger order than the squared mean mutation rate, these effects can potentially be substantial. Furthermore, we predict that mutation rate heterogeneity becomes increasingly significant as more loci are taken into account.

These results make it possible to quantify the effect of heterogeneity, given a distribution of mutation rate. Even if selection coefficients and inheritance patterns are unknown, in the two-locus deterministic model we have an upper bound on the relative increase in double mutant frequency given by the squared coefficient of variation (c^2) of the mutation rate distribution. However, very few studies to date have detected and quantified mutation rate heterogeneity. In one study with budding yeast, a model in which 35% of replications are "hypomutator" ($\sim 4 \times 10^{-8}$ mutations/bp/cell division) and 65% "hypermutator" ($\sim 4 \times 10^{-7}$ mutations/bp/cell division) provided the best fit to experimental data (Kennedy et al. 2015). This distribution has $c^2 = 0.39$ ("Methods—Quantifying Effects with Two Mutation Rates" section); thus, we predict a modest increase in the frequency of double mutants at two focal loci of at most 39%. Nonetheless, stronger genetic hypermutators are known to exist (supplementary text I.1) and could be present at highly variable frequencies in populations, depending on selective conditions and timing (Mao et al. 1997; Boe et al. 2000; Desai and Fisher 2011). For example, consider *E. coli* hypermutators with 200-fold elevated mutation rate (Lee et al. 2012). If present at 0.5% frequency, they would increase the population's mean mutation rate by less than 2-fold, and thus increase the frequency of double mutants by less than 4-fold via the mean, but up to $\sim$51-fold further through variance itself at fixed mean (cf. fig. 2). Although this choice of hypermutator frequency maximizes relative effect size [Equation (25)], taking hypermutators at a frequency of only 5×10^{-4} (Ninio 1991) still yields up to $\sim$17-fold increase in double mutant frequency due to variance. It is thus plausible that mutation rate heterogeneity plays a significant role in the acquisition of multi-point mutations in some populations, though the typical situation remains unclear.

Our analysis also elucidated the role of inheritance of mutation rate, by comparing cases where an individual's mutation rate is either identical to its parent's or drawn independently at random, while the population-level distribution is the same. More generally, the extent of parent-offspring correlation could fall on a broad spectrum between these two extremes, yielding intermediate effects of mutation rate variation. The more strongly mutation rate is correlated through a lineage, the greater the effect of variation, since stepwise as well as simultaneous accumulation of mutations can be boosted. A previous study concluded that simultaneous mutations make a negligible contribution to multilocus adaptation (Lynch and Abegg 2010), but only tested cases where single mutants were neutral or slightly deleterious. However, if intermediates are highly deleterious, simultaneous mutations play a crucial role. In this case, the extent of mutation rate inheritance is unimportant, so long as the population exhibits variability.

While we modeled asexually reproducing populations, we expect qualitatively similar but weaker effects to hold in sexual populations. Recombination can bring together mutations generated in different lineages, thus reducing the importance of hypermutators in accelerating the first appearance of multi-point mutants. In the longer term, recombination would counteract positive linkage disequilibrium, reducing the excess of multiple mutants generated by mutation rate heterogeneity.

Evolutionary Consequences

Given these findings, mutation rate heterogeneity could clearly play a role in multi-locus adaptation. We are aware of only two previous models of adaptation from *de novo* mutations under more specific forms of mutation rate heterogeneity. Aoki and Furusawa (2001) found through stochastic simulations that having two distinct, non-inherited mutation rates (versus mutation rate fixed to the mean) promotes adaptation on a rugged fitness landscape, but lacked analytical expressions to elucidate parameter effects. Lynch and Abegg (2010) developed analytical approximations for the waiting time until appearance of multi-point mutants that ultimately fix, which included contributions from genetic or transient hypermutators. Our findings suggest that the accelerated appearance of double mutants is a general consequence of mutation rate heterogeneity, whether or not differences are heritable. Thus, in particular, heterogeneity could help populations to cross fitness valleys, a classic problem in evolutionary biology (Wright 1932). By clustering mutations onto the same genetic background, it should also shift the regime in which clonal interference operates (cf. Desai et al. 2007). Nonetheless, the magnitude of these effects depends on the actual mutation rate distribution. Indeed, Lynch and Abegg (2010) found a limited role for mutators under their parameterization of frequency and effect size.

An important caveat in our discussion of adaptation at focal loci is that we have neglected deleterious mutations occurring at other loci; the availability of such mutations could imply that indefinitely increasing variance does not actually maximize overall adaptation rate. Lynch and Abegg (2010) modeled the cost of deleterious background mutations phenomenologically and concluded that it inhibited genetic mutators from making a significant contribution to adaptation. On the other hand, a non-heritably elevated mutation rate would reduce the accumulation of both beneficial and deleterious mutations relative to the heritable case. Altogether, similar to considerations for fixed mutation rate (Sniegowski et al. 2000), the level of variance maximizing adaptation will likely depend on the balance of beneficial and deleterious mutations available. Furthermore, on a rugged fitness landscape, Aoki and Furusawa (2001) observed a breakdown in the advantage of heterogeneity when mean mutation rate was high and the proportion of hypermutators was low, which forced all mutations into very few individuals. This finding suggests that optimal variance will also depend on mean mutation rate. Clarifying these effects would provide an interesting direction for further analytical work.

Our analysis of equilibrium as well as transient dynamics appears to be novel in the context of mutation rate heterogeneity, and points to implications for adaptation from standing genetic variation as well as from *de novo* mutations. We found that costly multi-point mutants are harbored at a higher long-term frequency in a heterogeneous population. If the environment changes and these genotypes become favorable, they could make an important contribution to adaptation (Barrett and Schluter 2008), subject to the same caveats regarding background mutations that remain deleterious.

In an applied context, these findings call for caution in the use of typical mutation rate estimates. According to our single-locus results, scoring a phenotype that can be conferred by a single point mutation will yield an estimate of the population's mean mutation rate, even if the rate varies among individuals. Similarly, pooling sequenced individuals and taking the total number of point mutations divided by the total number of examined sites yields an estimate of the mean mutation rate. While these estimates are reasonable in themselves, naive extrapolations assuming this mutation rate to be fixed will underestimate the chance of overcoming higher genetic barriers if the population is actually heterogeneous. This issue is particularly concerning for analyses of the likelihood that multi-drug resistant pathogens or cancerous cells "pre-exist" or are generated during drug treatment (Ribeiro and Bonhoeffer 2000; Komarova and Wodarz 2005; Colijn et al. 2011). In addition, while the significance of generally elevated mutation rate in cancerous cells has previously been pointed out (Loeb et al. 1974; Loeb 1991, 2001; Loeb et al. 2003), our results suggest that, even when controlling for changes detected in mean mutation rate, progression to cancer via accumulation of multiple mutations may occur faster than expected.

Finally, in a well-adapted population with only deleterious mutations available, mutation rate heterogeneity reduces long-term mutational load. We demonstrated analytically that this effect is quite general and independent of mutation rate inheritance. Taken together with the enhanced frequency of higher-order mutants, then, inter-individual heterogeneity allows a population both to maintain higher mean fitness and to explore genotype space more widely, which could be advantageous for facing future adaptive challenges. This possibility has previously been suggested verbally (Furusawa and Doi 1992; Drake et al. 2005; Combe et al. 2015), and uncovered in simulations of evolution on a rugged fitness landscape, in which heterogeneous populations tended to more efficiently find and remain on global fitness peaks (Aoki and Furusawa 2001). More broadly, other forms of heterogeneity—including variation of mutation rate through time or across the genome—have also been suggested to balance evolvability against fitness loss through deleterious mutations (Sniegowski et al. 2000; Galhardo et al. 2007).

These findings suggest that mutation rate variance, if determined by heritable mechanisms, could itself evolve. In the long term in a constant environment, where all mutations are expected to be deleterious, increasing variance will always

benefit the population by reducing mutational load. In the shorter term or in a changing environment, increasing variance (perhaps only up to an intermediate optimum) could have benefits by accelerating adaptation. However, we expect stronger selection on mutation rate itself as an individual trait, than on the mutation rate distribution as a population-level trait; a formal analysis of their joint evolution is beyond the scope of this study.

Interestingly, recent work has suggested evolutionary benefits of a mutation rate that depends on fitness (Ram and Hadany 2014; Belavkin et al. 2016), which implicitly induces inter-individual heterogeneity, but contrasts with our assumption that fitness does not affect mutation rate, and implies that the mutation rate distribution will change over time according to population composition. Furthermore, the identified evolutionary advantage is relative to a population in which mutation rate is fitness-independent and thus uniform, but not necessarily with the same mean. Due to these fundamental model differences, the advantages attributed to fitness-dependent mutation rates cannot be unambiguously extended to inter-individual heterogeneity in general. A better understanding of the potential evolutionary benefits provided by diverse forms of variability in mutation rates remains open to future work.

Empirical Approaches

Currently, empirical estimates of mutation rate distributions in populations are almost completely lacking. However, we see considerable potential to employ both well-established and novel techniques to quantify the extent of mutation rate heterogeneity due to either systematic or random effects.

If mutation rate is systematically affected by a particular variable (genetic or environmental), one can indirectly quantify its distribution in a population by combining an estimate of fixed or mean mutation rate under each condition (using standard methods, e.g., fluctuation tests) with a measurement or model of how this variable is distributed. Although many factors affecting mutation rate are qualitatively well established (supplementary text I, Supplementary Material online), estimates suitable for parameterizing population genetic models are often lacking. It has been a common practice, especially in past studies of stress-induced mutagenesis and characterizations of natural isolates, only to report the mean frequency of mutants after culture growth, as opposed to mutation rate per generation (see also Rosche and Foster 2000; MacLean et al. 2013). Thus, standard methods are far from exhausted in investigating relationships between mutation rate and relevant genetic and environmental variables in populations.

Another approach to quantifying mutation rate heterogeneity (from any source, including random effects) is to examine whether mutant counts or frequencies in a population deviate from the expectation under homogeneity. The data could come from experiments not originally conducted for this purpose, but should provide the following features: (1) cover a sufficiently long stretch of the genome (relative to mutation rate) such that multi-point mutants are likely to be found; (2) preserve linkage information among these mutated sites; (3) provide a sufficiently large, representative sample of individuals, or highly resolved population-level mutant frequencies; and (4) be conducted under conditions in which selection can essentially be ruled out. The last point arises because it is problematic to distinguish mutation rate heterogeneity from positive epistasis, as both yield an overrepresentation of multi-point mutants (supplementary text II.2.4.4, Supplementary Material online).

Thus, while models of a few loci of interest under selection are valuable for predicting evolutionary dynamics, as we did here, they are less useful for data analysis aiming to detect mutation rate heterogeneity. For this purpose, the most straightforward approach is to use experimental protocols minimizing selection and examine the longest possible stretch of the genome, pooling individuals simply according to number of mutations. Then an infinite-locus model with neutral mutations is appropriate for determining the expected distribution of mutant counts. Under the null model of fixed mutation rate, clearly a single round of replication will yield Poisson-distributed mutant counts ("Methods—Occurrence of Mutations" section, and Drake et al. 2005; Drake 2007; Elez et al. 2010; Kennedy et al. 2015). Actually, even over multiple generations, a Poisson distribution arises if mutations are neutral, or under the unrealistic restriction that they have equal and multiplicative fitness effects, but not when they exhibit epistasis (supplementary text II.2.4.4, Supplementary Material online).

An analysis of published mutant collections from a wide range of taxa found that higher-order mutants were often over-represented relative to a Poisson distribution, which the authors interpreted as evidence of mutation rate heterogeneity (Drake et al. 2005; Drake 2007). However, the data came from a wide range of experimental systems and it was unclear whether selection could be ruled out in all cases. Most of the experiments also relied on reporter genes, which have several pitfalls and are increasingly being replaced by whole genome sequencing (Long et al. 2016). The intriguing finding by Drake and colleagues thus encourages follow-up studies that carefully consider confounding factors and statistically quantify the extent of heterogeneity.

At first glance, the abundance of recent studies applying population-level deep sequencing appears to offer a wealth of genotype frequency data to test for mutation rate heterogeneity. However, selection will generally be acting in these populations; indeed the focus of many such studies is to provide insights into adaptation (e.g. Lang et al. 2013). Furthermore, the high error rates of next-generation sequencing (NGS) technologies (Beerenwinkel and Zagordi 2011) limit the detection of very rare mutations before their frequencies have been increased by positive selection. Promisingly, the recently developed "maximum-depth sequencing" (MDS) method has sufficiently high yield and low error rate to estimate *de novo* mutation rates at specific target loci (Jee et al. 2016). However, both standard NGS methods (Beerenwinkel and Zagordi 2011) and, so far, the MDS method (Jee et al. 2016) operate by sequencing short reads of at most a few hundred nucleotides. This eliminates linkage information between all but the closest sites and

thereby makes the chance of detecting multi-point mutants vanishingly small. Global haplotype reconstruction algorithms exist, but require sufficient genetic diversity to align overlapping reads (Beerenwinkel and Zagordi 2011). On the other hand, barcoding single DNA molecules before sequencing avoids loss of linkage information, with novel techniques attaining high throughput (Borgström et al. 2015). For the sequencing step, standard short-read methods (Borgström et al. 2015) or potentially MDS (Supplementary Information in Jee et al. 2016) can be adopted. Therefore, if selection in the source population of interest can be avoided, these new developments in sequencing methods appear to have potential to generate useable data to test for mutation rate heterogeneity through the frequencies of multi- versus single-point mutants.

Alternatively, techniques isolating individual cells and pinpointing mutations in single replications have recently been applied to investigate mutation rate heterogeneity in samples that are unbiased by selection against all but perhaps lethal mutations. These approaches enable enumeration of *de novo* mutations either by fluorescently labelling nascent mutation foci (in *E. coli*; Elez et al. 2010) or by comparing whole genome sequences of mother and daughter cells (in budding yeast; Kennedy et al. 2015), with the potential for extension to other organisms. The biggest hurdle is the cost and labor involved in obtaining a large enough sample size.

In parallel with experimental developments, rigorous statistical methods will be crucial. Ideally, empirical studies should not only detect deviations from the null model of fixed mutation rate, but also fit an alternative model to quantify a mutation rate distribution (cf. Kennedy et al. 2015). Formal power analyses, identifying the minimal statistically detectable effect size for a given sample size, would be valuable both to interpret existing data and to prospectively aid experimental design.

Importantly, factors besides inter-individual heterogeneity could also cause deviations from the null model. Besides the aforementioned issue of selection, other forms of mutation rate variation could conceivably affect mutant counts. Mutation rate that changes over time will not by itself yield deviations from a Poisson distribution, as long as population members within each generation have the same rate (supplementary text II.2.4.4, Supplementary Material online). Mutation rate that consistently varies across genome sites (Rogozin and Pavlov 2003; Lang and Murray 2008) will lead to an over-representation of intermediate mutants, i.e., an effect opposite to variation among individuals with uniformity across a genome (supplementary text II.2.4.4, Supplementary Material online). On the other hand, if a single mutational event introduces changes at multiple nearby nucleotides (Averof et al. 2000; Schrider et al. 2011), clearly multi-point mutants will be over-represented, though in contrast to our model, with non-random spacing. For the purposes of analyzing mutant counts, proximal multi-nucleotide changes may simply be counted as single "mutations" (Drake et al. 2005; Drake 2007). On the other hand, going beyond our present model, mechanisms yielding inter-individual heterogeneity may act only locally in the genome, and using information on mutation spacing could yield insights into these underlying processes (Hill et al. 2004). Generally, data analyses should consider this (non-exhaustive) range of alternative models, perhaps applying formal model selection, in order to draw robust conclusions regarding mutation rate heterogeneity.

Conclusions

Novel technologies and the falling cost of genome sequencing are opening exciting new avenues to test proposed mutational models and quantify the extent of mutation rate heterogeneity, in natural isolates as well as laboratory strains. These empirical advances, combined with statistical methods, could enable parameterization of models that evaluate the evolutionary consequences of this variability. Our findings suggest that variability of mutation rate among population members could enhance multi-locus adaptation both from beneficial *de novo* mutations and previously deleterious mutations in the standing genetic variation, as well as reducing the mutational load in a well-adapted population. The vast majority of population genetic analyses, fixing mutation rate to an estimate representing the mean, may thus underestimate the potential for adaptation.

Methods

We model a haploid, asexually reproducing population with non-overlapping generations, extending classic population genetic models to incorporate a mutation rate that varies among co-existing population members. We focus on genotype dynamics at one or more fitness-determining loci, and assume throughout that mutation rate neither depends on the genotype at the focal loci, nor has any direct fitness effect. There are, however, notable cases in which this assumption will not hold; for example, a correlation between replicative fitness and mutation rate, due to mutual dependency on speed of replication, has been demonstrated in viruses (Furió et al. 2005; Dapp et al. 2013), and environmental factors could affect both fitness and mutation rate.

Numerical results and plots were generated with R (R Core Team 2015), and some analytical results were derived with use of Mathematica (Wolfram Research, Inc. 2016).

Occurrence of Mutations

We assume that mutation rate is uniform across loci within one individual, and that given this rate, mutations occur independently among loci. We do not distinguish different mutant alleles at one locus, and we neglect back mutations. Then among n non-mutant loci in an individual with realized mutation rate u, the number of new mutations that arise "simultaneously" (i.e., in one generation) follows a binomial distribution, where the probability of j mutations is:

$$p_{nj}(u) = \binom{n}{j} u^j (1-u)^{n-j} \tag{8}$$

In the limit as $n \to \infty$ and $u \to 0$ such that $nu \equiv \lambda$, we obtain an "infinite-locus" model in which every new mutation

occurs at a unique site. Then the number of mutations occurring in an individual with per-genome mutation rate λ follows a Poisson distribution; that is, the probability of j mutations is:

$$p_j(\lambda) = e^{-\lambda}\lambda^j/j! \quad (9)$$

This model does not address the complexity of multi-step intracellular replication cycles in viruses, for which a Poisson-distributed number of mutations per genome is not necessarily expected after a single infection cycle (Duffy et al. 2008; Sanjuán et al. 2010).

Deterministic Model of Genotype Frequency Dynamics

We denote the frequency of genotype i at generation t by $x_i(t)$ and its relative fitness by w_i. Without loss of generality we take the wild type (carrying no mutations) to have relative fitness 1, while type i has $w_i = 1 - s_i$ (thus $s_i > 0$ for a deleterious mutant and $s_i < 0$ for a beneficial mutant). Population mean fitness is given by

$$\bar{w}(t) := \sum_i w_i x_i(t). \quad (10)$$

Generally, then, for any collection of types i and proportions p_{ij} of type i offspring from a type j parent, one can write a set of recursions:

$$x_i(t+1) = \sum_{\forall j} \frac{w_j p_{ij} x_j(t)}{\bar{w}(t)} \quad (11)$$

where census occurs after mutation and before selection. For a finite collection of types, these recursions can be rewritten as a matrix equation:

$$x(t+1) = \frac{1}{\bar{w}(t)} M x(t) \quad (12)$$

where $x(t)$ collects the frequencies of each genotype at time t into a vector, and M is the $2^n \times 2^n$ "mutation-selection matrix" where $M_{ij} = w_j p_{ij}$. These equations are valid even if total population size or absolute fitness values change over time, as long as the relative fitness values are constant (Day 2005, p. 278). Extending the model to heterogeneous mutation rate, with or without inheritance, essentially involves defining the appropriate matrices M (supplementary text II.2.1, Supplementary Material online).

By solving Equation (12) and then taking Taylor series expansions of $x(t)$ about zero of the mutation rate moments, we obtain approximations that hold for positive or negative selection coefficients at sufficiently small times (supplementary text II.2.2 and II.2.3, Supplementary Material online). In the two-locus model, in a population initially composed only of the wild type, the double mutant frequency in a homogeneous population takes the form:

$$x_{11}(t; \text{hom}) \approx \beta(t)\langle U\rangle^2 \quad (13)$$

where

$$\beta(t) = \frac{s_{01} + s_{10} - s_{01}s_{10}}{s_{01}s_{10}s_{11}} - \frac{(1-s_{01})^{t+1}}{s_{01}(s_{11}-s_{01})} - \frac{(1-s_{10})^{t+1}}{s_{10}(s_{11}-s_{10})} - \frac{(s_{01}+s_{10}-s_{01}s_{10}-2s_{11}+s_{11}^2)(1-s_{11})^t}{s_{11}(s_{11}-s_{01})(s_{11}-s_{10})}$$

when mutation rate is heterogeneous, in the non-inherited case,

$$x_{11}(t; \text{het}, H=0) \approx \beta(t)\langle U\rangle^2 + \left(1-(1-s_{11})^t\right)\frac{V}{s_{11}} \quad (14)$$

and in the perfectly inherited case,

$$x_{11}(t; \text{het}, H=1) \approx \beta(t)(\langle U\rangle^2 + V) \quad (15)$$

The equilibrium frequencies x^* are given by the eigenvectors of M, with equilibrium population mean fitness $\bar{w}^*$ given by the corresponding eigenvalues. When all mutants are deleterious, the dominant eigenvalue is associated with the polymorphic equilibrium (mutation-selection balance). The solutions are again approximated using Taylor series expansions (supplementary text II.2.2 and II.2.3, Supplementary Material online; table 1).

When comparing double mutant frequency across model cases, the absolute increase due to heterogeneity of the mutation rate is defined by

$$\Delta_{\text{abs}}(t) := x_{11}(t; \text{het}) - x_{11}(t; \text{hom}) \quad (16)$$

and the relative increase by

$$\Delta_{\text{rel}}(t) := (x_{11}(t; \text{het}) - x_{11}(t; \text{hom}))/x_{11}(t; \text{hom}) \quad (17)$$

such that the fold-change is $(1 + \Delta_{\text{rel}})$.

The infinite-locus model (supplementary text II.2.4, Supplementary Material online), based on Kimura and Maruyama (1966), assumes that a Poisson-distributed number of *de novo* mutations arises on a genome in each generation (as in the "Occurrence of Mutations" section) and fitness (w_i) is fully determined by the number of mutations carried (i). The special case in which mutations have equal and multiplicative effects takes $w_i = (1-s)^i$. We obtain equilibrium genotype frequencies by numerical iteration up to a specified tolerance (supplementary text II.2.4.3, Supplementary Material online).

Branching Process Model

A general multi-type branching process model incorporating mutation rate heterogeneity is laid out in the supplementary text II.3, Supplementary Material online. The illustrated results use a specific case of the two-locus model in which the wild type has exactly two offspring in total, while single mutants have either two offspring (neutral) or zero offspring (lethal). Among these offspring, mutations at the two loci occur probabilistically as described in the "Occurrence of Mutations" section.

We derive the following recursions for the probability $P_{2;0}(t)$ that a double mutant has not yet appeared in a lineage initiated by a single wild type individual

(supplementary text II.3.3, Supplementary Material online). Given non-inherited mutation rate with mean $\langle U\rangle$ and variance V, in the neutral single mutant case,

$$P_{2;0}(t+1;\langle U\rangle,V) = (((1-\langle U\rangle)^2+V)P_{2;0}(t;\langle U\rangle,V) + 2(\langle U\rangle(1-\langle U\rangle)-V)(1-\langle U\rangle)^{2^{t+1}-2})^2 \quad (18)$$

while in the lethal single mutant case,

$$P_{2;0}(t+1;\langle U\rangle,V) = (((1-\langle U\rangle)^2+V)P_{2;0}(t;\langle U\rangle,V) + 2(\langle U\rangle(1-\langle U\rangle)-V))^2 \quad (19)$$

If mutation rate is perfectly inherited, Equations (18) and (19) apply within a lineage with fixed mutation rate $U \equiv u_k$ by substituting $\langle U\rangle = u_k$ and $V = 0$.

For illustration, mutation rate equals u_h with probability q_h or u_l with probability $1-q_h$, and there are in total N wild-type progenitors in generation 0. In the non-inherited case, the probability that no double mutant has yet appeared in the entire population is then

$$P_{2;\text{pop}}(t;\text{het},H=0) = P_{2;0}(t;\langle U\rangle,V)^N \quad (20)$$

where $\langle U\rangle$ and V are given by Equation (23). When mutation rate is perfectly inherited, we take N_l progenitors with mutation rate u_l and N_h with rate u_h, such that $N_l+N_h=N$ and $N_h/N = q_h$ (thus restricting q_h to increments of $1/N$). Then

$$P_{2;\text{pop}}(t;\text{het},H=1) = P_{2;0}(t;u_h,0)^{N_h}P_{2;0}(t;u_l,0)^{N_l} \quad (21)$$

In the homogeneous case, all individuals have mutation rate fixed to $\langle U\rangle$, and we calculate

$$P_{2;\text{pop}}(t;\text{hom}) = P_{2;0}(t;\langle U\rangle,0)^N \quad (22)$$

As a scalar measure of the waiting time, we evaluate $T_{0.5}$, the first generation at which $P_{2;\text{pop}}(t) \leq 0.5$.

Quantifying Effects with Two Mutation Rates

If mutation rate simply takes on either of two values, u_h (high) with probability q_h or u_l (low) with probability $1-q_h$, where $u_h/u_l = \rho \geq 1$, particularly simple expressions are available for the mean and variance of mutation rate:

$$\langle U\rangle = u_l(1-q_h+\rho q_h) \quad (23a)$$

$$V = \frac{\langle U\rangle^2 q_h(1-q_h)(\rho-1)^2}{(1-q_h+q_h\rho)^2} \quad (23b)$$

Thus, the squared coefficient of variation, which provides an upper bound on the relative effect of heterogeneity (Δ_{rel}) on the frequency of double mutants in the deterministic two-locus model, is simply

$$c^2 = \frac{q_h(1-q_h)(\rho-1)^2}{(1-q_h+\rho q_h)^2} \quad (24)$$

For fixed ρ, it can readily be shown that c^2 takes on a maximal value of

$$\max_{q_h} c^2 = \frac{1+\rho^2}{4\rho} \quad (25)$$

when the fraction of individuals with the higher mutation rate is $q_h = \frac{1}{1+\rho}$. Since c^2 depends on the mutation rates only through their ratio, scaling up both u_h and u_ℓ by the same factor (e.g. by considering a larger target size as a locus) will not affect these results.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the ETH Zurich; the European Research Council under the 7th Framework Programme of the European Commission (PBDR: grant number 268540 to SB); and the Swiss National Science Foundation (grant number 155866 to SB). The authors thank members of the Theoretical Biology group at ETH Zurich for insightful comments on this work, particularly Dominique Cadosch and Antoine Frenoy for discussion of stress-induced mutagenesis. The authors also thank Antoine Frenoy and Oskar Hallatschek for comments on a draft of this manuscript, and the editors and anonymous reviewers for their insightful and constructive critique.

References

Al-Lazikani B, Banerji U, Workman P. 2012. Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotechnol*. 30:1–13.

Aoki K, Furusawa M. 2001. Promotion of evolution by intracellular coexistence of mutator and normal DNA polymerases. *J Theor Biol*. 209:213–222.

Averof M, Rokas A, Wolfe KH, Sharp PM. 2000. Evidence for a high frequency of simultaneous double-nucleotide substitutions. *Science* 287:1283–1286.

Baquero MR, Nilsson AI, Turrientes MC, Sandvang D, Galán JC, Martínez JL, Frimodt-Møller N, Baquero F, Andersson DI. 2004. Polymorphic mutation frequencies in *Escherichia coli*: emergence of weak mutators in clinical isolates. *J Bacteriol*. 186:5538–5542.

Barber LJ, Davies MN, Gerlinger M. 2015. Dissecting cancer evolution at the macro-heterogeneity and micro-heterogeneity scale. *Curr Opin Genet Dev*. 30:1–6.

Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol*. 23:38–44.

Beerenwinkel N, Zagordi O. 2011. Ultra-deep sequencing for the analysis of viral populations. *Curr Opin Virol*. 1:413–418.

Belavkin R, Channon V, Aston A, Aston E, Krašovec JR, Knight CG. 2016. Monotonicity of fitness landscapes and mutation rate control. *J Math Biol*. 73:1491–1524.

Bjedov I, Tenaillon O, Gérard B, Souza V, Denamur E, Radman M, Taddei F, Matic I. 2003. Stress-induced mutagenesis in bacteria. *Science* 300:1404–1409.

Björkholm B, Sjölund M, Falk PG, Berg OG, Engstrand L, Andersson DI. 2001. Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. *Proc Natl Acad Sci U S A*. 98:14607–14612.

Boe L. 1992. Translational errors as the cause of mutations in *Escherichia coli*. *Mol Gen Genet*. 231:469–471.

Boe L, Danielsen M, Knudsen S, Petersen JB, Maymann J, Jensen PR. 2000. The frequency of mutators in populations of *Escherichia coli*. *Mutat Res*. 448:47–55.

Borgström E, Redin D, Lundin S, Berglund E, Andersson AF, Ahmadian A. 2015. Phasing of single DNA molecules by massively parallel barcoding. *Nat Commun*. 6:7173.
Bürger R. 2000. The Mathematical Theory of Selection, Recombination and Mutation, 1st edn. Chichester: Wiley.
Cairns J. 1998. Mutation and cancer: the antecedents to our studies of adaptive mutation. *Genetics* 148:1433–1440.
Colijn C, Cohen T, Ganesh A, Murray M. 2011. Spontaneous emergence of multiple drug resistance in tuberculosis before and during therapy. *PLoS One* 6:e18327.
Combe M, Garijo R, Geller R, Cuevas JM, Sanjuán R. 2015. Single-cell analysis of RNA virus infection identifies multiple genetically diverse viral genomes within single infectious units. *Cell Host Microbe* 18:424–432.
Cover TM, Thomas JA. 2006. Elements of Information Theory, 2nd edn. Hoboken (NJ): John Wiley & Sons, Inc.
Dapp MJ, Heineman RH, Mansky LM. 2013. Interrelationship between HIV-1 fitness and mutation rate. *J Mol Biol*. 425:41–53.
Day T. 2005. Modelling the ecological context of evolutionary change: déjà vu or something new? In: Cuddington K and Beisner BE, editors, Ecological paradigms lost: routes of theory change, chapter 13. Burlington (MA): Academic Press (Elsevier). p. 273–309.
Denamur E, Matic I. 2006. Evolution of mutation rates in bacteria. *Mol Microbiol*. 60:820–827.
Denamur E, Bonacorsi S, Giraud A, Duriez P, Hilali F, Amorin C, Bingen E, Andremont A, Picard B, Taddei F, Matic I. 2002. High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. *J Bacteriol*. 184:605–609.
Desai MM, Fisher DS. 2011. The balance between mutators and nonmutators in asexual populations. *Genetics* 188:997–1014.
Desai MM, Fisher DS, Murray AW. 2007. The speed of evolution and maintenance of variation in asexual populations. *Curr Biol*. 17:385–394.
Drake JW. 1993. General antimutators are improbable. *J Mol Biol*. 229:8–13.
Drake JW. 2007. Too many mutants with multiple mutations. *Crit Rev Biochem Mol Biol*. 42:247–258.
Drake JW, Holland JJ. 1999. Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A*. 96:13910–13913.
Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. *Genetics* 148:1667–1686.
Drake JW, Bebenek A, Kissling GE, Peddada S. 2005. Clusters of mutations from transient hypermutability. *Proc Natl Acad Sci U S A*. 102:12849–12854.
Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*. 9:267–276.
Elez M, Murray AW, Bi LJ, Zhang XE, Matic I, Radman M. 2010. Seeing mutations in living cells. *Curr Biol*. 20:1432–1437.
Furió V, Moya A, Sanjuán R. 2005. The cost of replication fidelity in an RNA virus. *Proc Natl Acad Sci U S A*. 102:10233–10237.
Furusawa M, Doi H. 1992. Promotion of evolution: disparity in the frequency of strand-specific misreading between the lagging and leading DNA strands enhances disproportionate accumulation of mutations. *J Theor Biol* 157:127–133.
Galhardo RS, Hastings PJ, Rosenberg SM. 2007. Mutation as a stress response and the regulation of evolvability. *Crit Rev Biochem Mol Biol*. 42:399–435.
Gillespie SH, Basu S, Dickens AL, O'Sullivan DM, McHugh TD. 2005. Effect of subinhibitory concentrations of ciprofloxacin on *Mycobacterium fortuitum* mutation rates. *J Antimicrob Chemother*. 56:344–348.
Gillies RJ, Verduzco D, Gatenby RA. 2012. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat Rev Cancer* 12:487–493.
Goldberg DE, Siliciano RF, Jacobs WR. Jr. 2012. Outwitting evolution: fighting drug-resistant TB, malaria, and HIV. *Cell* 148:1271–1283.
Gonzalez C, Hadany L, Ponder RG, Price M, Hastings PJ, Rosenberg SM. 2008. Mutability and importance of a hypermutable cell subpopulation that produces stress-induced mutants in *Escherichia coli*. *PLoS Genet*. 4:e1000208.
Haigh J. 1978. The accumulation of deleterious genes in a population—Muller's Ratchet. *Theor Popul Biol*. 14:251–267.
Hill KA, Wang J, Farwell KD, Scaringe WA, Sommer SS. 2004. Spontaneous multiple mutations show both proximal spacing consistent with chronocoordinate events and alterations with p53-deficiency. *Mutat Res*. 554:223–240.
Jee J, Rasouly A, Shamovsky I, Akivis Y, R. Steinman S, Mishra B, Nudler E. 2016. Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature* 534:693–696.
Johnson T. 1999. The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. *Proc R Soc B* 266:2389–2397.
Kennedy SR, Schultz EM, Chappell TM, Kohrn B, Knowels GM, Herr AJ. 2015. Volatility of mutator phenotypes at single cell resolution. *PLoS Genet*. 11:e1005151.
Kimura M, Maruyama T. 1966. The mutational load with epistatic gene interactions in fitness. *Genetics* 54:1337–1351.
Knudson AG. 2001. Two genetic hits (more or less) to cancer. *Nat Rev Cancer* 1:157–162.
Kohanski MA, DePristo MA, Collins JJ. 2010. Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Mol Cell* 37:311–320.
Komarova NL, Wodarz D. 2005. Drug resistance in cancer: principles of emergence and prevention. *Proc Natl Acad Sci U S A*. 102:9714–9719.
Krašovec R, Belavkin RV, Aston JAD, Channon A, Aston E, Rash BM, Kadirvel M, Forbes S, Knight CG. 2014. Mutation rate plasticity in rifampicin resistance depends on *Escherichia coli* cell-cell interactions. *Nat Commun*. 5:3742.
Lang GI, Murray AW. 2008. Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* 178:67–82.
Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500:571–574.
LeClerc JE, Li B, Payne WL, Cebula TA. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274:1208–1211.
Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A*. 109:E2774–E2783.
Lengauer C, Kinzler KW, Vogelstein B. 1998. Genetic instabilities in human cancers. *Nature* 396:643–649.
Loeb KR, Loeb LA. 2000. Significance of multiple mutations in cancer. *Carcinogenesis* 21:379–385.
Loeb LA. 1991. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res*. 51:3075–3079.
Loeb LA. 2001. A mutator phenotype in cancer. *Cancer Res*. 61:3230–3239.
Loeb LA, Springgate CF, Battula N. 1974. Errors in DNA replication as a basis of malignant changes. *Cancer Res*. 34:2311–2321.
Loeb LA, Loeb KR, Anderson JP. 2003. Multiple mutations and cancer. *Proc Natl Acad Sci U S A*. 100:776–781.
Long H, Miller SF, Strauss C, Zhao C, Cheng L, Ye Z, Griffin K, Te R, Lee H, Chen CC, Lynch M. 2016. Antibiotic treatment enhances the genome-wide mutation rate of target cells. *Proc Natl Acad Sci U S A*. 113:E2498–E2505.
Lynch M, Abegg A. 2010. The rate of establishment of complex adaptations. *Mol Biol Evol*. 27:1404–1414.
MacLean RC, Torres-Barceló C, Moxon R. 2013. Evaluating evolutionary models of stress-induced mutagenesis in bacteria. *Nat Rev Genet*. 14:221–227.
Mansky LM, Cunningham KS. 2000. Virus mutators and antimutators: roles in evolution, pathogenesis and emergence. *Trends Genet*. 16:512–517.
Mansky LM, Temin HM. 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol*. 69:5087–5094.

Mansky LM, Le Rouzic E, Benichou S, Gajary LC. 2003. Influence of reverse transcriptase variants, drugs, and Vpr on Human Immunodeficiency Virus Type 1 mutant frequencies. *J Virol*. 77:2071–2080.

Mao EF, Lane L, Lee J, Miller JH. 1997. Proliferation of mutators in a cell population. *J Bacteriol*. 179:417–422.

Matic I, Radman M, Taddei F, Picard B, Doit C, Bingen E, Denamur E, Elion J. 1997. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* 277:1833–1834.

McCool JD, Long E, Petrosino JF, Sandler HA, Rosenberg SM, Sandler SJ. 2004. Measurement of SOS expression in individual *Escherichia coli* K-12 cells using fluorescence microscopy. *Mol Microbiol*. 53:1343–1357.

Miller JH. 1996. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. *Annu Rev Microbiol*. 50:625–643.

Morosini MI, Baquero MR, Sánchez-Romero JM, Negri MC, Galán JC, del Campo R, Pérez-Díaz JC, Baquero F. 2003. Frequency of mutation to rifampicin resistance in *Streptococcus pneumoniae* clinical strains: *hexA* and *hexB* polymorphisms do not account for hypermutation. *Antimicrob Agents Chemother*. 47:1464–1467.

Ninio J. 1991. Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics* 129:957–962.

Oliver A, Cantón R, Campo P, Baquero F, Blázquez J. 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* 288:1251–1253.

Prunier AL, Malbruny B, Laurans M, Brouard J, Duhamel JF, Leclercq R. 2003. High rate of macrolide resistance in *Staphylococcus aureus* strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. *J Infect Dis*. 187:1709–1716.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*, 3.3.1 edn. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ram Y, Hadany L. 2012. The evolution of stress-induced hypermutation in asexual populations. *Evolution* 66:2315–2328.

Ram Y, Hadany L. 2014. Hadany. Stress-induced mutagenesis and complex adaptation. *Proc R Soc B* 281:20141025.

Ribeiro R, Bonhoeffer S. 2000. Production of resistant HIV mutants during antiretroviral therapy. *Proc Natl Acad Sci U S A*. 97:7681–7686.

Richardson AR, Yu Z, Popovic T, Stojiljkovic I. 2002. Mutator clones of *Neisseria meningitidis* in epidemic serogroup A disease. *Proc Natl Acad Sci U S A*. 99:6103–6107.

Rogozin IB, Pavlov YI. 2003. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res*. 544:65–85.

Rosche WA, Foster PL. 2000. Determining mutation rates in bacterial populations. *Methods* 20:4–17.

Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral mutation rates. *J Virol*. 84:9733–9748.

Schrider DR, Hourmozdi JN, Hahn MW. 2011. Pervasive multinucleotide mutational events in eukaryotes. *Curr Biol*. 21:1051–1054.

Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387:703–705.

Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. 2000. The evolution of mutation rates: separating causes from consequences. *BioEssays* 22:1057–1066.

Suárez P, Valcárcel J, Ortín J. 1992. Heterogeneity of the mutation rates of Influenza A viruses: isolation of mutator mutants. *J Virol*. 66:2491–2494.

Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of mutator alleles in adaptive evolution. *Nature* 387:700–702.

Tenaillon O, Denamur E, Matic I. 2004. Evolutionary significance of stress-induced mutagenesis in bacteria. *Trends Microbiol*. 12:264–270.

Uphoff S, Lord ND, Okumus B, Potvin-Trottier L, Sherratt DJ, Paulsson J. 2016. Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. *Science* 351:1094–1097.

Watson ME, Jr, Burns JL, Smith AL. 2004. Hypermutable *Haemophilus influenzae* with mutations in *mutS* are found in cystic fibrosis sputum. *Microbiology* 150:2947–2958.

Wolfram Research, Inc. 2016. Mathematica, 10.4 edn. Champaign (IL): Wolfram Research, Inc.

Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: *Proceedings of the Sixth International Congress of Genetics*, p. 356–366.