Introduction

The most common definition of inbreeding is a preferential mating of closely related individuals. While there is nothing wrong with this definition, it does not provide the broad perspective needed when thinking about relatedness of genes in a population. The degree of relatedness of individuals or genes in a population depends on the size of the population. The way to think about this problem is to consider the size of a population required to provide separate ancestors for all the individuals. Take a population of diploid, sexual, organisms as an example. Every individual must have two parents, four grandparents, eight great-grandparents, and so on. With $t$ generation back in time, an individual will have $2^t$ ancestors. It is easy to see that it does not take too many generations before the number of individuals required for unrelated ancestors becomes larger than any real population could possibly meet. Hence, in a finite population any pair of individuals must be related to each other through one or more common ancestors in the past. This means that given random mating in both a large population and a small population, individuals in the small population will be more closely related (inbred) than those in a large population.

The practical consequence of inbreeding is that two potential mates will have shared a common ancestor at some point in their history and carry genes that were passed down to both individuals from that same common ancestor. Such genes are said to be identical by descent (IBD); i.e., they have originated by replication of one single gene in a previous generation. We measure the probability that any two alleles at a randomly chosen locus within a single individual are identical by descent as the coefficient of inbreeding ($F$). The coefficient of inbreeding refers to a single individual and expresses the degree of relationship between those individuals' parents.

Note that the coefficient of inbreeding will be different for individuals from different families, even if they come from a randomly mating population. This is because random mating of some pairs will be more closely related to random mating of other pairs.

You might have foreseen that it is possible to think about inbreeding in three different ways:

I. Individual inbreeding in a pedigree sense
II. Inbreeding as a population deviation from Hardy-Weinberg equilibrium
III. Inbreeding arising from a finite population size

I. Individual inbreeding in a pedigree sense

Perhaps the simplest way of viewing inbreeding is on the pedigrees of individuals. Below is the conventional representation of a pedigree, with the lines representing gametes contributed by parents to offspring, squares and circles representing males and females respectively. The example below shows a mating between first cousins.

![Pedigree Diagram]
$F_1$, the inbreeding coefficient for individual 1, can be easily computed for this pedigree by using the method of "path analysis" which has the following steps:

1. Find each path that alleles might take to become IBD.
2. Count the number of lines ($n$) in each path (path segments).
3. Compute the probability of the path. The probability for a complete path is $(1/2)^{n-1}(1 + F_{CA})$. $F_{CA}$ is the inbreeding coefficient of the common ancestor (CA).
4. Sum the probabilities over all possible paths (indexed by $i$ below)

$$F_1 = \sum_i (1/2)^{n_i-1}(1 + F_{CA})$$

Now let's take another look at our first-cousin pedigree. We can label those individuals that contribute to the inbreeding of individual designated I. We then identify the complete paths by which individual "I" can receive alleles that are IBD. For clarity we can separate those paths from each other and write them in a more simplified way. Now it is a simple task to count the number of path segments in each complete path and compute the probabilities. Both paths 1 and 2 have six segments.

Except for the path segments to the common ancestor, the probability that a particular allele is transmitted along a segment = $1/2$. There are two possible CAs in this pedigree (individual A and individual B). The CA always has two paths. The probability of transmitting the specified allele along the two paths connected to the CA is

$$\frac{1}{2} \text{ prob. that CA transmitted identical allele down both paths} + \left(\frac{1}{2}\right)F_{CA} \text{ prob. that CA transmitted diff alleles that were IBD}$$
We get a probability of $\frac{1}{4}$ for the same allele because there are 2 out of 4 ways to transmit the same allele. Let $x_1$ and $x_2$ be the two alleles. The four possible ways to transmit them down two paths is

<table>
<thead>
<tr>
<th>Path 1</th>
<th>Path 2</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_1$</td>
<td>$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$</td>
</tr>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$x_1$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>$x_2$</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

The first part of the above equation is due to IBD from $x_1 + x_1$ or $x_2 + x_2$:

$$\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

The second part comes from the probability that different alleles $x_1 + x_2$ or $x_2 + x_1$ are IBD:

$$(\frac{1}{4})F_{CA} + (\frac{1}{4})F_{CA} = (\frac{1}{2})F_{CA}$$

So the total probability of getting two IBD alleles from the CA is:

$$\frac{1}{2} + (\frac{1}{2})F_{CA} = (\frac{1}{2})(1 + F_{CA})$$

All the other path segments have a probability of $\frac{1}{2}$ because only transmission of identical alleles is allowed from that point on. Accommodating all the other segments in a path gives:

$$\left(\frac{1}{2}\right)^{n-1}(1 + F_{CA})$$

[same as above; that is a relief]

First let’s assume the ancestral population was completely outbred; $F_{CA} = 0$. In this case the inbreeding coefficient for individual I is:

$$F_1 = (1/2)^{n-1}(1 + 0) + (1/2)^{n-1}(1 + 0)$$

$$F_1 = 0.03125 + 0.03125 = 0.0625$$

Now let’s assume the ancestral population was inbred with; $F_{CA} = 0.375$. In this case the inbreeding coefficient for individual I is:

$$F_1 = (1/2)^{6-1}(1 + 0.375) + (1/2)^{6-1}(1 + 0.375)$$

$$F_1 = 0.043 + 0.043 = 0.09$$
II. Inbreeding as a population deviation from HW equilibrium

Now consider the effect of inbreeding on the frequencies of genotypes in a population. $F$ = the probability that two gametes are IBD due to inbreeding beyond random mating expectations in a local population.

<table>
<thead>
<tr>
<th>Male gametes</th>
<th>A (freq = $p$)</th>
<th>a (freq = $q$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>female gametes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A ($p$)</td>
<td>AA</td>
<td>$p[p(1-F) + F]$</td>
</tr>
<tr>
<td>a ($q$)</td>
<td>Aa</td>
<td>$q[p(1-F)]$</td>
</tr>
</tbody>
</table>

For AA in this proof, the probability that that the first gamete contains allele A is $p$. Now we must consider two sources for the second gamete containing the A allele. The first source is out-breeding. Under out-breeding alleles are the same states but NOT IBD, with a probability of $(1 - F)$. The second source of the A allele in the second gamete is due to inbreeding. Here the probability that the alleles are IBD = $F$.

$$f_{AA} = \frac{p}{\text{The prob that the first allele was an } A} \left[ \frac{p(1-F)}{\text{The prob A by random mating}} + \frac{F}{\text{The prob of A by mating with a relative}} \right]$$

Given the first gamete was A (with probability = $p$) and inbreeding has occurred (with probability = $F$), then the second gamete must have carried the A allele with probability = 1 (otherwise it could not have been inbreeding because otherwise alleles would not have been the same).

For AA, the frequency is:

$$p[p(1-F) + F]$$
$$p[p - pF + F]$$
$$p^2 - p^2F + pF$$
$$p^2 + pF(1 - p)$$
$$p^2 + pqF$$

Thus, in a population with an inbreeding coefficient = $F$, the genotype frequencies are:
So what does this tell us about genotype frequencies under Inbreeding? Let's look at an example where the frequency of the A allele = $p = 0.6$, and the frequency of the a allele = $q = 0.4$. For three different inbreeding coefficients we get the following genotype frequencies.

<table>
<thead>
<tr>
<th>$F$ = 0 (HW)</th>
<th>$F$ = 0.5</th>
<th>$F$ = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA = 0.36</td>
<td>AA = 0.48</td>
<td>AA = 0.6 $[= p]$</td>
</tr>
<tr>
<td>Aa = 0.48</td>
<td>Aa = 0.24</td>
<td>Aa = 0</td>
</tr>
<tr>
<td>aa = 0.16</td>
<td>aa = 0.28</td>
<td>aa = 0.4 $[= q]$</td>
</tr>
</tbody>
</table>

Inbreeding yields change in the genotype frequencies of the population, but does not alter the allele frequencies. Hence “losing variation to inbreeding” is only a loss of heterozygosity; there is no loss of allelic variation!

Disequilibrium among loci: Inbreeding slows the rate of approach to equilibrium among different loci. Linkage disequilibrium breaks down via recombination producing gametic types (i.e., configurations of alleles in gametes) that were not present in the parents. Consider an extreme case of inbreeding where heterozygotes are very rare. In such a case recombination will only rarely result in new gametic types because most of the loci are homozygous; hence the approach to equilibrium will be extremely slow.

The most extreme cases of linkage disequilibrium are found in predominantly self-fertilizing species of plants (e.g., barley [Hordeum vulgare] and wild oat [Avena barbata]). Barley is a species that will self-fertilize at a rate close to 99%. An agricultural population of barley was found to have a value of $D$ that was 66% of its theoretical maximum. The persistence of LD in barley is thought to be a consequence of the selfing mating pattern.

\[ F = (H_{\text{HW}} - H)/H_{\text{HW}} \]

We can rewrite this as:
\[ H = H_{\text{HW}} - (H_{\text{HW}} \times F) \]

Remember that \( H_{\text{HW}} = 2pq \)

\[ H = 2pq - 2pqF \]

\[ H = 2pq (1 - F), \text{ as above} \]

So, you can see that if we have a sample of a real population we can measure the observed frequency of heterozygotes and estimate the expected frequency under HW and thereby obtain an estimate of \( F \).

Note that population subdivision is a very important aspect of population genetics. Because population subdivision causes an inbreeding-like effect, it is useful to measure subdivision in terms of the decrease in the proportion of heterozygote genotypes. In structured populations, individuals do not mate at random; rather, they are more likely to mate with individuals of the same subpopulation than with individuals from a different subpopulation. We can take advantage of such a non-random mating structure to measure the extent of population subdivision by a quantity called the fixation index (denoted as \( F_{ST} \)). \( F_{ST} \) measures the reduction of heterozygosity of a structured population (i.e., non-random mating due to structure) relative to a random mating population (i.e., no structure).

\[ F_{ST} = \frac{(H_T - H_S)}{H_T} \]

where

\[ H_T = \text{The expected heterozygosity of an individual in a total population that is random mating} \]

\[ H_S = \text{The expected heterozygosity of an individual in a subpopulation that is random mating} \]

The use of \( F_{ST} \) and other such \( F \)-statistics is beyond the scope of this course. All \( F \)-statistics, sometimes also called hierarchical \( F \)-statistics, are types of “inbreeding coefficients” that differ only in the reference population. You should be aware that \( F \)-statistics are important and very popular tools for studying the dynamics of genetic variation in natural populations.

**III. Inbreeding arising from a finite population size**

Let’s set up an idealized population with the following characteristics:

1. A finite population with \( N \) individuals
2. Each individual produced equal numbers of sperms and eggs
3. Sperm and eggs unite at random

Note this type of population is not as strange as it might seem at first glance. A hermaphroditic marine invertebrate that broadcast spawns by shedding gametes into the ocean will have similar characteristics.

Let’s start generation zero \( (g_0) \) completely out bred, i.e., all alleles at a locus can be regarded as non identical. So given any one allele at a locus, the likelihood of a random union with an identical allele is equal to the likelihood of picking ones own allele from the gamete pool, and this is simply \( 1/2N \). This is the inbreeding coefficient of generation zero.

\[ F_0 = \frac{1}{2N} \]
Now let’s consider the second generation. There are now two ways that identical alleles can be united. The first way is by sampling one’s own gamete from the current generation of gametes \((g_1)\) just as was shown above (i.e., \(1/2N\)). The second is by sampling the non-self gamete pool (probability = \(1 - 1/2N\)) and getting a gamete that is IBD (probability = \(F_0\)).

\[
F_1 = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_0
\]

We can extend this to any number of subsequent generations, \(g\).

\[
F_g = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_{g-1}
\]

Thus the inbreeding coefficient for a finite population is made up of two parts. One part is the incremental increase in inbreeding \((1/2N)\) occurring within a generation, and the other part is the “remainder” of the inbreeding from the last generation. Hence we can think about the increase in inbreeding per generation (symbolized by \(\Delta F\)).

\[
\Delta F = \frac{1}{2N}
\]

At this point we have a big problem: true breeding structures rarely conform to our idealized population. Natural populations will not accumulate inbreeding as specified in the above formulas. To model natural populations, we will express the size of the population as an effective number of breeding individuals. We call this the EFFECTIVE POPULATION SIZE (symbolized by \(N_e\)).

\(N_e = \) the number of an ideal population which undergoes the same rate of increase in \(F\) as the actual (non-ideal) population. This is sometimes called the INBREEDING EFFECTIVE SIZE of the population. Hence, we can model the incremental change in inbreeding in a natural population by:

\[
\Delta F = \frac{1}{2N_e}
\]

and

\[
N_e = 1/2 \Delta F
\]

Any factor that affects the variance in reproductive success will impact the \(N_e\). Ecological factors will certainly affect fecundity (or fitness); hence such factors will have an impact on \(N_e\). Can you predict the consequences?
Let’s look at some cases where this is important.

1. **Unequal numbers in successive generations.** It is highly unlikely that the census size of the population will be similar among different generations. Droughts, floods, etc. ensure that over time the variance in N will be quite high. In this case the effective number \( N_e \) will be the harmonic mean on the numbers in each generation.

\[
\frac{1}{N_e} = \frac{1}{g} \left[ \frac{1}{N_1} + \frac{1}{N_2} + \frac{1}{N_3} + \ldots + \frac{1}{N_g} \right] \quad \text{(approx.)}
\]

Under a harmonic mean the generations with the smallest numbers have the most effect. This means that if you sample a present-day population with a very large census size, it could be exhibiting inbreeding at a rate consistent with a much smaller size than the census size because of a period in its history when it had a very small census size, say due to something like a drought. The process of a population passing thorough a period of a small size is called a BOTTLENECK. The effect of a small population size in the history of a population on genetic diversity is called the BOTTLENECK EFFECT. The expansion to a large census size will reduce the amount of new inbreeding (see the first term of our equation for \( F_g \) above), but it has no effect on the amount of previous inbreeding.

### Effective population size is dominated by historical lows and can be very much lower than current census size.

![Graph showing population crash and recovery](image.png)

An example of a species with an extreme boom-bust abundance cycle is the Atlantic stripes bass (*Morone saxatilis*). Interestingly, this species is well known as one of the most depauperate in genetic variation.
2. Different numbers of males and females. It is quite common for males and females to be present in significantly different numbers in a population. This is called the sex ratio, and the rate at which a population accumulates inbreeding is very sensitive to the sex ratio. The reason is that the sampling variance differs between the sexes (just like the sampling variances differed among generations in the above case). Let’s use the following symbols for counts of males and females:

\[ N_m = \text{the number of males in the population} \]

\[ N_f = \text{the number of females in the population} \]

\[ N = N_m + N_f \]

The effective number will be twice the harmonic mean of the two sexes.

\[ \frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f} \quad \text{(approx.)} \]

\[ N_e = \frac{4N_m N_f}{N_m + N_f} \quad \text{(approx.)} \]

These approximations are close enough except in the case when both \( N_m \) and \( N_f \) are small. The rate of inbreeding will depend on the number of the less numerous sex. For example, if over the long term a population maintained a harem system with very large numbers of females and only one male, the effective number would be close to 4.
3. Variance. When the variance of reproductive success is greater than expected under random mating the effective population size will be less than the census size. There are numerous other ways in which this can occur, such as non-random distribution of family sizes, overlapping generations, variation among individuals in life spans and age of sexual maturity, and natural selection.

The concept of \( N_e \) is very important in the disciplines of ecology and conservation biology. However, there are constraints and limitations. Many different genetic markers are available from which to obtain data for the measurement of \( N_e \) (mtDNA, allozymes, microsatellites, etc.). Because of different modes of inheritance and levels of variability there can be different levels of precision and power associated with the estimate of \( N_e \). There is also debate about the methods used to estimate \( N_e \) and the interpretation of such estimates. In particular, many methods make assumptions such as no natural selection, and no migration between populations; these assumptions might be incorrect in certain cases. Such issues are beyond the scope of this course, but would be covered well in a full course on population genetics.

### Table

<table>
<thead>
<tr>
<th>Species</th>
<th>( N_e/N )</th>
<th>Species</th>
<th>( N_e/N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puma</td>
<td>0.64</td>
<td>Moose</td>
<td>0.27</td>
</tr>
<tr>
<td>Florida Panther</td>
<td>0.25</td>
<td>Northern Elephant Seal</td>
<td>0.22</td>
</tr>
<tr>
<td>Rainbow trout</td>
<td>0.90</td>
<td>White-tailed Deer</td>
<td>0.52</td>
</tr>
<tr>
<td>Coho Salmon</td>
<td>0.24</td>
<td>Bighorn Sheep</td>
<td>0.44</td>
</tr>
<tr>
<td>Re-spotted newt</td>
<td>0.07</td>
<td>Grey Squirrel</td>
<td>0.59</td>
</tr>
<tr>
<td>WoodFrog</td>
<td>0.44</td>
<td>Black Bear</td>
<td>0.69</td>
</tr>
<tr>
<td>Red-cockaded Woodpecker</td>
<td>0.63</td>
<td>Grizzly Bear</td>
<td>0.28</td>
</tr>
<tr>
<td>Acorn Woodpecker</td>
<td>0.09</td>
<td>Wild Oats</td>
<td>0.51</td>
</tr>
<tr>
<td>Spotted Owl</td>
<td>0.39</td>
<td>White Spruce</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Note that these estimates were made in different studies and in some cases by using different methods. Comparison is not always straightforward. Data were obtained from review by Frankham (1995)
**Inbreeding depression**

In species that outcross, i.e., those that regularly avoid close inbreeding, inbreeding at a rate greater than expected under random mating will lead to offspring with reduced fitness. This effect is called **inbreeding depression**. Inbreeding depression is one of the oldest observations related to breeding, and as a result most human cultures have long prohibited close inbreeding. Interestingly, this prohibition has not always been strictly observed. Charles Darwin married his first cousin, Emma Wedgewood. Darwin was himself the grandson of Josiah Wedgewood. Health difficulties of Darwin’s own children have, in part, been attributed to inbreeding depression.

As we have seen previously, although a deleterious recessive allele can be quite rare in a population as a homozygous recessive, it can be found “hiding” in a surprisingly large number of heterozygotes. As a recessive allele, its effect on fitness is masked by the dominant allele. Inbreeding, by producing an excess of homozygote genotypes, lowers the mean fitness of inbred individuals by putting together these recessive alleles at more loci than in outbred individuals.

Let’s take a look at a single locus example in humans. **Cystic Fibrosis (CF)** is an extremely common recessive lethal genetic disease in North America. The CF allele causes excessive secretions of mucus from the lungs, pancreas, and other organs that lead to blockages in the lungs, digestive track, etc. These blockages eventually lead to pneumonia and other types of infection. If untreated most children with CF die between that ages of three and four years. The frequency of CF in the United States is about 1 in 2500.

Let’s use our knowledge of populating genetics to determine if inbreeding will lead to an increased chance of the CF allele appearing as a homozygous recessive.

**Random Mating:**

- Frequency of CF = $q = 1/2500$
- Risk of CF under random mating = $q^2 = 0.00000016$

**First Cousin Mating:**

Remember that $F$ for the offspring of a fist cousin mating = 1/16 (assuming that the great grandparents were unrelated)

- Risk of CF in a first cousin mating = $q[q(1-F) + F] = 0.00002515$

*The risk of CF in the offspring of a first cousin marriage is 157 times larger than in the offspring of a random mating.*
EFFECTS OF INBREEDING IN CAPTIVITY

Because the detrimental effects of close inbreeding are well known, inbreeding depression has been an issue of great concern in the conservation and management of endangered species. In most cases, the effects of inbreeding on endangered species have been studied in captive populations. It is obvious that managers of captive populations, most especially in cases of endangered species, actively work to make the environment less harsh than the natural environment of the organism. Because the ultimate goal of a captive breeding program is usually to replenish the natural environment by seeding it from captive populations, the consequences of evaluating inbreeding effects in a more benign environment must be considered.

Evaluating the effect of inbreeding in captivity on individuals later released into the natural environment is problematic. One study conducted by Jimenez et al. (1994) took a sample of mice (*Peromyscus leucopus*) from their natural environment and brought them into the lab where half were inbred \( F = 0.25 \) and half were outbred \( F = 0 \). Both populations of mice were re-released into their natural environment on three different occasions and their progress was followed for 10 weeks. The inbred individuals had a lower survivorship (56%) that the outbred individuals.

Another interesting observation was that the males from the inbred population lost a significant amount of body mass throughout the experiment whereas the non-inbred males did not. Survivorship is summarized below.

<table>
<thead>
<tr>
<th>Population</th>
<th>Environment</th>
<th>Non-inbred</th>
<th>Inbred</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) General</td>
<td>Captive</td>
<td>1</td>
<td>( \overline{w}_I )</td>
</tr>
<tr>
<td></td>
<td>Natural</td>
<td>( \overline{w}_N )</td>
<td>( \overline{w}_I \overline{w}_N )</td>
</tr>
<tr>
<td>(b) Mouse data</td>
<td>Captive</td>
<td>1</td>
<td>( \overline{w}_I = 0.935 )</td>
</tr>
<tr>
<td></td>
<td>Natural</td>
<td>( \overline{w}_N = 0.221 )</td>
<td>( \overline{w}_I \overline{w}_N = 0.207 )</td>
</tr>
</tbody>
</table>

The observed survival of inbred mice is 22% less than predicted by assessing inbreeding in captivity (0.046 verses 0.207). It seems that inbred individuals will have lower survivorship than expected by studies in captivity.
INBREEDING DEPRESSION IN A WILD POPULATION

To further investigate the relevance of inbreeding to fitness in natural populations, Slate et al. (2000) investigated the relationship between lifetime breeding success and inbreeding in a wild population of Red Deer (Cervus elaphus) on the Isle of Rum, Scotland. The results showed that relatively outbred deer, as measured by heterozygosity averaged over N loci had greater breeding success (number of offspring) over the course of their lifetimes than did more inbred deer in the same population. The implications for threatened populations are important; the likelihood of extinction might be increased if management decisions lead to reductions in breeding success due to inbreeding.

**Breeding success in Red Deer in relation to heterozygosity**

Standardized heterozygosity is the ratio of the individual’s heterozygosity to the mean heterozygosity at the same loci.

Adapted from Slate et al. 2000
Keynotes on Inbreeding Depression and Captive Breeding

- Inbreeding depression can significantly affect fitness.

- The effects of inbreeding on fitness will vary among species. The effects will be difficult to predict in specific cases.

- The effects of inbreeding are likely to be variable over populations, traits and environments.

- Species that typically have low effective population sizes in natural populations will have few deleterious alleles to contribute to inbreeding depression.

- The negative effects of inbreeding might be greater than predicted by the effect measured in the benign environments of captive populations.

- Detecting inbreeding depression is difficult, and methods have low power in many cases. Statistical power might be a particularly important issue in the benign environments of captive populations.

- Severe inbreeding depression does not mean that a population is beyond hope. Cases such as Speke’s gazelles (4 founders), Przewalski’s horses (13 founders) and Black footed ferrets (6 founders) demonstrate that stable and successful captive breeding programs can be established from very low numbers of founders.

- Sometimes alleles with large detrimental affects can be intentionally purged from a captive population in an attempt to reduce inbreeding depression. While this is clearly possible in theory, the often cited successful application of this management strategy to captive populations of Speke’s gazelles has recently been questioned.