Evolution Module

6.5 Inbreeding (Revised)

Bob Gardner and Lev Yampolski

Integrative Biology and Statistics (BIOL 1810)

Fall 2007

Note. In this section, we introduce some ideas from graph theory and use them to quantify the level of inbreeding of an individual.

Note. Informally, a *graph* is a mathematical structure, which is usually represented by a collection of dots, called *vertices*, some of which are connected by line segments (or curves), called *edges*. This idea should not be confused with the use of the term "graph" in the setting of graphing a function.

Definition. A graph G is a set V of vertices,

$$V = \{v_1, v_2, \dots, v_n\}$$

and a set E of unordered pairs of elements of V, called *edges*.

Note. The edge consisting of vertices v_1 and v_2 is denoted (v_1, v_2) or (v_2, v_1) . We can draw pictures of a graph as follows. A graph G with vertex set $V = \{0, 1, 2, 3\}$ and edge set $E = \{(0, 1), (1, 2), (2, 3), (3, 0)\}$ can be represented as:



Definition. A complete graph on v vertices, denoted K_v , is a graph with vertex set V which satisfies |V| = v, and with edge set $E = \{(v_i, v_j) \mid v_i, v_j \in V, i \neq j\}.$

Example. Here are some examples of complete graphs:



Exercise. How many edges does K_v have (in terms of v)?

Definition. A *walk* in a graph is a sequence

$$W = v_0 e_1 v_1 e_2 v_2 \cdots v_{n-1} e_n v_n$$

which starts and ends with a vertex and alternates between vertices and edges, such that each edge in the sequence has as its two endpoints the vertices before it and after it in the sequence. If the vertices of the walk are distinct (i.e., all different) then the walk is called a *path*. **Example.** Two different walks from vertex 0 to vertex 4 in K_5 are:



The first walk is a path, but the second one is not (since it repeats vertex 2).

Definition. A graph G in which there is a path from any vertex to any other vertex is a *connected graph*.

Note. Graphs which are not connected can be broken into connected components. Consider:



Definition. A cycle on n vertices, denoted C_n , is a graph with vertex set $V = \{v_1, v_2, \dots, v_n\}$ and edge set

$$E = \{(v_1, v_2), (v_2, v_3), \cdots, (v_{n-1}, v_n), (v_n, v_1)\}.$$

Example. A 5-cycle, C_5 , is:



Exercise. How many edges does a cycle on n vertices have?

Definition. Let graph G_1 have vertex set V_1 and edge set E_1 , and let graph G_2 have vertex set V_2 and edge set E_2 . Then graph G_1 is a *subgraph* of graph G_2 if $V_1 \subset V_2$ and $E_1 \subset E_2$.

Exercise. Does this graph have a subgraph which is a cycle?



Definition. A graph which has no subgraphs which are cycles is *acyclic*. An acyclic connected graph is a *tree*.

Example. There are 3 trees on 5 vertices:



Example. What are the 6 trees on 6 vertices?

Exercise. How any edges does a tree on v vertices have?

Note. We are primarily interested in applying graphs to genealogies in this section. We do so by letting the individuals of the geneaology be represented by vertices and the relationship of parent-offspring be represented by edges. So the traditional genealogy in represented as a graph as follows:



Note. In calculating the inbreeding coefficient, we want to focus on only two individuals in a population. Suppose, for example, that the two red individuals in the genealogy above mate to produce an offspring. Since the two red parents are related (they share a pair of grandparents, and so are cousins), the resulting offspring is inbreed. Now the graph above is cluttered with individuals who do not affect the relatedness of the two red individuals. So we crop off the unnecessary individuals by keeping the common ancestors (the grandparents) and the intervening individuals which link the red individuals to the common ancestors. We get the following:



Adding the offspring (in green) of the red individuals produces the graph:



We now find all cycles in the graph which contain the offspring (in green) and a common ancestor (in blue). There are two such cycles (in blue):



Note. As seen previously, inbreeding has the effect on a population of reducing heterozygosity (relative to the level of heterozygosity expected from Hardy-Weinberg), though it will not, by itself, change allele frequencies.

Note. For an inbred individual, we define the inbreeding coefficient F as the probability that the two alleles of a locus chosen at random are *identical by descent* (*IBD*). By identical by descent, we mean that the two alleles are descendants of the same allele, with one copy being passed from an ancestor down a lineage to the father and then to the individual, and the other copy being passed from the same ancestor down the lineage to the mother and then to the individual. Since an individual is inbreed, by definition, if the mother and father are related, we will calculate F by following the passage of the allele in question from the ancestor to the father the mother, and finally to the individual.

Note. Consider the offspring of two half-cousins (that is, the offspring of two individuals who share a single grandparent). The graph representing the relevant genealogy is:



We start by assuming that individual A is not inbreed. The probability that A passes the same allele through gametes to individuals B and C is 1/2. The probability that B passes the allele inherited from A on to D is 1/2. Similarly, the probability that C passes the allele inherited from A on to E is 1/2. Next, the probability that the allele D inherited from B is passed on to I is 1/2, and the probability that the allele E inherits from C is passed on to I is 1/2. We now multiply these probabilities (of independent events) to get the probability that individual I has inherited the same allele at the given locus along the two lines of descent as: $(1/2)^5 =$ 1/32. Therefore the inbreeding coefficient of individual I is $F_I = 1/32$. Notice that each step in the above process is represented by a vertex in the unique path joining the parents of I.

Note. Next, let's assume that individual A is inbred with inbreeding coefficient F_A . Then as before, the probability that A passes on the same allele to both of it's offspring is 1/2. However, since A is inbreed, it is possible that both of A's alleles at the relevant locus are already identical by descent. The probability of this, by definition, is F_A . So if A passes one of its alleles to one offspring and passes the other allele to the other offspring (an event with probability 1/2), then the offspring of A will still have alleles which are identical by descent. Hence the probability that both of A's offspring get alleles which are identical by descent is $1/2 + F_A \times 1/2 = (1 + F_A)$ F_A /2. We then compute the probabilities of the alleles being passed on to individual I as above. This time, the inbreeding coefficient for I is $F_I = (1/2)^4 \times (1+F_A)/2 = (1/2)^5(1+F_A).$

Note. If the parents of an individual have more than one common ancestor, we simply compute the contribution to the inbreeding coefficient of the offspring by summing along each of the paths from the parents through a common ancestor, and then sum. For example, consider the inbreeding coefficient of the offspring of cousins. The cousins share two common ancestors (their grandparents) and, as seen above, this produces two paths between the cousins which pass through a single common ancestor:



If the grandparents are not inbreed, then the inbreeding coefficient for individual I is $F_I = 1/32 + 1/32 = 1/16$. If the two common ancestors, denoted GF and GM, are inbred

with inbreeding coefficients F_{GF} and F_{GM} respectively, then the inbreeding coefficient of I is

$$F_I = (1/2)^5 (1 + F_{GF}) + (1/2)^5 (1 + F_{GM}).$$

Note. Now let's consider a more complicated genealogy. The following is based on work done by Sewall Wright concerning the breeding of cattle (Wright 1922). Consider the genealogy:



We wish to find the inbreeding coefficient of individual RG, F_{RG} . To do so, we need to find the common ancestors of the parents of RG, namely RDG and PR. Tracing through the graph, we find that the common ancestors are CE and LR. Notice that there are two different paths between the parents

which pass through CE and two different paths which pass through LR. This leads to the following four paths:



Next, we count the number of vertices in each path from RDG to PR. This yields two paths of length 4 and two paths of length 7. So we have $F_{RG} = (1/2)^4 + (1/2)^4 + (1/2)^7 + (1/2)^7 = 0.140625$.

Excercise. Suppose two unrelated individuals have two offspring, a first generation male and female. Next, the first generation male and female mate with each other to produce a second generation male and female. The second generation male and female produce a third generation male and female, and so forth. This is repeated mating of siblings.



Find the inbreeding coefficient for the second generation siblings, F_2 , and the third generation siblings, F_3 . What is the inbreeding coefficient of the *n*th generation siblings, F_n , in terms of the inbreeding coefficients of the previous generations. (HINT: This one requires mathematical induction and is tricky! The answer is $F_n = (1/2)F_{n-1} + (1/4)F_{n-2} + (1/4)$.)