

# BIOB480/BIOE548 notes 10/10/2024

## Introduction

- Reminder: HW5 and take-home quiz by EOD Friday.
- Questions on HW6?

## Inbreeding

An individual is inbred when its parents share one or more common ancestors. Because inbreeding is unavoidable in small populations, it is a topic of major importance in conservation genetics. Inbreeding reduces reproductive fitness by increasing homozygosity (though more identical by descent pairings), which exposes deleterious recessive alleles. We measure inbreeding with the inbreeding coefficient  $F$ , which scales from 0 (completely outbred) to 1 (completely inbred).

We can start with a simple model of self fertilization in a diploid lineage, where a parent (with genotype  $A_1A_2$ ) produces gametes in proportions  $\frac{1}{2}A_1$  and  $\frac{1}{2}A_2$ . There is thus a  $\frac{1}{2} * \frac{1}{2}$  probability its offspring will be homozygous for  $A_1$ , a  $\frac{1}{2} * \frac{1}{2}$  probability it will be homozygous for  $A_2$ , and a  $\frac{1}{2} * \frac{1}{2} + \frac{1}{2} * \frac{1}{2} = \frac{1}{2}$  probability it will be heterozygous (either  $A_1A_2$  or  $A_2A_1$ ). Its inbreeding coefficient is the summed probabilities both its alleles are IBD, which in this case is equal to the sum of the probabilities of both homozygous genotypes, as there is no other ancestral source of alleles:  $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ .

We can expand this model to two unrelated grandparents producing two children who have a child with each other (so-called “full sib” mating). Labeling each grandparental allele independently regardless of state to indicate descent gives us genotypes  $A_1A_2$  and  $A_3A_4$ . The child’s inbreeding coefficient is the sum of the probability its two alleles are derived from the same grandparental copy. For example, there is a  $\frac{1}{2}$  probability the first grandparent passes on  $A_1$  to the first parent, and a  $\frac{1}{2}$  probability that same copy is passed on to their child, which we multiply for a  $\frac{1}{4}$  overall probability of inheriting the first copy of  $A_1$ . Similarly, the probability of the first grandparent passing  $A_1$  on to the second parent and that second parent passing it on to their child is also  $\frac{1}{4}$ .

We thus end up with a  $\frac{1}{16}$  the child of a full sib mating event to be homozygous for  $A_1$  with alleles that are identical by descent. Their full inbreeding coefficient is the sum of all such possible IBD combinations, which in this case is  $F = \frac{1}{16}(\text{ for } A_1A_1) + \frac{1}{16}(\text{ for } A_2A_2) + \frac{1}{16}(\text{ for } A_3A_3) + \frac{1}{16}(\text{ for } A_4A_4) = \frac{4}{16} = \frac{1}{4}$

The calculation above is based on finding a “loop” in a pedigree: a route where a single allele copy can be passed on through descendants before combining (to become identical by descent) in a common ancestor. This forms the basis for calculating the pedigree inbreeding coefficient,  $F_P$  (or  $F_X$ , where  $X$  is an individual on the pedigree). The inbreeding coefficient was derived by Sewall Wright (1889 - 1988), one of the three major founders of population genetics. Wright, who spent his career at University of Chicago and University of Wisconsin, Madison following his Ph.D. at Harvard, had broad interests in animal breeding, evolution, and statistics. One of his major contributions is the path coefficient, which is a method for assessing the correlation between two nodes on a directional graph. In conservation genetics, this lets us determine the proportion of ancestry passed on from an ancestor to an offspring  $n$  generations apart, while accounting for inbreeding (more on that in a second). Path coefficients can range from from  $\frac{1}{2}$  (e.g., father / mother to child) to  $\rightarrow 0$  (e.g., the common ancestor of all humanity to you):

$$P_{AO} = 2^{-n} \cdot \sqrt{\frac{1 + f_A}{1 + f_O}}$$

For example, grandparents contribute an average of 1/4th of their genes to their grandchildren ( $n = 2$  generations apart), assuming no inbreeding:

$$P_{AH} = 2^{-2} \cdot \sqrt{\frac{1+0}{1+0}} = \frac{1}{4} \cdot 1 = \frac{1}{4}$$

Path coefficients can be used to determine the *coefficient of relationship*, or the proportion of genetic material two individuals  $B$  and  $C$  share given each common ancestor they share (here denoted by  $A_i$ ). Here, we keep track of the number of ancestors between  $B$  and  $A$  with  $n$ , and  $C$  and  $B$  with  $m$ :

$$r_{BC} = \sum_i P_{A_i B} \cdot P_{A_i C} = \sum_i 2^{-n_i} \cdot \sqrt{\frac{1+f_{A_i}}{1+f_B}} \cdot 2^{-m_i} \cdot \sqrt{\frac{1+f_{A_i}}{1+f_C}} = \sum_i \left(\frac{1}{2}\right)^{(n_i+m_i)} \frac{1+f_{A_i}}{\sqrt{(1+f_{B_i})(1+f_{C_i})}}$$

For example, the relationship coefficient between two non-inbred siblings will be the product of the path coefficients of between each child and their father plus the product of the path coefficients of each child and their mother. Since  $f_{\#}$  is 0 in all cases, the fraction on the right side reduces to 1, and we only need to worry about the first half of the equation (i.e.,  $r_{BC} = \sum_i (\frac{1}{2})^{(n_i+m_i)}$ ):

$$r_{DE} = P_{AD} \cdot P_{AE} + P_{BD} \cdot P_{BE} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

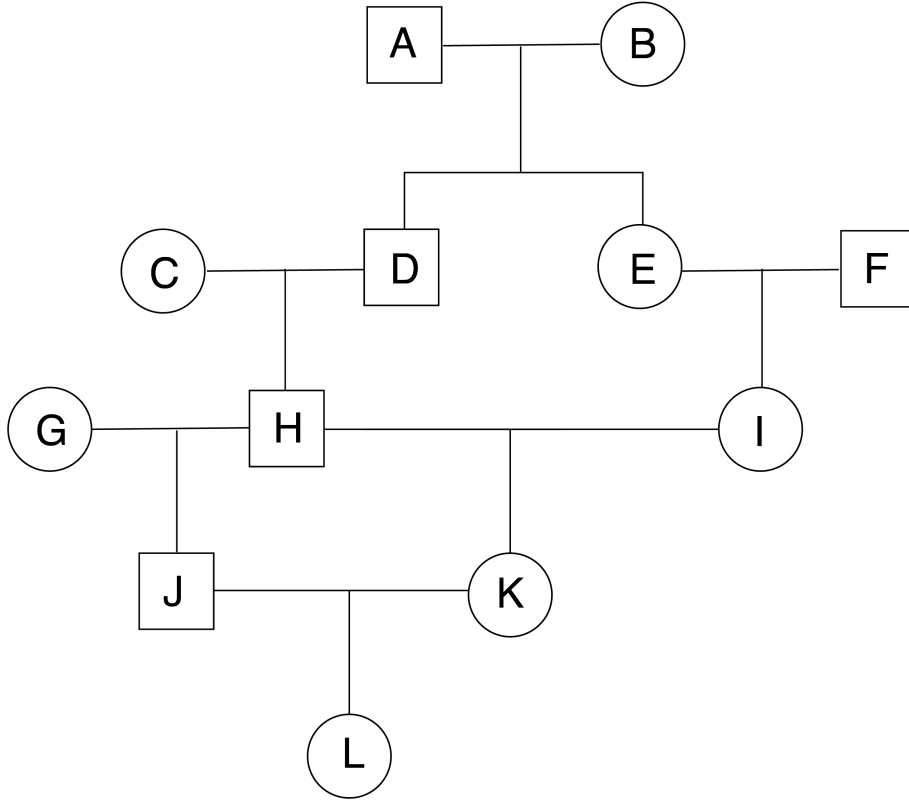
The inbreeding coefficient for any individual  $X$  is the product of the relationship coefficients of two individuals, which gives us the proportion of the gene that is likely to be identical by descent:

$$f_X = \sum_i \left(\frac{1}{2}\right)^{(n_i+m_i)} \frac{1+f_{A_i}}{\sqrt{(1+f_{B_i})(1+f_{C_i})}} \left(\frac{1}{2}\right)^{(m_i+n_i)} \frac{1+f_{A_i}}{\sqrt{(1+f_{C_i})(1+f_{B_i})}}$$

This unwieldy equation simplifies nicely:

$$F_X = \sum \frac{1}{2}^{n_i+m_i+1} \cdot (1+F_A)$$

In this equation,  $n_i$  refers to the number of generations between  $X$  and the common ancestor through one side of the pedigree, while  $m_i$  comes from the number of generations between  $X$  and the common ancestor through the *other* side of the pedigree. (The “+1” comes from multiplying  $\frac{1}{2}^{(n_i+m_i)} \cdot \frac{1}{2}^{(n_i+m_i)}$ —this is the same as the increase in the exponent in  $2^1 \cdot 2^2 = 2^3$ . You may find it easier to think of the entire exponent as  $n-1$ , which is simply the number of links in the entire loop from  $X$  to a common ancestor and then back the other side. We’ll use this notation for the rest of the class.) This coefficient is then multiplied by  $1 +$  the inbreeding coefficient of the common ancestor ( $F_A$ ) and then summed over all loops by which an allele could be IBD. For example, consider the following pedigree, where squares indicate males, circles indicate females, and lines between males and females lead to offspring in the next generation (i.e., further down the figure):



We can calculate the inbreeding coefficient of individual  $K$  as follows:

$$F_K = \sum \frac{1^{n-1}}{2} \cdot (1 + F_A) = \frac{1^5}{2} \cdot (1 + 0) + \frac{1^5}{2} \cdot (1 + 0) = 0.0625$$

Note that there are two routes in which an allele can be IBD in  $K$  (highlight the common ancestor of the two copies with a bar):  $H\bar{D}\bar{A}\bar{E}I$  and  $H\bar{D}\bar{B}\bar{E}I$ , accounting for an origin in her great-grandfather or great-grandmother, respectively. A slightly more complicated situation is calculating the inbreeding coefficient for  $L$ :

$$F_L = \sum \frac{1^{n-1}}{2} \cdot (1 + f_A) = \frac{1^6}{2} \cdot (1 + 0) + \frac{1^6}{2} \cdot (1 + 0) + \frac{1^3}{2} \cdot (1 + 0) = 0.15625$$

Now we have loops  $JH\bar{D}\bar{A}\bar{E}IK$ ,  $JH\bar{D}\bar{B}\bar{E}IK$ , and  $JHK$ . However, there is still no inbreeding among potential common ancestors, so  $F_A$  remains 0. It can be helpful to organize this information into a table, especially as the number of loops grows:

Loop	$n$	$F_{CA}$
$JH\bar{D}\bar{A}\bar{E}IK$	6	0
$JH\bar{D}\bar{B}\bar{E}IK$	6	0
$JHK$	3	0