

### OLM 11.3. Genetic drift and inbreeding

The genetic differentiation of small populations is often characterized by Wright's  $F$  statistics, because inbreeding depression depends on the increasing probability of two alleles identical by descent to come together in the same individual due to the homogenizing effect of genetic drift. We introduce the concept of the coefficient of inbreeding in this OLM, we show how it increases generation by generation in a population of finite size, and connect it to the variance of allele frequencies discussed in OLM 11.1.

Two alleles of a genetic locus are considered identical by descent if both are the copies of the same allele present in an ancestral population (at  $t=0$ ). Even though all alleles are descendants of the same ancestral allele after all, this concept is still useful, because it allows for the calculation of the probability of an individual being inbred given the lineages of descent (family trees), i.e., it makes it possible to calculate the probability that both alleles in a locus are the copies of the same ancestral allele. This means that by arbitrarily fixing the ancestral generation the level of inbreeding can be determined for the population at any later time. The probability of finding alleles identical by descent on a locus of a randomly chosen individual in the population is called the inbreeding coefficient of the population and is denoted by  $F$ .

We wish to determine the change of  $F$  from generation to generation in a diploid population of  $N$  individuals. Assuming that the pairing of gametes is completely random (with selfing allowed) the probability of two copies of the same ancestral gene coming together in an offspring is

$$\frac{1}{2N} \cdot \frac{1}{2N} \cdot 2N = \frac{1}{2N} \quad (11.3.1)$$

and the probability of the complementary event (that the two alleles descended from two different parental genes) is

$$1 - \frac{1}{2N} \quad (11.3.2)$$

However, even in this case the two alleles of the same locus may be copies of the same ancestral allele with a probability  $F_{t-1}$ . Therefore, the proportion of individuals at generation  $t$  whose alleles are identical by descent ( $F_t$ ) is

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1} \quad (11.3.3)$$

The expected relative frequency of individuals with alleles not identical by descent at generation  $t$  is

$$1 - F_t = 1 - \frac{1}{2N} - \left(1 - \frac{1}{2N}\right) F_{t-1} \quad (11.3.4)$$

which, after rearrangement, takes the form

$$1 - F_t = (1 - F_{t-1}) \left(1 - \frac{1}{2N}\right). \quad (11.3.5)$$

Since  $F_0=0$ ,

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t. \quad (11.3.6)$$

This relation is a familiar one, since according to Eq. (11.1.12),

$$V_t = p_0 q_0 \left(1 - \left(1 - \frac{1}{2N}\right)^t\right) \quad (11.3.7)$$

that is,

$$F_t = \frac{V_t}{p_0 q_0} \quad (11.3.8)$$

The inbreeding coefficient may take values between 0 and 1. It is zero in an infinitely large panmictic population, and it is 1 if all populations have a single fixed allele of the same ancestry, i.e., they are completely inbred.

In a large population consisting of reproductively isolated subpopulations the proportion of heterozygotes is always smaller than expected from the corresponding Hardy-Weinberg equilibrium (this is called the Wahlund effect). The measure of the differentiation is the Wright index of fixation,  $F_{ST}$ . It is calculated from the average heterozygosity ( $H_S$ ) of  $S$  subpopulations and the heterozygosity expected in the pooled population at Hardy-Weinberg equilibrium ( $H_T$ ) as

$$F_{ST} = \frac{H_T - H_S}{H_T} \quad (11.3.9)$$

The expected proportion of heterozygotes ( $H_T$ ) in the randomly mating pooled population can be easily calculated from the average allele frequencies ( $\bar{p}$  and  $\bar{q}$ ):

$$H_T = 2\bar{p}\bar{q} = 2\bar{p}(1 - \bar{p}) = 2(1 - \bar{p}^2) = 2 - 2\bar{p}^2 \quad (11.3.10)$$

and the same formula can be applied to calculate the heterozygosity of subpopulation  $i$ :

$$H_i = 2 - 2p_i^2 \quad (11.3.11)$$

from which we get

$$H_S = \frac{\sum_i^S H_i}{s} = 2 - 2 \frac{\sum_i^S p_i^2}{s} \quad (11.3.12)$$

Substituting these values into the formula of the Wright index:

$$F_{ST} = \frac{\bar{p}^2 - \frac{\sum_i^S p_i^2}{s}}{\bar{p}\bar{q}} \quad (11.3.13)$$

The numerator of this formula is an estimation for  $V$  based on the allele frequencies observed in the subpopulations;  $\bar{p}$  and  $\bar{q}$  are the estimations for the initial allele frequencies ( $p_0$  and  $q_0$ ). This means that  $F_{ST}$  can be used for estimating the inbreeding coefficient (Eq. (11.3.8)).

A basic introduction to F-statistics is provided by the textbook of Conner and Hartl (2004).

## References

Conner, J.K. and Hartl, D.L., 2004 *A Primer of Ecological Genetics*, Sinauer pp.240