

## Chapter 4

# Mutation

*In short, mutations are accidents, and accidents will happen.*

Alfred Sturtevant

### 4.1 Mutation as an evolutionary force

So far we have assumed there is allelic variation in the population but did not consider where it comes from. Such genetic variation is incipiently introduced into a population through mutation. This is the spontaneous production of heritable changes in the DNA sequence of an individual and it results in the creation of new alleles. To understand the evolutionary consequences of mutation, we again focus our attention on a situation where only two alleles can segregate in the population: the A allele and the B allele. We denote by  $\mu_A$  the mutation rate of allele A, which is the probability that, during reproduction, a parent carrying the A allele produces an offspring carrying allele B, hence allele A has mutated into allele B. Let us denote by  $\mu_B$  the mutation rate of the B allele, which is the probability that an offspring to a B parent carries a A allele.

With these mutation rates we can now evaluate the change  $\Delta p$  in the frequency of the A allele over one demographic time period in a haploid population. We assume that every individual has equal fitness (i.e.,  $w_A = w_B$ ) so that there is no natural selection. When this is the case, the change  $\Delta p$  in the frequency of A over one demographic time step; namely, from the parental to the offspring generation is given by

$$\Delta p = \mu_B(1 - p) - \mu_A p. \quad (4.1)$$

This can be understood as follows. A fraction  $1 - p$  of individuals in the parental generation carry allele B and thus produce offspring that have mutated into A with probability  $\mu_B$ , which increases the frequency of allele A by  $\mu_B(1 - p)$ . A fraction  $p$  of individuals in the

population carry allele A and produce offspring that have mutated into B with probability  $\mu_A$ , which thus decreases the frequency of A by  $-\mu_A p$ .

Eq. (4.1) shows that mutation alone results in allele frequency change, even in the absence of natural selection. When the frequency of A is close to zero ( $p \approx 0$ ),  $\Delta p$  is determined by  $\mu_B$ . As mutation rates cannot be negative ( $\mu_B \geq 0$ ), allele A tends to increase in frequency in this case. By contrast, when the frequency of A is close to 1 ( $p \approx 1$ ), the change in allele frequency due to mutation is  $-\mu_A$ . This change can only be negative meaning that A will decrease in frequency. Hence, the boundary points ( $p = 0$  and  $p = 1$ ) are unstable in the presence of mutation. The change in allele frequency is in the direction of the interior of the state space and thus results in an increase in variability.

Eq. (4.1) also informs us that the evolutionary dynamic will settle into an equilibrium when  $\Delta p = 0$ . This equilibrium can be obtained by solving  $\mu_B(1 - p^*) - \mu_A p^* = 0$  for  $p^*$ , which produces

$$p^* = \frac{\mu_B}{\mu_A + \mu_B}. \quad (4.2)$$

This is the ratio of mutation towards A to the total mutation rate.

For this model, allele frequency can actually be expressed explicitly as a function of time, by using explicit time indices and writing eq. (4.1) as  $p_{t+1} = p_t + \mu_B(1 - p_t) - \mu_A p_t$ . Iterating this equation (like in eq. 2.6 Appendix 2 of Chapter 2), and setting the initial frequency to  $p_0$ , we obtain that the frequency  $p_t$  of the mutant at time  $t$  is

$$p_t = \underbrace{\frac{\mu_B}{\mu_A + \mu_B}}_{p^*} + \frac{(\mu_A p_0 - \mu_B(1 - p_0))(1 - \mu_A - \mu_B)^t}{\mu_A + \mu_B}. \quad (4.3)$$

Owing to the fact that  $-1 < (1 - \mu_A - \mu_B) < 1$  for non-zero mutation rates (i.e, for  $\mu_A \neq 0$  and  $\mu_B \neq 0$ ) we have that  $(1 - \mu_A - \mu_B)^t \rightarrow 0$  as  $t \rightarrow \infty$ . Thus, the frequency of allele A eventually converges asymptotically towards  $p^*$ , for any initial frequency  $p_0$ .

Two things should be noted from the above equations:

- Mutation is a “linear force”. That is, the change in allele frequency is a linear function of allele frequency  $p_t$  (eq. 4.1) and not some complicated non-linear function (this is a general result holding regardless of the number of alleles in the population).
- Mutation is a diversifying force. In the long run, the evolutionary dynamic converges towards a single global equilibrium where the different alleles coexist (this is also a general result, holding for an arbitrary number of alleles, as long as one allows for any allele to be eventually reached from any other allele, possibly through several mutation events).

## 4.2 The interaction between selection and mutation

### 4.2.1 Allele frequency change

We saw that both selection and mutation independently result in allele frequency change (respectively eq. 3.9 in Chapter 3 and eq. 4.1). They are thus both **evolutionary forces**. We now investigate the interdependency between these two evolutionary forces. To this aim, we seek to evaluate the change in allele frequency  $\Delta p = p' - p$  over one demographic time step by decomposing it into a sum of two terms

$$\Delta p = \Delta p_s + \Delta p_\mu. \quad (4.4)$$

Here,  $\Delta p_s$  is the change in allele frequency due to natural selection on faithfully transmitted alleles and  $\Delta p_\mu$  is the change in allele frequency due to mutation.

What are these two components equal to? Since, by definition,  $\Delta p_s$  is the change due to selection on a faithfully transmitted allele, it is directly given by eq. (3.9), whereby

$$\Delta p_s = \frac{p(1-p)s}{\bar{w}}. \quad (4.5)$$

However, we cannot simply use eq. (4.1) for  $\Delta p_\mu$ , because different alleles have different fitnesses. Indeed, as mutation is the probability for a parent to produce offspring that bear a different allele, the change in allele frequency due to mutation must take into account reproductive success. To obtain the expression for  $\Delta p_\mu$ , we note that the proportion of individuals in the offspring generation that descend from A parents is  $w_A p / \bar{w}$  (eq. 3.6), and each such offspring has mutated away from the parental type with probability  $\mu_A$ , which makes a loss by amount  $\mu_A w_A p / \bar{w}$  in the frequency of allele A. A complementary proportion  $w_B(1-p)/\bar{w}$  of individuals in the offspring generation descend from B parents, and each such offspring carries A with probability  $\mu_B$ , which results in a gain by amount  $\mu_B w_B(1-p)/\bar{w}$  in the frequency of A. Combining these, we thus have that in the presence of natural selection the change in the frequency of allele A due to mutation is

$$\Delta p_\mu = \mu_B \frac{w_B(1-p)}{\bar{w}} - \mu_A \frac{w_A p}{\bar{w}}, \quad (4.6)$$

which reduces to eq. (4.1) if there are no differences in fitness ( $w_A = w_B = \bar{w}$ ).

If we now substitute eqs. (4.5)–(4.6) into eq. (4.4), the change in A frequency over one demographic time period can be written as

$$\Delta p = \frac{p(1-p)s}{\bar{w}} + \frac{\mu}{\bar{w}}, \quad (4.7)$$

where

$$s = w_A - w_B \quad (4.8)$$

is the selection coefficient (recall eq. 3.9) and

$$\mu = \mu_B w_B (1 - p) - \mu_A w_A p \quad (4.9)$$

will be called the **mutation flow rate** of allele A, which is either positive, negative, or zero. Two things should be noted about eq. (4.7):

- It describes total evolutionary change. The expected dynamics of any allele can be put into this simple form, regardless of the biological complexity of the situation at hand (i.e., it applies under constant, density-dependent, and frequency-dependent selection).
- The change due to selection,  $p(1 - p)s$ , can be counteracted by the change due to mutation,  $\mu$ , so that the population can be prevented to evolve towards expressing the allele resulting in highest fitness. Below, we will look at a dramatic example of this effect.

### 4.2.2 Mutation-selection balance

We will now see that for selection to be the dominant evolutionary force, the mutation rate needs to be not too large. To investigate this, let us assume a situation of frequency and density-independent (i.e., constant) selection so that the fitnesses,  $w_A$  and  $w_B$ , of the two competing alleles are constant. We assume that allele A is favored over allele B, meaning  $s > 0$ . We also consider that the A allele is the optimal allele, and we assign it a fitness of one  $w_A = 1$  and assume that B has a lower fitness  $w_B = 1 - s$ . Let us further assume that only one-way mutations occur, from A to B (so that  $\mu_A > 0$  and  $\mu_B = 0$  in eq. 3.10). Such mutated alleles that beget a reduction of fitness relative to that of the optimal allele are called **deleterious mutations**. From these assumptions, the mean fitness in the population  $\bar{w} = pw_A + (1 - p)w_B$  (eq. 3.7) is  $\bar{w} = 1 - (1 - p)s$  and the change in allele frequency due to selection (eq. 4.5) is

$$\Delta p_s = \frac{p(1 - p)s}{1 - (1 - p)s}, \quad (4.10)$$

while the change due to mutation (eq. 4.6) is

$$\Delta p_\mu = -\frac{\mu_A w_A p}{\bar{w}} = -\frac{\mu_A p}{1 - (1 - p)s}. \quad (4.11)$$

The addition of these two components,  $\Delta p_s + \Delta p_\mu$ , implies that the total change in allele frequency (eq. 4.7) is

$$\Delta p = \frac{p(1 - p)s - p\mu_A}{1 - (1 - p)s}. \quad (4.12)$$

The numerator shows that the change in allele frequency depends on two opposing evolutionary forces: selection that tends to increase the frequency of the A allele and mutation that tends to decrease it.

The equilibrium allele frequency satisfies  $\Delta p = 0$ , therefore to obtain it explicitly we have to solve  $p^*(1 - p^*)s - p^*\mu_A = 0$  for  $p^*$ , which gives

$$p^* = 1 - \frac{\mu_A}{s}. \quad (4.13)$$

This is the frequency of A at a mutation-selection equilibrium. It increases with an increase in the selection coefficient and decreases with an increase in the mutation rate. This is a stable equilibrium and will be reached for any positive initial frequency  $p_0 > 0$  of the A in the population<sup>1</sup>. Provided that the force of selection is stronger than that of mutation  $s > \mu$ , there will be a positive frequency of A in the population, otherwise the frequency goes to zero.

### 4.2.3 Genetic load

We just saw that at a mutation-selection equilibrium where the allele A has a selective advantage of  $s$  and mutates back to B at rate  $\mu_A$ , the equilibrium frequency of A is  $p^* = 1 - \mu_A/s$  (eq. 4.13). With this frequency, we can now compute the **genetic load** of the population, which under the regime of constant selection is defined as

$$L = \frac{w_{\max} - \bar{w}}{w_{\max}}, \quad (4.15)$$

where  $w_{\max}$  is the fitness of an individual carrying the allele that has the largest fitness in the population and mean fitness  $\bar{w}$  is evaluated at the equilibrium frequency  $p^*$ . The genetic load thus measures the decrease in mean fitness in the population relative to the reference allele, which results in the largest fitness for its carrier among the alternative alleles present in the population.

In the above example, the allele with the largest fitness is A ( $w_{\max} = w_A$ ), thereby, recalling that  $w_A = 1$ ,  $\bar{w} = 1 - (1 - p^*)s$ , the genetic load can be written as

$$\begin{aligned} L &= \frac{w_A - \bar{w}}{w_A} \\ &= 1 - (1 - (1 - p^*)s) \\ &= (1 - p^*)s \\ &= \left(\frac{\mu_A}{s}\right)s \\ &= \mu_A, \end{aligned} \quad (4.16)$$

where the fourth equality is obtained by substituting eq. (4.13). This shows that at a mutation-selection equilibrium, the genetic load is exactly equal to the mutation rate from the fittest allele to the other one:

$$L = \mu_A, \quad (4.17)$$

---

<sup>1</sup>This can be seen from the explicit time dependent solution of eq. 4.12, which can be obtained by iteration and yields

$$p_t = \left(1 - \frac{\mu_A}{s}\right) \frac{1}{1 + \frac{(1-p_0-\mu_A/s)}{p_0} \left(\frac{1-s}{1-\mu_A}\right)^t}. \quad (4.14)$$

If  $\mu > s$  and  $t \rightarrow \infty$  or  $p_0 = 0$ , the denominator of the second fraction becomes infinitely large, which implies  $p \rightarrow 0$ . Otherwise this denominator converges to one.

which, remarkably, is independent of the coefficient of selection  $s$ . This feature, whereby the genetic load can be interpreted as the mutation rate towards less fit alleles, also holds under sexual reproduction and an arbitrary number of alleles, if only approximately.