Stress-induced mutagenesis and complex adaptations

Yoav Ram

and

Lilach Hadany

corresponding author, [lilach.hadany@gmail.com](mailto:lilach.hadany@gmail.com)

Dept. of Molecular Biology and Ecology of Plants, Tel Aviv University, Tel-Aviv 69978, Israel

Last update: 10 June 2013

# Summary (max 200)

# Keywords (3-6)

population genetics; mathematical modeling; evolvability; stress-induced variation; fitness landscape

# Short title

Stress-induced mutagenesis and complex adaptations

# Introduction

Stress-induced mutagenesis, the phenomenon in which stressed or maladapted individuals increase their mutation rate, has been demonstrated in numerous species, both prokaryote and eukaryote (1–3). The phenomenon is considered by many to have a meaningful impact on *evolvability* - the capacity of individuals and populations to adapt to environmental changes [REFS]. In a previous work we showed that stress-induced mutagenesis is favored by natural selection over constant rate mutagenesis in asexual populations and that it increases the mean fitness of populations due to the increased generation of beneficial mutations in maladapted individuals (4). Here, we focus instead on the effect of stress-induced mutagenesis on the evolution of complex traits.

Complex traits, coded by multiple genes, present an open evolutionary question, first described by Sewall Wright in 1931 (5): *if different alleles are separately deleterious but jointly advantageous, how can a population evolve from one co-adapted gene complex to a better one?* Or, in terms of adaptive landscapes (6), how can a population cross a fitness valley and shift from one adaptive peak to a higher one?

Wright suggested the "shifting-balance theory of evolution" (7), which is based on the division of the population into small sub-populations and relies on genetic drift and migration as complementary processes to mutation and selection. This solution is valid (8–10) but seems to be limited to specific parameter ranges (11–14). As a result, there is a disagreement if the "shifting-balance theory" is an important process in evolution (15,16).

Here, we analyze a population genetic model of an asexual population undergoing an adaptive peak shift in a rugged adaptive landscape. We derive analytical expressions that suggest that stress-induced mutagenesis increases the population adaptation rate and show the results of stochastic simulations that validate our analytic expressions. We then discuss possible predictions of our model and how it relates to the literature.

# Material and methods

## Analytical model

Consider two loci with alleles *A/a* and *B/b* and a population at a mutation-selection balance (MSB) in an environment in which *ab* is the wildtype with a fitness value of *1*, single mutants (*Ab* and *aB*) have a fitness value of *1-s*, with *s* as the selection coefficient, and double mutants (*AB)* have a fitness value of *(1-s)2*. This corresponds to a multiplicative fitness function *(1-s)m* where *m* is the number of deleterious mutations the individual has.

Mutation from *a* to *A* and from *b* to *B* occurs with a probability *µ* at reproduction (we neglect back-mutation) - *µ* is therefore the site-specific beneficial mutation rate. In addition, the number of new deleterious mutations that occur across the genome at reproduction follows a Poisson distribution with an average *U* - the genomic deleterious mutation rate. Likely, there is a direct relation between *U* and *µ* (we use *µ=U/5000*), but having two separate parameters helps to distinguish between the two effects of mutation: the generation of beneficial mutations (*µ*) and deleterious mutations (*U*). Individuals with stress-induced mutation and a fitness lower than 1 hypermutate, thereby increasing both mutation rates *τ*-fold. To incorporate the effects of random genetic drift into the model, we denote the population size by *N*.

At the MSB, the frequency of wildtype (*ab*) individuals is , the frequency of single mutants (*Ab* and *aB* combined) is and the frequency of double mutants (*AB*) is .

We are interested in the capacity of the population to adapt to an environmental change in which the fitness of the double mutant *AB* changes from *(1-s)2*  to *1+sH*, where *H* scales the advantage of *AB* in comparison with the disadvantage of single mutants. Figure 1a presents an illustration of the adaptive landscape of the analytical model.

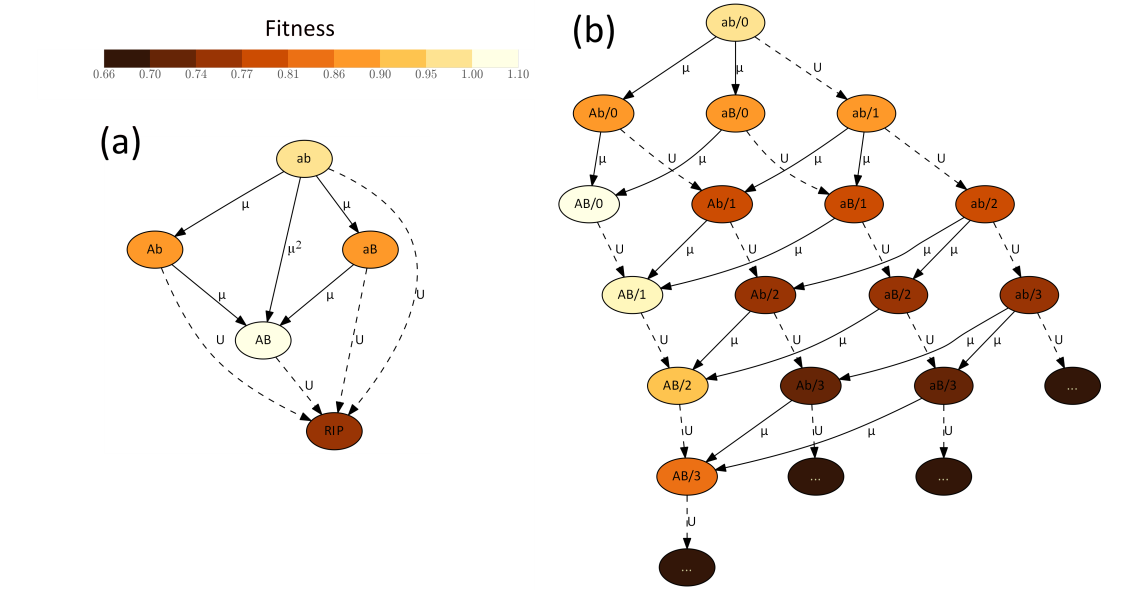


Figure 1 – Adaptive landscape illustration. Each node represent a specific genotype. Node labels specify the alleles at the *A/*a and *B/*b loci and the number of additional deleterious alleles after the forward slash (only in (b) and only as much as 3 deleterious mutations are shown to keep the figure simple). Arrows define the direction of mutation (solid for site-specific beneficial mutations with rate µ, dashed for genomic deleterious mutations with rate *U*). Node colour indicates the fitness of a genotype, from pale brown for optimal fitness to dark brown for lower fitness, see the colourbar. Parameters used: *s=0.01*, *H=2*. (a) In the analytical model individuals with deleterious mutations are considered evolutionary dead-ends (*RIP*) anddo not contribute to adaptation.(b) In the stochastic model we make no such assumption. In this panel only single mutations are shown for simplicity.

There are several constraints on the main parameters:

1. The above MSB approximations are only valid when *µ/s<1* or *µ<s*.
2. If *N*(*µ/s)2>1* then there are double mutants in the population at the MSB and therefore adaptation will be rapid and will not require new mutations.
3. If *Nµ/s<1* then there are no single mutants at the MSB and double mutatns must be generated by a double mutation in a wildtype individual. Therefore, increasing the mutation rate of individuals with fitness below 1 will have a much smaller effect than if single mutants were abundant.
4. If we assume that individuals that accumulated deleterious mutations are "evolutionary dead-ends" (figure 1a) and cannot be the origin of adaptation, then the fraction of such individuals must be small - *U/s<1*. This replaces the first constraint above because *µ* is much smaller than *U*.

To summarize the above constraints:

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |

For the bacteria species *Escherischia coli* estimations of the selection coefficient and mutation rates are *s=0.01*-*0.3* (17,18), *U=0.003-0.0004* (19,20) and *µ =10-6-10-9* (18). Taking the conservative alternative, we get a constraint on the population size - *105 ≤ N ≤ 10*7.

In the following derivations we assume both of the above constraints.

## Stochastic model

To validate our analytic approximations, we developed a Wright-Fisher simulation model with mutation, selection and random genetic drift. The main differences between the analytical model and the stochastic simulations are: (i) The simulations incorporate genetic drift by randomly sampling each generation from the previous one using a multinomial distribution. (ii) Individuals with deleterious mutations are not "evolutionary dead-ends" - individuals are allowed to accumulate up to 25 deleterious mutations (figure 1b). (iii) Simulations start with an *ab* mutation-free population. After reaching MSB, the environment is changed so that *AB* is advantageous. Therefore, the stochastic model assumes nothing about the distribution of deleterious mutations at the MSB. (v) We ran simulations in which the population size is lower or higher than in constraint (2).

We wrote the simulations with Python (<http://www.python.org>) using NumPy (<http://www.numpy.org>) and SciPy (<http://www.scipy.org>). The source code for the simulations is available on GitHub (<https://github.com/yoavram/ruggedsim>.git).

# Results

The adaptation process can be divided to two distinct processes: the appearance of a double mutant *AB* and the fixation of a double mutant *AB*. The first process mainly depends on mutation and is therefore much longer than the second one, which depends on the interplay between selection and genetic drift. In the following sections we provide expressions of the expected waiting time for appearance, fixation and adaptation. Expressions derived via first-order approximations are marked by a star (\*). The full derivations of the following expressions are given in the *Electronic Supporting Material*.

## Appearance of a double mutant

Because there are no double mutants (*AB*) at the time of the environmental change, double mutants can appear either via a double mutation in a wildtype individual (*ab*) or via a single mutation in a single mutant (*Ab* or *aB*). At the MSB the number of deleterious mutations follows a Poisson distribution (21,22). Therefore, the frequencies of mutation-free wildtype *ab* and single mutants aB and *Ab* are roughly and . Therefore, the probability *q* that a newborn is a double mutant can be summarized by:

|  |  |
| --- | --- |
|  | (3) |

However, if mutation is stress-induced, then the mutation rate of single mutants is increased *τ*-fold and the appearance probability is:

|  |  |
| --- | --- |
|  | (4) |

Note that stress-induction increases the transition from single mutants to other types, but does not significantly change the MSB frequency of single mutants, because this frequency is mainly determined by the mutation rate of the wildtype which does not hypermutate.

The above expressions can be approximated to the following simpler expressions:

|  |  |
| --- | --- |
|  | (5) |
|  | (6) |

Taking the derivative

we can see that if then increasing *τ* increases the appearance probability of the double mutant.

## Fixation of a double mutation

Assuming that the advantage of the double mutant is considerable (for example, *H>1)* and that the population size is large (constraint (2) ensures that), a double mutant has two possible fates after its appearance: fixation or extinction. Following Eshel (23) the fixation probability *ρ* of the double mutant is

|  |  |
| --- | --- |
|  | (7) |

where *α* is the fitness of the double mutant relative to the population mean fitness:

|  |  |
| --- | --- |
|  | (8) |

and assuming that fitness is measured by the number of progeny which is Poisson distributed. Here, mutation incurs a fitness cost of *e-U* – this is the fraction of progeny that do not have deleterious mutations. This factor cannot be ignored because the mutation rate is variable in the population.

Because the frequency of double mutants is very low at the stage where they are subject to possible extinction by genetic drift, the population mean fitness can be calculated without considering double mutants. Therefore, the value we use is the mean fitness of the population at the MSB. Without stress-induced mutation, this evaluates to (21). Therefore, and

|  |  |
| --- | --- |
|  | (9) |

Assuming that is small we get

|  |  |
| --- | --- |
|  | (10) |

which is a classic result in population genetics (23). However, as we have shown before (4), the mean fitness of a population with stress-induced mutagenesis can be different from *e-U* because of the variable mutation rate. Here, the mean fitness with stress-induced mutagenesis can be calculated by separating the population to the mutation-free fraction which has fitness *1* and the rest of the population with a fraction of . Within the mutation-loaded fraction, individuals have at least one deleterious mutation and additional mutations are Poisson distributed with expectation because these individuals are hypermutating. Therefore the mean fitness of this fraction is . Taken together, the mean fitness of a population with stress-induced mutagenesis is

|  |  |
| --- | --- |
|  | (11) |

Pluging this in (7) and (8) gives a different fixation probability for populations with stress-induced mutagenesis:

|  |  |
| --- | --- |
|  | (12) |

This can be further simplified and approximated to:

|  |  |
| --- | --- |
|  | (13) |

Because and this expression suggests that - stress-induced mutagenesis increases the fixation probability of the double mutant. This effect occurs because stress-induced mutagenesis causes maladapted individuals to accumulate more deleterious mutations then adapted individuals, creating a wider fitness distribution, and a higher relative fitness for adapted individuals. This effect increases with the mutation rate *U* and with the mutation rate fold-increase *τ* and decreases with selection *s*, but it does not depend on the site-specific beneficial mutaition rate *µ*. However, note that the derivative of with respect to *τ* is , so increasing *τ* only has a mild effect on the increase of .

## Adaptation rate

From the probability *q* that in a population without double mutants a random newborn is a double mutant we can derive the probability that some double mutants appear in the next generation: . The constraint (2) guarantees that *Nq* is very small and this probability can be approximated by *Nq*.

Once a double mutant appears it has a probability *𝜌* to go to fixation.

The time for adaptation *Ta* can be approximated by the waiting time for a double mutant which will go to fixation *Tw*. This is true as long as fixation is much faster than mutation (guaranteed by *µ*2*<2* which is a weaker constraint than (1)). *Tw* follows a geometric distribution with probability *Nq𝜌* and therefore the expected time for adaptation can be approximated by:

|  |  |
| --- | --- |
|  | (14) |

Without stress-induced mutation, we plug in Eqs. (10) and (5) into (12) and get:

|  |  |
| --- | --- |
|  | (15) |

With stress-induced mutation we plug in Eqs. (6) and (13) into (14) and get:

|  |  |
| --- | --- |
|  | (16) |

Comparing these two expression, a sufficient condition for faster adaptation with stress-induced mutagenesis is

|  |  |
| --- | --- |
|  | (17) |

For example, in *E. coli* the genomic mutation rate *U* is estimated to be between 0.003 (19) and 0.0004 (20), which sets the upper limit on *τ* to be between 333 and 2,500. The values of *τ* which were documented in lab (24) and wild strains (25) is lower than that – *1<τ<100*.

## Simulation results

A comparison of the analytical approximations with the simulation results are given in figure 2. The approximations fit the simulation results very well.

The starting point of all the lines is at *τ=1* which represents populations without stress-induced mutagenesis. Hence, both the approximations and the simulation results agree that stress-induced mutation increases the adaptation rate, and that this effect increases with *τ*.



Figure 2 – Stress-induced mutagenesis reduces the adaptation time. The figure shows a comparison of the analytical approximations in dashed blue and the simulation results in black circles. (a) Fold-change in waiting time for appearance of a double mutant with stress-induced mutagenesis compared to without. Note that a 10-fold increase in mutation rate results in a ~10-fold decrease in waiting time. (b) Fold-change in the number of double mutant appearances before fixation. The increase in fixation probability (Eq. 12) is to small to be seen on this scale. (c) The waiting time for adaptation – appearance and fixation of a double mutant – in generations, as a function of the fold-increase in mutation rate in stressed individuals. Note how a minor increase in mutation rate results in a considerable decreae in adaptation time. Parameters used: selection coefficient *s=0.05*, double mutant advantage *H=2*, genomic mutation rate *U=0.0004*, locus specific mutation rate *µ=U/5000*, population size *N=106*.

# Discussion

# Acknowledgments

We thank …

The work was funded by …

# Data accessability

The simulations source code and the source code used to analyse the data and generate the are available as a *git* repository on GitHub at <https://github.com/yoavram/ruggedsim.git>.

# References

# Electronic supporting material

## Fixation probability with stress-induced mutation

We derive the fixation probability of a double mutant *𝜌* in a population with stress-induced mutation. The paramters are defined in the Analytical model section.

Pluging in the population mean fitness to the relative fitness of the double mutant we get

Pluging that in the fixation probability gives the final result:

## First-order approximations

Here we derive the first-order approximations from the full analytical approximations described in the Methods

Analytical model section. The term "first-order" approximations is used here to describe the approximation of analytical expressions by linear expression or polynomials of the first degree. For example, the expression can be written as a Tayloer series . Therefore, when *U* is very small – for bacteria it is estimated to be between 10-4 and 10-2­­ – the linear expression *1-U* is a good approximation for .

We will denote these first-order approximations by an asterix (\*) added to the parameter symbol.

### Appearance of a double mutant

Starting with Eq. (3) for populations without stress-induced mutation:

The last step assumes that *2s* is much larger than *s2* and *sU* ismuch larger than *2µ.* Rearranging the last expression gives us

For a population with stress-induced mutation the first-order approximation is based on the full expression in Eq. (4):

The last approximation assumes that *Us* is smaller than *U* and that *τU* is much larger than *µ/s.*

The last approximation assumed that *2τ>s* and 2*τ2>1*, because *τ>1* and probably even *τ≥10*. Rearranging the last expression gives us the first order approximation for populations with stress-induced mutation:

Note that by setting and acknowledging that , can be derived from .

### Fixation of a double mutant

The fixation probability without stress-induced mutation is described and approximated in Eqs. (8) and (9).

With stress-induced mutation, the fixation probability (Eq. (11)) can be approximated by:

## Comparison of adaptation rate

From Eqs. (16) and (17) we can derive the adaptation rate with stress-induced mutation in term of the rate without stress-induced mutation:

Now if and because the second term is positive then we can infer that the rate with stress-induced mutation is faster than without. This condition can ve rewritten:

Using the quadrate formula this translates to:

Because *U* is very small, *1-2U* is well apprxomated by *1*, and *1/U2* is much larger than *2*, so this can be approximated by