

## Population Genetics II

This second unit considers the effects of **mutation**, **drift**, **inbreeding**, and **migration** (gene flow); some of these will be considered in combination with selection. As before, it assumes a single diploid genetic locus, with two alleles **A** and **a**, with relative frequencies **p** and **q** (with  $q = 1 - p$ ). Selection is assumed to be **constant** over time, and **density-** and **frequency-independent**.

### Mutation – selection balance

Selection will eliminate a deleterious (harmful) allele from the population (unless the heterozygote has the highest fitness). Mutation, however, can produce new copies of this allele. These two offsetting processes can give an equilibrium frequency of the deleterious allele.

#### *Selection*

The selection model is the set of genotype fitnesses, which will be described as

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

Given this model, the change in the frequency of the A allele caused by selection is given by

$$\Delta_s p = \frac{pq s}{\bar{w}} [ph + q(1 - h)]$$

Since the *a* allele is deleterious we can assume it is rare, i.e. *q* is near 0 and *p* is near 1. Almost all individuals will be AA homozygotes with relative fitness of 1, so we can also assume  $\bar{w}$  is near 1.  $\Delta_s p$  therefore (approximately) simplifies to

$$\Delta_s p \approx qs[h + q(1 - h)] = qsh + q^2s(1 - h)$$

Since *q* is small,  $q^2$  is very small and can be ignored, giving, as an approximation

$$\Delta_s p = qsh$$

#### h = 0

If the *a* allele is fully recessive ( $h = 0$ ) the preceding approximation is not useful. In this case, we can still assume *p* and  $\bar{w}$  are near 1, but now

$$\Delta_s p = \frac{pq s}{\bar{w}} [ph + q(1 - h)] \approx qs[0 + q \cdot 1] = q^2s$$

## ***Mutation***

Since the  $a$  allele is much less common than the  $A$  allele, mutation from  $a$  to  $A$  will be very rare compared to mutation from  $A$  to  $a$ , so the former can be ignored. The rate of mutation from  $A$  to  $a$  will be approximately constant (independent of allele frequencies, assume  $p$  is near 1). The rate of decrease in  $p$  (i.e. the rate of creation of new  $a$  alleles) therefore can be described simply as

$$\Delta_m p = -\mu$$

where  $\mu$  is the mutation rate from  $A$  to  $a$  (per time step), and thus is the rate of decrease in  $p$  due to mutation.

## ***Equilibrium***

At equilibrium these two processes must balance:

$$\Delta_s p + \Delta_m p = 0$$

For the case with  $h \neq 0$ , the equilibrium is (approximately)

$$\hat{q}sh - \mu = 0$$

$$\hat{q} = \frac{\mu}{sh}$$

The denominator of this term is the reduction in fitness of the  $Aa$  heterozygote, below the fitness of the  $AA$  homozygote. Since  $q$  is small, the  $a$  allele will occur almost entirely in the heterozygote, so this term in effect measures the selection against the  $a$  allele. The equilibrium thus is the ratio of the rate of creation of  $as$  and the strength of selection that is eliminating them.

### $h = 1$

In the case of a fully dominant  $a$  allele ( $h = 1$ ), the equilibrium above becomes  $\hat{q} = \frac{\mu}{s}$

### $h = 0$

In the case of a fully recessive  $a$  allele, the equilibrium is (approximately)

$$\Delta_s p + \Delta_m p = 0$$

$$\hat{q}^2 s - \mu = 0$$

$$\hat{q}^2 = \frac{\mu}{s}$$

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

This equilibrium again depends on the ratio of the rate of creation of  $a$  alleles and the strength of selection eliminating them, but in this case it scales with the square-root of the ratio; this reflects that selection only acts on the  $aa$  allele which is present at the very low frequency of  $q^2$ .

### *Conclusion*

In the absence of heterozygote advantage or disadvantage (i.e. with  $0 \leq h \leq 1$ ), when selection alone would eliminate the deleterious allele, mutation will offset this effect and lead to an equilibrium at which the deleterious allele will persist. (This equilibrium will be stable, though this won't be shown here.)

The equilibrium frequency of the deleterious allele will be somewhere between  $\mu/s$  and  $\sqrt{\mu/s}$ , depending on the fitness of the heterozygote. The first of these bounds will be the smaller value, and will occur when the deleterious allele is fully dominant; the second, larger bound will occur when the deleterious allele is fully recessive. It of course makes sense that a recessive deleterious allele would be present at higher frequency than a dominant deleterious allele, since selection rarely acts against it ( $aa$  homozygotes being rare).

Mutation rates typically are in the range of  $10^{-6}$  to  $10^{-4}$  per generation. At one extreme, with the lower of these mutation rates and with a dominant and lethal allele ( $h=1, s=1$ ), the equilibrium frequency would be  $10^{-6}$ . A fully recessive allele, even if fully lethal ( $h=0, s=1$ ), however, could have a frequency as high as  $\sqrt{10^{-4}} = 0.01$  (assuming the higher mutation rate). Of course if the  $a$  allele is only mildly deleterious ( $s$  near 1) its equilibrium frequency will be much higher, but in that case the approximations used above will not be accurate.

## Drift

Genetic drift is the random fluctuation in allele frequencies caused by chance events (during Mendelian segregation, fertilization, survival and reproduction) in finite populations. Although it has no direction with respect to particular alleles, its effect is to reduce genetic diversity as alleles randomly become rare and eventually are lost from the population, and also to produce random differences among different populations.

### *Random sampling of alleles*

Imagine a population with a constant abundance of  $N$  individuals, and thus  $2N$  total gene copies. At generation  $t$  the frequency of the  $A$  allele is  $p_t$ . We then sample  $2N$  gene copies at random to go into the next generation. The number of  $A$  alleles in the next generation, let's call it  $X$ , thus will have a binomial distribution with  $n = 2N$  "trials" and probability of "success" of  $p_t$ ; the frequency of  $A$  in the next generation will be  $p_{t+1} = X / 2N$ .

From the mean and variance of a binomial distribution, we get the results that

- the mean ("expected value") of  $p_{t+1}$  is  $p_t$ : **there is no systematic change in allele frequencies**
- the variance of  $p_{t+1}$  is  $Var(p_{t+1}) = \frac{p(1-p)}{2N}$  :

**the variance is inversely proportional to the population size.**

(This variance can be thought of as being the variability across replicate populations of the same size  $N$  and starting  $p_t$ , or across different unlinked loci in one population if all started with  $p_t$ , or as measuring the unpredictability of  $p_{t+1}$  in a given population.)

The same process will occur at the next generation, but with the new allele frequency  $p_{t+1}$  taking the place of  $p_t$ ; this new allele frequency being different among replicate populations, the variability among the populations will be even greater in the next generation.

After  $T$  generations (starting with  $p_0$ ), the variance of the allele frequency  $p$  will be

$$Var(P_T) = p_0(1-p_0) \left[ 1 - \left( 1 - \frac{1}{2N} \right)^T \right]$$

At each generation there is some probability of one of the alleles being lost entirely, and the population becoming "fixed" for the other allele. Obviously once this happens there is no further random drift (assuming no mutation or immigration to reintroduce the lost allele). The probabilities of which allele will become fixed are directly proportional to their current frequency: at generation  $t$ , the probability that  $A$  will eventually become fixed in the population is simply  $p_t$ .

After very many generations virtually every one of the replicate populations will have become fixed for either  $A$  or  $a$ ; the variance approaches  $Var(p) = p_t(1-p_t)$  which simply reflects the variability across populations in which allele became fixed.

### ***Inbreeding coefficient***

If two alleles are descended from the exact same gene copy in a previous generation they are said to be “identical by descent” or “autozygous.” The extent of autozygosity in a population is measured by the “inbreeding coefficient”  $F$ , which is the probability that two randomly chosen alleles will be autozygous. (Inbreeding is mating between relatives, but it need not be due to assortative mating; in the present consideration of drift, inbreeding is simply a result of finite population size.)

Assume a population of  $N$  diploid hermaphrodites in which mating (including self-fertilization) is completely random. (Also assume there is no selection, mutation, or migration.)

Two randomly chosen gene copies in generation  $t$  can be identical by descent in two ways:

- they could be copies of the same gene copy in the previous generation; the probability of this is  $1 / 2N$ ;

(Think of picking two parental alleles. The first one can be any allele in the population. Then pick the second: since there are  $2N$  total alleles, the probability it is the same as the first pick is  $1 / 2N$ .)

- they could be copies of alleles in the previous generation that were already identical by descent; the probability of this is  $[1 - (1 / 2N)]F_{t-1}$ .

(This is the product of the probability the alleles are not copies of the same parental allele,  $1 - (1 / 2N)$ , and the probability they were already identical by descent, i.e.  $F_{t-1}$ .)

Combining these, the inbreeding coefficient in generation  $t$  will be

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_{t-1}$$

Thus **any population of finite size becomes more inbred over time**, and this occurs **more rapidly when  $N$  is small**.

This latter conclusion can be seen more easily by considering the probability that two randomly chosen alleles are not identical by descent,  $G = 1 - F$ , which is a measure of the genetic variability in the population.

$$G_t = \left(1 - \frac{1}{2N}\right)G_{t-1} \quad (\text{see below for derivation})$$

$$\begin{aligned}
G_t &= 1 - F_t \\
&= 1 - \left[ \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1} \right] \\
&= 1 - \frac{1}{2N} - \left(1 - \frac{1}{2N}\right) F_{t-1} \\
&= \left(1 - \frac{1}{2N}\right) - \left(1 - \frac{1}{2N}\right) F_{t-1} \\
&= \left(1 - \frac{1}{2N}\right) (1 - F_{t-1}) \\
&= \left(1 - \frac{1}{2N}\right) G_{t-1}
\end{aligned}$$

If  $N$  is constant, this equation is like that for geometric population growth, with  $1 - (1 / 2N)$  the finite rate of growth (or in this case, of decrease). Thus after  $T$  generations, starting from  $G_0$ ,

$$G_T = [1 - (1 / 2N)]^T G_0$$

If  $N$  is very large,  $G$  decreases slowly: the rate of decrease  $1 - (1 / 2N)$  is close to 1. Conversely, if  $N$  is small, the loss can be rapid. In the extreme, with  $N = 1$ ,  $G$  decreases by 1/2 every generation. A “rule” sometimes applied in conservation biology is that  $N$  needs to be at least 50 to maintain genetic diversity; 50 is simply the size at which  $G$  will decrease by 1% each generation.

### *Effective population size $N_e$*

The preceding results are based on several extreme assumptions, including that

- every individual has the same reproductive output,
- the population size is constant in time,
- there are no sexes (organisms are hermaphrodites),
- mating is random within the entire population, and
- generations do not overlap (so individuals cannot mate with their ancestors).

When these assumptions are not true, the rate of loss of diversity due to drift will be different than that presented above; in most cases, the loss will be more rapid than in the idealized population assumed above.

The **effective size  $N_e$**  of a real population is the **size of an idealized population** (i.e. meeting all the assumptions of the previous section) **that would experience genetic drift** — i.e. would lose diversity due to random effects — **at the same rate as the real population is experiencing**. In other words, effective population size is a measure of the rate of loss of diversity due to drift,

rather than directly a measure of abundance. It typically will be smaller than the number of adults in the population, though in some cases (notably, outbreeding) it can be larger.

Some factors which reduce  $N_e$  below  $N$  are the following:

Two sexes:

Still assuming equal reproductive success, constant abundances, panmictic mating, etc., but now with mating only between males and females, of which there are  $N_m$  and  $N_f$

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

If the sex ratio is equal ( $N_m = N_f$ ),

$$N_e = N_m + N_f = N$$

If the sex ratio is unequal,  $N_e$  approaches the  $N$  of the less abundant sex.

Variance in reproductive success:

If reproductive output varies among individuals, with variance  $\sigma^2_R$  (but all other assumptions of the idealized model above are met),

$$N_e = \frac{4N - 2}{\sigma^2_R - 2}$$

The greater the variability in reproductive success, the lower  $N_e$  will be.

Temporal variation in abundance:

If the size of the actual population varies over time, with  $N_i$  the abundance in generation  $i$  (but all other assumptions of the idealize model being met),

$$N_e = 1 / \left( \frac{1}{t} \sum_{i=1}^t \frac{1}{N_i} \right)$$

The effective population size thus is the harmonic mean of the abundances, which is dominated by the small values.

## Inbreeding

As defined above, the coefficient of inbreeding  $F$  is the probability that two alleles chosen at random from some population will be identical by descent (autozygous). Causes of inbreeding can include, in addition to genetic drift, any factor which leads to matings among relatives being more frequent than they would be with random mating. One very common such factor is that limitations to dispersal cause individuals to tend to live near — and thus mate with — their relatives.

By itself (that is, in the absence of drift, selection, etc.) non-random mating does not change allele frequencies. It does, though, affect genotype frequencies, causing them to depart from Hardy-Weinberg equilibrium: mating among relatives increases the frequency of homozygotes.

### *Genotype frequencies with inbreeding:*

If a fraction  $F$  of a population is autozygous and the remaining fraction  $1 - F$  are the result of random mating (so their genotype frequencies are at Hardy-Weinberg equilibrium), genotype frequencies will be

genotype	frequency among autozygous individuals	frequency among non-allozygous individuals	overall frequency
AA	$p$	$p^2$	$pF + p^2(1 - F)$
Aa	0	$2pq$	$2pq(1 - F)$
aa	$q$	$q^2$	$qF + q^2(1 - F)$

If  $H$  is the frequency of heterozygotes (and  $p$  and  $q$  are known from the genotype frequencies), the inbreeding coefficient becomes

$$F = \frac{2pq - H}{2pq} = 1 - \frac{H}{2pq}$$

$H = 2pq(1 - F)$ $H = 2pq - 2pqF$ $2pqF = 2pq - H$ $F = \frac{2pq - H}{2pq}$
--

i.e.  $F$  is 1 minus the ratio of the actual frequency of heterozygotes to the heterozygote frequency if the population were at Hardy-Weinberg equilibrium.

In the section above on drift, the coefficient  $G$  was defined as the probability two randomly selected alleles are not autozygous, equal to  $1 - F$ . From the result just shown, we see that

$$G = \frac{H}{2pq}$$

We can also think of the quantity  $2pq$  as the initial heterozygosity,  $H_0$  (since it is the frequency of heterozygotes in a population at Hardy-Weinberg equilibrium, before drift or inbreeding have acted). So then

$$G_t = \frac{H_t}{H_0} \quad \text{and} \quad F_t = 1 - \frac{H_t}{H_0} = \frac{H_0 - H_t}{H_0}$$

so we can interpret  $F$  as the fraction of initial (Hardy-Weinberg) heterozygosity that has been lost.

Furthermore, by the same argument as used for  $G$  in the section on drift, it can be shown that when only drift is acting,

$$H_T = [1 - (1 / 2N)]^T H_0$$

## Mutation – drift balance

When only drift is acting,  $F$ , the probability two alleles are identical by descent, changes according to

$$F_t = \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right) F_{t-1}$$

(where  $N$  has been replaced by  $N_e$  to allow for deviations from the assumptions used to derive this relationship in the section above on drift).

If one of the alleles has just mutated, however, they will no longer be identical. If the mutation rate (per allele) is  $\mu$ , the probability neither of these alleles has mutated is  $(1 - \mu)^2$ . Therefore the probability that two random alleles are identical by descent and remain identical then will be

$$F_t = \left[ \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right) F_{t-1} \right] (1 - \mu)^2$$

Because mutation acts to create genetic variability, while drift eliminates variability, an equilibrium will be reached at which their rates are balanced. This equilibrium is

$$\hat{F} = \frac{(1 - \mu)^2}{2N_e - (2N_e - 1)(1 - \mu)^2}$$

At equilibrium,  $F_t = F_{t-1} = \hat{F}$ . Thus

$$\hat{F} = \left[ \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right) \hat{F} \right] (1 - \mu)^2$$

$$\hat{F} = \frac{1}{2N_e} (1 - \mu)^2 + \left(1 - \frac{1}{2N_e}\right) \hat{F} (1 - \mu)^2$$

$$\hat{F} = \frac{1}{2N_e} (1 - \mu)^2 + \left(1 - \frac{1}{2N_e}\right) (1 - \mu)^2 \hat{F}$$

$$\hat{F} - \left(1 - \frac{1}{2N_e}\right) (1 - \mu)^2 \hat{F} = \frac{1}{2N_e} (1 - \mu)^2$$

$$\left[ 1 - \left(1 - \frac{1}{2N_e}\right) (1 - \mu)^2 \right] \hat{F} = \frac{1}{2N_e} (1 - \mu)^2$$

$$\hat{F} = \frac{\frac{1}{2N_e} (1 - \mu)^2}{1 - \left(1 - \frac{1}{2N_e}\right) (1 - \mu)^2}$$

$$\hat{F} = \frac{(1 - \mu)^2}{2N_e \left[ 1 - \left(1 - \frac{1}{2N_e}\right) (1 - \mu)^2 \right]}$$

$$\hat{F} = \frac{(1 - \mu)^2}{2N_e - (2N_e - 1)(1 - \mu)^2}$$

Since  $\mu$  is small,  $\mu^2$  is tiny and terms including it can be ignored. Then the equilibrium level of inbreeding is approximately (see below for derivation)

$$\hat{F} \approx \frac{1}{4N_e\mu + 1}$$

Since heterozygosity  $H = 1 - F$ , the equilibrium level of heterozygosity is

$$\hat{H} \approx \frac{4N_e\mu}{4N_e\mu + 1}$$

Therefore if  $N_e\mu \gg 1$ ,  $\hat{H}$  will be high (near 1), while if  $N_e\mu \ll 1$ ,  $\hat{H}$  will be small (near 0). For instance, assuming an intermediate value of  $10^{-5}$  for  $\mu$ ,  $N_e$  of 100,000 will give  $\hat{H} = 0.8$ , while  $N_e$  of 1000 (certainly not small!) will give  $\hat{H}$  of only 0.038.

$$\begin{aligned}
\hat{F} &= \frac{(1 - \mu)^2}{2N_e - (2N_e - 1)(1 - \mu)^2} \\
&= \frac{1 - 2\mu + \mu^2}{2N_e - (2N_e - 1)(1 - 2\mu + \mu^2)} \\
&= \frac{1 - 2\mu + \mu^2}{2N_e - 2N_e(1 - 2\mu + \mu^2) + (1 - 2\mu + \mu^2)} \\
&= \frac{1 - 2\mu + \mu^2}{2N_e - 2N_e + 4N_e\mu - 2N_e\mu^2 + 1 - 2\mu + \mu^2} \\
&= \frac{1 - 2\mu + \mu^2}{4N_e\mu - 2N_e\mu^2 + 1 - 2\mu + \mu^2} \\
&\approx \frac{1 - 2\mu}{4N_e\mu + 1 - 2\mu}
\end{aligned}$$

This last denominator is approximately  $(1 + 4N_e\mu)(1 - 2\mu)$  :

$$\begin{aligned}
(1 + 4N_e\mu)(1 - 2\mu) &= 1 + 4N_e\mu - 2\mu - 8N_e\mu^2 \\
&\approx 1 + 4N_e\mu - 2\mu
\end{aligned}$$

So

$$\begin{aligned}
\hat{F} &\approx \frac{1 - 2\mu}{4N_e\mu + 1 - 2\mu} \\
&\approx \frac{1 - 2\mu}{(1 + 4N_e\mu)(1 - 2\mu)} \\
&= \frac{1}{1 + 4N_e\mu}
\end{aligned}$$

## Population subdivision and gene flow

In real populations with spatial subdivision, the pattern and rates of dispersal among the subpopulations of course can be very complex. For modeling the effects of dispersal, though, a few idealized spatial patterns typically are used:

- island – mainland: the focus is on one subpopulation receiving immigrants from a much larger “mainland” population, which is assumed to not be affected by migration.
- island (or island – pool): two or more subpopulations exchange migrants with each other, with migration rates similar between any two subpopulations; there is no constant “mainland” population.
- stepping stone: three or more subpopulations exchange migrants, but with migration rates between any two subpopulations depending on the distance (1- or 2-dimensional) between them; there is no constant “mainland” population.

In the following I will use island–mainland and island models.

### *Neutral alleles*

#### Island-mainland model:

Assume an island–mainland structure, and that in each generation a fraction  $m$  of the breeding individuals in the island subpopulation are new immigrants from the mainland. Let  $p_m$  represent the frequency of allele  $A$  in the mainland population (which is constant, since migration from the island to the mainland is negligible). Then if  $p_t$  is the frequency of the  $A$  allele prior to the immigration, the new frequency will be

$$p_{t+1} = (1 - m)p_t + mp_m$$

Rearranging this gives the change in  $p$ :

$$\begin{aligned}\Delta p &= p_{t+1} - p_t \\ &= (1 - m)p_t + mp_m - p_t \\ &= mp_m - mp_t \\ &= m(p_m - p_t)\end{aligned}$$

Not surprisingly, the rate of change is proportional to both the rate of immigration ( $m$ ) and the magnitude of the difference in allele frequencies between the mainland source population and the island subpopulation.

This system reaches equilibrium ( $\Delta p$  equals 0) only when the allele frequency in the island subpopulation has become identical to that in the mainland population:

$$\hat{p}_t = p_m$$

### Island model:

I won't show it here formally, but equilibrium is reached when all island subpopulations have the same allele frequencies:

$$\hat{p}_i = \bar{p}$$

where  $\bar{p}$  is the mean frequency of the  $A$  allele in the entire ensemble of subpopulations. Clearly if any island subpopulation differed from the overall average, the allele frequency in the emigrants leaving it would differ from the frequency in the immigrants entering it, and its local frequency would be changing.

### ***F<sub>st</sub> and the Wahlund effect***

If a population is composed of multiple local subpopulations with limited gene flow among them, then even if each subpopulation is at Hardy-Weinberg equilibrium, the population as a whole will not be. This is called the "Wahlund effect." One way to measure it is as follows.

If the mean allele frequency of the  $A$  allele in the entire population is  $\bar{p}$  (and that of the  $a$  allele is  $\bar{q}$ ), the expected frequency of heterozygotes, if the entire population was panmictic and at Hardy-Weinberg equilibrium, would be  $2\bar{p}\bar{q}$ . Then the Wahlund effect is defined by the relationship

$$H = 2\bar{p}\bar{q} - 2V_q \quad \text{or} \quad V_q = \bar{p}\bar{q} - (H/2)$$

Another way to measure the genetic effect of the population subdivision is by the probability that two genes taken at random from two different subpopulations are identical by descent, which Sewall Wright termed  $F_{ST}$ . This can be contrasted with the inbreeding coefficient within a single subpopulation ( $F_{IS}$ ) and that within the entire population ( $F_{IT}$ ). (The subscripts indicate "Subpopulation within Total," "Individual within Subpopulation," and "Individual within Total.") These three measures are related as

$$F_{ST} = \frac{F_{IT} - F_{IS}}{1 - F_{IS}}$$

Furthermore,  $F_{ST}$  and the Wahlund measure are closely related (since they measure the same thing):

$$F_{ST} = \frac{V_p}{\bar{p}\bar{q}}$$

## Gene flow – drift balance

Gene flow decreases the variation among subpopulations, but drift increases it. The equilibrium level of variability can be found by balancing the rates of these two factors, exactly as was done with mutation and drift above.

If the migration rate is low, the probability that an allele migrating into a subpopulation will be identical by descent with an allele already in the subpopulation will be very small, and can be ignored. The probability two randomly chosen alleles will be autozygous therefore changes according to the model developed previously for drift, adjusted by the probability neither of the two alleles is a new immigrant. That is, with  $m$  the migration rate,

$$F_t = \left[ \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right) F_{t-1} \right] (1 - m)^2$$

Note that this is exactly the same as the model used earlier for mutation–drift balance, but with the migration rate  $m$  in place of the mutation rate  $\mu$ . Thus the equilibrium is approximately (see section above on mutation–drift balance for derivation)

$$\hat{F} \approx \frac{1}{4N_e m + 1}$$

Using the Wahlund equation for the effect of subdivision, this equilibrium can also be expressed in terms of the equilibrium variance in allele frequency among the subpopulations, as

$$\hat{V}_p \approx \frac{\bar{p}(1 - \bar{p})}{4N_e m + 1}$$

From either of these expressions for the equilibrium, we see that the variability among subpopulations, and the resulting inbreeding (deficiency of heterozygotes), decreases in proportion to the product  $N_e m$ .

Since  $m$  is the fraction of the subpopulation that has just immigrated,  $Nm$  is the number of immigrants per generation. The term  $N_e m$ , with  $N_e$  rather than  $N$ , generally will be smaller than the raw number of immigrants, but can be thought of as the contribution of immigrants to the effective population size. (Note that it is assumed immigrants and residents do not differ, on average, in fitness, including in reproductive output.)

## Selection – gene flow balance

From previously, the change in allele frequencies due to selection is

$$\Delta_s p = \frac{pq_s}{\bar{w}} [ph + q(1-h)] \quad \text{or} \quad \Delta_s q = - \left\{ \frac{pq_s}{\bar{w}} [ph + q(1-h)] \right\}$$

From above, the change in the frequency of the deleterious allele  $a$  in the subpopulation,  $q_i$ , caused by immigration at rate  $m$  from a mainland population with frequency  $q_m$  is

$$\Delta_m q_i = m(q_m - q_i)$$

At equilibrium these two rates will cancel out, i.e.

$$\Delta \hat{q}_i = \Delta_s \hat{q}_i + \Delta_m \hat{q}_i = 0$$

or

$$m(q_m - \hat{q}_i) - \frac{\hat{p}_i \hat{q}_i s}{\hat{w}_i} [\hat{p}_i h + \hat{q}_i (1-h)] = 0$$

Solving for this equilibrium in the general case is complicated, but simple results can be gotten as approximations in various cases. First, assume that selection is weak ( $s$  is small) and there is partial dominance ( $0 < h < 1$ ) rather than heterozygote advantage or disadvantage. Also assume the deleterious allele is rare ( $q_i$  is small), since selection is acting against it. Then:

- if immigration far outweighs selection, i.e.  $m \gg s$ ,

$\hat{q}_i$  approaches  $q_m$ , i.e. the subpopulation converges on the mainland population.

- if selection far outweighs immigration, i.e.  $m \ll s$ ,

$\hat{q}_i \approx \frac{mq_m}{s}$  i.e. a fraction of the mainland frequency, with the fraction being the ratio of the rate of addition of  $a$  alleles by immigration ( $m$ ) and the elimination of  $a$  alleles by selection ( $s$ ).

- in the intermediate case with the rate of immigration and the strength of selection comparable, i.e.  $m \approx s$ ,

$\hat{q}_i \approx \sqrt{q_m}$ , which is intermediate between the two previous results.