Electronic Supplementary Material for

"Stress-Induced Mutagenesis and Complex Adaptation"

Yoav Ram and Lilach Hadany

# Figure reproduction

All the figures were produced using [Python](https://www.python.org/) [1] with the [NumPy](http://www.numpy.org/) [2], [SciPy](http://scipy.org/scipylib/index.html) [3], [Matplotlib](http://matplotlib.org/) [4], [Pandas](http://pandas.pydata.org/) [5], [scikits-bootstrap](https://github.com/cgevans/scikits-bootstrap), [statsmodels](http://statsmodels.sourceforge.net/), and [mpltools](http://tonysyu.github.io/mpltools/) packages. The code is available as an [IPython](http://ipython.org/) Notebook [6] on the [Proceedings of the Royal Society B](http://rspb.royalsocietypublishing.org/) website as an Electronic Supplementary Material. The notebook includes the analytic approximations as Python functions and uses the simulation raw data which is necessary for Figures 2, 3, S1, S2, and F1.

The raw data is available on [Dryad](http://datadryad.org/) (DOI: [10.5061/dryad.3066j](http://dx.doi.org/10.5061/dryad.3066j)).

# Appendix A: Appearance of a double mutant

In the following analysis we assume that and – see model overview for details. This also means that and . The probability *q* that a random offspring in the next generation is *AB* given there are no *AB* in the current generation can be approximated by:

.

Using the above assumptions, this resolves to:

.

Taking the derivative with respect to *U* and denoting (*g* can be thought of as the number of non-specific loci in the genome):

.

So *q* increases with *U* because the right hand side is guaranteed to be true under the assumption .

For a population with SIM

.

The last approximation assumes that *.* Rearranging the last result, we find the approximation

|  |  |
| --- | --- |
| . |  |

Taking the derivative with respect to *τ*,

,

because *q*, *U*, and *τ* are all positive. So the condition guarantees that *qSIM*increases with *τ*, and it is also sufficient for (not shown).

# Appendix B: Fixation of a double mutant

Following Eshel [7], the fixation probability *ρ* of the double mutant *AB* is

,

where *α* is the fitness of the double mutant relative to the population mean fitness and assuming that fitness is measured by the average number of progeny which is Poisson distributed:

.

Here, we only consider progeny without new deleterious mutations; their fraction is *e-U*. This factor cannot be ignored because there is variation in mutation rates within the population (see "minor technical point" by Johnson and Barton [8]). At this stage, double mutants are still very rare, so we can use the population mean fitness at the MSB. The population mean fitness can be approximated by (see Online Appendix D). Therefore,

Assuming *sH* is small () we can approximate this by

.

# Appendix C: Fixation of a double mutant with SIMe

With SIMe the mutation rate of *ab* is *τU* while that of *AB* is only *U*. We assume the population reached a MSB before the fixation of *AB* because convergence to MSB [9] is much faster than adaptation. Following the derivation in Appendix B, the relative fitness of SIMe is

.

Plugging that in the fixation probability,

.

This can be simplified by a 1st order approximation for :

.

Because , the right hand side is greater than 1. Therefore,

.

Because the appearance with SIMe is the same as with CM, the adaptation rate with SIMe can be written as

.

# Appendix D: Mean fitness at the mutation-selection balance

Denote the frequency of individuals with *x* deleterious alleles by *fx*. The frequency of such individuals in the next generation *f'x* is given by

|  |  |  |
| --- | --- | --- |
|  | , |  |

where *mx,y* is the transition probability from *y* deleterious alleles to *x* deleterious alleles and is the population mean fitness.

The term *mx,y* consist of the fitness of individuals with *y* deleterious alleles, *ωy*, and the probability that the precise number of mutations occurred. Specifically, if *y≥x* then exactly *y-x* beneficial mutation must occur; if *y≤x* then exactly *x-y* deleterious mutations must occur:

.

Using the probability mass function of a Poisson distribution, we can expand the above equation to

,

where *ωy* is the fitness of individuals with *y* deleterious alleles, is the population mean fitness (), *δ* and *β* are the fraction of mutations that are deleterious and beneficial, respectively (*δ+β*=1 and 0≤*β*<*δ*≤1*)*, and *Uy*is the average number of new mutations per generation in an individual with *y* deleterious alleles.

This can be written as a matrix equation by multiplying the frequencies vector *f* and the mutation-selection matrix *M*:

At the mutation-selection balance (MSB), *f\** solves the equation (a star \* denotes equilibrium quantities)

.

Without beneficial mutations (*δ*=1 and β=0), the above equation simplifies to

and *M* is a triangular matrix. In this case the population mean fitness can be found by solving the equation for *f0*:

,

which means that the population mean fitness is equal to the product of the fitness of mutation-free individuals and the probability that a mutation-free individual does not mutate. If and (constant uniform mutation rate) then [10] and by the forward substitution method the frequencies vector is

,

that is, the number of deleterious mutations per individual is Poisson distributed with average *U/s* [11]. With constitutive mutagenesis (CM), the population mean fitness at the MSB is : it decays exponentially as a function of *τ* the mutation rate fold increase. In contrast, stress-induced mutagenesis (SIM), as shown by Agrawal [12], does not change the population mean fitness with respect to normal mutagenesis (NM). This is because the least loaded individuals (*x*=0), with fitness *ω*0=1, also have the lowest mutation rate, *U*, and therefore the population mean fitness is .

With beneficial mutations (β>0), the matrix *M* is a positive matrix, and by the *Perron-Frobenius Theorem* (Otto and Day 2007, p. 709) is the largest eigenvalue of *M* and *f\** is its unique positive eigenvector with .

This eigenvalue problem is hard to solve analytically, however, by neglecting elements outside the main three diagonals of *M* we have shown before [14] that the population mean fitness increases with the mutation rate of individuals with a below–average fitness:

.

Nevertheless, this framework allows the numerical calculation of the population mean fitness for finite *n*-by-*n* mutation-selection matrices by defining *n* such that . The mean fitness of populations with different mutational strategies is then calculated by manipulating *Ux*.

Figure D1 shows that is a good approximation to the population mean fitness (because ), and that SIM slightly increases the population mean fitness with respect to NM; a sufficient condition is that the mutation rate of individuals with below average fitness is increased [14]. Since we assume that ,then . Therefore, for SIM to increase the population mean fitness it must increase the mutation rate in individuals with at least one deleterious mutation.

****

**Figure D1 – Mean fitness at the mutation-selection balance with stress-induced mutagenesis.** The brightness represents the fitness advantage of stress-induced mutagenesis over normal mutagenesis at the mutation-selection balance. The x-axis is the fraction of mutations that are beneficial *β*. The y-axis is the mutation rate fold increase under stress *τ*. "X" marks the parameter set *β*=1/5000 and *τ*=10, in which the fitness advantage of SIM is ~5⋅10-9.

# Appendix E: Possible relationships between stress and mutation

In the main text we used a threshold relationship between stress and mutation: if fitness drops below a threshold (<1 for SIM, ≤1 for SIMe), the mutation rate increases *τ*-fold. But the relationship between stress and mutation can be more complex. For example, Agrawal [12] has used a continuous relationship defined by a curvature parameter *k*. This relationship defines the mutation rate for an individual with fitness *ω*, baseline mutation rate *U*, and a maximal mutation rate fold increase *τ* as

.

When *k* approaches 0 this expression approaches *U*, corresponding to the NM strategy. When *k* approaches infinity this expression approaches eq. 1(1), corresponding to the SIM threshold strategy. See Figure E1 for a plot of these continuous relationships for various values of *k*.

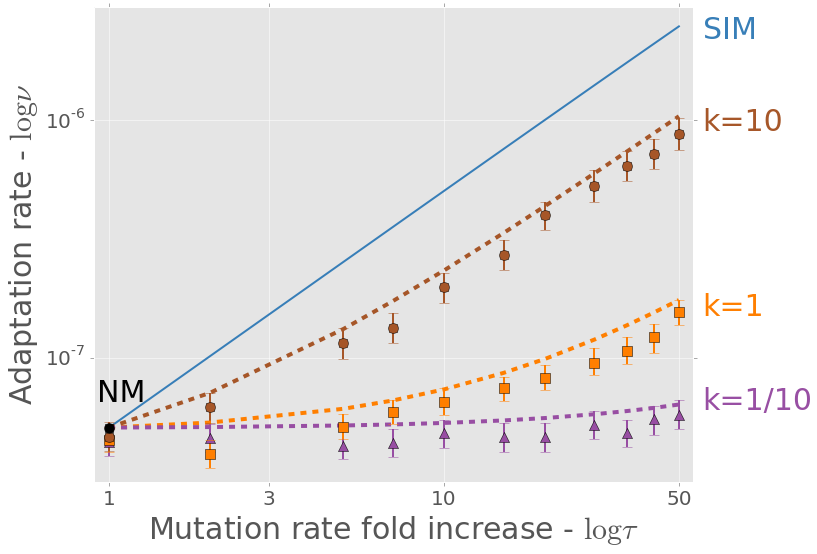
Figure E2 shows the adaptation time for three continuous strategies (*k*=1/10, 1, and 10). Remarkably, the dynamics of a continuous strategy can be approximated by a threshold strategy by matching the mutation rates of single mutants:

.

This is equivalent to using a threshold strategy with mutation rate increase *τ*-(*τ*-1)(1-*s*)*k*. The dashed lines in Figure E2 demonstrate this approximation. The continuous strategies can be approximated by threshold strategies because the main factor determining the adaptation rate is the mutation rate increase of the wildtype and the single mutants (*ab*, *aB*, and *Ab*). This is because individuals with more than a single mutation do not have a significant contribution to adaptation.



**Figure E1– Different relationships between stress and mutation.** The figure shows continuous relationships between fitness (x-axis) and mutation rate (y-axis) in solid lines and threshold relationships in dashed lines. The threshold relationship is defined in section 2 of the main text. The continuous relationships are defined in Supporting Text S4. Each panel shows a pair of relationships, with *k* increasing from 1/10 (convex relationship), to 1 (linear relationship) to 10 and 100 (concave relationships). Each continuous relationship is compared with a threshold relationship that has the same mutation rate for wildtype (*ab/0*) and single mutants (*Ab/0*, *aB/0*, *ab/1*). Figure S5 shows that the adaptation rate with such threshold relationship approximates the adaptation rate with a continuous relationship.



**Figure E2 – Complex adaptation with continuous relationship stress-induced mutagenesis.** The figure shows the adaptation rate *ν* as a function of the mutation rate increase *τ* (both in log scale). The solid line is an analytical approximation of SIM (same as in Figure 2). Markers are the results of simulations of adaptation with SIM with different continuous relationships between fitness and mutation rate. These continuous relationships are defined by a mutation rate fold increase *τ*=10 and a curvature parameter *k*=10, 1, and 1/10, top to bottom**.** Each dashed line is an approximation of a continuous SIM using a SIM threshold strategy with *τ*=4.61, 1.45, and 1.05, top to bottom. The fit between the dashed lines and the corresponding markers suggests that a threshold strategy captures the adaptive dynamics. Error bars represent 95% confidence interval of the mean (at least 1,000 simulations per point; computed with bootstrap with 1,000 samples per point). Parameters (see Table 1): *U*=0.0004, *s*=0.05, *β*=0.0002, *H*=2, *N*=106.

# Appendix F: Competitions between mutational strategies

We also simulated direct competitions between the different mutational strategies (NM, CM, and SIM). In these competitions, half of the population alters its mutational strategy to an invading strategy at the time of the environmental change. Each simulation provides a sample of the frequency of the invading strategy after the appearance and subsequent fixation or extinction of the double mutant *AB*. If the average final frequency is significantly lower or higher than 50% we consider the invading strategy disfavored or favored by natural selection over the initial strategy. Statistical significance was calculated using a 1-sample 2-tailed t-test.

Figure F1 summarizes the competitions. CM clearly loses to both SIM and NM (first and second panels from the right). SIM is significantly advantageous over NM when the mutation rate increase is large enough (*τ*>2; 2-tail t-test, P<0.0015).

These results show that the evolutionary advantage of SIM at the population-level corresponds to an individual-level advantage and can lead to the evolution of stress-induced mutagenesis by natural selection, even when constitutive mutagenesis is strongly disfavored. This is consistent with previous results in smooth fitness landscapes [14].

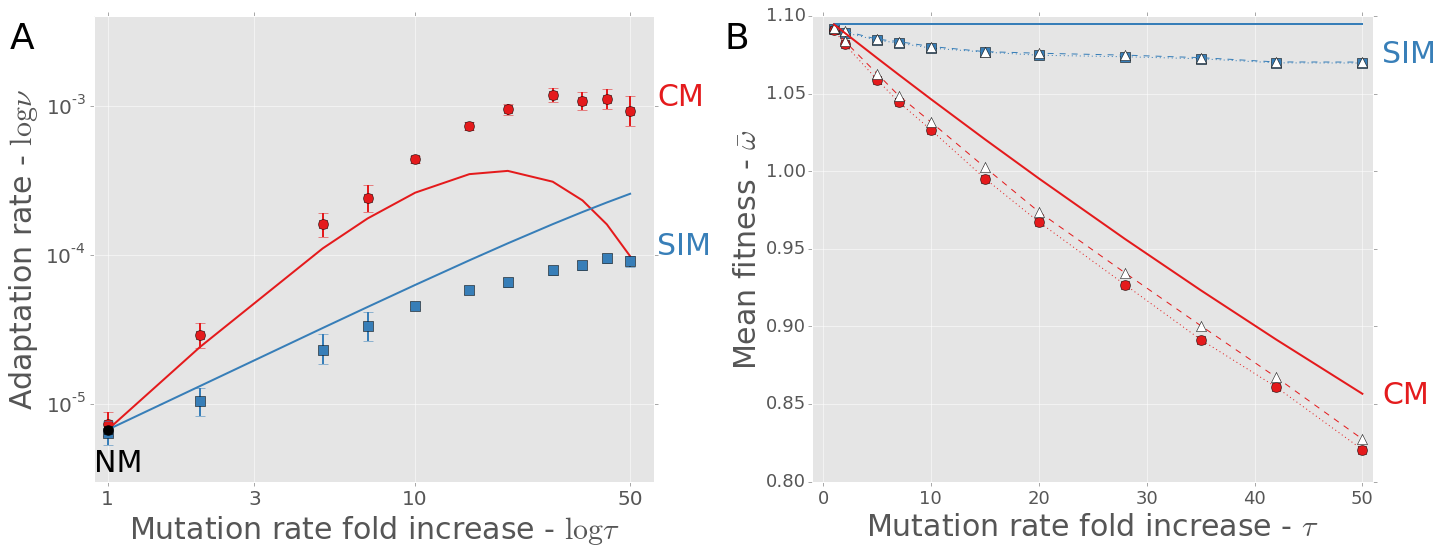


**Figure F1 – Direct competitions between three mutational strategies.** The figure shows the average final frequency of (from right to left): stress-induced mutagenesis (SIM) vs. constitutive mutagenesis (CM); CM vs. normal mutagenesis (NM); SIM vs. NM; and NM vs. NM (control). Initial frequencies are always 0.5. Several mutation rate fold increases are shown on the x-axis. SIM defeats CM and is significantly advantageous over NM when *τ*>2 (2-tail t-test, P<0.0015). CM losses to NM and SIM (P≈0). Therefore, SIM is favored by selection over both NM and CM. Changing roles between resident and invader didn't affect the results (not shown). Error bars represent the standard error of the mean (at least 100 simulations per point). Parameters (see Table 1): *U*=0.0004, *s*=0.05, *β*=0.0002, H=2, *N*=106.

# Appendix G: Strong mutation regime

What happens when mutation is as strong as selection? Figure G1A shows results for *s*=10*U*. When the average number of deleterious alleles per individual *τU/s* is over one, adaptation with CM is likely to occur on a deleterious background. Because our approximation neglects adaptation on deleterious backgrounds, it underestimates the adaptation rate (Figure G1A). Note that although the adaptation rate continues to increase with *τ*, the population carries more deleterious alleles after adaptation, resulting in a lower population mean fitness (Figure G1B) and eventually a lower fixation probability and adaptation rate (Figure G1A).

With SIM, the average number of deleterious alleles per individual *U/s* does not increase with *τ*, because mutation-free individuals (*ab/0*) do not hypermutate. As in the case of weak mutation, when *τ*>10 the double mutant can appear on a deleterious background, resulting in hitch-hiking and a lower fixation probability, and causing our approximation to overestimate the adaptation rate (Figure G1A).

****

**Figure G1 – Adaptation with strong mutation.** When the deleterious mutation rate is high – here *U*=*s*/10 – the adaptation process can lead to hitch-hiking of deleterious alleles with the beneficial double mutant. **(A)** The adaptation rate *ν* as a function of the mutation rate increase *τ* (both in log scale). A black circle for normal mutagenesis (NM; *τ*=1); red solid line and circles for constitutive mutagenesis (CM); blue solid line and squares for stress-induced mutagenesis (SIM). Lines are analytic approximations. Markers are the means of stochastic simulations results. Error bars represent 95% confidence interval of the mean (at least 1,000 simulations per point; computed with bootstrap with 1,000 samples per point). Parameters (see Table 1): *U*=0.005, *s*=0.05, *β*=0.0002, *H*=2, *N*=106. **(B)** The population mean fitness after successful fixation of the beneficial double mutant as a function of the mutation rate increase *τ*. Solid lines are analytic approximations neglecting adaptation from deleterious background (*e-τU(1+sH)*); dotted lines with filled squares (SIM) and circles (CM) are the means of stochastic simulation results; dashed lines with white triangles are predictions based on the genotype on which *AB* appeared in the simulations, including MSB but disregarding the effects of drift during the fixation process (which only has a significant effect with CM due to higher mutation rates in the wildtype). Error bars are too small to see. Same parameter values as in panel A.

# Supporting Figures



**Figure S1 – Waiting time for the appearance of a double mutant** as a function of the mutation rate fold increase *τ*. Normal mutagenesis (NM) is *τ*=1; constitutive mutagenesis (CM) in red; stress-induced mutagenesis (SIM) in blue. Lines are analytic approximations (eqs. 2-3 in main text). Markers are means of simulation results - black circles for the standard simulations, white triangles for alternative simulations in which *AB* cannot appear on deleterious backgrounds. The standard error of the mean was too small to show. At least 1,000 simulations per point. Both axes are in log scale. The appearance time decreases as a function of *τ*2 and *τ* with CM and SIM, respectively. Appearance time is slightly longer if *AB* only appears on unloaded background (white triangles) which explains the difference between the analytic approximations and the simulation results for SIM in Figure 2. Parameters are the same as in Figure 2.



**Figure S2 – Fixation probability** **of the double mutant *AB*** as a function of the mutation rate fold increase *τ* with three mutational strategies: constitutive mutagenesis (CM; top panels in red), stress-induced mutagenesis (SIM; middle panels in blue) and stress-induced mutagenesis with environmental stress (SIMe; bottom panels in green; see section 3.5 in main text). Dashed lines are analytic approximations; black error bars represent simulation results with 95% confidence interval of the mean (at least 1,000 simulations per point; computed with bootstrap with 10,000 samples per point); solid lines are the logistic regression lines computed from the simulation results. The three left panels are results of the standard simulations. The three right panels are results of simulations in which *AB* cannot appear on deleterious backgrounds - in these cases there is no significant difference between the simulation results and our analytic approximations (compare solid and dashed lines; regression slope tests with α=0.05). However, if *AB* can appear on a deleterious background (left panels) then its fixation probability is lower [8]. For example, the fixation probability of *AB* with a single deleterious mutation is . In addition, the figure shows that SIMe­ has a higher fixation probability than CM and SIM: the green lines, representing SIMe, are always higher than the red and blue lines representing CM and SIM. Parameters are the same as in Figure 2.

# References

1. Van Rossum, G. & others 2007 Python Programming Language. In *USENIX Annual Technical Conference*,

2. Van der Walt, S., Colbert, S. C. & Varoquaux, G. 2011 The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **13**, 22–30. (doi:10.1109/MCSE.2011.37)

3. Jones, E., Oliphant, T., Peterson, P. & others 2001 SciPy: Open source scientific tools for Python.

4. Hunter, J. D. 2007 Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95. (doi:10.1109/MCSE.2007.55)

5. McKinney, W. 2010 Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference* (eds S. van der Walt & J. Millman), pp. 51–56.

6. Perez, F. & Granger, B. E. 2007 IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **9**, 21–29. (doi:10.1109/MCSE.2007.53)

7. Eshel, I. 1981 On the survival probability of a slightly advantageous mutant gene with a general distribution of progeny size - a branching process model. *J. Math. Biol.* **12**, 355–362. (doi:10.1007/BF00276922)

8. Johnson, T. & Barton, N. H. 2002 The effect of deleterious alleles on adaptation in asexual populations. *Genetics* **162**, 395–411.

9. Gordo, I. & Dionisio, F. 2005 Nonequilibrium model for estimating parameters of deleterious mutations. *Phys. Rev. E* **71**, 18–21. (doi:10.1103/PhysRevE.71.031907)

10. Kimura, M. & Maruyama, T. 1966 The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–51.

11. Haigh, J. 1978 The accumulation of deleterious genes in a population - Muller’s Ratchet. *Theor. Popul. Biol.* **14**, 251–267. (doi:10.1016/0040-5809(78)90027-8)

12. Agrawal, A. F. 2002 Genetic loads under fitness-dependent mutation rates. *J. Evol. Biol.* **15**, 1004–1010. (doi:10.1046/j.1420-9101.2002.00464.x)

13. Otto, S. P. & Day, T. 2007 *A biologist’s guide to mathematical modeling in ecology and evolution*. Princeton University Press.

14. Ram, Y. & Hadany, L. 2012 The evolution of stress-induced hypermutation in asexual populations. *Evolution* **66**, 2315–28. (doi:10.1111/j.1558-5646.2012.01576.x)