Summary
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?

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

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

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

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

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

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

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

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

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

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

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

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

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

(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.

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

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

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

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

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

$$H_0 = 0.42$$
$$H_t = 2p_0q_0(1 - F_t) = 2(0.3)(0.7)(1 - 0.147) = 0.358$$
$$\sigma^2_{q20} = p_0q_0F_{20} = (0.3)(0.7)(0.147) = 0.031$$

(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 $u = 10^{-3}$?

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

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

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

$$F_{ST} = \frac{1}{4N((m)\frac{n}{n-1}^2 + (u)\frac{n}{n-1}) + 1} = \frac{1}{4 * 100((0.01)\frac{10}{9}^2 + (10^{-3})\frac{10}{9}) + 1} = 0.157$$

$$F_{ST} = \frac{1}{4N((m)\frac{n}{n-1}^2) + 1} = \frac{1}{4 * 100((0.01)\frac{10}{9}^2) + 1} = 0.168$$

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

(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?

$$F_{ST} = \frac{H_T - H_S}{H_S} = \frac{0.32 - 0.21}{0.32} = 0.344$$

$$F'_{ST} = \frac{F_{ST}}{\frac{(n-1)(1-H_S)}{n-1+H_S}} = \frac{0.344}{\frac{9(1-0.21)}{9+0.21}} = 0.445$$

$$D = \frac{H_T - H_S}{1 - H_S} \frac{n}{n-1} = \frac{0.32 - 0.21}{1 - 0.21} \frac{10}{9} = 0.155$$

$$F_{ST} = \frac{H_T - H_S}{H_S} = \frac{0.89 - 0.83}{0.89} = 0.0.067$$

$$F'_{ST} = \frac{F_{ST}}{\frac{(n-1)(1-H_S)}{n-1+H_S}} = \frac{0.067}{\frac{9(1-0.83)}{9+0.83}} = 0.433$$

$$D = \frac{H_T - H_S}{1 - H_S} \frac{n}{n-1} = \frac{0.89 - 0.83}{1 - 0.83} \frac{10}{9} = 0.392$$

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

- https://pubmed.ncbi.nlm.nih.gov/26059970/
- Tested whether the realized proportion of the genome that is IBD is predicted better by pedigree inbreeding coefficients or genomic (marker-based) measures of inbreeding
- Inbred individuals have lower heterozygosity due to fractions of loci that are identical by descent, which arises if two gene copies both originate from one copy in a common ancestor of the parents
- measures of inbreeding try to predict the proportion of the genome that is IBD
- F_P is the "best" measure for inbreeding and uses known common ancestors of parents, also assumes that pedigree founders are unrelated and non-inbred
- Problems with F_P: measures of IBD can vary among individuals within a pedigree, the standard deviation of IBD is high in species with few chromosomes/short genetic maps, and F_P cannot account for inbreeding caused by ancestors not included in a pedigree
- Using molecular data, IBD can be estimated as the increase in individual homozygosity relative to HW expected homozygosity (F_H) , or use mapped genetic markers to identify IBD tracts or ROH (F_{ROH} estimates the IBD as the proportion of genome within ROH

- Tested various measures of inbreeding using simulated data, along with the assumption that population founders were unrelated and non-inbred
- Re-ran all tests to assess whether results are different if inbreeding is due to recent ancestors
- LD pruning: reduce LD because F_{ROH} is estimated after pruning to avoid detecting fixed ROH from ancient population processes
- Used 3 estimators of IBD (sliding window approach in PLINK, excess number of ohmozygotes in an individual relative to mean number expected under random mating, and a ML estimator using HMM
- Comparison of difference inbreeding measures: assessed the proportion of variance in IBD explained by each measure from linear regression on IBD
- Pedigree-based estimators had low precision, F_{ROH} and F_H had good precision with large number of SNPs, and molecular-based estimators were always more precise than pedigree-based when large number of SNPs were used
- Pedigree-based was more biased (underestimated IBD), but so were marker-based estimators when using few SNPs, F_H overestimated
- correlation values for pedigree estimates were lower using shorter genetic maps, molecular methods had good precision in partially isolated populations with shorter genome maps
- pedigree-based methods were also imprecise when inbreeding was only from recent ancestors
- marker based methods performed similarly when inbreeding was from recent ancestors or not
- F_P was downwardly biased and this can be explained by inbreeding in the founder population which is unaccounted for. Moving the base population back in time can reduce this bias.
- All marker-based estimators were precise when a large number of SNPs were used.
- the number of SNPs necessary for estimating inbreeding depends on whether we are interested in inbreeding due to distant ancestors, marker heterozygosity, the proportion of missing genotypes, and the variance in IBD
- Detecting IBD tracts depends non SNP density
- Marker-based is most advantageous when genetic maps are short
- Limitations of the study: assumed no crossover interference, assume pedigrees and genotypes were error free, and did not simulate inbreeding depression

Pedigree free genomic kinship estimates for heritability in the wild

- https://pubmed.ncbi.nlm.nih.gov/24917482/

Genomics for quantitative genetics in the wild

- https://www.sciencedirect.com/science/article/abs/pii/S0169534717302343