Genetic and statistical sampling



Experiment by Buri (1956)



Fig. 3.3. Distributions of gene frequencies in 19 consecutive generations among 105 lines of *Drosophila melanogaster*, each of 16 individuals. The gene frequencies refer to two alleles at the 'brown' locus (bw^{75} and bw), with initial frequencies of 0.5, The height of each black column shows the number of lines having the gene frequency shown on the scale below, previously fixed lines being excluded. (*After Buri, 1956.*)

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

If $\theta_0 = 0$, $\theta_t = 1 - \left(1 - \frac{1}{2N}\right)^t$

will give information about t, the time since the populations diverged.

Total variance of allele frequencies

In the total sampling framework, different individuals can not be regarded as having been sampled independently. Even if there is random mating within populations, there are frequency differences between populations. In statistical language, the component of variance between populations is given by the covariance of individuals within populations. The genetic sampling process, evolutionary forces, now plays a role, and the expectation for alleles from different individuals is written as

$$E(x_{ij}x_{i'j'}) = P_{\frac{4}{A}}$$

$$Var(\tilde{P}_{A}) = (P_{\frac{4}{A}} - P_{A}^{2}) + \frac{1}{2n}(P_{A} + P_{AA} - 2P_{\frac{4}{A}})$$

$$P_{AA} = P_{A}^{2} + P_{A}(1 - P_{A})F$$

$$P_{\frac{4}{A}} = P_{A}^{2} + P_{A}(1 - P_{A})\theta$$

$$Var(\tilde{P}_{A}) = P_{A}(1 - P_{A})\left(\theta + \frac{F - \theta}{n} + \frac{1 - F}{2n}\right)$$
group coancestry coefficient θ_{L}



In a random mating population

$$Var\left(\tilde{P}_{A}\right) = P_{A}(1 - P_{A})\theta + P_{A}(1 - P_{A})\frac{1 - \theta}{2n}$$
$$= P_{A}(1 - P_{A})\theta$$