

## OLM 11.4. Spatial differentiation of populations and migration

The composition of small populations may differ substantially even without selection, due to drift as shown in OLM 11.1. The balancing effect of dispersal and the consequential gene flow (migration) between subpopulations that decreases the differences between them should not be overlooked either. We will study the effect of the balance between drift and dispersal in this OLM.

In a large population consisting of reproductively isolated subpopulations the average allele frequencies do not change in time, but the variance of the frequency distribution increases due to drift. The gene flow between the subpopulations decreases the variance, so the system tends to an equilibrium:

$$V_{t+1} = V_t = \hat{V}. \quad (11.4.1)$$

First we study the effect of immigration on the differences among local allele frequencies. Let us assume for now that immigration may occur from any of the subpopulations different from the one accepting the disperser. Thus the allele frequency among the dispersers is close to that in the pooled population, which is  $p_0$ . Assuming that in a randomly chosen subpopulation the proportion of alleles coming from outside is  $m$ , the allele frequency in generation  $t$  is

$$p_t = mp_0 + (1 - m)p_{t-1} \quad (11.4.2)$$

Subtracting  $p_0$  from both sides and rearranging yields a formula suggesting that migration reduces the difference between the allele frequencies of the subpopulation and the pooled population by a factor  $1-m$  in each generation:

$$p_t - p_0 = (1 - m)(p_{t-1} - p_0) \quad (11.4.3)$$

The variance of relative allele frequencies ( $V_t$ ) introduced in OLM 11.1 is the expectation of  $(p_t - p_0)^2$ ; therefore, dispersal would decrease the variance of the allele frequency in the absence of drift as

$$V_t = E\{(p_t - p_0)^2\} = (1 - m)^2 E\{(p_{t-1} - p_0)^2\} = (1 - m)^2 V_{t-1} \quad (11.4.4)$$

That is, the larger the dispersal the faster the genetic difference between the subpopulations decreases.

We have shown in OLM 11.1 that genetic drift increases the variance of allele frequencies at a pace depending on the size of the subpopulations:

$$V_{t+1} = V_t - \frac{V_t - p_0 q_0}{N} \quad (11.4.5)$$

Considering both genetic drift and allele dispersal we get

$$V_{t+1} = (1 - m)^2 \left( V_t - \frac{V_t - p_0 q_0}{2N_e} \right) \quad (11.4.6)$$

In equilibrium we may substitute  $V_t$  and  $V_{t+1}$  with the equilibrium value of the variance, so that

$$\hat{V} = (1 - m)^2 \left( \hat{V} - \frac{\hat{V} - p_0 q_0}{2N_e} \right) \quad (11.4.7)$$

Rearrangement yields

$$\hat{V} = \frac{(1-m)^2 p_0 q_0}{2N_e - (1-m)^2 2N_e + (1-m)^2} \quad (11.4.8)$$

Assuming  $m$  to be small we may get a simple approximate formula in two steps. Applying the

$$(1 - m)^2 = 1 - 2m + m^2 \approx 1 - 2m \quad (11.4.9)$$

approximation we obtain

$$\hat{V} = \frac{(1-2m)p_0 q_0}{4mN_e + (1-2m)} \quad (11.4.10)$$

The second step is the application of  $1 - 2m \approx 1$  approximation, so that we get the formula known from textbooks:

$$\hat{V} \approx \frac{p_0 q_0}{1 + 4N_e m} \quad (11.4.11)$$

Eq. 11.4.11 is often used in its alternative formulation built on Wright's fixation index ( $F_{ST}$ ) instead of the variance of allele frequency (cf. OLM 11.3):

$$F_{ST} = \frac{\hat{V}}{p_0 q_0} = \frac{1}{1 + 4N_e m} \quad (11.4.12)$$

Eq. 11.4.11 can be used to estimate the dispersal rate by assessing the number of dispersing alleles per generation and determining the variance of allele frequency between subpopulations. This will obviously be a crude estimate, partly because of the approximations in the calculations above and the specific assumptions of the model, partly because the population may not be in a drift-dispersal equilibrium yet. Moreover, the estimation of the dispersal rate is sensitive to the effective population size which is difficult to assess in practice, because it requires knowledge regarding past fluctuations in actual size.