Stress-induced mutagenesis, adaptability and adaptedness

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: June 18, 2013

# Model

We consider a population of *N* haploid asexual individuals. The number of new mutations at replication is Poisson distributed with an average of *U* mutations per genome. A mutation is deleterious or beneficial with probabilities *δ* and *β* such that *δ*+*β=1*. The effect of deleterious mutations on fitness are multiplicative (i.e., independent), such that the fitness of an individual with *x* deleterious mutations is *(1-s)x*, where *s* is the selection coefficient which we assume is higher than the mutation rate: *s>U*. Unless otherwise mentioned, beneficial mutations have an opposite effect, essentially reducing the number of deleterious mutations in the individual. Mutational strategies are defined by two parameters: the fold increase in mutation rate, *τ*, and the minimum number of deleterious mutations sufficient to induce hypermutation, *π*. The three prototypical strategies are: normal mutagenesis (NM), with *π=0* and *τ=1*, where there is no increase in mutation rates; constitutive mutagenesis (CM) with *π=0* and *τ>1*, where all individuals increase their mutation rate by *τ*; and stress-induced mutagenesis (SIM), with *π>0* and *τ>1*, where only individuals with at least π deleterious mutations increase their mutation rate by *τ*.

We develop four distinct models: (i) mutation-selection balance in a constant environment, (ii) adaptive evolution of a one-locus trait in a smooth fitness landscape, (iii) adaptive evolution of a double-locus trait in a rugged fitness landscape, and (iv) the loss of the fittest genotype by drift, aka *Muller's ratchet*. We use a mixture of analytic approximations and stochastic simulation to compare the effect of different mutational strategies (NM, CM and SIM) in these models.

## Mutation-selection balance

Denote the frequency, fitness and mutation rate of individuals with *x* deleterious mutations by , and , and the population mean fitness by . The frequency of individuals with *x* deleterious mutations in the next generation can therefore be described by:

.

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

The MSB distribution of *x* *f\** fulfills:

*M* is a positive matrix, and therefore by the *Perron-Frobenius Theorem* (Otto and Day 2007, p. 709) is the largest eigenvalue of *M* and *f\** is its unique

n-negative eigenvector with .

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

.

So *M* is a triangle matrix and its largest eigenvalue is the largest main diagonal element: . If and then the frequencies vector has been shown to be , that is, the number of deleterious mutations per individual is Poisson distributed with an average *U/s* (Haigh 1978).

With beneficial mutations this eigenvalue problem is harder to solve analytically. By neglecting elements outside the main three diagonals we have shown before (Ram and Hadany 2012) that:

.

However, this framework allows to easily calculate the population mean fitness numerically for finite *n*-by-*n* matrices (by defining *n* such that ).

## Adaptation in a simple fitness landscape

Consider a population at a MSB in which a new adaptation is now available. Assume this new adaptation requires a single beneficial mutation provides the same fitness advantage as other beneficial mutations. This scenario is equivalent to increasing the number of deleterious mutation in all the population by one after reaching an MSB.

### Appearance of the beneficial mutation

The probability that a random individual in the next generation will have the adaptation but will not have any additional deleterious mutations:

.

Assuming we neglect the effect of beneficial mutations on the MSB and assume a Poisson distribution of deleterious mutations in the population (Haigh 1978), as well as (this will be justified later):

,

where is the modified Bessel function of the first kind.

By neglecting the less significant terms in the above sum () we can approximate this by:

.

### Fixation of the beneficial mutation

After the beneficial mutation appears it can either go to fixation by selection or to extinction by drift. Following Eshel (1981) The probability of fixation is:

where *α* is the growth rate of individuals with the beneficial mutation (and without any deleterious mutations). This can be calculated as:

,

where is the population mean fitness before adaptation. Therefore:

.

Or, assuming :

.

### Total adaptation time

From the probability *q* that a random individual in the next generation has the beneficial mutation we can derive the probability that the beneficial mutation appears in the next generation: . Assuming that *Nq* is very small (that is, *N* is not too large, or otherwise adaptation will always occur instantly), this probability can be approximated by *Nq*.

Once the beneficial mutation appears it has a probability 𝜌 to go to fixation.

The time for adaptation *T* can be approximated by the waiting time for a beneficial mutation which will go to fixation *Tw*. This is true as long as fixation is much faster than mutation. *Tw* follows a geometric distribution with probability *Nqρ* and therefore the expected time for adaptation can be approximated by:

.

## Adaptation in a rugged fitness landscape

We consider a population at a MSB in which the wildtype genotype is *ab* and its fitness is 1. Site-specific mutations change *a* to *A* and *b* to *B* at reproduction with probability *µ* (without back-mutations). Individuals with SIM and fitness lower than 1 increase their mutation rate *τ*-fold.

We are interested in the capacity of the population to adapt to an environmental change after which the double mutant *AB* is the optimal genotype, with fitness *1+sH* (*H* is the double mutant relative advantage). We develop two models. The first is analytic (Figure 1a), in which we assume that: (i) individuals with deleterious alleles do not contribute to the adaptation process and (ii) that the number of deleterious mutations per individual at the MSB is Poisson distributed. The second model is a stochastic Wright-Fisher simulation with selection, mutation and drift (Figure 1b), in which we: (i) let individuals with accumulated deleterious mutations contribute to adaptation, and (ii) let the MSB evolve from a mutation-free population.



**Figure 1 – Adaptive landscape illustration.** Nodes represent genotypes: the alleles a or A and b or B, and in panel b also the number of deleterious alleles across the genome. Mutagenesis is induced in stressed genotypes (fitness < 1), indicated by ellipses, while adapted genotypes (fitness > 1) do not hypermutate. Solid arrows represent site-specific mutations at the *a/A* and *b/B* loci, which occur with probability *µ*. Dashed arrows represent deleterious mutations across the genome which are Poisson distributed with an average *U*. Node colour represents fitness (see colourbar), from pale brown for the fittest genotype (*1+sH*, where *s=0.05* is the selection coefficient and *H=2* is the double mutant advantage) to dark brown for genotypes with accumulated deleterious mutations (*(1-s)x*, where *x* is the number of deleterious mutations). (a) In the analytic model genotypes with deleterious mutations are considered "evolutionary dead-ends" (RIP) and do not contribute to adaptation. (b) In the stochastic model individuals can accumulate up to 25 deleterious mutations (the figure only shows as much as three). Multiple mutations can occur concurrently but are not shown for simplicity of the illustration.

### Constraints on the parameter space

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 . Therefore, there are several constraints on the parameter range:

1. The above MSB approximations are only valid when *U/s<1* or *U<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 mutants 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 is equal to the first constraint.

These constraints are summarized by:

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

Table 1 summarizes the model parameters with estimated values for *Escherischia coli*. Taking the conservative estimations, the population size *N* must be between105and107, but the constraint can also be met for other combinations of the parameter values.

Table 1 – Model parameters and estimated values for bacteria

|  |  |  |  |
| --- | --- | --- | --- |
| Symbol | Name | Estimate | References |
| *s* | Selection coefficient | 0.001-0.03 | (Kibota and Lynch 1996; Gordo, Perfeito, and Sousa 2011) |
| *H* | Double mutant advantage | 1-10 | (Gordo, Perfeito, and Sousa 2011) |
| *U* | Genomic deleterious mutation rate | 0.0004-0.003 | (Drake et al. 1998; Wielgoss et al. 2011) |
| *µ* | Site-specific beneficial mutation rate | U/5000 | (Gordo, Perfeito, and Sousa 2011) |
| *τ* | Fold-increase in mutation rate | 1-100 | (Bjedov et al. 2003; Hall and Henderson-Begg 2006) |
| *N* | Population size | 104-1010 |  |

### 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* (Figure 1a). We assume that at the MSB the number of deleterious mutations per individual follows a Poisson distribution (Haigh 1978). Therefore, the frequencies of mutation-free wildtype *ab* and single mutants *aB* and *Ab* are and . The probability *q* that a random individual in the next generation is a double mutant, given there are no double mutants and neglecting individuals with deleterious mutations is:

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

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 simplified by using first-order approximations. Starting with Eq. (3) for populations without SIM:

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

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

For a population with SIM 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.* Now,

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 SIM:

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

Note that by setting and because , is consistent with .

### Fixation probability with stress-induced 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 a similar development as in (‎1.2.2), we find that:

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

Assuming that is small we get

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

In this case, the fixation probability with SIM equals that of NM and CM because the mutation rate of the wildtype *ab* equals that of the double mutant *AB*.

# Results

## Mutation-selection balance

Without beneficial mutations and with a mutation rate constant in time and uniform across the population, the mean fitness equals (Kimura and Maruyama 1966). Constitutive mutagenesis causes the population mean fitness to exponential decay as a function of the mutation rate fold-increase *τ*. Stress-induced mutagenesis, as was numerically shown by Agrawal (Agrawal 2002), does not change the population mean fitness with respect to NM, because the fittest individuals, with fitness *1*, also have the lowest mutation rate, *U*, and therefore the largest eigenvalue of the mutation-selection matrix is .

With beneficial mutations, this is still a good approximation, but the actual value of the population mean fitness is slightly higher than . As we have shown before (Ram and Hadany 2012), CM decreases the population mean fitness with respect to NM, but SIM can increase the population mean fitness with respect to NM. A sufficient condition is that the mutation rate of individuals with below average fitness is increased. Because we assume that *,* then , so setting *π=1* ensures that the condition is met and therefore SIM increases the population mean fitness. We explore the magnitude of this increase for specific parameter ranges in Fig. X.

## Adaptation in a simple fitness landscape

### Appearance of the beneficial mutation

Assuming SIM increases the mutation rate in all individuals without the beneficial mutation, there is no difference between SIM and CM, and both have a higher probability to generate the beneficial mutation.

### Fixation of the beneficial mutation

In NM and CM all individuals have the same mutation rate, so:

which is a classical result in population genetics (Eshel 1981).

In SIM, if we assume that all individuals without the beneficial mutation increase their mutation rates by *τ*:

.

Because as long as , this means that and that fixation of a beneficial mutator has a higher probability with SIM. Also, as *τ* increases, the fixation probability increases - the derivative w.r.t *τ* is positive and for *s=0.05*, *U=0.0004*, SIM with *τ=10* and *τ=10* increase the fixation probability from 0.1 to 0.174 and 0.107.

Using the first-order Taylor expansion we can further approximate this by:

.

### Total adaptation time

Given the expression for *q* and *ρ*, we can approximate the adaptation time of NM (*τ=1*) and CM (*τ>1*) by:

.

With SIM the expected adaptation time is approximately:

.

This is much higher than the expected adaptation time of NM because of the faster appearance of the beneficial mutation and moderately higher than the time for CM, because of the increased fixation probability.

## Adaptation in a rugged fitness landscape

The adaptation rate *ν* is the inverse of the expected adaptation time which can be approximated, similar to (‎1.2.3), by *1/Nρq*:

There are several interpretations we can make from these expressions. First, comparing SIM to NM, we can see that as long as *τU<1*, the adaptation rate with SIM is higher than with NM. Second, as long as *τU<1/2*, the adaptation rate with SIM increases with *τ*. Third, as long as *τU<2*, adaptation with CM is faster than with SIM.

Because the genomic mutation rates in microbes is generally on the order of 10-4-10-2(see Table 1), this sets the limits on *τ* to 100-10,000 for SIM to increase the adaptation rate.