

## Lecture 6

# Inbreeding and Crossbreeding

Bruce Walsh. [jbwalsh@u.arizona.edu](mailto:jbwalsh@u.arizona.edu). University of Arizona.

*Notes from a short course taught June 2006 at University of Aarhus*

*The notes for this lecture were last corrected on 23 June 2006. Please email me any errors.*

### CHANGES IN THE MEAN AND VARIANCE UNDER INBREEDING

#### 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 (Lecture 4). In an individual inbred to amount  $F$ , a randomly-chosen locus has both alleles IBD with probability  $F$  and hence is a homozygote.

To compute the genotypic probabilities under inbreeding, suppose we chose a locus at random. Denote the frequency of allele  $A_1$  by  $p$  and the  $\text{freq}(A_2)$  by  $q$ . With probability  $F$  the two alleles at our target locus are IBD, and hence this locus is always homozygous, with  $\text{freq}(A_1A_1) = p$  and  $\text{freq}(A_2A_2) = q = 1 - p$ . 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

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

while the mean under inbreeding can be expressed as

$$\mu_F = \mu_0 - 2Fpqd \tag{6.1a}$$

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

$$\mu_F = \mu_0 - 2F \sum_{i=1}^k p_i q_i d_i = \mu_0 - B F \quad (6.1b)$$

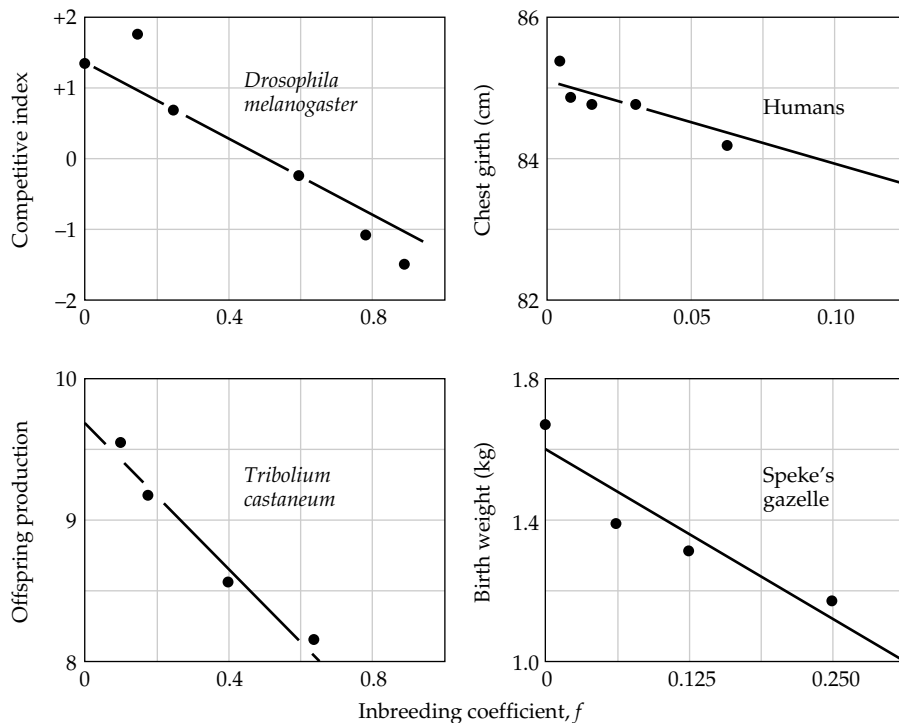
where  $B = 2 \sum p_i q_i d_i$  is the reduction in the mean under complete inbreeding ( $F = 1$ ).

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 trait value of the heterozygote is closest to the trait value of the largest homozygote).
- 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 illustrate:



### Computing the Inbreeding Depression Coefficient, $B$

In many cases, lines cannot be completely inbred 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 using a series of lines under partial inbreeding.

Recalling that

$$\mu_F = \mu_o - BF \quad (6.2)$$

it immediately follows that the slope of the regression of the population mean  $\mu_F$  on the inbreeding coefficient  $F$  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 (polynomial) function of  $F$ . Hence, if epistasis is absent, we expect a linear regression of mean on  $F$  to be an adequate 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 fraction of overdominant loci can 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.

### Minimizing the Rate of Inbreeding

Given the detrimental effects of inbreeding, breeders and zoo keepers go to great lengths to minimize the rate of increase in  $F$ . Inbreeding occurs in any finite population, but the larger the effective population size  $N_e$ , the slower the effects of drift. One standard way for maximizing  $N_e$  is to ensure that all individuals make an equal contribution of offspring to the next generation, as the effective population decreases as the variance in offspring number increases. Maximal  $N_e$  occurs when each male and female in the population leaves exactly the same number of offspring. Alas, in many breeding situations there is a very skewed sex ratio, often due to constraints of reproductive biology (a bull can leave effectively an infinite number of offspring, while a cow typically has only one a year).

Gowe et al. (1959) suggest that when the sex ratio of contributing parents is  $r$  females to each male, that every male should contribute (exactly) one son and  $r$  daughters, while every female should leave one daughter and also with probability  $1/r$  contribute a son. Wang (1997) improved on this by imposing that a female contributing a son does not contribute a daughter, but instead that another female from the same male family contributes two daughters.

### 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 ( $\sigma_A^2 = \sigma_G^2$ ) present in the base population:

	General	$F = 1$	$F = 0$
Between Lines	$2F\sigma_A^2$	$2\sigma_A^2$	0
Within Lines	$(1 - F)\sigma_A^2$	0	$\sigma_A^2$
Total	$(1 + F)\sigma_A^2$	$2\sigma_A^2$	$\sigma_A^2$

Inbreeding increases genetic variance between lines and decreases genetic variance within lines. With dominance, the expressions are not simple functions of the base population genetic variances, but rather depend on gene frequencies, so there is no simple expression for the re-distribution of variance. Indeed, when non-additive variance is present, the additive variance can actually increase (over some range) with  $F$ .

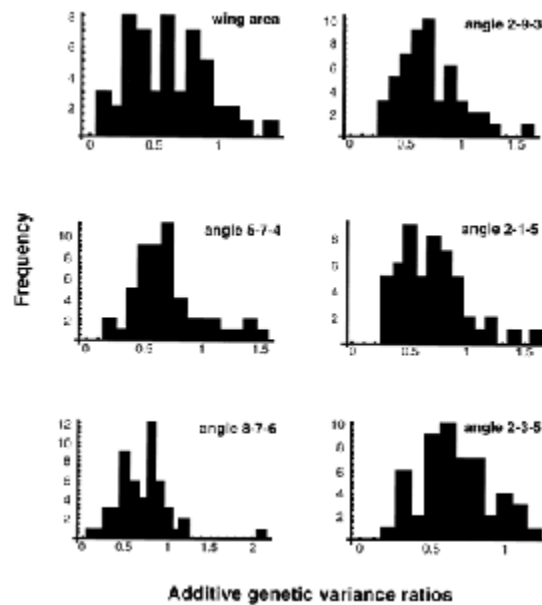
The heritability within any one inbred line (assuming only additive variance) is

$$h_t^2 = \frac{(1 - F_t)\sigma_A^2}{(1 - F_t)\sigma_A^2 + \sigma_E^2} = h_0^2 \frac{1 - F_t}{1 - h_0^2 F_t}$$

The right-most expression is expressed in terms of heritability  $h_0^2$  in the base population.

### Example: Effect of Inbreeding on the Additive Genetic Variance

Whitlock and Fowler (1999, *Genetics* 152:345-353) 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  $\sigma_A^2$  in inbred lines to  $\sigma_A^2$  in the base population.



On average,  $\sigma_A^2$  in the inbred lines was 60-71% of original  $\sigma_A^2$  (value of 1 in the plots), depending on the trait. The reduction in  $\sigma_A^2$  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,  $\sigma_A^2$  for a given trait varied greatly among lines, with a few lines even showing significant increases in  $\sigma_A^2$ . 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  $\sigma_m^2$ . 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

$$\sigma_A^2 = \sigma_G^2 = 2N_e\sigma_m^2$$

With  $\sigma_m^2 = 10^{-3}\sigma_E^2$  and  $N_e = 2$  for full sib inbreeding,  $\sigma_G^2$  at mutation-drift equilibrium is  $4 \times 10^{-3}\sigma_E^2$ , and the heritability is

$$h^2 = \frac{\sigma_A^2}{\sigma_G^2 + \sigma_E^2} = \frac{4 \times 10^{-3}\sigma_E^2}{4 \times 10^{-3}\sigma_E^2 + \sigma_E^2} = 0.004$$

which is trivial.

Mutation also contributes to the increase in variance between sublines derived from a common inbred line. The variance among lines (i.e., the variance of the line means) from new mutation after  $t$  generations is

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

At equilibrium the rate of divergence is  $2\sigma_m^2$  per generation and the total divergence expected is  $2t\sigma_m^2$ , which is not negligible.

### HYBRIDIZATION AND CROSSBREEDING

**Hybridization**, the crossing of two (or more) lines or populations is a common feature of applied breeding and is also of considerable interest to evolutionary biologists. There are two distinct reasons to consider crosses between populations. First, breeders cross populations in an attempt to combine the best features of each line/population/species. This can often fail, in quite spectacular fashion. Consider the rabbage, the  $F_1$  between a radish and a cabbage. The idea was to combine a radish root with a cabbage head. Unfortunately, the result was a radish head on a cabbage root. In this case, it would appear that the favorable genes for each species were either recessive and/or the result of strong epistatic interactions.

The second reason to consider cross breeding is **heterosis** (or **hybrid vigor**), wherein (unlikely the poor rabbage), the  $F_1$  is superior in some trait to either parent. Heterosis can occur between a cross of two lines that otherwise seem nearly identical. In such cases, the purpose of creating hybrids is not to combine advantageous traits, but rather to uncover favorable genes that are masked by either dominance and/or epistasis. Heterosis present in an  $F_1$  is usually reduced (often considerably) in an  $F_2$ , mirroring the decrease in heterozygosity from the  $F_1$  to the  $F_2$ . Given these facts, there has been much effort (especially by plant breeders) to identify and create inbred lines that cross well (such lines are said to show **nicking**) to produce favorable  $F_1$  hybrids. Lines are often classified into **heterotic groups**, wherein individuals from different groups when crossed show significant heterosis, while crosses between individuals within a group show little or no heterosis.

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  $\mu_F = \mu_0 - BF$ . If all these lines are crossed at random,  $F = 0$ , and the mean of the crossbreds =  $\mu_0$ , the mean of the outbred population.

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

### Types of Crosses

The number of different possible crosses one can make is limited only by the energy and imagination of the worker and the reproductive biology of the organism under consideration. The most common is a **single cross** (often denoted **SC**), the  $F_1$  between two lines. With a collection of lines in hand, how does the breeder decide which to cross to produce the best-performing SC? One approach is simply to make all  $n(n - 1)/2$  single crosses among all  $n$  lines, a **diallel design**. The problem is that the full diallel design involves a considerable amount of crosses—45 possible SCs involving a collection of 10 lines, 190 SCs for 20 lines, 1225 SCs for 100 lines. If one could choose a subset of all lines, then a full diallel is certainly reasonable, but what criteria should be used?

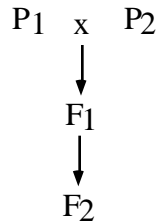
A strategy for selecting parental lines is suggested from the analysis of a full diallel, where one can estimate the **general combining ability** (or **GCA**) of each line and the **specific combining ability** (**SCA**) for each cross (Sprague and Tatum, 1942). The GCA of a line is the difference of its average performance in hybrids relative to the grand mean of all line crosses (LW Chapter 20). The expected mean of a cross is the sum of the maternal and paternal general combining abilities,  $GCA(P) + GCA(M)$ . Since the GCA from a random individual (being a deviation from the grand mean) is zero, if only one parent is known, the expected offspring mean is just the GCA from that parent. The deviation between the predicted and actual values estimates the specific combining ability for that particular cross. The GCA is in effect the breeding value of a line, and (as with breeding values) it is a function of population of lines chosen. Hence, the mean value of the single cross  $l_i \times l_j = GCA_i + GCA_j + SCA_{ij}$ . Thus, if one knew something about the GCA's of the lines in our initial collection, a quite reasonable starting point is to perform a diallel using those with the highest GCAs. While it is certainly true that the best line may involve a cross between two lines with very low GCAs, but a very large SCAs, using the GCAs nonetheless offers a good compromise starting point.

The question is how to obtain the GCAs for a series of lines without making all the controlled crosses for a full diallel. Two of the most common approaches are the **topcross design**, in which a particular line (the **testor**) serves as the common male (pollen) parent and the GCAs of female plants are estimated by the performance of their seed. The other approach is a **polycross design** wherein females are allowed to be randomly fertilized with pollen from all the other lines. Again, the elite lines are chosen by those with the best-performing offspring. Under either design, care must be taken to avoid selfing. While manual pollen control can be used if the flowers are sufficiently large and the number of flowers per plant and the number of crosses is small, a much more efficient (and elegant) approach is to use female plants that are genetically male-sterile.

As we will shortly see, plant and animal breeders often work with more complex hybrids than single crosses. **Triple** or **three-way crosses (3W)** crosses involve an  $F_1$  crossed to a third line, e.g.,  $A \times (B \times C)$ , which is often written as  $A \cdot BC$ . The resulting individuals have 50% of their genes from line  $A$ , and 25% from each of lines  $B$  and  $C$ . **Four-way (4W)** or **double crosses (DC)** are the crosses of two different  $F_1$ 's,  $AB \cdot CD = (A \times B) \times (C \times D)$  so that the hybrid is the result of equal contributions from four different lines. [ While the term double cross is very common in the literature, we will interchangeably use it with four-way cross, as it is easy for the casual reader to assume that a double-cross involves fewer lines than a triple cross, while this distinction is clear using three- vs. four-way crosses. ] **Modified cross** involve two closely related populations (say  $A$  and  $A^*$ ). For example, a **modified single cross** is  $(A \times A^*) \times B$ , while a **modified triple cross** is  $(A \times A^*) \times (B \times C)$ . **Multiple cross** are also used, such as the cross using two (different) four-way cross individuals as parents, which involves eight lines.

### Heterosis: Change in the Mean Under Crossbreeding

Consider the cross between two particular parental strains ( $P_1$  and  $P_2$ ). 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 from the midparental value, so that for the  $F_1$ ,

$$H_{F_1} = \mu_{F_1} - \frac{\mu_{P_1} + \mu_{P_2}}{2} \quad (6.3)$$

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

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

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 the contribution to heterosis from this particular locus as

$$H_{F_1} = \mu_{F_1} - \frac{\mu_{P_1} + \mu_{P_2}}{2} = (\delta p)^2d \quad (6.4a)$$

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

$$H_{F_1} = \sum_{i=1}^n (\delta p_i)^2 d_i \quad (6.4b)$$

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 absent in the other (so that  $|\delta_i| = 1$ ). Note that this occurs with crosses between completely inbred lines. Likewise,  $H = 0$  if  $\delta = 0$ .
- $H$  is specific to each particular cross.  $H$  must be determined empirically, since we do not know the relevant loci nor their allele 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  $\text{freq}(A_1) = (p + p + \delta)/2 = p + \delta/2$ . Since the  $F_2$  is formed by random mating, the genotype frequencies are in HW equilibrium with allele frequency  $p + \delta/2$ , and the  $F_2$  mean becomes

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

A little algebra shows that

$$H_{F_2} = \mu_{F_2} - \frac{\mu_{P_1} + \mu_{P_2}}{2} = \frac{(\delta p)^2 d}{2} = \frac{H_{F_1}}{2} \quad (6.5)$$

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 frequencies.

## Agricultural Importance of Heterosis

Heterosis is extremely important in world agriculture. 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.

**Table 6.1.** Estimates of the world-wide contribution of heterosis to both yield and land savings. 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	$55 \times 10^6$	$13 \times 10^6$ ha
Sorghum	48	40	19	$13 \times 10^6$	$9 \times 10^6$ ha
Sunflower	60	50	30	$7 \times 10^6$	$6 \times 10^6$ ha
Rice	12	30	4	$15 \times 10^6$	$6 \times 10^6$ ha

## Hybrid Corn

Corn (maize) is the third most important crop in the world (behind wheat and rice), and the most important crop in the United States as measured by acres planted and total yield. Almost all of the US crop is **hybrid corn**, the seed from crosses of elite inbred lines (there are small amounts of **open pollinated (OP)** corn, but these are largely restricted to use by small independent farmers). The notion of hybrid corn traces back to Shull (1908, 1909), who noted that, because of the significant high parent heterosis seen in many corn crosses, the objective of corn breeders should be to find and maintain the best parental lines for hybrids. By maintaining these elite inbred lines and crossing them as needed, genetically identical hybrid seed can be produced year after year.

The problem with implementing the widespread usage of hybrid seed was that the early inbred lines used by breeders typically had very low seed set. In a (single) cross between two inbred lines, the seed parent is inbred, and although the resulting hybrid seed produce plants with superior

yield, each seed parent produces only a small number of seeds. Thus, poor fertility of the initial inbred lines resulted in rather few hybrid seed per plant. Jones (1918, 1922) suggested that instead of using an inbred line as the seed parent, one instead uses a hybrid line, with the hybrid seed being produced by a double cross instead of a single cross. Since the seed parent is a hybrid (a single cross between two inbred lines), it should show heterosis in seed production, resulting superior seed set. This suggestion directly opened up the vast commercial potential of hybrid corn. For example, in 1925 Henry A. Wallace (a farmer who later became Vice-President of the United States) formed the Pioneer Hi-Bred Corn Company of Iowa. From the 1930's to the 1960's, most US hybrid corn was produced by double crosses. However, it was soon noted that the seed from single crosses usually out perform double and three-way crosses. As breeders were able to generate inbred lines with higher yield, hybrid corn based on single crosses became both commercially feasible and also desirable, given the increased yields relative to double crosses. Since the 1970's, most hybrid corn in the US is the result of single crosses.

Hybrid corn also offers one of the best cautionary tails for plant breeders – the great **Southern Corn leaf Blight** (SCLB) epidemic from 1970-1971 (Tatum 1971, Ullstrup 1972). This was one of the most damaging plant epidemics in history — in terms of food energy destroyed it was much larger than the potato blight epidemic in the 1840's that produced widespread famine (Ullstrup 1972). The roots of the epidemic in the United States trace back to what (at the time) appeared to be an elegant genetic solution to a major labor problem in the production of hybrid corn. Corn can self-fertilize, an undesirable feature as it reduce the fraction of hybrid seeds on a plant. To prevent this, corn breeders manually detassled plants, removing the pollen-shed organ and allowing pollen only to be donated from the parents of interest. Typically one row of pollen plants is planted for every four to six (detassled) rows of seed parents, so that a pollen parent is within two or three rows of any seed parent (any seed set on the pollen plants is ignored). The discovery of a cytoplasmic factor (a mitochondrial-encoded gene) that produced sterility in males seemed to be an elegant solution around the hassle of detassling (Jones and Everett 1949, Rogers and Edwardson, 1952). Plants with this T (for Texas, the origin of the line) cytoplasm are denoted by **Tcms**. Seed parents with Tcms cytoplasm produce normal and viable seeds, but no pollen, and hence to not have to be detassled. As a result of the very convenient feature, by 1970, almost 85% of hybrid corn seed in the US contained Tcms. In 1969, a previous unknown strain of the fungus *Helminthosporium maydis* was detected in a few areas in the midwest corn growing region of the central US. Plants contained Tcms were hyper-susceptible to this strain, and the result was over a billion dollar loss (in 1970's dollars) and major angst in both the private and public agricultural sectors. Corn is typically a rather disease-free crop (indeed, Ullstrup 1972 notes that in 1970, the US Department of Agriculture considered corn so healthy that it did not employ any full-time corn pathologists in the US corn belt), so this outbreak was even more of a shock and lead to a considerable focus on increasing the genetic diversity in crops.

Finally, it should be remarked that heterosis in maize is not just restricted to crosses between inbred lines. Richley (1922) reported the results from 244 crosses of open pollinated varieties, finding that 82.4% had yields above the midparental average, and 55.7% of the crosses exceeded the best parent.

### Hybridization in Other Crops

A major reason that hybrid corn has been commercially successful is because of its reproductive structures, with pollen being produced on a long tassel, allow for relatively easy manual control of pollination. For many species with perfect flowers (i.e., both male and female parts), manual control of pollination is just not economically feasible because flowers are too numerous, too small, or both. Male-sterile genes are critical to the creation of hybrids in such species, as is chemical sterilization of pollen. In contrast with field crops, horticultural crops have much more favorable economic conditions for the commercial exploitation of heterosis. Here, flowers are typically large

and easily handled (unlike the grass-family flower heads of many important field crops). Equally important, the yield from each individual plant has significant commercial value (Dodds 1955), so that individual manipulation of single plants often is economically feasible. Some examples of the successful use of hybridization in horticultural crops include tomatoes, egg plant, cucurbits, and onions.

### Crossbreeding in Animals: General Concepts

As the mule illustrates, the importance of between-species hybrids in animal breeding goes back to prehistoric times. The aggressive utilization of crossbreeding to exploit heterosis (as opposed to crosses simply to combine desirable features from two different lines) followed Sewall Wright's (1922) extremely influential publication on crossbreeding (and inbreeding) in guinea pigs.

Animal breeders often distinguish between **individual** and **maternal heterosis**. Individual heterosis is *enhanced performance in a hybrid individual*, while maternal heterosis is *enhanced maternal performance* (such as increased litter size and higher survival rates of offspring). Maternal heterosis is often comparable, and can be greater than, individual heterosis (e.g., Table 6.2). For example, fertility in sheep is improved more by using crossbred ewes (resulting in a 9% maternal heterotic effect) than by using crossbred lambs from purebred ewes (only a 2-3% individual heterotic effect). Maternal effect heterosis is roughly twice as individual heterosis for several traits in European cattle (*Bos taurus*).

**Table 6.2.** Estimate of individual  $h^I$  and maternal heterosis  $h^M$  in sheep (estimation of  $h^I$  and  $h^M$  is discussed later in the lecture notes). Results presented as percentage of parental means.  $n_I$  and  $n_M$  indicate the number of estimates used for the reported individual and maternal values. **Prolificacy** is the litter size at birth.

Trait	$n_I$	Mean $h^I$ (in %)	$n_M$	Mean $h^M$ (in %)
Birth weight	42	3.2	12	5.1
Weaning weight	56	5.0	27	6.3
Preweaning growth rate	19	5.3		
Postweaning growth rate	10	6.6		
Yearling weight	18	5.2		
Ovulation rate			4	-2.0
Fertility	20	2.6	30	8.7
Prolificacy	20	2.8	31	3.2
Birth-weaning survival	29	9.8	25	2.7
Lambs per ewe	20	5.3	25	11.5
Lambs reared per ewe	20	15.2	25	14.7
Total weight lambs/ewe	24	17.8	25	18.0
Carcass traits	7	$\simeq 0$		

Maternal and individual heterosis effects can be combined by using crossbred dams. For example, for total weight of lambs rear per mated ewe has an 18% individual heterotic advantage in a crossbred offspring and an addition 18% advantage (from maternal heterosis) when crossbred ewes are used in place of purebred ewes. This combining of maternal and individual heterotic effects is one reason why three-way crosses are common in animal breeding, generally by crossing a male from line *A* with a hybrid female (from a  $B \times C$  cross). This strategy exploits maternal heterosis in the female, with the sire line often chosen for its contribution to some production trait.

In theory, one could also consider **paternal heterosis**, increased performance due to paternal effects in hybrid sires, but there has been little evidence to date of this being a significant effect.

For example, Bradford et al. (1963) compared purebred versus crossbred sire performance, finding no major differences between them except for a slight (4%) elevation of lamb survival, where the hybrid exceeded both parents.

### **Crossbreeding in Animals: Heterosis in *Bos indicus* × *Bos taurus* Hybrids**

Some interesting results on crossbreeding are found in the literature examining crosses between European (*Bos taurus*) and tropical (*Bos indicus*) cattle (reviewed in McDowell 1985). *Bos taurus* breeds have been selected for significant genetic improvement in a number of traits (such as milk and meat yield), but are adapted for only temperate climates. *Bos indicus* breeds are found in tropical countries, and while they show fairly little genetic improvement, they are well adapted to the local environment. Crosses between *taurus* and *indicus* have been performed in the hope of generating higher performing hybrids that are also adapted to tropical environments.

*Bos indicus* × *B. taurus* hybrids are usually superior to indigenous breeds (i.e., the local breed of *B. indicus*) in milk yield and fitness measures such as calving age and interval. Trail et al. (1985) examined crosses between exotic *B. taurus* breeds (Angus and Red Poll) and indigenous *B. indicus* breeds (Ankole, Boran and Zebu) in Africa. Crossbred (exotic X indigenous) dams showed superior maternal performance over straightbred indigenous dams. However, while maternal effects were apparently superior, the progeny of straightbred (Boran) dams were actually heavier at 24 months than the progeny of exotic × crossbred dams, suggesting that the 3/4 exotic composition is not as favorable for individual performance as 1/2 exotic, 1/2 indigenous. Interesting, three-breed crosses between an improved breed sire and a cross-bred dam (a second improved breed × an indigenous breed) generally tend to do poorer than two-breed crosses. This is contrary to the general superiority of three breed crosses of *B. taurus* in temperate areas (McDowell 1985).

*Bos indicus* × *B. taurus* hybrids show higher levels of heterosis than observed in crosses among breeds within each species. This is perhaps not unexpected, as allele frequencies likely have diverged more between species than between the breeds within a species. Everything else being equal (which, of course, is never assured), crosses involving widely allele frequencies differences would show increased heterosis. Interesting, heterosis in crosses of *B. indicus* breeds is higher than is observed in crosses of *B. taurus* breeds. For example, Gregory et al. (1985) found that material heterosis in crosses within *B. indicus* lines was intermediate between levels produced in *B. indicus* × *B. taurus* hybrids and crosses among *B. taurus* breeds. One explanation is that there is more gene frequency divergence among the various indigenous *B. indicus* breeds than among the *B. taurus* breeds.

While *indicus* × *taurus* crosses do indeed show significant heterosis, they can also have unanticipated economic disadvantages. McDowell (1985) points out, regarding the hybrids, that "some farmers have reservations that could influence national breeding programs. The two-breed crossbred male is not as temperamental as the favored draft breed, e.g., Haryana in India, and does not move as rapidly for plowing or in performing cartage. The smaller hump of the crossbred is not well suited to handle the traditional wooden yoke. For those reasons the crossed male has a price discrimination against it in the draft market." Further, crossbreeds typically require supplemental feeding, otherwise they can become nutritionally stressed (McDonnell 1981, 1985).

### **Synthetics and Rotational Crossbreeding**

To maximally exploit heterosis, we ideally would use only F<sub>1</sub> individuals, as the heterotic advantage decreases in the F<sub>2</sub>, e.g., Equation 6.4b. In much of plant breeding, **terminal crosses** are used, wherein the final hybrid individuals are the endpoints and do not reproduce further. For example, a company sells a farmer F<sub>1</sub> seed, and this seed is generated anew each generation, with the F<sub>1</sub> plants themselves not allowed to reproduce. While such schemes can work in plants with their enormous reproductive potential, they are more difficult in animals. Consider a three-way cross of an *A* sire to a *B* × *C* dam (with *C* being the grandam). While artificial fertilization only requires

modest numbers of sires from the  $A$  and  $B$  lines/populations, in large farm animals, the number of offspring is on order of the number of dams. Thus, to produce  $n$  triple cross hybrid offspring requires on the order of  $2n$  dams (the grandam to produce the  $B \times C$  dam, and this dam herself). One potentially very important application of whole-animal cloning would be the creation of cloned lines from exception dams showing outstanding maternal heterosis.

Two approaches have been suggested as potential compromises between the cost of continually providing  $F_1$ 's and the decrease in performance of  $F_2$ . The first is the use of **synthetics**, where  $n$  parental lines are chosen and a random-mating population formed by first making all  $n(n-1)/2$  pairwise intercrosses between the lines. Sewall Wright gave an approximate expression for the decrease in heterosis in a synthetic as

$$F_2 = F_1 - \frac{F_1 - \bar{P}}{n} \quad (6.6a)$$

where  $\bar{P}$  is the average mean of all parents and the  $F_1$  value is the average of all pairwise crosses between lines. Since  $F_1 - \bar{P} = H$  is the heterosis, this can be expressed as

$$H_{F_2} = F_2 - \bar{P} = \left( F_1 - \frac{H}{n} \right) - \bar{P} = (F_1 - \bar{P}) - \frac{H}{n} = H \left( 1 - \frac{1}{n} \right) \quad (6.6b)$$

The idea behind synthetics is that, the larger  $n$ , the smaller the loss of  $F_1$  heterosis.

A second approach is **rotational crossbreeding**, which continually recycles hybrid individuals. Here, hybrids from the previous generations are crossed (in rotation) to pure lines. For example, a three-bred rotational would use  $A \times B$  as the first generation. In generation two, dams from the first generation are crossed to line  $C$ . In generation three, dams from generation two are crossed to sires from line  $A$ , and the rotation continues over all three lines in subsequent generations. This approach represents a compromise between trying to maintain maximal heterozygosity within a line without having to regenerate the line anew each generation.

Carmon et al. (1956) examined the average (asymptotic) performance of rotational crossbreeding under the assumption of no epistasis. For a two-line ( $A, B$ ) rotation, the predicted mean is

$$\hat{R}_2 = \bar{z}_{AB} - \frac{\bar{z}_{AB} - \bar{P}_2}{3}, \quad \text{where} \quad \bar{P}_2 = \frac{\bar{z}_A + \bar{z}_B}{2} \quad (6.7a)$$

Here  $\bar{P}$  is the average of the two parental lines and  $\bar{z}_{AB}$  the mean value of their cross. Note that the mean performance under rotational crossbreeding is less than the single cross performance. For a three-line ( $ABC$ ) rotational cross, the predicted mean is

$$\hat{R}_3 = \overline{SC}_3 - \frac{\bar{z}_{AB} - \bar{P}_3}{7}, \quad \text{where} \quad \overline{SC}_3 = \frac{\bar{z}_{AB} + \bar{z}_{AC} + \bar{z}_{BC}}{3} \quad (6.7b)$$

where  $\overline{SC}_3$  is the average of the three single crosses between these three lines and  $\bar{P}_3$  the average of the three parental lines.

For a four-line rotation, the order of the rotation matters. Letting the rotation be  $A, B, C, D$ , the predicted long-term performances

$$\hat{R}_4^{(A,B,C,D)} = \overline{SC}_4 - \frac{\overline{SC}_{na} - \bar{P}_4}{15}, \quad \text{where} \quad \overline{SC}_{na} = \frac{\bar{z}_{AC} + \bar{z}_{BD}}{2} \quad (6.7c)$$

As above,  $\bar{P}_4$  is the mean of the original lines,  $\overline{SC}_4$  is the mean of all six possible single-crosses between the four lines, and  $\overline{SC}_{na}$  the average of the two single crosses of non-adjacent lines in the rotations.

As an application of the above expressions, consider the following data for various crosses of Devon and Brahman cattle (from Kidder et al. 1964): The midparent  $\bar{P}$ ,  $F_1$ , two-breed rotational crossbred R, synthetic S, and the backcross (BC) have the following means

Trait	Means				
	$\bar{P}$	$F_1$	R	S	BC
Weaning weight	154.2	180.5	178.3	170.1	181.4
12-month weight	210.5	246.8	232.2	212.3	233.6
18-month weight	274.9	315.7	296.6	276.6	295.3
12-18 month weight gain	64.4	68.9	64.4	64.4	61.7

Note from these data that the rank of performance is  $F_1 > R > S > \bar{P}$ . How well does Equation 6.7 predict the rotational crossbred performance? Here the  $F_1$  corresponds to  $\bar{z}_{AB}$ , so that the predicted equilibrium value is

$$\hat{R}_2 = F_1 - \frac{F_1 - \bar{P}_2}{3}$$

For example, for weaning weight

$$\hat{R}_2 = 180.5 - \frac{180.5 - 154.2}{3} = 171.7$$

which is 96% of the observed value. Similarly, the predicted values (and fraction of the actual values) for 12, 18, and gain are, respectively, 234.7 (101%), 302.1 (102%), and 67.5 (104%). Hence, for these data, there is a slight tendency to overestimate the true mean.

Turning to Equation 6.6 to predict the synthetic performance,  $n = 2$  and hence for weaning weight

$$\hat{S}_2 = 180.5 - \frac{180.5 - 154.2}{2} = 167.4$$

which is 98% of the observed value. Similarly, the predicted values (and fraction of the actual values) for 12, 18, and gain are, respectively, 228.65 (107%), 295.3 (107%), and 66.7 (104%).

### Estimating the Amount of Heterosis in Maternal Effects

The basic approach for incorporating maternal effects was given in a specific case (three-line crosses) by Magee and Hazel (1959), while a more general treatment was presented by Dickerson (1969). Maternal heterotic effects are judged to be of sufficient importance that significant effort is usually made to use crossbred dams, while sires are very often purebreds. Crossbred dams are very often chosen on the basis of favorable reproductive traits, while sires are often picked for other traits (such as size or carcass traits).

The model used to estimate maternal effects and heterosis is as follows: We can consider the mean value of a line as consisting of an average direct (or individual) effect  $g^I$ , a maternal genetic effect  $g^M$  expressed through the mother, and even a potential grand-maternal effect expressed through the dam  $g^{M'}$ . For example, the mean value for line A is

$$\mu_A = \mu + g_A^I + g_A^M + g_A^{M'} \quad (6.8a)$$

as the dam in this cross is from line A and the mother of this dam (the granddam) is also from line A. Crossbred offspring potentially experience an additional heterotic effect  $h$ . Hence, the expected mean in a cross with line A as a sire and B as a dam is

$$\mu_{AB} = \mu + \frac{g_A^I + g_B^I}{2} + g_B^M + g_B^{M'} + h_{AB}^I \quad (6.8b)$$

as the individual genetic value of an  $AB$  individual is the average of the two lines plus any additional heterotic direct effect  $h_{AB}^I$ . In this cross, both the dam and granddam are from line  $B$ . Conversely, the expected mean of the reciprocal cross (now with  $A$  as the dam) is

$$\mu_{BA} = \mu + \frac{g_A^I + g_B^I}{2} + g_A^M + g_A^{M'} + h_{AB}^I \quad (6.8c)$$

From Equation 6.8a-c, it follows that an estimate of the individual (direct) heterotic effect is the obvious one (Nitter 1978),

$$\frac{\mu_{AB} + \mu_{BA}}{2} - \frac{\mu_{AA} + \mu_{BB}}{2} = h_{AB}^I \quad (6.8d)$$

as the maternal effects cancel. Likewise, the difference in reciprocal crosses

$$\mu_{BA} - \mu_{AB} = (g_A^M + g_A^{M'}) - (g_B^M + g_B^{M'}) \quad (6.8e)$$

provides an estimate of the difference in maternal + grandmaternal genetic effects for the two lines.

If the dam is crossbred, then she has the potential of heterotic maternal effects  $h^M$ . Likewise, if the granddam is crossbred, there are also potential grandmaternal heterotic effects  $h^{M'}$ . The final complication is that if  $h_{AB}$  is the heterotic (individual or maternal) contribution in the  $F_1$ , heterosis can decrease in the  $F_2$  due to recombination breaking up favorable gene combinations. This is incorporated into the general model via a **recombinational loss** term,  $r_{AB}$ . As an example of putting all these pieces together, consider the three-way cross using a crossbred dam,

$$\mu_{C \cdot AB} = \frac{2g_C^I + g_A^I + g_B^I}{4} + \frac{h_{CA}^I + h_{CB}^I}{2} + \frac{g_A^M + g_B^M}{2} + h_{AB}^M + g_B^{M'} + \frac{r_{ab}^I}{2} \quad (6.9a)$$

Here since the dam is a crossbred, there is a potential heterotic component ( $h_{AB}^M$ ), while the granddam of the  $AB$  dam is  $B$ . The final piece is that the individual heterotic component in the  $F_1$  (the  $AB$ ) cross may be potential degraded in the  $F_2$ , which is accounted for by the  $r_{ab}^I$  term. Table 6.3 summarizes these coefficients for a variety of crosses. For example, estimates of the various effects can be obtained by suitably weighted combination of line means. Combining 6.8b and 6.9a gives (Nitter 1978),

$$\mu_{C \cdot AB} - \frac{\mu_{CA} + \mu_{CB}}{2} = h_{AB}^M + \frac{r_{ab}^I}{2} \quad (6.9b)$$

Hence, if recombination effects are small (these would be absent in the absence of epistasis), Equation 6.9b allows for a direct estimate of maternal heterotic effects.

Table 6.3 summarizes the weights on the various heterotic components we are trying to estimate for a number of crosses.

**Table 6.3.** Coefficients for direct (or individual) genetic effect  $g^I$ , genetic maternal effects  $g^M$  and genetic grandmaternal effects  $g^{M'}$ , their heterotic counterparts ( $h$ ) and recombination corrections  $r$  for a variety of crosses. After Dickerson (1969).

Cross	Order	$g^I$	$h^I$	$r^I$	$g^M$	$h^M$	$r^M$	$g^{M'}$	$h^{M'}$
P <sub>1</sub>	A	A			A			A	
F <sub>1</sub>	A·B	$\frac{A+B}{2}$	AB		B			B	
	B·A	$\frac{A+B}{2}$	AB		A			A	
3W	(C·D)·A	$\frac{C+D+2A}{4}$	$\frac{CA+DA}{2}$	$\frac{cd}{2}$	A			A	
	C·(A·B)	$\frac{2C+A+B}{4}$	$\frac{CA+CB}{2}$	$\frac{ab}{2}$	$\frac{A+B}{2}$	AB		B	
	C·(B·A)	$\frac{2C+A+B}{4}$	$\frac{CA+CB}{2}$	$\frac{ab}{2}$	$\frac{A+B}{2}$	BA		A	
B <sub>1</sub>	A·(A·B)	$\frac{3A+B}{4}$	$\frac{AB}{2}$	$\frac{ab}{2}$	$\frac{A+B}{2}$	AB		B	
B <sub>1</sub>	A·(B·A)	$\frac{3A+B}{4}$	$\frac{AB}{2}$	$\frac{ab}{2}$	$\frac{A+B}{2}$	AB		A	
F <sub>2</sub>	(A·B) <sup>2</sup>	$\frac{A+B}{2}$	$\frac{AB}{2}$	ab	$\frac{A+B}{2}$	AB		B	
F <sub>3</sub>	(A·B) <sup>3</sup>	$\frac{A+B}{2}$	$\frac{AB}{2}$	ab+	$\frac{A+B}{2}$	$\frac{AB}{2}$	ab	$\frac{A+B}{2}$	AB

## Lecture 6 Problems

1. Suppose that four genes, each with two alleles affecting body weight have been identified and are segregating in a random-bred strain of mice. The mean weights for the three genotypes (in grams) for each of the four genes are as follows:

Gene A	Gene B	Gene C	Gene D
$A_1A_1$ 22	$B_1B_1$ 22	$C_1C_1$ 22	$D_1D_1$ 22
$A_1A_2$ 21	$B_1B_2$ 22	$C_1C_2$ 20	$D_1D_2$ 19
$A_2A_2$ 20	$B_2B_2$ 20	$C_2C_2$ 20	$D_2D_2$ 19
freq( $A_2$ ) 0.5	freq( $B_2$ ) 0.5	freq( $C_2$ ) 0.1	freq( $D_2$ ) 0.2

Calculate for each gene the inbreeding depression caused by it in the progeny of one full-sib mating, when the allele frequencies in the random-bred population are as shown. Recall that  $F = 0.25$  after one generation of full sib mating. What is the total inbreeding depression? What is the value of  $B$ ?

2. Suppose highly inbred lines of mice are crossed in pairs, and the adult body weights of the two lines crossed and of the  $F_1$  are as shown below. Calculate: (1) the amount of heterosis shown by each cross and (2) the expected body weights in the  $F_2$  generation if random breeding within each crosses was continued. Assume that there are no maternal effects on adult body weight.

	Line A	Line B	$F_1$
(i)	18	16	20
(ii)	22	16	21
(i)	18	18	24

3. Suppose you have four lines (A through D) and make all pairwise crosses between them. The resulting parental and  $F_1$  means are as follows:

Line	A	B	C	D
A	10	20	14	16
B		14	15	17
C			12	22
D				15

The above table is read as follows: Diagonal elements are the line means, i.e., 10 for A, 12 for C, while the off-diagonal elements are the means of the crosses, e.g., a  $A \times C$  cross has mean 14, etc.

- a: What is the expected mean of a four-line synthetic?
- b: What is the expected mean of a 4-way rotation with order A:B:C:D?
- c: What is the expected mean of a 4-way rotation with order A:C:B:D?
- d: What is the expected mean of a 3-way rotation with A, B, and C?
- e: What is the expected mean of a rotation with C and D?

## Solutions to Lecture 6 Problems

1. The contribution to inbreeding depression from each loci is  $2Fp_iq_id_i$

	Gene A	Gene B	Gene C	Gene D
$d_i$	0	1	-1	-1.5
$2p_iq_id_i$	0	0.5	-0.18	-0.48

$$B = \sum 2p_iq_id_i = -0.16, \quad F \sum 2p_iq_id_i = -0.04$$

Hence, the inbreeding depression is 0.04.

2. Recall that Heterosis  $H = F_1 - (\mu_A + \mu_B)/2$ , while only half of this is retained in the  $F_2$ , with the mean of the  $F_2$  being  $H/2 + (\mu_A + \mu_B)/2$

	Line A	Line B	$F_1$	$H$	$F_2$ mean
(i)	18	16	20	3	18.5
(ii)	22	16	21	2	20
(i)	18	18	24	6	21

3. a: Here  $n = 4$ ,

$$\bar{P} = (10 + 14 + 12 + 15)/4 = 12.75$$

$$F_1 = (20 + 14 + 16 + 15 + 17 + 22)/6 = 17.33$$

giving

$$F_2 = F_1 - \frac{F_1 - \bar{P}}{n} = 17.3 - \frac{17.3 - 12.75}{4} = 16.19$$

- b: Recall Equation 6.7c,

$$\hat{R}_4^{(A,B,C,D)} = \overline{SC}_4 - \frac{\overline{SC}_{na} - \bar{P}_4}{15}, \quad \text{where } \overline{SC}_{na} = \frac{\bar{z}_{AC} + \bar{z}_{BD}}{2}$$

Here,  $\bar{P}_4 = 12.75$ ,  $\overline{SC}_4 = 17.33$  and

$$\overline{SC}_{na} = \frac{14 + 17}{2} = 15.5$$

giving  $\hat{R}_4^{(A,B,C,D)} = 17.15$ .

- c: Here

$$\overline{SC}_{na} = \frac{\bar{z}_{AB} + \bar{z}_{CD}}{2} = 21$$

giving  $\hat{R}_4^{(A,C,B,D)} = 16.78$

- d: Recalling Equation 6.7b,

$$\hat{R}_3 = \overline{SC}_3 - \frac{\bar{z}_{AB} - \bar{P}_3}{7}, \quad \text{where } \overline{SC}_3 = \frac{\bar{z}_{AB} + \bar{z}_{AC} + \bar{z}_{BC}}{3}$$

Here  $\overline{SC}_3 = 19.33$ ,  $\bar{P}_3 = 13$ , giving  $\hat{R}_3 = 18.33$ .

- e: Recall Equation 6.7a,

$$\hat{R}_2 = \bar{z}_{AB} - \frac{\bar{z}_{AB} - \bar{P}_2}{3} = 22 - \frac{22 - 13.5}{3} = 19.17$$

Note that this is higher than the predicted means under the above four proposed crosses.