

# Population Genetics I

The material presented here considers a single diploid genetic locus, with two alleles **A** and **a**; their relative frequencies in the population will be denoted as **p** and **q** (with  $q = 1 - p$ ).

## Hardy-Weinberg equilibrium (review)

Requirements:

1. no selection
2. no mutation
3. no stochastic (random) effects, i.e. infinitely large population
4. no emigration or immigration
5. random mating

If these conditions hold, **allele frequencies will not change** and genotype frequencies will be

$$AA = p^2, \quad Aa = 2pq, \quad aa = q^2 \quad (1)$$

## Effect of selection

In the following selection is assumed to be **constant** over time, and **density-** and **frequency-independent**: each genotype has a constant fitness, regardless of the generation, the population density, and the relative frequencies of the genotypes.

The main results will be:

1. equilibria are:
  - a) fixation of one allele (*i.e.*  $p = 1$  or  $p = 0$ ), or
  - b) a mixture (polymorphism), if the heterozygote is either
    - more fit than either homozygote (in which case the polymorphism is stable), or
    - less fit than either homozygote (in which case the polymorphism is unstable).
2. the rate of evolution (change in mean fitness) is (approximately) proportional to the genetic variance times the strength of selection (Fisher's "fundamental theorem of natural selection")
3. selection maximizes (approximately) the mean fitness of the population
4. the rate of change in the relative frequency of a rare allele is slower for a recessive allele than a dominant allele

*Selection model*

$p_t$  = relative frequency of **A** allele in current generation

$q_t$  = relative frequency of **a** allele in current generation

$w_{AA}$  = relative survival of **AA** genotype

$w_{Aa}$  = relative survival of **Aa** genotype

$w_{aa}$  = relative survival of **aa** genotype

Relative frequencies in the next generation:

First, multiply each genotype frequency (in the current generation) by the relative survival:

$$\text{AA: } p_t^2 w_{AA}$$

$$\text{Aa: } 2p_t q_t w_{Aa}$$

$$\text{aa: } q_t^2 w_{aa}$$

But the total size of the population will have changed, so these need to be rescaled to be relative frequencies, with  $p_{t+1} + q_{t+1} = 1$ .

Average survival will be

$$\bar{w}_t = p_t^2 w_{AA} + 2p_t q_t w_{Aa} + q_t^2 w_{aa} \quad (2)$$

The new relative genotype frequencies will be the frequencies given above divided by this average survival, i.e.

$$\text{AA: } p_t^2 w_{AA} / \bar{w}_t, \text{ etc.}$$

Since each individual has two copies of this locus, the frequency of the **A** allele in the next generation will be (with  $N$  being the population size):

$$\begin{aligned} p_{t+1} &= \frac{2N \cdot \text{frequency of AA} + N \cdot \text{frequency of Aa}}{2N} \\ &= \text{frequency of AA} + \frac{1}{2}(\text{frequency of Aa}) \\ &= \frac{p_t^2 w_{AA} + \frac{1}{2} \cdot 2p_t q_t w_{Aa}}{\bar{w}} \\ &= p_t(p_t w_{AA} + q_t w_{Aa}) / \bar{w} \end{aligned} \quad (3)$$

Now define  $w_{A;t}$  as the mean fitness of **A** alleles, given the current  $p_t$ :

$$w_{A;t} = (\text{probability a given A allele is in AA}) \times (\text{fitness of AA}) \\ + (\text{probability a given A allele is in Aa}) \times (\text{fitness of Aa}) \quad (4)$$

$$w_{A;t} = p_t w_{AA} + q_t w_{Aa} \quad (5)$$

So now the relative frequency of the **A** allele in the next generation can be expressed as

$$p_{t+1} = \frac{p_t \cdot w_{A;t}}{\bar{w}} \quad (6)$$

The ratio  $w_{A;t}/\bar{w}$  can be interpreted as the mean relative fitness of **A** alleles (in generation  $t$ ), and is the growth rate of  $p$ . The preceding equation relating  $p_{t+1}$  and  $p_t$  is analogous to the equation for geometric growth ( $N_{t+1} = \lambda N_t$ ) except that in this genetic equation the mean relative fitness (i.e. growth rate) is not constant: it depends on  $p_t$ .

### Change in relative frequencies

Define  $\Delta_p$  as the change in  $p$  from generation  $t$  to  $t+1$ , so

$$\Delta_p = p_{t+1} - p_t = \frac{p_t \cdot w_{A;t}}{\bar{w}} - p_t \\ = p_t \left( \frac{w_{A;t}}{\bar{w}} - 1 \right) \quad (7) \\ = \frac{p_t}{\bar{w}} (w_{A;t} - \bar{w})$$

Temporarily dropping the  $t$  subscripts for reduce clutter, and substituting from above for  $w_{A;t}$  and  $\bar{w}_t$ , this gives

$$\Delta_p = \frac{p}{\bar{w}} [(p \cdot w_{AA} + q \cdot w_{Aa}) - (p^2 \cdot w_{AA} + 2pq \cdot w_{Aa} + q^2 \cdot w_{aa})] \quad (8)$$

Gathering together terms involving each of the  $w$ s, and simplifying the  $p$ s and  $q$ s gives

$$\begin{aligned}
\Delta_p &= \frac{p}{\bar{w}}[(p-p^2) \cdot w_{AA} + (q-2pq) \cdot w_{Aa} - q^2 \cdot w_{aa}] \\
&= \frac{p}{\bar{w}}[p(1-p) \cdot w_{AA} + q(1-2p) \cdot w_{Aa} - q^2 \cdot w_{aa}] \\
&= \frac{p}{\bar{w}}[pq \cdot w_{AA} + q(1-p-p) \cdot w_{Aa} - q^2 \cdot w_{aa}] \\
&= \frac{p}{\bar{w}}[qp \cdot w_{AA} + q(q-p) \cdot w_{Aa} - q^2 \cdot w_{aa}]
\end{aligned} \tag{9}$$

A  $q$  from each of the summands in the brackets can be pulled out of the brackets:

$$\Delta_p = \frac{pq}{\bar{w}}[p \cdot w_{AA} + (q-p) \cdot w_{Aa} - q \cdot w_{aa}] \tag{10}$$

A slight rearrangement gives the final expression:

$$\Delta_p = \frac{pq}{\bar{w}}[p \cdot (w_{AA} - w_{Aa}) + q \cdot (w_{Aa} - w_{aa})] \tag{11}$$

### Equilibrium

The equilibrium frequency of the  $A$  allele,  $\hat{p}$ , will be the value of  $p$  at which  $\Delta_p$  equals 0, *i.e.* there is no change in frequencies. From equation (11) above for  $\Delta_p$ , we can see that there are two “trivial” equilibria: if either  $p = 0$  or  $p = 1$  (so  $q = 0$ ),  $\Delta_p = 0$ . This makes sense: if only one allele is present in the population (and we are assuming there is no mutation), there can be no change in allele frequencies.

If there is a non-trivial equilibrium with both alleles present, it will be the value of  $p$  at which the term in brackets in equation (11) equals 0:

$$\begin{aligned}
\hat{p} \cdot (w_{AA} - w_{Aa}) + (1 - \hat{p}) \cdot (w_{Aa} - w_{aa}) &= 0 \\
\hat{p} \cdot [(w_{AA} - w_{Aa}) - (w_{Aa} - w_{aa})] + (w_{Aa} - w_{aa}) &= 0 \\
\hat{p} \cdot [w_{AA} + w_{aa} - 2w_{Aa}] &= -(w_{Aa} - w_{aa}) \\
\hat{p} &= \frac{w_{Aa} - w_{aa}}{2w_{Aa} - (w_{AA} + w_{aa})} \\
\hat{p} &= \frac{w_{Aa} - w_{aa}}{(w_{Aa} - w_{AA}) + (w_{Aa} - w_{aa})}
\end{aligned} \tag{12}$$

For  $\hat{p}$  to be greater than 0, the numerator and denominator of the final expression in (12) must have the same sign (both positive or both negative). For  $\hat{p}$  to be less than 1, the two terms in the

denominator must have the same sign, so that the absolute value of the denominator is greater than that of the numerator. Combining these conditions gives:

$0 < \hat{p} < 1$  requires either

1.  $w_{Aa} > w_{AA}$  and  $w_{Aa} > w_{aa}$ : the heterozygote has the greatest fitness

or

2.  $w_{Aa} < w_{AA}$  and  $w_{Aa} < w_{aa}$ : the heterozygote has the lowest fitness

In words: under the assumptions of this simple model, **a polymorphism — both alleles remaining in the population — requires either heterozygote advantage or heterozygote disadvantage.** (It will be seen below that the latter is unstable, but the former is stable.)

*Rate of change in p and change in average fitness*

From equation (2) above,  $\bar{w}_t = p_t^2 w_{AA} + 2p_t q_t w_{Aa} + q_t^2 w_{aa}$ . Expressing the  $q$ s in terms of  $p$  (and again dropping the  $t$  subscripts), this is

$$\begin{aligned}\bar{w} &= p^2 w_{AA} + 2p(1-p)w_{Aa} + (1-p)^2 w_{aa} \\ &= p^2 w_{AA} + 2p w_{Aa} - 2p^2 w_{Aa} + w_{aa} - 2p w_{aa} + p^2 w_{aa}\end{aligned}\tag{13}$$

We can then take the derivative of this to get an expression for how the average fitness of the population changes as the allele frequencies change:

$$\begin{aligned}\frac{d}{dp}\bar{w} &= 2p w_{AA} + 2w_{Aa} - 4p w_{Aa} + 0 - 2w_{aa} + 2p w_{aa} \\ &= 2[p w_{AA} + w_{Aa} - 2p w_{Aa} - w_{aa} + p w_{aa}] \\ &= 2[p(w_{AA} - w_{Aa}) + (1-p)(w_{Aa} - w_{aa})]\end{aligned}\tag{14}$$

Now note that this expression for  $\frac{d}{dp}\bar{w}$  is 2 times the bracketed term in equation (11) for  $\Delta_p$ . So we get

$$\Delta_p = \frac{pq}{2\bar{w}} \cdot \frac{d}{dp}\bar{w}\tag{15}$$

To interpret this result, note that the product  $pq$  is proportional to the variance in allele frequencies (from the binomial formula) and  $\frac{d}{dp}\bar{w}$  measures the strength of selection: how strongly fitness varies with allele frequency. Equation (15) then says that

**the rate of evolution is proportional to the product of  
genetic variability x strength of selection**

The result in equation (15) also tells us that (under the assumptions of this simple model) **natural selection will maximize mean fitness**  $\bar{w}$ . To see this, note that at the non-trivial equilibrium, with both  $p$  and  $q$  greater than 0,  $\Delta_p = 0$  requires that  $\frac{d}{dp}\bar{w} = 0$ ; this in turn implies that  $\bar{w}$  is at a maximum (actually slope of 0 could also indicate a minimum, but this is not the case here).

### *Dominance*

So far selection was defined simply by the three genotype fitnesses,  $w_{AA}$ ,  $w_{Aa}$  and  $w_{aa}$ , which were left entirely unspecified. To explore the effects of dominance, we can specify these three fitnesses using two parameters, one to represent the difference in fitness between the two homozygotes and the second to represent the degree of dominance, *i.e.* the fitness of the heterozygote. Specifically, let

$$\begin{aligned}w_{AA} &= 1 \\w_{Aa} &= 1 - (h \cdot s) \\w_{aa} &= 1 - s\end{aligned}$$

Let's assume  $s$  is positive, so that the  $a$  allele is deleterious ( $aa$  has lower fitness than  $AA$ ). The magnitude of  $s$  measures this deleterious effect.

The parameter  $h$ , together with  $s$ , determines the fitness of the heterozygote. If  $h = 0$ , the heterozygote has fitness 1, the same as the  $AA$  homozygote: the  $A$  allele is completely dominant. Conversely if  $h = 1$ , the fitness of the heterozygote is the same as that of the  $aa$  homozygote ( $1-s$ ): the  $a$  allele is completely dominant. If  $0 < h < 1$ , the heterozygote's fitness is somewhere between those of the homozygotes: there is incomplete dominance. If  $h = \frac{1}{2}$  exactly, the alleles have additive effects: the heterozygote fitness is the average of the two homozygotes' fitnesses. If  $h < 0$ , the heterozygote's fitness is greater than 1, and thus greater than that of the  $AA$  homozygote; this is called "overdominance." Similarly, if  $h > 1$ , the heterozygote has lower fitness than the  $aa$  homozygote (and of course also the  $AA$  homozygote); this is "underdominance."

Substituting these expressions for the genotype fitnesses into the expression for the change in allele frequency  $p$  (equation 11) gives:

$$\begin{aligned}\Delta_p &= \frac{pq}{\bar{w}}[p \cdot (w_{AA} - w_{Aa}) + q \cdot (w_{Aa} - w_{aa})] \\&= \frac{pq}{\bar{w}}[p \cdot (1 - (1 - hs)) + q \cdot ((1 - hs) - (1 - s))] \\&= \frac{pq}{\bar{w}}[p hs + qs(1 - h)] \\&= \frac{pqs}{\bar{w}}[ph + q(1 - h)]\end{aligned}\tag{16}$$

The quantity in this equation outside the brackets is never negative, and equals 0 only if either only one allele is present ( $p = 0$  or  $1$ ) or there is no selection ( $s = 0$ ). The sign of  $\Delta_p$  therefore depends on the sign of the term in brackets in equation (1). Several situations are possible, depending on the value of  $h$  (*i.e.* on the degree and nature of dominance):

1. If  $h < 0$ , *i.e.* overdominance or heterozygote advantage:

When  $p$  is large, the  $ph$  term will dominate and the bracketed term will be negative:  $p$  will decrease.

Conversely, when  $p$  is small, the  $q(1-h)$  term will dominate and the bracketed term will be positive:  $p$  will increase.

This constitutes stabilizing selection for a polymorphism (both alleles present, *i.e.*  $0 < p < 1$ ): if  $A$  alleles are rare ( $p$  small), selection increases their frequency (increases  $p$ ), and if  $A$  alleles are at very high frequency ( $p$  large), selection reduces them (reduces  $p$ ).

2. If  $h > 1$ , *i.e.* underdominance or heterozygote disadvantage:

When  $p$  is large, the  $ph$  term will dominate and the bracketed term will be positive:  $p$  will increase.

Conversely, when  $p$  is small, the  $q(1-h)$  term will dominate and the bracketed term will be negative:  $p$  will decrease.

This constitutes disruptive selection: there is a polymorphic equilibrium, but it is unstable. If  $p$  is above the equilibrium, selection will cause it increase further, until the  $A$  allele becomes fixed in the population; conversely, if  $p$  is below the equilibrium, selection will cause it to decrease further and eventually the  $a$  allele will become fixed.

3. If  $0 \leq h \leq 1$ :

$[ph + q(1 - h)] > 0$ , so  $\Delta_p > 0$  for all values of  $p$  (other than  $p = 0$  or  $1$ ).

In this case, representing the entire spectrum from complete dominance of  $A$  to complete dominance of  $a$ , there is directional selection for  $A$ :  $p$  always increases, and will go to fixation ( $p = 1$ ).

a) If  $h$  is 0, or very small ( $A$  is dominant):

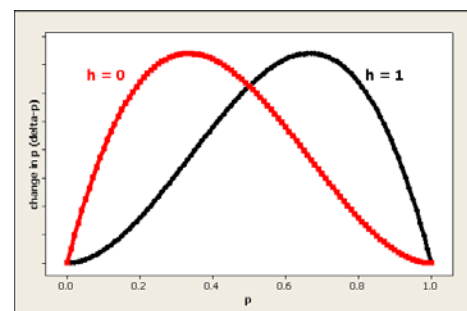
$$[ph + q(1 - h)] \cong q \text{ so } \Delta_p \approx \frac{pq^2s}{\bar{w}} = \frac{p(1-p)^2s}{\bar{w}}$$

In this case, as shown by the red curve in the figure, the response to selection is fastest —  $\Delta_p$  is largest — when  $p$  is smallish, *i.e.*  $A$  is rare.

b) If  $h$  is 1, or nearly so ( $a$  is dominant):

$$[ph + q(1 - h)] \cong p \text{ so } \Delta_p \approx \frac{p^2qs}{\bar{w}} = \frac{p^2(1-p)s}{\bar{w}}$$

In this case (black curve), the response to selection is fastest when  $p$  is largish, *i.e.*  $A$  is common.



The rate of evolution in these latter cases, with  $0 \leq h \leq 1$ , is fastest when the rare allele is the dominant one. The rare allele will occur mostly in heterozygotes. If it is dominant, its fitness effect still will be expressed even in the heterozygotes, and selection can act on it (for it, if advantageous, against it if deleterious). If, though, the rare allele is recessive, those heterozygotes — and thus the rare recessive allele also — will have similar fitness to the common allele. With little variation in fitness, evolution will be slow. In essence, a rare recessive allele “hides” from selection because its deleterious effect occurs only in the very rare homozygote recessives.

Putting these effects another way: If the advantageous allele is dominant it will increase quickly from rarity (since it is expressed in the heterozygotes even while it is rare), but if the advantageous allele is common, elimination of the deleterious recessive allele will be slow (since it rarely is expressed). Conversely, if the advantageous allele is recessive, its increase from rarity will be slow but once it becomes common, it will go to fixation (the deleterious allele will be eliminated) rapidly.