INBREEDING

When matings take place between relatives, the pattern is referred to as inbreeding.

There are three common areas where inbreeding is observed:
- mating between relatives
- small populations
- hermaphroditic plants and animals

Inbreeding tends to reduce diversity within populations (but may increase diversity, or at least the variance, among populations).

Decrease of Heterozygosity Under Inbreeding

In a complete selfing population, the number of heterozygotes at any given locus decreases.

Let us take gene A with two alleles:

<table>
<thead>
<tr>
<th>t = 0</th>
<th>(1/4)AA</th>
<th>(1/2) Aa</th>
<th>(1/4)aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>selfing</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>t = 1</td>
<td>AA</td>
<td>(1/4)AA</td>
<td>aa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1/2)Aa</td>
<td>(1/4)aa</td>
</tr>
</tbody>
</table>

Genotype frequencies at t= 1 can then be calculated:

\[ f(AA) = \sum [f(genotype at t = 0)] X [probability of getting AA from each genotype] \]

\[ = [f(AA|t=0)][P(AA|t=1)] + [f(Aa|t=0)][P(AA|t=1)] + [f(aa|t=0)][P(AA|t=1)] \]

\[ = (1/4)(1) + (1/2)(1/4) + (1/4)(0) \]

\[ = 3/8 \]

\[ f(Aa) = 2/8 \]

\[ f(aa) = 3/8 \]

Note that under inbreeding, the genotype frequencies are not in HW proportions (mating is nonrandom).

Inbreeding coefficient

Inbreeding levels quantified by the *inbreeding coefficient F*.

This coefficient takes advantage of decrease in heterozygosity associated with inbreeding:

\[ F = (H_0 - H_I)/H_0 \]

\[ H_0 \] is expected heterozygosity under random mating

\[ H_I \] is heterozygosity under inbreeding.
Note that

\[ H_i = H_0 - H_0 F = H_0 (1 - F) \]

For example, when \( p = q = 1/2 \), under HWE \( H_0 = 1/2 \).
If we have one round of selfing, at \( t = 1 \),

\[ H_i = 1/4, \]

\[ F = (H_0 - H_i)/H_0 \]

\[ F = (0.5 - 0.25)/(0.5) = 0.5. \]

• Probability definition of \( F \)

\( F \) can also be regarded as a measure of the probability that two alleles in a population are identical by descent or IBD.

Eg. Let us take alleles from a population of gametes

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0.25</td>
</tr>
<tr>
<td>Aa</td>
<td>0.50</td>
</tr>
<tr>
<td>aa</td>
<td>0.25</td>
</tr>
</tbody>
</table>

• Relationship of \( F \) to Population Size

Let's say that at \( t = 0 \)
a (diploid) population has \( N \) individuals = \( 2N \) gametes.
At \( t = 1 \),

the probability that progeny of this population will have alleles that are identical by descent is

\[ F_1 = P(\text{getting allele A1}) \times P(\text{second allele that is IBD to A1}) \]

\[ = 1 \times (1/2N) = 1/2N \]
At $t = 2$
$F_2$ = "new" + "old" inbreeding
\[ \downarrow \]
fraction of alleles that were IBD at $t = 1$
Assuming that the population size stays the same each generation ($N$),
$F_2 = \frac{1}{2}N + (1 - \frac{1}{2}N)F_1$

In general, in an ideal population
$F_t = \frac{1}{2}N + (1 - \frac{1}{2}N)F_{t-1}$
\[ = \Delta F + (1 - \Delta F)F_{t-1} \]
\[ \Delta F = (F_t - F_{t-1})/(1 - F_{t-1}) \]
This last expression is a general expression, but only in an ideal population is $\Delta F = \frac{1}{2}N$. 

• Cumulative Inbreeding

We don’t usually express inbreeding on a per generational base. Rather, we express inbreeding as cumulative inbreeding over \( t \) generations compared to a base, non-inbred population \( (F_0 = 0) \).

\[
F_1 = \Delta F + (1 - \Delta F)F_0 = \Delta F \\
F_2 = \Delta F + (1 - \Delta F)F_1 = \Delta F + (1 - \Delta F)\Delta F \\
= 2\Delta F - (\Delta F)^2 = 1 - (1 - \Delta F)^2
\]

In general,

\[
F_t = 1 - (1 - \Delta F)^t
\]

The quantity \( \Delta F \) is the rate of inbreeding, and is equal to \( 1/2N \) for an ideal population.

• Relationship between Inbreeding and Random Genetic Draft

Both inbreeding and random genetic drift occur in finite populations, and the two processes are related.

for random genetic drift the consequences are

• dispersion of gene frequency among lines
• genetic uniformity within lines
• variance of gene frequency among lines can be related to the rate of inbreeding within lines

For an ideal population, variance of gene frequency within sublines is

\[
\sigma_{st}^2 = \text{variance of gene frequencies under drift} \\
= p_0q_0/2N \\
= p_0q_0\Delta F
\]

Cumulated variance of gene frequency between lines

\[
\sigma_{qt}^2 = p_0q_0[1 - (1 - 1/2N)]' \\
= p_0q_0F_t,
\]

Genotype Frequencies

<table>
<thead>
<tr>
<th>Genotype</th>
<th>initial freq</th>
<th>drift</th>
<th>inbreeding</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1A1</td>
<td>( p_0^2 )</td>
<td>( +\sigma_{qt}^2 )</td>
<td>( +p_0q_0F_t )</td>
</tr>
<tr>
<td>A1A2</td>
<td>( 2p_0q_0 )</td>
<td>( -2\sigma_{qt}^2 )</td>
<td>( -p_0q_0F_t )</td>
</tr>
<tr>
<td>A2A2</td>
<td>( q_0^2 )</td>
<td>( +\sigma_{qt}^2 )</td>
<td>( +p_0q_0F_t )</td>
</tr>
</tbody>
</table>

↑ initial frequencies in ideal population (HWE)

\( F_t = 0 \) for random mating
\( F_t = 1 \) for complete inbreeding.
Inbreeding in Pedigreed Populations

The preceding estimates of inbreeding, given by

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

are an average inbreeding coefficient for a population.

If we actually knew who was mating with who, we can actually calculate \( F_I \), the inbreeding coefficient of a particular individual I. This is not an estimate, but is an exact figure.

This opportunity arises when one has pedigreed populations or pedigree data.

To compute \( F_I \),

1. Find all common ancestors in a pedigree. Alleles in individual I can only be IBD if they were inherited through both its parents via a common ancestor.

2. Trace all genetic paths up through one parent and down another.

\[ F_I = \left(\frac{1}{2}\right)^i(1 + F_A) \]

In general, \( F_I = \left(\frac{1}{2}\right)^i(1 + F_A) \) where \( i \) is the number of individuals in the genetic path.

For example, in the preceding example, the genetic path is DBACE and \( i = 5 \).

Assuming that the ancestor is not inbred, then \( F_A = 0 \), and \( F_I = \left(\frac{1}{2}\right)^5 \).
Why “1/2(1+F_A)” around the common ancestor A?

Consider two alleles \( \alpha_1 \) and \( \alpha_2 \)

The probability that any combination of alleles

\( \alpha_1, \alpha_2; \quad \alpha_1, \alpha_1; \quad \alpha_2, \alpha_2; \quad \alpha_2, \alpha_1 \)

are transmitted to the following generation is 1/4.

AND

\( \alpha_1, \alpha_1 \) these two alleles are IBD
\( \alpha_2, \alpha_2 \) these two alleles are IBD
\( \alpha_2, \alpha_1 \) IBD, only if \( \alpha_2, \alpha_1 \) are IBD, and only if individual A is autozygous - \( F_A \)

Prob. is therefore \( 1/2(1+F_A) = 1/4 + 1/4 + 1/4F_A + 1/4F_A \)

When dealing with more complex pedigrees, you calculate \( F_1 \) by adding up all the \( F_i \)'s through all common ancestors of the parents.

\( \text{Pedigree} \quad \text{Paths: GDACE} \quad \text{GBACE} \)

\( F_1 = F_{\text{GDACE}} + F_{\text{GBACE}} \)

\( = (1/2)^2(1 + F_A) + (1/2)^2(1 + F_B) \)

If \( F_A = F_B = 0 \), then

\( F_1 = (1/2)^2 + (1/2)^2 = 0.0625 \)
**EFFECTIVE POPULATION SIZE**

The effects of genetic drift on variance in gene frequency and on increase of homozygous genotypes, can be expressed in terms of the rate of inbreeding, $\Delta F$.

\[
\sigma^2_h = p_0 q_0 \Delta F \quad \text{(one generation of drift)}
\]

\[
\sigma^2 = p_0 q_0 [1 - (1 - \Delta F)^t] = p_0 q_0 \Delta F_t \quad \text{($t$ generations of drift)}
\]

\[
f(A_1A_1) = p_0^2 + p_0 q_0 \Delta F_t
\]

These expressions are true for any population structure.

In the idealized population is $\Delta F = 1/2N$, where $N$ is the actual (census) number of individuals.

$\Delta F$, the rate of inbreeding, is thus $1/(2N_e)$,

where $N_e$ is calculated for specific departures from the ideal model.

---

**Reality Check**

In real populations, $N \neq$ number of breeding individuals nor do all individuals contribute the same number of offspring.

the effective population size, $N_e$, takes into account population structure and history.

$N_e$ is usually less than $N$, the census number of individuals.

For humans,

$N \sim 6,000,000,000$

$N_e \sim 500,000$

---

**Effective population size under different demographic scenarios**

- Unequal numbers in successive generations

In many natural populations, populations size oscillates over time and changes between successive generations.

\[
\frac{1}{N_e} = \frac{1}{t} \left( \frac{1}{N_1} + \frac{1}{N_2} + \ldots + \frac{1}{N_t} \right)
\]

where $N_1$, $N_2$, $N_t$ are the actual sizes of the population over $t$ generations.
For example, 

<table>
<thead>
<tr>
<th>Gen</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
</tr>
<tr>
<td>4</td>
<td>10000</td>
</tr>
</tbody>
</table>

In this example, $N_e = 36$.

generations with the smallest numbers have greatest effect on overall rate of inbreeding

most inbreeding occurs with small population size, and the total inbreeding is the "old" inbreeding + "new" inbreeding.

• Unequal numbers of males and females

Many populations consist of fewer breeding males than females. If there are fewer of one sex breeding every generation, this amounts to a kind of bottleneck because 1/2 the alleles must come from each sex.

If $N_m$ and $N_f$ are the numbers of males and females, respectively, each generation:

$$
\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f}
$$

$$
N_e = \frac{4N_mN_f}{(N_m + N_f)}
$$

$N_e$ is twice the harmonic mean of the numbers in the two sexes, and thus tends towards the number of the less numerous sex.

For example, a herd of cattle kept with 4 bulls and 100 cows each generation.

$N_m = 4$, $N_f = 100$

$N_e$ of 15.385

In contrast, a herd maintained with 4 bulls and 200 cows

$N_m = 4$, $N_f = 200$

$N_e$ of 15.686.

The rate of inbreeding decreases very little ($1/2N_e$) by doubling the cow number, because of the bottleneck through the male gene pool.

• Non-random distribution of family size

family size is the number of progeny per parent or pair of parents that survive to become breeding individuals.

In ideal population, each individual contributes equally to the next generation,

family size per pair of parents = 2.

Because each pair of parents produces a very large number of zygotes of which only two survive, survival of zygotes in the ideal population follows a Poisson distribution (the distribution of rare events).

The mean of a Poisson distribution is equal to its variance, so in the ideal population the variance of family size = 2.
In real populations,

Mean family size $\sim 2$.

Variance in family size ($\sigma_k^2$) $> 2$

Reason?
Unequal contribution of parents to the next generation

differing fertilities among the parents
differing viabilities among the zygotes.

A greater proportion of the following generation will be the offspring of a smaller number of parents, so the effective number of parents is less than the actual number.

When the mean family size is 2,

$$N_e = \frac{4N}{2 + \sigma_k^2}$$

where $\sigma_k^2$ is the variance in family size

Under different scenarios

$\sigma_k^2 = 2$  \hspace{1cm} N$_e$ = N

$\sigma_k^2 > 2$  \hspace{1cm} N$_e$ < N

$\sigma_k^2 < 2$  \hspace{1cm} N$_e$ = 2N

(The latter case can be achieved with controlled breeding of laboratory (or zoo) populations, in which equal numbers of males and females are picked from each family to be parents of the next generation).