Lecture 6

Inbreeding and Crossbreeding

Changes in the Mean

1. Inbreeding

Inbreeding (mating of related individuals) often results in a change in the mean of a trait compared with its value in a random-mating population. Its importance is that inbreeding is generally harmful and reduces fitness. In particular, inbreeding often causes a reduction of the mean value for quantitative traits associated with reproduction and viability.

Inbreeding is intentionally practiced to:

- create genetic uniformity of laboratory stocks
- produce stocks for crossing (animal and plant breeding)

Inbreeding is unintentionally generated:

- by keeping small populations (such as is found at zoos). Genetic drift is a special case of inbreeding. The smaller the population, the quicker inbreeding accumulates.
- during selection (which has the effect of reducing the population size relative to the no-selection case).

The critical parameter for describing inbreeding is F, *the probability that the two alleles at a locus in an individual are identical by descent.* In an individual inbreed to amount F, a randomly-choosen locus has both alleles IBD with probability F and hence is a homozygote.

To deduce how inbreeding changes the mean value of a quantitative trait, consider a large number of inbred lines ($F > 0$) that were derived from an initial base population. The initial gene frequencies of alleles A_1 and A_2 at a single locus affecting the trait are p_0, q_0; these gene frequencies are expected to remain the same averaged over all inbred lines (we denote these averages by p and $q = 1 - p$).

To compute the genotypic probabilities under inbreeding, suppose we chose a locus at random. With probability F the two alleles are IBD, and hence this locus is always homozygous, with freq(A_1A_1) = p and freq(A_2A_2) = $q = 1 - p$. If the alleles are not IBD, then the genotypic frequencies follow Hardy-Weinberg. Thus, the expected genotypic frequencies under inbreeding become

Genotype	Alleles IBD	Alleles not IBD	Population frequency
A_1A_1	$F \cdot p$	$(1-F)p^2$	$p^2 + Fpq$
A_2A_1	0	$(1-F)2pq$	$(1-F)2pq$
A_2A_2	$F \cdot q$	$(1-F)q^2$	$q^2 + Fqq$

If the genotypes A_1A_1, A_1A_2, A_2A_2 have values of $a, d, -a$, then the mean under inbreeding becomes

$$\begin{aligned}\mu_F &= a \cdot (p^2 + Fpq) + d \cdot (1-F)2pq - a \cdot (q^2 + Fqq) \\ &= a(2p-1) + 2(1-F)pqd\end{aligned}$$

Noting that the mean character value in a random mating population ($F = 0$) is

$$\mu_0 = a(2p-1) + 2pqd,$$

the mean under inbreeding can be expressed as

$$\mu_F = \mu_0 - 2Fpqd$$

More generally, if there are k loci, then the mean is

$$\mu_F = \mu_0 - 2F \sum_{i=1}^{k} p_i \, q_i \, d_i = \mu_0 - B \, F$$

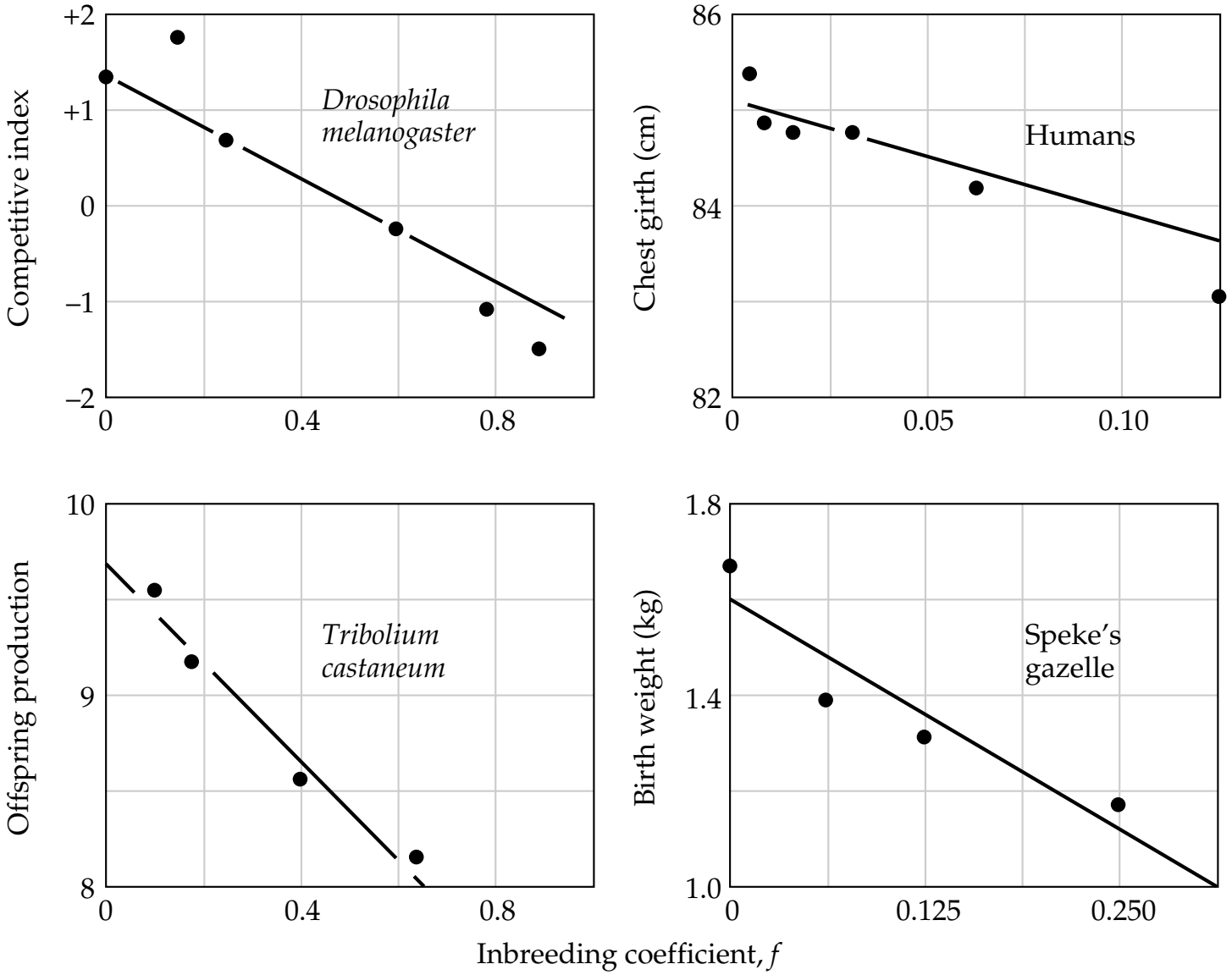
where $B = 2 \sum p_i \, q_i \, d_i$ is the reduction in the mean under complete inbreeding ($F = 1$).

Hence,

- there will be a change of mean value under inbreeding only if $d \neq 0$, i.e., dominance is present.
- for a single locus, if $d > 0$, inbreeding will decrease the mean value of the trait. If $d < 0$, inbreeding will increase the mean.
- with multiple loci, a decrease in the mean under inbreeding (**inbreeding depression**) requires **directional dominance**, with the dominance effects d_i tending to be positive.
- the magnitude of the change of mean on inbreeding depends on gene frequency, and is greatest when $p = q = 0.5$

Inbreeding Depression in Fitness Traits

Fitness-related traits (such as viability, offspring number and body size) often display inbreeding depression, as the following examples illustrates:



Computing the Inbreeding Depression Coefficient, B

In many cases, lines canno tbe completely inbreed due to either time constraints and/or because in many species lines near complete inbreeding are nonviable. In such cases, one must estimate the inbreeding depression from the changes in lines for a series of lines under partial inbreeding.

Recalling that

$$\mu_f = \mu_o + BF$$

it immediately follows that the slope of the regression of the population mean μ_f on the inbreeding coefficient estimates the inbreeding depression coefficient B.

The above equation is true if loci combine additively (no epistasis), in which case the change in mean should be directly proportional to F (i.e., a linear function of F), the inbreeding coefficient. If epistasis is present, the change in mean can be a nonlinear (polynominal) function of F. Hence, if epsitasis is absent, we expect a linear regression of mean on F to be an adequfit fit of the data.

Why do traits associated with fitness show inbreeding depression?

Two competing hypotheses have been proposed:

- **Overdominance Hypothesis:** Genetic variance for fitness is caused by loci at which heterozygotes are more fit than both homozygotes. Inbreeding decreases the frequency of heterozygotes, increases the frequency of homozygotes, so fitness is reduced. Since some inbred lines have means for fitness traits equal to the base population, this explanation cannot be generally true.
- **"Dominance" Hypothesis:** Genetic variance for fitness is caused by rare deleterious alleles that are recessive or partly recessive; such alleles persist in populations because of recurrent mutation. Most copies of deleterious alleles in the base population are in heterozygotes. Inbreeding increases the frequency of homozygotes for deleterious alleles, so fitness is reduced.

While the dominance hypothesis is sufficient to account for inbreeding depression, even a very small factor of overdominant loci have a major effect on the level B. Hence, even though most loci that contribute to

inbreeding depression may due to uncovering of deleterious recessives, the bulk of the contribution to inbreeding depression could theoretically come from a much smaller fraction of overdominant loci.

2. Line Crossing or Crossbreeding

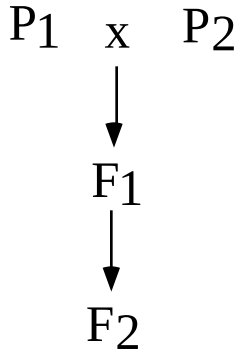
When inbred lines are crossed, the progeny show an increase in mean for characters that previously suffered a reduction from inbreeding. This increase in the mean over the average value of the parents is called **hybrid vigor** or **heterosis**. Fitness lost on inbreeding is restored by crossing.

To see how heterosis is inbreeding depression in reverse, imagine a large number of inbred lines derived from an outbred base population in which $F = 0$. The mean in each line declines with inbreeding, and the mean of all inbred lines is $\mu_F - \mu_0 - B\,F$. If all these lines are crossed at random, $F = 0$, and the mean of the crossbreds = μ_0, the mean of the outbred population.

Heterosis can also arise in crosses between outbred (i.e., randomly mating lines), as we detail below.

Single Crosses

Consider the cross between two particular parental strains (P_1 and P_2), which may have no known common origin. In this case the heterosis depends on the difference in gene frequency between the lines, and the amount of heterosis changes from the F_1 to the F_2. Suppose the crossing scheme is:



We will define heterosis as the deviation mean from the midparental value, so that for the F_1,

$$H_{F_1} = \mu_{F_1} - \frac{\mu_{P_1} + \mu_{P_2}}{2}$$

Let the allele frequencies for a diallelic locus in populations 1 and 2, be p and $p + \delta p$, respectively. We assume the genotypes in P_1 and P_2 are in Hardy-Weinberg proportions (which also hold if the lines are completely inbred), giving the means as

$$\mu_{P_1} = (2p - 1)a + 2p(1 - p)d$$
$$\mu_{P_2} = \mu_{P_1} + 2(\delta p)a - 2(\delta p)^2 d$$

In the F_1, the probability of (say) an A_1A_2 locus is the probability of receiving an A_1 from P_1 and an A_2 from P_2 ($p\,[1 - (p + \delta p)\,]$) or an A_2 from P_1 and an A_1 from P_2 ($[\,1-p\,]\,[\,p+\delta p\,]$). Considering the other two genotypes gives the mean of the F_1 (expressed in terms of the means for P_1) as

$$\mu_{F_1} = \mu_{P_1} + (\delta p)a$$

giving a (mid-parental) heterosis (for this particular locus) of

$$H_{F_1} = \mu_{F_1} - \frac{\mu_{P_1} + \mu_{P_2}}{2} = (\delta p)^2 d$$

Hence, for this locus to show heterosis ($h > 0$), we require both a difference in allele frequencies between the populations ($\delta p \neq 0$) and positive dominance ($d > 0$). Note immediately that overdominance ($d > a$) is not required for heterosis. Summing over all loci, the heterosis produced by dominance for this cross becomes

$$H_{F_1} = \sum_{i=1}^{n} (\delta p_i)^2 \, d_i$$

Hence,

- heterosis depends on dominance. $d = 0$ = no inbreeding depression and no heterosis. As with inbreeding depression, directional dominance is required for heterosis.

- H is proportional to the square of the difference in gene frequency between populations. H is greatest when alleles are fixed in one population and lost in the other. $H = 0$ if $\delta = 0$.

- H is specific to each particular cross. H must be determined empirically, since we do not know the relevant loci and their gene frequencies.

Heterosis in F_2

The F_2 generation is derived by mating the F_1 at random. The gene frequencies in the F_1 are the average of the two parents, so that freq(A_1) $= (p + p + \delta)/2 = p + \delta/2$. Since the F_2 is formed by random mating, and the genotype frequencies are in HW equilibrium with allele frequency $p + \delta/2$, giving the F_2 mean as

$$\mu_{F_2} = a([p + \delta/2]^2 - [1 - p - \delta/2]^2) + d2[p + \delta/2][1 - p - \delta/2]$$

A little algbrea shows that

$$H_{F_2} = \mu_{F_2} - \frac{\mu_{P_1} + \mu_{P_2}}{2} = \frac{(\delta p)^2 d}{2} = \frac{H_{F_1}}{2}$$

so that in the F_2, only half the advantage of the F_1 hybrid is preserved. Since (presumably) random mating also occurs in subsequent generations, the heterosis in future generations is the same as the F_2 heterosis, as the allele frequencies do not change and genotypes remain in Hardy-Weinberg frequecies.

Agricultural importance of heterosis

Heterosis is extremely important in world agricultural. Crosses often show **high-parent heterosis**, wherein the F_1 not only beats the average of the two parents (**mid-parent heterosis**), it exceeds the value of the best parent.

The importance of high-parent heterosis is illustrated by the following estimates of the world-wide contribution of heterosis to both yield and land savings. Here the percent hybrid advantage is the yield increase of the hybrid over the best single variety. (After Duvick 1999).

Crop	% planted as hybrids	% Hybrid yield advantage	Annual added yield Percent	tons	Annual Land savings
Maize	65	15	10	55×10^6	13×10^6 ha
Sorghum	48	40	19	13×10^6	9×10^6 ha
Sunflower	60	50	30	7×10^6	6×10^6 ha
Rice	12	30	4	15×10^6	6×10^6 ha

Change of Variance With Inbreeding

Inbreeding causes a re-distribution of genetic variance within and between lines. For completely additive loci, this can be expressed in terms of the genetic variance ($V_A = V_G$) present in the base population:

	General	$F = 1$	$F = 0$
Between Lines	$2FV_A$	$2V_A$	0
Within Lines	$(1-F)V_A$	0	V_A
Total	$(1+F)V_A$	$2V_A$	V_A

Analogous to the single locus case, inbreeding increases genetic variance between lines and decreases genetic variance within lines. With dominance, the expressions are not simple multiples of the base population genetic parameters, and depend on gene frequency, so there is no general solution for the re-distribution of variance.

The heritability within any one inbred line (assuming only additive variance) is

$$h_t^2 = \frac{(1-F_t)V_A}{(1-F_t)V_A + V_E}$$

or

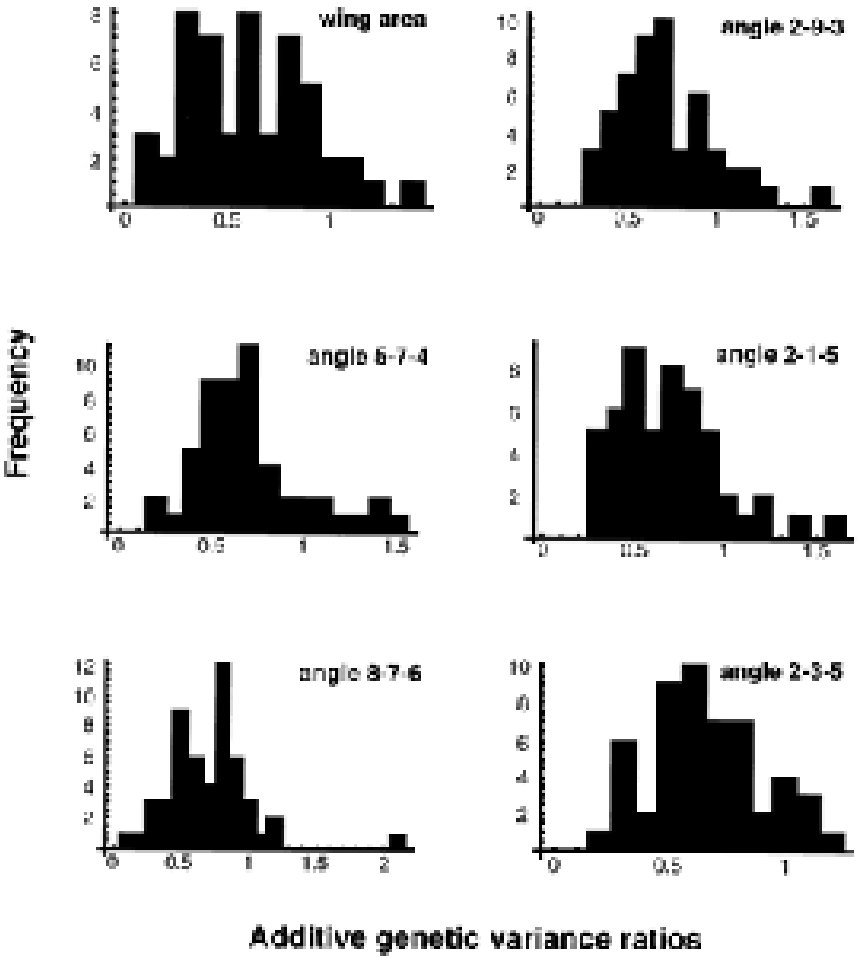
$$h_t^2 = h_0^2 \, \frac{1-F_t}{1-h_0^2 \, F_t}$$

when expressed in terms of heritability in the base population.

Example: Effect of Inbreeding on the Additive Genetic Variance

(M. Whitlock and K. Fowler, 1999, Genetics 152:345-353)

Whitlock and Fowler created 52 inbred lines of *Drosophila melanogaster* by passing each through a bottleneck of one pair ($F \simeq 0.25$). Parent-offspring regressions were used to estimate additive genetic variances for a set of wing dimensions in each line, as well as in the outbred base population from which the lines were derived. Results were expressed as the ratio of V_A in inbred lines to V_A in the base population.



On average, V_A in the inbred lines was 60-71% of original V_A, depending on the trait. The reduction in V_A was thus slightly greater than the theoretically

predicted 25%. The authors attributed this to additional inbreeding that took place after the one-pair bottlenecks.

Notably, V_A for a given trait varied greatly among lines, with a few lines even showing significant increases in V_A. This is not surprising, because allele frequencies in each line will change randomly.

Change of Variance with Inbreeding and Mutation

Inbred lines will never completely lose all genetic variance, because new mutational variance is introduced each generation at rate V_M. In the long term the genetic variance within an inbred line will reach an equilibrium level at which the variance gained each generation from mutation is exactly balanced by the variation lost by inbreeding. We consider the particular case where inbreeding is caused by genetic drift in a finite population, in which case the accumulation of inbreeding scales as $1/(2N_e)$, where N_e is the effective population size.

Assume:

- Strictly neutral mutations
- Strictly additive mutations
- Symmetrical distribution of mutational effects

Then at equilibrium

$$V_A = V_G = 2N_e V_M$$

Note that this is the same as that expected within a selection line in mutation-drift equilibrium. With $V_M = 10^{-3} V_E$ and $N_e = 2$ for full sib inbreeding,

V_G at mutation-drift equilibrium is $4 \times 0^{-3} V_E$, and the heritability is

$$h^2 = \frac{V_A}{V_G + V_E} = \frac{4 \times 0^{-3} V_E}{4 \times 0^{-3} V_E + V_E)} = 0.004$$

which is trivial.

Mutation also contributes to the increase in variance between sublines derived from a common inbred line. The variance among lines from new mutation after t generations is

$$V_B = 2V_M[\, t - 2N_e(1 - e^{-t/2N_e})]$$

At equilibrium the rate of divergence is $2V_M$ per generation and the total divergence expected is $2tV_M$, which is not negligible.