

Problems

- (1) In a set of populations with 20 reproductive individuals each, the initial allele frequencies of a biallelic locus are $p_0 = 0.3$ and $q_0 = 0.7$. If the populations are kept under ideal conditions, what will be the expected values of the heterozygosity, the variance of allele frequencies and inbreeding at generations $t = 5, 20$ and 100 ?

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

$$F_5 = 1 - \left(1 - \frac{1}{2 * 20}\right)^5 = 0.119$$

$$F_{20} = 1 - \left(1 - \frac{1}{2 * 20}\right)^{20} = 0.397$$

$$F_{100} = 1 - \left(1 - \frac{1}{2 * 20}\right)^{100} = 0.920$$

$$\sigma_t^2 = p_0 q_0 F_t$$

$$\sigma_5^2 = 0.3 * 0.7 * 0.119 = 0.025$$

$$\sigma_{20}^2 = 0.3 * 0.7 * 0.397 = 0.083$$

$$\sigma_{100}^2 = 0.3 * 0.7 * 0.920 = 0.193$$

$$H_t = H_0 \left(1 - \frac{1}{2N}\right)^t$$

$$H_0 = 1 - \sum_{i=1}^n p_i^2 = 1 - [(0.3)^2 + (0.7)^2] = 1 - 0.58 = 0.42$$

$$H_5 = 0.42 \left[1 - \left(\frac{1}{2 * 20}\right)^5\right] = 0.370$$

$$H_{20} = 0.42 \left[1 - \left(\frac{1}{2 * 20}\right)^{20}\right] = 0.253$$

$$H_{100} = 0.42 \left[1 - \left(\frac{1}{2 * 20}\right)^{100}\right] = 0.033$$

- (2) In problem 4.2 we considered a plant species whose natural reproduction is a combination of autogamy and allogamy in equal proportions. From a large population of this species a line with census size $N = 20$ was founded, which was maintained for 10 generations in the greenhouse by allogamous reproduction.

$$\alpha = \frac{\beta}{2 - \beta} = 0.5 / (2 - 0.5) = 0.333$$

$$F_{10} = 1 - \left(1 - \frac{1}{2N}\right)^{10} = 1 - \left(1 - \frac{1}{40}\right)^{10} = 0.224$$

$$1 - F_{IT} = (1 - F_{IS})(1 - F_{ST})$$

$$1 - (1 - F_{IS})(1 - F_{ST}) = F_{IT} = 1 - (1 - 0.333)(1 - 0.224) = 0.482$$

F_{IS} is inbreeding due to non-random mating in a sub-population. F_{ST} is inbreeding due to population subdivision.

- (3) In problem 4.1 it was assumed that the locus considered is neutral. Suppose that one of the alleles of the locus is deleterious with a purging coefficient of $d = 0.2$. What would then be the expected values of the heterozygosity, the variance of allele frequencies and the inbreeding for this locus in generation $t = 20$?

See recursivePurgedInbreedingCoefficient.r for an mplementation of equation 4.26 which computes g_{20} , the purged inbreeding coefficient at generation 20, as 0.147.

$$H_0 = 0.42$$

$$H_t = 2p_0q_0(1 - F_t) = 2(0.3)(0.7)(1 - 0.147) = 0.358$$

$$\sigma_{q20}^2 = p_0q_0F_{20} = (0.3)(0.7)(0.147) = 0.031$$

- (4) In a population subdivided into $n = 10$ sub-populations of constant census size $N = 100$, among which migration takes place with a rate $m = 0.01$ per generation following an island model, a molecular marker whose mutation rate is $u = 10^{-6}$ is analysed. (a) What is the value of the differentiation indices of F_{ST} and D , considering or not mutation for the marker? (b) What would these values be if the marker mutation rate were $u = 10^{-3}$?

$$F_{ST} = \frac{1}{4N((m)\frac{n}{n-1})^2 + (u)\frac{n}{n-1} + 1} = \frac{1}{4N((0.01)\frac{10}{9})^2 + (10^{-6})\frac{10}{9} + 1} = 0.168$$

$$F_{ST} = \frac{1}{4N((m)\frac{n}{n-1})^2 + 1} = \frac{1}{4N((0.01)\frac{10}{9})^2 + 1} = 0.168$$

$$D \approx \frac{1}{1 + \frac{m}{u(n-1)}} = \frac{1}{1 + \frac{0.01}{10^{-6}(9)}} = 0.001$$

$$F_{ST} = \frac{1}{4N((m)\frac{n}{n-1})^2 + (u)\frac{n}{n-1} + 1} = \frac{1}{4N((0.01)\frac{10}{9})^2 + (10^{-3})\frac{10}{9} + 1} = 0.157$$

$$F_{ST} = \frac{1}{4N((m)\frac{n}{n-1})^2 + 1} = \frac{1}{4N((0.01)\frac{10}{9})^2 + 1} = 0.168$$

$$D \approx \frac{1}{1 + \frac{m}{u(n-1)}} = \frac{1}{1 + \frac{0.01}{10^{-6}(9)}} = 0.001$$

- (5) In an experimental fish population maintained in 10 tanks, diversity for a group of molecular markers (SNPs) is analysed obtaining an average expected heterozygosity in each tank of $H_S = 0.21$ and a total heterozygosity for the whole set of tanks $H_T = 0.32$. (a) Calculate the differentiation index F_{ST} , the corrected index relative to its maximum value and the D index. (b) The analysis is repeated with microsatellite markers obtaining values of $H_S = 0.83$ and $H_T = 0.89$. Are the results of both types of markers consistent with each other?

Self Assessment

Papers

Expected vs. realized kinship of full-sibs

- <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.0020041>
- Estimating additive and dominance genetic variance is based on expected proportion of genes shared by different types of relatives
- Genome-wide coverage of markers enables the estimation of parameters like heritability using the realized degree of identity-by-descent sharing
- this study estimated heritability of height using genome-wide identity-by-descent sharing
- this study showed that it is feasible to estimate genetic variance from within-family segregation, permitting the estimation of genetic variation for disease susceptibility free from confounding, non-genetic factors, also allow partitioning of genetic variation into additive and non-additive components
- Equating observed phenotypic covariance to degree of genetic relationships from pedigree data

- Coefficient of relationship: expected proportion of alleles that are identical-by-descent between relatives (determines additive genetic covariance between relatives)
- Additive effects: for a given sibling pair, π = genome-wide mean IBD is the sum of proportion shared from paternal and maternal contributions

Estimated inbreeding using genomics versus pedigrees

Pedigree free genomic kinship estimates for heritability in the wild

Genomics for quantitative genetics in the wild