

Appendix: analytical derivations

In this appendix, we detail the analytical derivations leading to the equations given in the main text, in four sections. First, in section I. “Settings and general results”, we briefly recall the derivations from Martin et al. (2013) leading to the evolutionary rescue (ER) probability. Then, in section II. “Application to Fisher’s Geometric Model (FGM)”, we apply this framework to the FGM, yielding the ER probability in the SSWM regime. In section III. “Small mutational effects approximation (SME)”, we derive explicit approximations under the assumption of weak mutation effects (limit as $\lambda/r_{max} \rightarrow 0$), illustrated by numerical examples (**Supplementary Figs.3-4**). Then, in section “IV. Key properties of the model” we use these approximations to provide simple insights into key properties of the model: proportion of ER caused by standing variance, characteristic stress and stress window over which ER changes from highly likely to highly unlikely.

All along $\mathbb{E}_X(\cdot)$ denotes an expectation taken over the distribution of X (which can be multivariate), and $X|Y$ denotes a random variable X conditional on Y . Most computations are checked in a Mathematica® notebook (**Supplementary file S1**), provided as a .cdf file that can be ran using a freely available “CDF player” from the Wolfram website.

I. Settings and general results

We recall the general model and approximations described in Martin *et al.* (2013), which apply to Fisher’s Geometrical Model (FGM), in the limit of Strong Selection Weak Mutation (SSWM). We call ‘new environment’ the environment imposed at the onset of stress ($t = 0$): it induces a decay of the population under study, at $t = 0$. The environment in which the population was before the onset of stress is called ‘previous environment’.

1. General stochastic demography: Each genotype i present at $t = 0$ or later produced by mutation (via a Poisson process) is characterized by the parameters (r_i, σ_i) of a Feller diffusion approximating its stochastic demography. Let $N_i(t)$ be the size of the genotypic class i at time t . The growth rate r_i and reproductive variance σ_i give the expectation and variance of the change, over dt , in $\log N_i(t)$ (which is $\Delta N_i/N_i$), given $N_i(t)$. We have $r_i = \mathbb{E}(\Delta N_i|N_i)/$

30 $(dt N_i)$ over some time interval (infinitesimal in the diffusion limit), and $\sigma_i = V(\Delta N_i | N_i) /$
31 $(dt N_i)$. In the diffusion limit, the population size of the genotypic class is characterized by the
32 stochastic differential equation $dN_i(t) = r_i N_i(t) dt + \sqrt{\sigma_i N_i(t)} dB_t$ where B_t is a standard
33 Brownian motion.

34

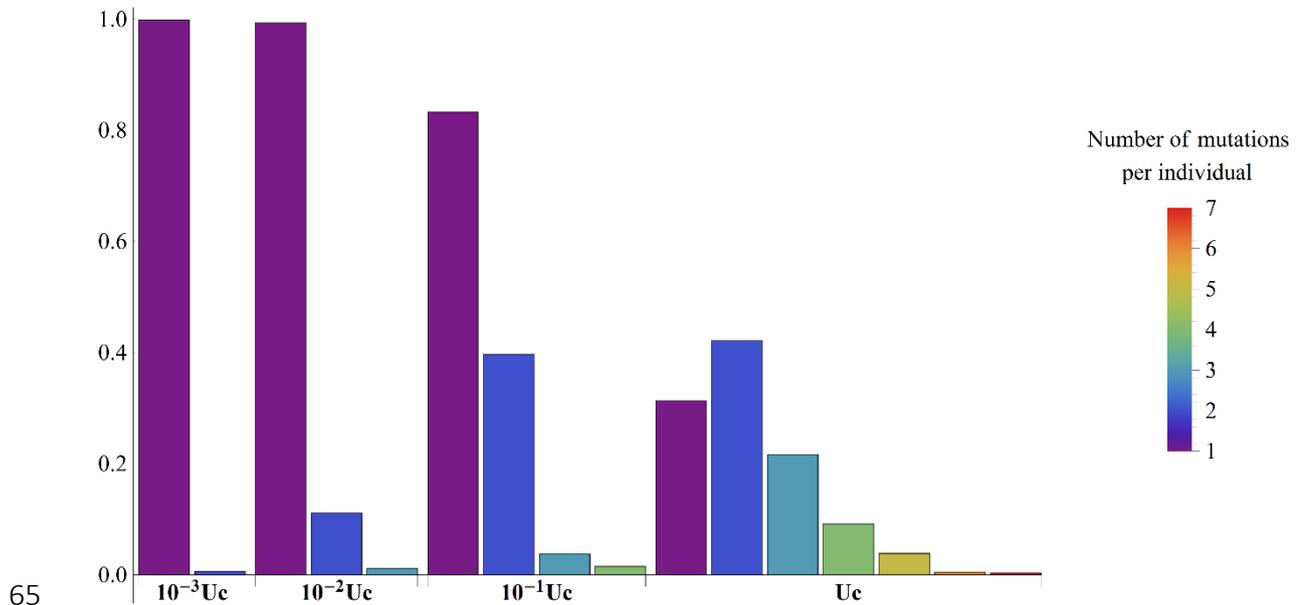
35 **2. Application to a discrete time model with Poisson offspring distribution:** In our case, we
36 consider, as an example, discrete non-overlapping generations where the subclass consisting
37 of individuals of genotype i produces $N_i(t + 1) \sim \text{Poisson}(W_i N_i(t))$ offspring over one
38 generation, where W_i is the absolute Darwinian (i.e., multiplicative) fitness of this genotype. In
39 this case $r_i = \mathbb{E}(N_i(t + 1) - N_i(t)) / N_i(t) = W_i - 1$, while $\sigma_i = V(N_i(t + 1) - N_i(t)) /$
40 $N_i(t) = W_i$ are two constant coefficients for any genotype. The diffusion limit applies when
41 the demographic changes per generation are small, which requires $W_i \rightarrow 1$. In this case we
42 retrieve a Feller diffusion where $r_i = W_i - 1 \rightarrow \log(W_i)$ is the absolute Malthusian fitness of
43 genotype i , while $\sigma_i = W_i \rightarrow 1$ is constant across genotypes (Martin *et al.* 2013).

44 Extinction of the population occurs if none of the genotypes present or produced over the
45 course to extinction avoids extinction: following classic notation, non-extinction (over infinite
46 time) is denoted “establishment”. The probability of establishment, for a lineage started in
47 single copy, with growth rate r and stochastic variance σ in the new environment, is $\pi(r) =$
48 $(1 - e^{-2r/\sigma})\Theta(r)$ where $\Theta(\cdot)$ is the Heaviside theta function ($\Theta(x) = 0$ if $x \leq 0$ and $\Theta(x) =$
49 1 if $x > 0$) and the reproductive variance $\sigma \approx 1$ (as in our simulation example mentioned
50 above considering a Poisson reproduction).

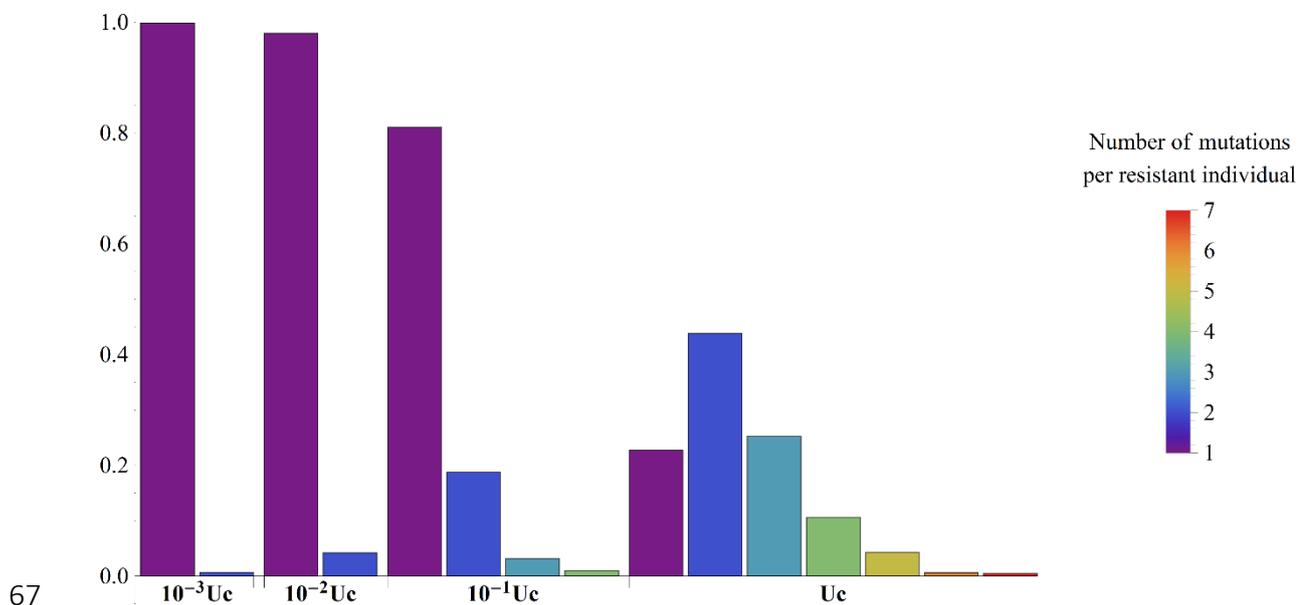
51

52 **3. Strong selection weak mutation (SSWM):** In this work we use a SSWM approximation
53 (Gillespie 1983; McCandlish and Stoltzfus 2014). We consider that the mutation rate is low
54 relative to the strength of selection, so that rescue mutations are typically single step mutants
55 (sampled from the pool of possible mutants), be it present before the onset of stress (from
56 standing variance, hereafter ‘SV’) or arising after it (*de novo* hereafter ‘DN’). We thus only
57 consider rescue from a single allele, which is randomly drawn among all possible alleles,
58 weighted by their probability to produce a rescue (illustrated in **Supplementary Fig.1**). Under
59 the SSWM approximation, all possible rescue events (SV or DN, from any given allele) arise as

60 alternative Poisson events (Martin *et al.* 2013). The overall probability of extinction is thus
 61 simply the zero class of a Poisson distribution and the probability of evolutionary rescue (ER) is
 62 that of the complementary event ('no extinction'). The key to describing the process is
 63 therefore to derive the rate of these Poisson events, over all possible mutations in the model
 64 considered.



66 (A) Proportion of individuals carrying different numbers of *de novo* mutations in rescued populations.



68 (B) Proportion of individuals carrying different numbers of *de novo* resistant mutations in rescued populations.

69 **Supplementary Figure 1:** Proportion of individuals carrying different numbers of *de novo* mutations (all mutations
 70 in (A) or only resistant mutations in (B)) in rescued populations, from exact simulations. These proportions were
 71 evaluated in those populations having been rescued, at the time where they reached our stop criterion, that is

72 when their current mean growth rate and population size imply a very low probability of future extinction, even
 73 in the absence of future adaptation (see simulation methods in the main text). The distributions are given for
 74 different mutation rates (x-axis), given relative to the critical mutation rate ($U_c = n^2 \lambda/4 = 0.1$ here) below which
 75 the SSWM assumption should hold. Populations were initially composed of a single clone, with $N_0 = 10^5$, $n = 4$,
 76 $\bar{s} = 0.01$, $r_{max} = 1.5$, $r_D = (0.042, 0.129, 0.23, 0.32)$ from right to left in each panel. Whenever $U \ll U_c$,
 77 rescued populations mostly consist of the wild type (0 mutation, not shown here) and single mutants (purple bars).
 78 As U approaches U_c , a substantial proportion of multiple mutants is found in late rescued populations (fewer
 79 when focusing on the resistant mutants that are the potential rescuers of the populations). Note however that
 80 this illustration does not ascertain whether these multiple mutants are the cause of rescue or not.

81

82 **4. Rescue from *de novo* mutation:** We first consider rescue starting from a clonal population (of
 83 inoculum size N_0), so that any rescue event is caused by *de novo* mutations ('DN'). We define
 84 the rate of 'DN' rescue events ω_{DN} per individual present at the onset of stress, such
 85 that $N_0 \omega_{DN}$ is the parameter of the Poisson number of 'DN' rescue events. Ignoring
 86 stochasticity in the decay dynamics of the wild-type (large $N_0 r_D$), the total number, over time,
 87 of rescue mutations for those mutants with growth rate r (more precisely, within the
 88 infinitesimal class $[r, r + dr]$) is $N_0 U \pi(r) f(r) dr / r_D$. Here $U \pi(r) f(r) dr$ is the rate, per
 89 capita per unit time, of mutations towards this class, weighted by their probability $\pi(r)$ to
 90 generate a rescue. After integrating over the whole distribution of possible resistant mutants
 91 we get (following Martin *et al.* 2013):

$$\omega_{DN} = U \frac{\mathbb{E}_r(\pi(r))}{r_D} = \frac{U}{r_D} \int_0^{r_{max}} (1 - e^{-2r/\sigma}) f(r) dr , \quad (A1)$$

92 where the random variable r denotes mutant growth rates in the new environment and $f(r)$
 93 their probability density function. The corresponding probability of extinction is then $P_{ext} =$
 94 $e^{-N_0 \omega_{DN}}$ (Martin *et al.* 2013).

95

96 **5. Rescue from standing variance:** In an alternative scenario, the initial population (of size N_0) is
 97 previously at equilibrium in the non-stressful 'previous environment'. In the SSWM regime, the
 98 rescue process results from the contributions from two independent processes: rescue caused
 99 by a mutant appearing after the onset of stress and rescue caused by a mutant already present
 100 at the onset of stress. The numbers of each event overall are approximately Poisson distributed,

101 with rates $N_0 \omega_{DN}$ (given in Eq.(A1)) and $N_0 \omega_{SV}$, respectively. The latter depends on the joint
 102 distribution of the cost c of resistance mutations in the previous environment, and of their
 103 growth rate r in the new environment. The cost c (in the previous environment) is equal to
 104 minus the selection coefficient, relative to the dominant background, of random mutations
 105 arising in this very background (Martin *et al.* 2013); we then have:

$$\omega_{SV} = U \mathbb{E}_{c,r} \left(\frac{\pi(r)}{c} \right) = U \mathbb{E}_r \left(\pi(r) \mathbb{E}_c \left(\frac{1}{c} \mid r \right) \right). \quad (\text{A2})$$

106 This Poisson approximation applies under the SSWM regime (Martin *et al.* 2013), regardless of
 107 whether the population is initially at stochastic mutation-selection-drift balance at constant
 108 size N_0 , or at quasi-deterministic mutation-selection balance (at some size $N \gg N_0$) followed
 109 by a bottleneck at the onset of stress (to reach size N_0). The extinction probability in the
 110 presence of both initial standing variance and *de novo* mutations is $P_{ext} = e^{-N_0 (\omega_{DN} + \omega_{SV})}$, and
 111 because rescue events are Poisson distributed, the proportion ϕ_{SV} of rescue from standing
 112 variants (over all rescue events) is

$$\phi_{SV} = \frac{\omega_{SV}}{\omega_{DN} + \omega_{SV}}. \quad (\text{A3})$$

113

114 II. Application to Fisher's Geometric Model (FGM)

115 In our case, the joint distribution of c and r emerges from the FGM: growth rates (both in the
 116 previous and new environment) are quadratic functions of phenotype, around an environment-
 117 dependent optimum. We recall that we directly give the results in the case of a Poisson
 118 offspring number as mentioned in **section I subsection 2** (where $\sigma \approx 1$ for all genotypes). In the
 119 following, we derive the above quantities in this context.

120

121 **1. Definitions:** Define n the number of dimensions of the fitness landscape (number of traits
 122 under stabilizing selection) and λ , the variance of mutational effects. In our model, λ varies with
 123 both the strength of selection and the effect of mutations. To see this, consider an isotropic
 124 model on some arbitrary phenotypic space: we can define the Malthusian fitness of a
 125 phenotype \mathbf{z} as $r(\mathbf{z}) = r_{max} - \lambda_S \|\mathbf{z}\|^2 / 2$ with λ_S the strength of stabilizing selection on each

126 trait. Each mutation creates a random perturbation $\mathbf{dz} \sim N(\mathbf{0}, \lambda_M \mathbf{I}_n)$ with \mathbf{I}_n the identity
 127 matrix in \mathbb{R}^n and λ_M the variance of mutational effects on each trait. As we focus only on fitness
 128 here, we can consider the model in a scaled phenotypic space, with phenotypes $\mathbf{x} = \sqrt{\lambda_S} \mathbf{z}$ so
 129 that $r(\mathbf{x}) = r_{max} - \|\mathbf{x}\|^2/2$ and $\mathbf{dx} \sim N(\mathbf{0}, \lambda \mathbf{I}_n)$ where $\lambda = \lambda_S \lambda_M$. Thus, we need not
 130 separate selective and mutational scalings, and simply ‘measure’ phenotypic traits in
 131 convenient units of ‘selection strength’. The mean fitness effect of random mutations is
 132 $|\mathbb{E}(s)| = \lambda n/2$, in absolute value (Martin and Lenormand 2015). Therefore, λ directly gives a
 133 measure of mutant selective effects per trait.

134 It proves handy to define the scaled variable $y = r/r_{max} \in [0,1]$ and the corresponding scaled
 135 decay rate $y_D = r_D/r_{max}$ (which are also used in the main text). For the sake of compactness,
 136 in the derivations of this appendix only, we also define a scaled height of the fitness
 137 peak $\rho_{max} = r_{max}/\lambda$ (scaled by the variance of mutational effects λ) and a measure of
 138 dimensionality $\theta = n/2$.

139

140 **2. Distribution of mutant growth rates:** The initial clone lies at fitness distance $s_0 = r_{max} + r_D$
 141 from the optimum in the new environment, which, together with n and λ , fully determines the
 142 distribution of mutant selection coefficients (and hence growth rates r). The distribution of
 143 selection coefficients $s = r + r_D$ of random mutants relative to the ancestor has known exact
 144 form for the isotropic FGM (eq.(3) in Martin and Lenormand 2015). From it, the distribution of
 145 growth rates r among random mutants, within the new environment, is readily obtained. It has
 146 stochastic representation

$$147 \quad r = r_{max} - \frac{\lambda}{2} \chi_n^2(2 \rho_{max} (1 + y_D))$$

148 where $\chi_n^2(v)$ is a non-central chisquare with n degree of freedom and non-centrality
 149 parameter v . The scaled growth rate $y = r/r_{max} \in [0,1]$ has stochastic representation $y =$
 150 $1 - \chi_n^2(2 \rho_{max}(1 + y_D)) / (2\rho_{max})$ with corresponding probability density function (see also

151 **Supplementary file S1)**

$$f_y(y) = e^{-\rho_{max}(2+y_D-y)} \rho_{max}^\theta (1-y)^{\theta-1} \frac{{}_0F_1(\theta, \rho_{max}^2(1+y_D)(1-y))}{\Gamma(\theta)}, \quad (A4)$$

152 where ${}_0F_1(\dots)$ is the confluent hypergeometric function.

153

154 **3. Rescue from *de novo* mutation:** The rate of ER from de novo mutations in Eq.(A1) can be
155 equivalently computed by integrating over the distribution of the scaled growth rate $y =$
156 r/r_{max} . The non-extinction probability, given y , is $\pi(y) = (1 - e^{-2 r_{max} y})\Theta(y)$ where $\Theta(\cdot)$ is
157 the Heaviside theta function ($\Theta(x) = 1$ for $x \in \mathbb{R}^+$ and $\Theta(x) = 0$ for $x \in \mathbb{R}^-$):

$$\omega_{DN} = U \frac{\mathbb{E}_y(\pi(y))}{r_D} = \frac{U}{r_D} \int_0^1 (1 - e^{-2 r_{max} y}) f_y(y) dy, \quad (A5)$$

158 where $f_y(y)$ is given by Eq.(A4). This integral can readily be computed numerically.

159

160 **4. Rescue from standing variance ($n \geq 2$):** For rescue from standing variants ('SV'), the
161 distribution of the cost, in the previous environment, of mutations with growth rate r (or scaled
162 growth rate y), in the new environment, must also be known (see Eq.(A2)). In the SSWM
163 approximation, we neglect the effect of standing background variation on the distribution of
164 fitness effects of mutations generated before the onset of stress. We thus consider that the
165 joint distribution of (c, r) is the one generated if all mutants arose from the dominant genotype
166 in the previous environment, which is optimal in this environment (same as the initial clone in
167 the DN rescue problem. For the FGM, the distribution of c has a known form, conditional on
168 the effect of the mutant ($s = r + r_D$) in the new environment. Consider a background optimal
169 in the previous environment, not too close to the optimum of the new environment and
170 with $n \geq 2$. The conditional cost then has a simple stochastic representation (from eq. (9) in
171 Martin and Lenormand 2015): $c|s \sim c_{min} + \gamma$, where $\gamma \sim \Gamma(\theta - 1/2, \lambda)$ is a gamma deviate
172 and $c_{min} = 2s_0 - s - 2s_0\sqrt{1 - s/s_0}$ is a constant, with $s_0 = r_{max} + r_D$. Expressed in terms of
173 scaled growth rates y ($s = y r_{max} + r_D$), we have $c_{min} = c_{min}(y) = (2 + y_D -$
174 $2\sqrt{(1 + y_D)(1 - y)} - y) r_{max}$ and we have $c|y \sim c_{min}(y) + \gamma$. The cost $c_{min}(y)$ can be
175 simply interpreted as an "incompressible cost": the minimum cost that mutants within the
176 class $[y, y + dy]$ must pay, because they cannot get close to the new optimum without moving
177 away from the former one. The stochastic component γ describes the variation in distance to
178 the former optimum, of those mutants within the class $[y, y + dy]$, i.e. lying on the subspace
179 of phenotypes equally distant to the new optimum. This component happens to be
180 independent of y , which simplifies our derivations.

181 From Martin *et al.* (2013), the key to predict ER from standing variance is the harmonic mean
 182 of $c|y$ (among random mutants with effect y), which we denote $c_H(y)$. From the stochastic
 183 representation of $c|y$ described above, it is given by (see also **Supplementary file S1**):

$$c_H(y) = \frac{1}{\mathbb{E}_c\left(\frac{1}{c} \mid y\right)} = \frac{1}{\mathbb{E}_\gamma\left(\frac{1}{\gamma + c_{min}(y)}\right)} = \frac{\lambda e^{-v(y)}}{E_{\theta-1/2}(v(y))}, \quad (\text{A6})$$

$$\text{with } v(y) = c_{min}(y)/\lambda = \rho_{max} \left(2 + y_D - 2\sqrt{(1 + y_D)(1 - y)} - y\right)$$

184 where $\mathbb{E}_\gamma(\cdot)$ is taken over the distribution of $\gamma \sim \Gamma(\theta - 1/2, \lambda)$ and $E_k(z) = \int_1^\infty e^{-zt}/t^k dt$
 185 is the exponential integral function. Using Eq.(A6), Eq.(A2) yields the rate of ER from standing
 186 variance:

$$\omega_{SV} \approx U \mathbb{E}_y\left(\frac{\pi(y)}{c_H(y)}\right) = \frac{U}{\lambda} \int_0^1 (1 - e^{-2r_{max}y}) e^{v(y)} E_{\theta-1/2}(v(y)) f_y(y) dy. \quad (\text{A7})$$

187

188 **5. Case $n = 1$:** Eq. (9) in Martin and Lenormand (2015) applies for all $n \geq 2$, but not if $n = 1$.
 189 In this case, the geometry of the landscape is more constrained, and the cost distribution is
 190 simplified. We can use elementary computations (see **Supplementary file S1**) from the results
 191 of (Martin and Lenormand 2015) to see that the conditional cost $c|y$ is a constant $c|y =$
 192 $c_{min}(y) = \lambda v(y)$, so that its harmonic mean is also $c_H(y) = \lambda v(y)$. Intuitively this result
 193 arises because, in one dimension, there is no freedom in the position of a mutant relative to its
 194 ancestor and to the optimum in the new environment (all three phenotypes must be aligned).
 195 The ER rate from standing variance, in this case, is simply

$$\omega_{SV} \approx U \mathbb{E}_y\left(\frac{\pi(y)}{c_H(y)}\right) = \frac{U}{\lambda} \int_0^1 (1 - e^{-2r_{max}y}) \frac{f_y(y)}{v(y)} dy. \quad (\text{A8})$$

196

197 **III. Small mutational effects approximation (SME)**

198 Eqs. (A5) and (A7) provide a mathematical framework to predict rescue from both *de novo*
 199 mutants and standing variants, in the SSWM regime. However, they do not provide simple
 200 closed form expressions (the integrals must be computed numerically). To gain more analytical

201 insight, we rely on a further approximation: we look for limit expressions for these rates, as the
 202 variance of mutational effects λ becomes small relative to the maximal growth rate r_{max} . As
 203 we have seen above (**Section II subsection 1**), λ measures the average mutant selective effect
 204 per trait. Therefore, we denote these limits “small mutation effect (SME) approximations”, in
 205 the sense that we let $\lambda/r_{max} = \rho_{max}^{-1} \rightarrow 0$ or $\rho_{max} \rightarrow \infty$, holding all other parameters fixed.
 206 This approximation implies that most resistant mutants grow much less than the optimal
 207 phenotype: they remain far from the optimum of the new environment so that $y = r/r_{max} \ll$
 208 1.

209

210 **1. Approximate probability density function of y :** As the SME implies both that $1 - y = O(1)$
 211 while $\rho_{max} \gg 1$, the second argument in the hypergeometric function in Eq.(A4) is large: $h =$
 212 $\rho_{max}^2(1 + y_D)(1 - y) = O(\rho_{max}^2) \gg 1$ for any y_D . Therefore, we can use an asymptotic
 213 expansion, when $|h| \rightarrow \infty$, for this function: ${}_0F_1(\theta, h)/\Gamma(\theta) \approx h^{1/4-\theta/2} e^{2\sqrt{h}}/(2\sqrt{\pi})$
 214 (Wolfram Research 2001). Plugging this into Eq.(A4) yields a simplified expression for the pdf
 215 of the scaled mutant growth rate distribution (see **Supplementary file S1**):

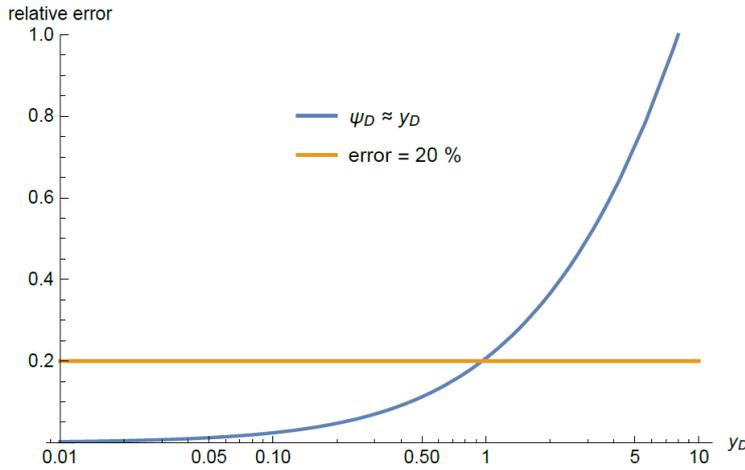
$$f_y(y) \approx \frac{\sqrt{\rho_{max}}}{2\sqrt{\pi}} (1 + y_D)^{1/4-\theta/2} e^{-v(y)} (1 - y)^{\theta/2-3/4}, \quad y \in [0,1], \quad (\text{A9})$$

216 where $v(y) = \rho_{max}(2 + y_D - 2\sqrt{(1 + y_D)(1 - y)}) - y$ was defined in Eq.(A6). Convergence
 217 to this limit is faster with (i) stronger stress (z increases with y_D) and (ii) lower dimensionality.
 218 It is roughly exact, for any ρ_{max} , when $\theta = 1/2$ (see **Supplementary file S1**).

219

220 **2. Change of variables $y \rightarrow \psi$:** The distribution in Eq.(A9) takes a more compact form by using
 221 a bijective change of variable, which corresponds to an alternative measure of the mutant
 222 growth rate y . More precisely, we consider $\psi = \psi(y) = 2(1 - \sqrt{1 - y})$ as a measure of
 223 growth rate, so that, conversely, $y = y(\psi) = \psi(1 - \psi/4)$. The scaled growth rate of the
 224 initial clone is $-y_D$ which yields a corresponding decay rate $\psi_D = |\psi(-y_D)| = 2(\sqrt{1 + y_D} -$
 225 $1)$. The transformation is bijective and strictly increasing ($\psi'(y) = 1/\sqrt{1 - y} > 0$), from $y \in$
 226 $[0,1]$ to $\psi \in [0,2]$. A linear approximation $\psi_D \approx y_D + o(y_D)$ yields a relative error $\leq 20\%$ for
 227 all $y_D \in [0,1]$, as illustrated in **Supplementary Figure 2** below.

228



229

230 **Supplementary Figure 2:** Relative error implied by the approximation $\psi_D = y_D$.

231

232 The probability density function $f_\psi(\cdot)$ of the transformed variable ψ , based on the
 233 approximate probability density function of y in Eq.(A9), is $y'(\psi)f_y(y(\psi))$ yielding:

$$f_\psi(\psi) = y'(\psi)f_y(y(\psi)) \approx \frac{\sqrt{\rho_{max}}}{2\sqrt{\pi}} e^{-\rho_{max} q(\psi)} \left(\frac{1 - \psi/2}{1 + \psi_D/2} \right)^{\theta - 1/2} . \quad (A10)$$

$$q(\psi) = \frac{1}{4} (\psi + \psi_D)^2 , \quad \psi \in [0,2]$$

234 This form makes it more visible how the SME corresponds to small y (here to small $\psi = y +$
 235 $o(y)$). Indeed, as ρ_{max} gets larger, the probability density function in Eq.(A10) is dominated by
 236 $e^{-\rho_{max} q(\psi)}$, which falls off sharply with ψ , so that most ψ values are small.

237

238 **3. Approximate ER rate ω_{DN} from *de novo* mutations:** Using the ψ -scale, the ER rate in Eqs.(A5)
 239 is amenable to the so-called Laplace method of approximation for integrals (Breitung 1994).
 240 Broadly speaking, this method studies integrals over some domain for ψ , involving integrands
 241 of the form $h(\psi) e^{-\rho q(\psi)}$ (with some functions $q(\cdot)$ and $h(\cdot)$, independent of ρ). As $\rho \rightarrow \infty$,
 242 such integrals are dominated by terms in the vicinity of the minimum of $q(\cdot)$, over the
 243 integration domain. They can thus be computed approximately, by (i) using the leading order
 244 of $h(\cdot)$ (and possibly $q(\cdot)$, although we do not require this) around this minimum, and by (ii)
 245 integrating over any domain that proves handy, away from the minimum.

246 Eq.(A5), once expressed in terms of ψ , with probability density function given in Eq.(A10), is of
 247 this form. We have $\pi(y(\psi)) = 1 - e^{-2 r_{max} \psi (1-\psi/4)}$ so that Eq. (A5) can be written:

$$\omega_{DN} = \frac{U}{r_D} \mathbb{E}_\psi(\pi(y(\psi))) \approx \frac{U}{r_D} \frac{\sqrt{\rho_{max}}}{2\sqrt{\pi}} \int_0^2 h(\psi) e^{-\rho_{max} q(\psi)} d\psi \quad , \quad (A11)$$

$$h(\psi) = \left(\frac{1 - \psi/2}{1 + \psi_D/2} \right)^{\theta - 1/2} (1 - e^{-2 r_{max} \psi (1-\psi/4)})$$

248 where the function $q(\cdot)$ (Eq.(A10)) has a unique minimum, over $\psi \geq 0$, at $\psi = 0$. Therefore,
 249 an approximation to the integral in Eq.(A11), as $\rho_{max} \rightarrow \infty$, is obtained by approximating $h(\cdot)$
 250 by its leading order around $\psi = 0$: $h(\psi) = h_*(\psi) + o(\psi)$, with $h_*(\psi) = 2 r_{max} \psi (1 +$
 251 $\psi_D/2)^{1/2-\theta}$. Plugging this into the integral, expressing r_D as $r_D = \psi_D (1 + \psi_D/4) r_{max}$, and
 252 computing the integral over $\psi \in [0, \infty]$ yields (see **Supplementary file S1**):

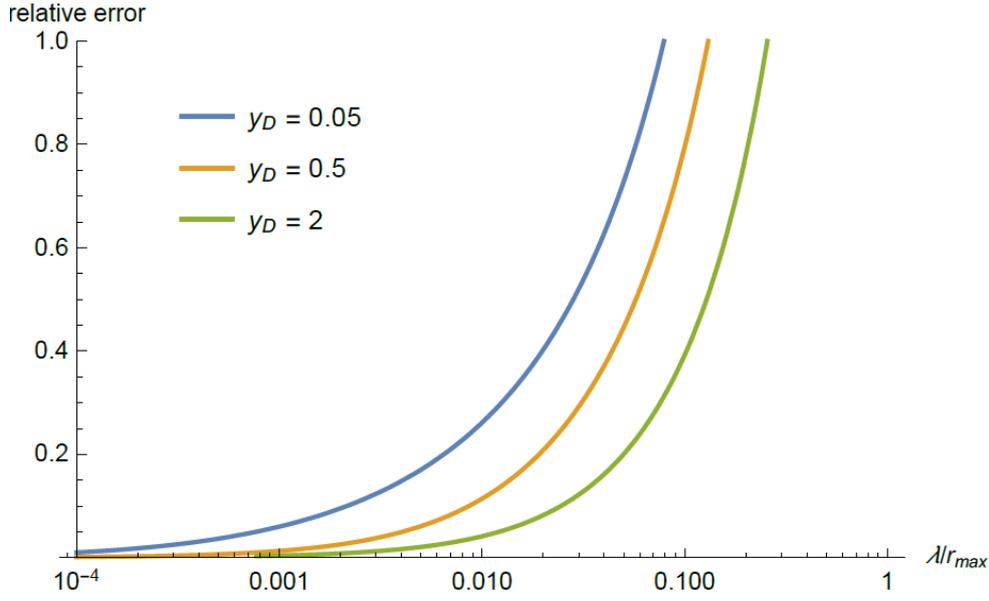
$$\omega_{DN} \xrightarrow{\rho_{max} \rightarrow \infty} \omega_{DN}^* = \frac{U}{r_D} \frac{\sqrt{\rho_{max}}}{2\sqrt{\pi}} \int_0^\infty h_*(\psi) e^{-\rho_{max} q(\psi)} d\psi$$

$$\omega_{DN}^* = U \frac{(1 + \psi_D/2)^{1/2-\theta}}{1 + \psi_D/4} g(\alpha) \quad , \quad (A12)$$

$$\text{with } \alpha = \psi_D^2 \frac{\rho_{max}}{4} \quad \text{and} \quad g(\alpha) = \frac{e^{-\alpha}}{\sqrt{\pi \alpha}} - \text{erfc}(\sqrt{\alpha})$$

253 where $\text{erfc}(\cdot)$ is the complementary error function. It can be checked numerically that the
 254 'exact' rate ω_{DN} (Eq.(A5)) indeed converges to this limit as $\lambda/r_{max} \rightarrow 0$ (i.e. as $\rho_{max} \rightarrow \infty$). This
 255 is illustrated in **Supplementary Fig.3**: the convergence to ω_{DN}^* , as $\lambda/r_{max} \rightarrow 0$, is faster for
 256 higher stress levels (higher y_D).

257



258

259 **Supplementary Figure 3:** relative error between ω_{DN}^* and ω_{DN} (here $\theta = 2$ and $r_{max} = 0.5$), for different stress
 260 levels (scaled decay rates y_D indicated in legend).

261

262 **4. Effect of FGM parameters on the rate of rescue from *de novo* mutations:** Here we detail how
 263 each of the FGM parameters (r_D, λ, n, r_{max}) qualitatively affects the rate of rescue from *de novo*
 264 mutations. First, we note that both ψ_D and α (Eq.(6) of the main text) are increasing functions
 265 of r_D , while $g(\cdot)$ (Eq.(7) of the main text) is a decreasing function of α and the factor
 266 $(1 + \psi_D/2)^{1/2-\theta}/(1 + \psi_D/4)$ (Eq.(7) of the main text) is a decreasing function of ψ_D . Overall,
 267 when r_D increases, the rate of ER (Eq.(A12)) decreases, and so does the ER probability. Second,
 268 the only effect of decreasing the variance of mutational effects (λ) is to increase α and thus to
 269 decrease the ER probability. Third, the effect of dimensionality ($\theta = n/2$) is straightforward:
 270 increasing n decreases the factor $(1 + \psi_D/2)^{1/2-\theta}$, thus decreasing the ER probability. Finally,
 271 the effect of the fitness peak height (r_{max}) is less obvious from the formula, as increased r_{max}
 272 decreases ψ_D but increases ρ_{max} (and hence potentially α).

273 In fact, from the definitions in Eq.(A12) and replacing by $\psi_D = 2(\sqrt{1 + y_D} - 1)$, $\partial_{r_{max}} \alpha =$
 274 $-(\sqrt{1 + y_D} - 1)^2 / (\sqrt{1 + y_D} \lambda) < 0$. Thus, increased r_{max} decreases both α and ψ_D and thus
 275 increases the ER probability.

276

277 **5. Approximate ER rate ω_{SV} from standing variance:** We can follow the same approach used to
 278 approximate ω_{DN} to compute an approximation for ω_{SV} in Eq.(A7). Expressed in terms of ψ ,

279 we find that $v(y(\psi)) = q(\psi) \rho_{max}$ (with $v(\cdot)$ from Eq.(A6) and $q(\cdot)$ from Eq.(A10)). The
 280 integral in Eq.(A7) thus becomes:

$$\omega_{SV} = U \mathbb{E}_\psi \left(\frac{\pi(y(\psi))}{c_H(v(y(\psi)))} \right) \approx \frac{U}{\lambda} \frac{\sqrt{\rho_{max}}}{2\sqrt{\pi}} \int_0^2 h(\psi) E_{\theta-1/2}(\rho_{max} q(\psi)) d\psi, \quad (A13)$$

281 with $h(\cdot)$ given in Eq.(A11).

282 The approximation is in two steps. In a first step, we find an asymptotic expression for
 283 the exponential integral function as $\rho_{max} \rightarrow \infty$, via the Laplace method. By the definition of
 284 this function, we have $E_{\theta-1/2}(\rho_{max} q(\psi)) = \int_1^\infty e^{-\rho_{max} q(\psi) u} u^{1/2-\theta} du$. With $0 \leq \psi \leq 2$,
 285 we have $\alpha \leq \rho_{max} q(\psi) \leq \alpha(2 + \psi_D/2)^2$ where $\alpha = \psi_D^2 \rho_{max}/4$ as given in Eq. (A12).
 286 Assume that $\rho_{max} \rightarrow \infty$, but further conditioning on ψ_D non-vanishing, e.g. letting $\lambda \rightarrow 0$ with
 287 r_D and r_{max} held constant. These criterions guarantee that α is large, so that $\rho_{max} q(\psi) \geq \alpha$
 288 is large too. We can then use the Laplace method as $\rho_{max} q(\psi) \rightarrow \infty$, to approximate the
 289 integral $E_{\theta-1/2}(\rho_{max} q(\psi))$. We apply the approximation around the minimum of the
 290 exponential term in the integrand ($e^{-\rho_{max} q(\psi) u}$), over the integration domain $u \in [1, \infty]$,
 291 namely around $u = 1$. Using the approximation $u^{1/2-\theta} \approx e^{(\theta-1/2)(u-1)}$ in the vicinity of $u =$
 292 1 , we get the following approximation for the exponential integral term:

$$E_{\theta-1/2}(\rho_{max} q(\psi)) \xrightarrow{\rho_{max} \psi_D^2/4 \rightarrow \infty} \frac{e^{-\rho_{max} q(\psi)}}{\theta - 1/2 + \rho_{max} q(\psi)}. \quad (A14)$$

293 In a second step, we now plug this asymptote into Eq.(A13). We retrieve the required form for
 294 our integral to apply the Laplace method as in **section III subsection 3** for the whole expression
 295 of ω_{SV} :

$$\omega_{SV} \xrightarrow{\rho_{max} \psi_D^2/4 \rightarrow \infty} \frac{U}{\lambda} \frac{\sqrt{\rho_{max}}}{2\sqrt{\pi}} \int_0^2 \eta(\psi) e^{-\rho_{max} q(\psi)} d\psi, \quad (A15)$$

$$\eta(\psi) = \frac{h(\psi)}{\theta - 1/2 + \rho_{max} q(\psi)}$$

296 We thus apply the exact same method as in Eq. (A12) with the leading order for $\eta(\psi)$,
 297 when $\rho_{max} \rightarrow \infty$, in the vicinity of $\psi = 0$, given by $\eta_*(\psi) = 2 r_{max} \psi (1 + \psi_D/2)^{1/2-\theta}/(\alpha +$
 298 $\theta - 1/2)$. The resulting asymptotic approximation for ω_{SV} , as $\rho_{max} \rightarrow \infty$ (with non-

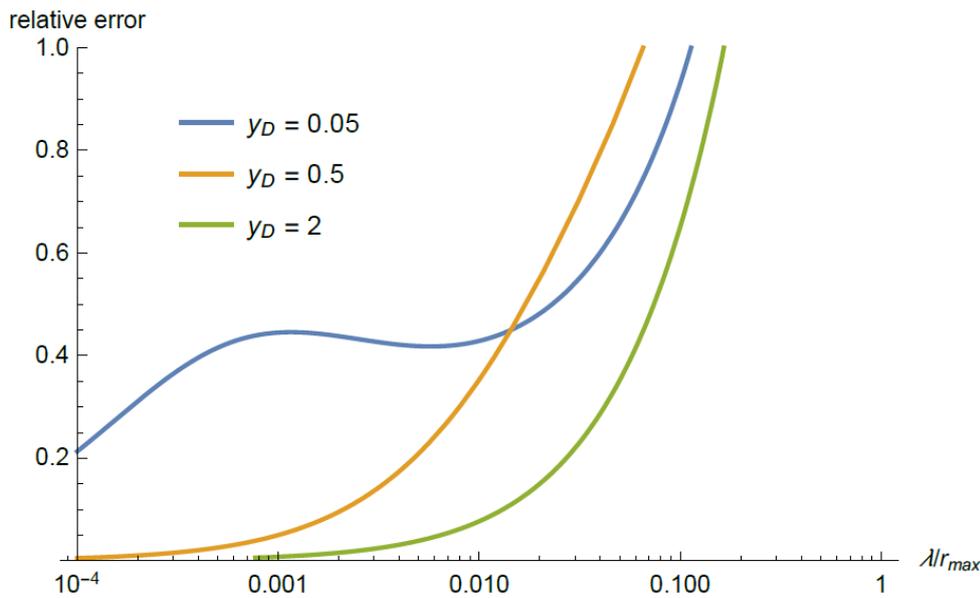
299 vanishing ψ_D), then satisfies the following relationship with the asymptotic ER rate from *de*
 300 *novo* mutations (ω_{DN}^* in Eq. (A12)):

$$\omega_{SV} \xrightarrow{\rho_{max} \psi_D^2/4 \rightarrow \infty} \omega_{SV}^* = \frac{U}{\lambda} \frac{\sqrt{\rho_{max}}}{2\sqrt{\pi}} \int_0^\infty \eta_*(\psi) e^{-\rho_{max} q(\psi)} d\psi \quad (A16)$$

$$\omega_{SV}^* = \omega_{DN}^* \frac{1 + \psi_D/4}{\epsilon/\psi_D + \psi_D/4} \text{ with } \epsilon = \frac{\theta - 1/2}{\rho_{max}}$$

301 The convergence of the ER rate ω_{SV} to its SME approximation (ω_{SV}^* , Eq.(A16)) is illustrated in
 302 **Supplementary Fig.4**. The convergence pattern is slightly more complex than for ω_{DN} , especially
 303 at low stress levels (e.g. with $y_D = 0.05$ in the figure): indeed, this SME limit now requires
 304 both $\rho_{max} \rightarrow \infty$ and ψ_D non-vanishing (i.e. $\alpha \rightarrow \infty$). Logically, convergence is again faster with
 305 higher stress levels (higher y_D).

306



307

308 **Supplementary Figure 4:** same as **Supplementary Fig.3** but for the relative error between ω_{SV} and ω_{SV}^* .

309

310 IV. Some key properties of the model

311 **1. Characteristic stress level:** The relationship between the parameters characterizing the stress
 312 (r_{max}, r_D, λ) and the rate of rescue shows a sharp drop from no extinction to nearly certain
 313 extinction. We here derive a heuristic characterization of this behavior, for *de novo* rescue
 314 (rescue from standing variance is discussed in a later section). We start by the heuristic behavior

315 suggested by Eq.(A12), in the limit of mild decay (Eq. 7b, main text): whenever $\psi_D \ll 2$, $\omega_{DN}^* \approx$
 316 $U g(\alpha)$. Second, we further simplify the model by taking a series expansion of $g(\alpha) \approx$
 317 $e^{-\alpha} \alpha^{-3/2} / 2\sqrt{\pi}$, when α is large. This approximation yields a relative error of less than 20% as
 318 long as $\alpha \geq 7$. Note that this approximation can apply (i.e., alpha can be that large) even for
 319 mild stress $\psi_D \ll 2$, as long as ρ_{max} is large enough.

320 We define α_p of level p by the set of parameter values such that $\alpha = \alpha_p$ and $P_R = p$ a
 321 given ER probability. Under the approximate heuristic derived above, α_p is characterized
 322 by $p = P_R \approx 1 - \exp(-N_0 U g(\alpha_p))$, with $g(\alpha) \approx e^{-\alpha} \alpha^{-3/2} / 2\sqrt{\pi}$. This implies $g(\alpha_p) =$
 323 $-\log(1 - p) / N_0 U$, and inversion of $g(\cdot)$ then yields

$$\alpha_p \approx \frac{3}{2} \mathcal{W} \left(\left(\frac{2}{\pi} \right)^{1/3} \frac{1}{3} \left(\frac{N_0 U}{\log(1/(1-p))} \right)^{2/3} \right), \quad (\text{A17})$$

324 where $\mathcal{W}(\cdot)$ is Lambert's ('productlog') function. A linear regression of $\mathcal{W}(x)$ vs. $\log(x) - 1$
 325 (checked by visual inspection, see **Supplementary file S1**) suggests that, over a biologically
 326 relevant range $\in [10, 10^{12}]$: $\mathcal{W}(x) \approx 0.9(\log(x) - 1)$. This yields the approximation (see
 327 numerical check in **Supplementary file S1**)

$$\alpha_p \approx 0.9 \left(\log(N_0 U) - \log \left(\log \left(\frac{1}{1-p} \right) \right) \right) - 3. \quad (\text{A18})$$

328 A characteristic stress level α_c can be defined as the value of α where the ER probability is
 329 50%: $\alpha_c = \alpha_{1/2}$. It characterizes the level of stress about which rescue drops from highly likely
 330 to highly unlikely. Setting $p = 1/2$ in Eq.(A18), the characteristic stress is approximately

$$\alpha_c \approx 0.9 \log(N_0 U) - 2.7. \quad (\text{A19})$$

331

332 **2. Self-consistency at large $N_0 U$:** We have used both (i) a large α approximation and (ii) a
 333 small $\psi_D \ll 2$ approximation (Eq. 7b) to derive the characteristic stress in Eq.(A19). We argue
 334 that they are self-consistent as long as $N_0 U$ is large.

335 We have (Eq.(A19)) $\alpha_c \approx 0.9 \log(N_0 U) - 2.7$, which is indeed large provided $N_0 U$ is large. We
 336 have seen that approximating $g(\alpha) \approx e^{-\alpha} \alpha^{-3/2} / 2\sqrt{\pi}$ should be reasonably accurate (<20%
 337 relative error) as long as $\alpha_c \geq 7$, which corresponds to $N_0 U \geq 5 \cdot 10^4$, a condition quite easily

338 met in microbial experiments, for example. Furthermore, as $\alpha_c = \rho_{max}(\psi_D^c)^2/4$ (Eq.(A12)), the
 339 corresponding characteristic ψ_D is equal to $\psi_D^c = 2\sqrt{(0.9 \log(N_0U) - 2.7)/\rho_{max}}$ (Eq.(A19)
 340 and is indeed negligible relative to 1 as long as $N_0U \ll 20 e^{\rho_{max}}$. This second criterion is in fact
 341 so easily met that it does not constrain the results; for example, with $\rho_{max} = 50$, the
 342 requirement is simply that $N_0U \ll 10^{25}$! Overall, it appears that Eq.(A19) is self-consistent
 343 whenever N_0U is large.

344

345 **3. Characteristic stress window:** Around the characteristic stress, the ER probability falls off
 346 more or less sharply. We define a characteristic stress window of level q over which P_R drops
 347 from $1/2 + q$ to $1/2 - q$. As an illustration, we use $q = 0.25$, so that the window characterizes
 348 the drop from 75% to 25% ER. This window can be directly computed from Eq.(A17) as $\Delta\alpha =$
 349 $\alpha_{1/4} - \alpha_{3/4}$. It is also approximately given by the inverse of the slope of the ER probability
 350 with α , at $\alpha = \alpha_c = \alpha_{1/2}$, namely: $\Delta\alpha \approx 2q/|P'_R(\alpha_c)| = 1/(2|P'_R(\alpha_c)|)$, with $q = 1/4$.
 351 Letting $P_R(\alpha) = 1 - e^{-N_0Ug(\alpha)}$ (Eq. 7b, main text), and using the large α approximation to $g(\cdot)$
 352 ($g(\alpha) \approx e^{-\alpha} \alpha^{-3/2}/2\sqrt{\pi}$), we have $g'(\alpha) \approx -g(\alpha)(1 + 3/(2\alpha))$, so that $P'_R(\alpha_p) = (1 -$
 353 $p) \log(1 - p) (1 + 3/2\alpha_p)$ for any level p . Setting $p = 1/2$ so that $\alpha = \alpha_c$ we have

$$\Delta\alpha \approx \frac{1/2}{|P'_R(\alpha_c)|} \approx \frac{2 \alpha_c}{\log 8 + \log 4 \alpha_c} \approx \frac{\alpha_c}{1 + 0.7 \alpha_c}. \quad (\text{A20})$$

354 The width of the window can be scaled by the value of the characteristic stress α_c around which
 355 the drop occurs, in order to characterize how sharp the drop is, as is done in Eq. (9) of the main
 356 text. Obviously, this simple heuristic, based on a linear approximation for P_R , gets more
 357 accurate over narrower windows, e.g. it is very accurate for describing the decay from 70% to
 358 30%, and less accurate for describing the decay from 95% to 5%.

359

360 **4. Proportion of rescue from standing variance:** The result in Eq. (A16) shows the relationship
 361 between ω_{SV}^* and ω_{DN}^* in the SME approximation. Then, from Eq.(A3), the proportion ϕ_{SV}
 362 converges to a simple limit ϕ_{SV}^* :

$$\phi_{SV}^* = \frac{\omega_{SV}^*}{\omega_{DN}^* + \omega_{SV}^*} = \frac{1 + \psi_D/4}{\epsilon/\psi_D + 1 + \psi_D/2} \quad (\text{A21})$$

363 Obviously, in the limit where $\epsilon \ll \psi_D$ (with ϵ in Eq.(A16)), the proportion ϕ_{SV}^* simplifies
 364 to $1/2 + 1/(2 + \psi_D)$, being always above 50%, and decreasing with higher stress-scaled decay
 365 rates.

366 For non-vanishing ϵ/ψ_D , as ψ_D varies, the proportion ϕ_{SV}^* reaches a maximum at the unique
 367 positive ψ_D where $\partial\phi_{SV}^*/\partial\psi_D = 0$, which is at $\psi_D = \psi_D^0 = \epsilon + \sqrt{\epsilon(4 + \epsilon)}$. At that point, its
 368 value is

$$\max \phi_{SV}^* = \phi_{SV}^*(\psi_D^0) = \frac{1 - \epsilon/2 - \sqrt{\epsilon(1 + \epsilon/4)}}{1 - 2\epsilon} = 1 - \sqrt{\epsilon} + o(\sqrt{\epsilon}). \quad (\text{A22})$$

369

370 **5. Stability of ϕ_{SV}^* across a range of stress:** In fact, as P_R drops sharply with ψ_D (or y_D), the
 371 proportion ϕ_{SV}^* happens to be fairly stable across stress levels, if stress only affects decay rates.
 372 To see this, we can derive the curvature, as ψ_D varies, of ϕ_{SV}^* with respect to the log of the
 373 extinction probability $|\log P_E|$, in the presence of *de novo* mutation and standing variance. We
 374 study this curvature around the value of ψ_D where the proportion is maximal, namely around
 375 $\psi_D = \psi_D^0$. It is equivalent and proves convenient to study this curvature by defining the two
 376 quantities as functions of α , and studying the curvature when α varies. Because $\psi_D^0 = O(\epsilon)$,
 377 we can use the small ψ_D approximation (Eq. 7b) $\omega_{DN}^* \approx Ug(\alpha)$. From Eq. (A16) and using $\psi_D =$
 378 $\sqrt{\alpha/\rho_{max}}$, we can write the ratio of ER rates as

$$\frac{\omega_{SV}^*}{\omega_{DN}^*} = 2 \frac{\sqrt{\alpha \rho_{max}}}{\alpha + \theta - 1/2} \left(1 - \frac{\psi_D}{4}\right), \quad (\text{A23})$$

379 Again, as we study this ratio in the vicinity of $\psi_D^0 = O(\epsilon)$, we can ignore the factor $1 - \psi_D/4$
 380 in Eq.(A23) and compute the proportion of ER from standing variance as:

$$Y(\alpha) = \phi_{SV}^* \approx \frac{2\sqrt{\alpha \rho_{max}}}{\alpha + \theta - 1/2 + 2\sqrt{\alpha \rho_{max}}}. \quad (\text{A24})$$

381 This is maximal at $\alpha = \theta - 1/2$, which is consistent with the expression for $\alpha = \rho_{max}\psi_D^2/4$
 382 when using the leading order for $\psi_D^0 = 2\sqrt{\epsilon} + o(\sqrt{\epsilon})$. The total ER rate from de novo mutations
 383 plus standing variance ($\omega^* = \omega_{DN}^* + \omega_{SV}^*$) is also simplified, once we ignore the factor $1 -$

384 $\psi_D/4$ in Eq.(A23). This yields a relatively simple form for the log of the extinction probability
 385 (recalling that $P_E = e^{-N_0\omega^*}$):

$$X(\alpha) = |\log P_E| \approx N_0 U (\omega_{DN}^* + \omega_{SV}^*) \approx N_0 U g(\alpha) \left(1 + \frac{2\sqrt{\alpha \rho_{max}}}{\alpha + \theta - 1/2} \right). \quad (A25)$$

386 The two quantities $(X(\alpha), Y(\alpha))$ define a parametric curve as α varies, with a maximum in $Y(\cdot)$
 387 at $\alpha = \theta - 1/2$. The curvature of $Y = \phi_{SV}^*$ with $X = \log P_E$, at this point α_0 , is given by (see
 388 e.g. Goldman 2005)

$$\kappa = \frac{X'(\theta - 1/2)Y''(\theta - 1/2) - Y'(\theta - 1/2)X''(\theta - 1/2)}{(X'(\theta - 1/2)^2 + Y'(\theta - 1/2)^2)^{3/2}}, \quad (A26)$$

389 using the expressions for $X(\cdot)$ and $Y(\cdot)$ in Eqs(A24) and (A25), we get

$$\kappa = \frac{e^{2(\theta-1/2)} \pi (\theta - 1/2)^{5/2} \sqrt{\rho_{max}}}{N_0^2 U^2 (\sqrt{\theta - 1/2} + \sqrt{\rho_{max}})^4}. \quad (A27)$$

390 This curvature is obviously very small, of order $1/(N_0^2 U^2 \rho_{max}^{3/2})$. This flatness of the
 391 proportion ϕ_{SV}^* with $|\log P_E|$ (extinction probability on a log-scale), shows formally that it is
 392 almost unaffected by changes in decay rates, over a wide range of stress levels, spanning
 393 several orders of magnitude of change in P_E (or equivalently P_R).

394

Bibliography

- Breitung K. W., 1994 *Asymptotic approximations for probability integrals*. Springer-Verlag, Berlin ; New York.
- Gillespie J. H., 1983 Some Properties of Finite Populations Experiencing Strong Selection and Weak Mutation. *Am. Nat.* 121: 691–708.
- Goldman R., 2005 Curvature formulas for implicit curves and surfaces. *Comput. Aided Geom. Des.* 22: 632–658.
- Martin G., Aguilée R., Ramsayer J., Kaltz O., Ronce O., 2013 The probability of evolutionary rescue: towards a quantitative comparison between theory and evolution experiments. *Phil Trans R Soc B* 368: 20120088.
- Martin G., Lenormand T., 2015 The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. *Evolution* 69: 1433–1447.

McCandlish D. M., Stoltzfus A., 2014 Modeling Evolution Using the Probability of Fixation: History and Implications. *Q. Rev. Biol.* 89: 225–252.

Wolfram Research, 2001 Confluent hypergeometric function $0F1$: Series representations.