

# 12

## Selection and Crossbreeding

### I: Basic Concepts

*This is a special draft version for the Nordic Summer Course in Animal Breeding, in 2003. A good deal of material specific to plants has been deleted from this draft, and thus the numbers for equations, examples, and figures skip around a fair bit*

**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 desirable 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 (unlike 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. It is this latter aim that is our main focus in this chapter. 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.

Our treatment of selection and crossbreeding starts by considering some of the types of crosses that can be used, followed by a short discussion of the role of hybrids in modern agriculture. We then turn to the nuts-and-bolts of crosses: the genetic variances between lines, estimating the effects of lines for combining prediction of lines means, and rotational crossbreeding. Our discussion continues in the next chapter, where we discuss selection for improved crossing lines, creation of synthetic and composite populations, and conclude by discussing methods to detect elite alleles in lines.

## 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** (LW Chapter 20). 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 GCAs 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.

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 (e.g., Sprague and Jenkins 1943), such as the cross using two (different) four-way cross individuals as parents, which involves eight lines.

## HETEROISIS

Schull (1952) coined the term **heterosis** to refer in an increase in the vigor or other agriculturally-related traits (which are usually components of fitness such as yield). While the basic concept of heterosis is straight-forward, it can be defined a number of different ways, depending on what reference population is used (Lamkey and Edwards 1999). The most standard usage is **mid-parent heterosis**, defined as the  $F_1$  mean exceeding the average mean of the two parental lines,

$$h = \mu_{F_1} - \frac{\mu_{P_1} + \mu_{P_2}}{2} = \mu_{F_1} - \mu_{\bar{P}} \quad (12.1)$$

Often the mean of the  $F_1$  exceeds the mean of the best parent, and such **high-parent heterosis** is highly sought after by breeders.

### Dominance and Heterosis

While epistasis can generate heterosis, most of the focus in the literature has been on dominance-generated heterosis. No doubt a major reason for this is that dominance is much easier to model, and expressions for predicted line cross means are rather simple when epistasis is negligible. To see under what conditions dominance generates heterosis, consider the cross between two populations,  $P_1$  and  $P_2$ . For on a diallelic locus underlying the trait of interest, let the genotypes **AA** : **Aa** : **aa** have genotypic values of  $a$  :  $d$  :  $-a$ , and let  $p$  and  $p + \delta_p$  be the frequency of allele **A** in populations 1 and 2 (respectively). A convenient notation will be to denote the means of the  $P_1$  and  $P_2$  parental populations by  $\mu_{11}$  and  $\mu_{22}$ , and the  $F_1$  by  $\mu_{12}$ . Assuming the parental populations are in Hardy-Weinberg, and following Willham (1970, and Willham and Pollak 1985) for their  $F_1$  cross, these means become

$$\begin{aligned} \mu_{11} &= (2p - 1)a + 2p(1 - p)d \\ \mu_{22} &= \mu_{11} + 2\delta_p a - 2\delta_p^2 d \\ \mu_{12} &= \mu_{11} + \delta_p a \end{aligned} \quad (12.2a)$$

The resulting (mid-parental) heterosis (for this particular locus) becomes

$$h = \mu_{12} - \frac{\mu_{11} + \mu_{22}}{2} = \delta_p^2 d \quad (12.2b)$$

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

$$h = \sum_{i=1}^n \delta_{p_i}^2 d_i \quad (12.2c)$$

Dominance thus contributes to heterosis provided it is directional (the  $d_i$ 's tend to be positive) and their are differences in the allele frequencies between population. Further, heterosis is expected to increase with the allele frequency differences in the populations being crossed. In the most extreme case, crossing inbred lines,  $h = \sum d_i$ .

Under what conditions does the mean of the  $F_1$  exceed that of both parents? Summing over all loci, we can rewrite Equation 12.2a as

$$\mu_{12} - \mu_{11} = \sum_{i=1}^n \delta_{p_i} a_i, \quad \text{and} \quad \mu_{12} - \mu_{22} = 2 \sum_{i=1}^n \delta_{p_i}^2 d_i - \sum_{i=1}^n \delta_{p_i} a_i$$

Thus, the hybrid exceeds the best parent (high parent heterosis) when

$$2 \sum_{i=1}^n \delta_{p_i}^2 d_i > \sum_{i=1}^n \delta_{p_i} a_i > 0 \quad (12.3a)$$

Note that while overdominance ( $d_i > a_i$ ) facilitates Equation 12.3, is it not required. For completely inbred lines, Equation 12.3a reduces to

$$2 \sum_{i=1}^n d_i > \sum_{i=1}^n a_i > 0 \quad (12.3b)$$

An important point is that the  $F_1$  is *not* in Hardy-Weinberg equilibrium, as the mating is non-random (an individual from population one is always crossed to one from population two). The frequency of heterozygotes in the  $F_1$  is

$$\begin{aligned} \Pr(\mathbf{Aa}) &= \Pr(\mathbf{A} \text{ for Pop 1}) \Pr(\mathbf{a} \text{ for Pop 2}) + \Pr(\mathbf{a} \text{ for Pop 1}) \Pr(\mathbf{A} \text{ for Pop 2}) \\ &= p(1 - [p + \delta_p]) + (1 - p)(p + \delta_p) \\ &= 2\bar{p}(1 - \bar{p}) + \frac{\delta_p^2}{2} \end{aligned} \quad (12.4)$$

where  $\bar{p} = p + \delta p/2$  is the frequency of  $\mathbf{A}$  in the hybrid population. Hence, the excess of heterozygotes in an  $F_1$  relative to a random-mating population with the same allele frequency is  $(\delta p)^2/2$ . If the allele frequencies are close, this excess is small, while its maximal value of 1/2 occurs with crosses between completely inbred lines (100 percent heterozygotes vs. 50 percent heterozygotes under random mating). After one generation of random mating (the  $F_2$ ), the amount of heterosis is expected to decrease, reflecting a decrease in the percentage of heterozygotes. In particular, the mean value in the  $F_2$  is  $\mu_{F_2} = (2\bar{p} - 1)a + 2\bar{p}(1 - \bar{p})d$ , and applying Equation 12.2a plus some simplification yields

$$\mu_{F_2} - \frac{\mu_{11} + \mu_{22}}{2} = \frac{\delta p^2 d}{2} = \frac{h}{2} \quad (12.5a)$$

Hence

$$\mu_{F_2} - \mu_{F_1} = \frac{h}{2} \quad (12.5b)$$

so that half the initial heterosis (from dominance) is lost in the  $F_2$ . However, since the  $F_2$  is in Hardy-Weinberg equilibrium, future generations are expected to have the same mean and no further reduction in heterosis occurs.

Several authors have extended this basic theory. Cress (1966) examines crosses where more than two alleles are segregating, concluding that negative heterosis is not unlikely, even when overdominance is present. Only in the case of complete dominance will the heterosis be positive (or zero) for all potential vectors of allele frequencies differences between the crossed populations. The effects of epistasis on heterosis have been examined by Willham and Pollak (1985) and Goodnight (1999). Finally, traits acting in a multiplicative fashion can also generate heterosis, even in the absence of dominance (Richley 1942, Williams 1959, Garifus 1959), and this has been examined by Schnell and Cockerham (1992).

### Inbreeding vs. Panmictic Heterosis

Equation 12.2 (showing the effect of dominance on heterosis) assumes that both parental populations are in Hardy-Weinberg. When this is not the case (for example, under inbreeding), the effects of heterosis can be confounded with the effects of inbreeding depression (Lamkey and Edwards 1999). Suppose each parental population is a collection of completely homozygous lines formed by continually selfing members of some initial (random mating) base population. [Note that this situation is different from a cross of two pure lines, where the genotypes are identical within each line and each line is in Hardy-Weinberg, and the above equations hold.] If  $p$  is the frequency of allele **A** in the initial base population of  $P_1$ , then after complete inbreeding,  $p$  of the individuals will be **AA**, while the rest  $(1 - p)$  are **aa**. Hence the mean of  $P_1$  is just

$$\mu_{11}^{(I)} = ap - a(1 - p) = a(2p - 1)$$

Likewise,

$$\mu_{22}^{(I)} = a(2[p + \delta_p] - 1)$$

where we have used  $I$  to denote the mean under complete inbreeding. Since the  $F_1$  resulting from the cross between these two populations has mean

$$\mu_{F_1} = ap(p + \delta_p) + dp(1 - [p + \delta_p]) + d(1 - p)(p + \delta_p) - a(1 - p)(1 - [p + \delta_p])$$

the heterosis, based on inbred populations as the parental reference, is

$$h_I = F_1 - \frac{\mu_{11}^{(I)} + \mu_{22}^{(I)}}{2} = d\delta_p + d2p(1 - [p + \delta_p])$$

This can be alternately expressed as,

$$h_I = \frac{d\delta_p^2}{2} + 2d\bar{p}(1 - \bar{p}) \quad (12.6)$$

Lamkey and Edwards (1999) refer to  $h_I$  as the **inbred midparent heterosis**, and Equation 12.2b as the **panmictic midparent heterosis**. There are two components to  $h_I$ , a contribution from the panmictic midparent heterosis and a contribution from the removal of inbreeding depression. To see this second contribution, note that the difference between the random mating and completely inbred contributions of a locus from population one is

$$\Delta\mu_{11}^{(I)} = \mu_{11} - \mu_{11}^{(I)} = ap^2 + 2dp(1 - p) - a(1 - p)^2 - a(2p - 1) = 2dp(1 - p)$$

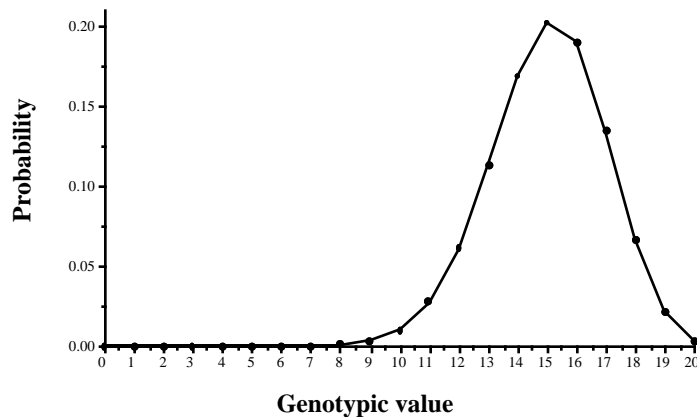
Lamkey and Edwards use the term **baseline heterosis** to denote the contribution in the  $F_1$  simply from the recovery of the effects of inbreeding depression reducing the parental means. Baseline heterosis,  $h_B$  is simply the difference between the panmictic and inbred values,

$$h_B = h - h_I = 2d\bar{p}(1 - \bar{p}) - \frac{d\delta_p^2}{4} \quad (12.7)$$

Since this contribution arises entirely from differences between the random mated and fully inbred values of the parental populations this, we can equivalently express this as

$$h_B = \frac{\Delta\mu_{11}^{(I)} + \Delta\mu_{22}^{(I)}}{2} = \frac{2dp(1 - p) + 2d[p + \delta_p](1 - [p + \delta_p])}{2}$$

which reduces to Equation 12.7.



**Figure 12.2.** The  $F_2$  distribution for 20 completely dominant loci of equal effect ( $d = a = 1$ ). Starting with a cross of pure lines, the probability that a locus contains at least one dominant allele is  $3/4$ , giving a mean value of  $20(3/4) = 15$ , which is also the mode. Note this mean is below the  $F_1$  mean of 20. Although the distribution is indeed skewed, the investigator is unlikely to observe those individuals with a genotypic value of 8 or less (unless the sample size is enormous). Focusing on the most likely classes of observed individuals (9 to 20), this restricted distribution shows little skew. After Collins (1921).

### Dominance vs. Overdominance and Epistasis

While dominance provides a convenient explanation of heterosis, several objections were initially raised to the strict dominance (i.e., no overdominance,  $d < a$ ). First, if no overdominance were involved, one would expect to find  $F_2$  individuals that were as good as the  $F_1$ , and this was generally not seen. Example 12.1 shows, one is unlikely to have sufficient sample size to find such individuals if the number of loci contributing to heterosis is moderate to large. Another objection was that the  $F_2$  distribution should be skewed, with the mode greater than the mean, while  $F_2$  populations generally show very little skew. Collins (1921) points out that the later is also a sample effect, see Figure 12.2. J

---

**Example 12.1.** Just how many  $F_2$  progeny do we have to screen to have a reasonable (say 50%) chance of recovering the largest genotypic value (i.e., one that matches the  $F_1$ ). We assume complete dominance, so that the presence of a single dominant allele is sufficient to maximize the contribution from a given locus (this is the most favorable setting). With a pure line cross, then with probability  $3/4$  at least one favorable allele is present in the  $F_2$ . Assuming unlinked loci, the probability of favorable alleles at  $n$  loci is  $(3/4)^n$ . To have a 50% probability that at least one individual has favorable alleles at all  $n$  loci requires sampling  $m$  individuals, where  $m$  satisfies

$$1 - [1 - (3/4)^n]^m = 0.5$$

This follows as  $(1 - (3/4)^n)$  is the probability that a single individual does not contain favorable alleles at all loci and this expression to the  $m$ th power is the probability that none of the  $m$  individuals do. Solving gives

$$m_{0.5} = \frac{\ln(0.5)}{\ln[1 - (3/4)^n]}$$

More generally, to have probability  $100\alpha$  requires a sample size of

$$m_\alpha = \frac{\ln(1 - \alpha)}{\ln[1 - (3/4)^n]}$$

For example, for  $n = 15, 20, 30,$  and  $40$  loci,  $m_{0.5} = 52, 218, 921, 3881,$  and  $68,924$ . Hence, while there is a reasonable chance of recovering the best genotype with the number of loci is modest (20 or less), this very quickly becomes impossible with more than 30 loci. Further confounding these numbers is that not only must a plant be present in the sample, but we also must be able to assess its genotype with very high confidence, requiring replication across a number of environments (see Chapter 10).

---

## AGRICULTURAL EXPLOITATION OF HETEROISIS

The exploitation of hybrid vigor in agriculture traces back at least 5,000 years to the Sumerians (Clutton-Brock 1992), who produced mules by crossing horses (*Equus caballus*) with donkeys (*S. asinus*). Modern agriculture is highly shaped by hybrids, with the heterosis resulting is significantly increased yield in crops. Such gains direct translate into fewer acres that must be farmed to obtain the same total yield, and these land savings are by no means trivial (Table 12.1). Another critical, yet often overlooked, benefit of hybrids to modern agriculture is uniformity, which allows for greater efficiencies in harvesting, such an greatly increased mechanization (Goldman 1999).

**Table 12.1.** 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

---

### 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 Wright's (1922) extremely influential publication on crossbreeding (and inbreeding) in guinea pigs. General reviews of heterosis in animals are given by Gowen (1952), Sang (1956), and Sheridan (1981).

Animal breeders often distinguish between **individual** and **maternal heterosis** (e.g., Nitter 1978). 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 12.2). For example, Nitter (1978) reports that 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). Cundiff et al. (1974a, 1975b) found that maternal effect heterosis was roughly twice as individual heterosis for several traits in European cattle (*Bos taurus*).

Maternal and individual heterosis effects can be combined by 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 (Nitter 1978). 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* × *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, 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.

**Table 12.2.** Estimate of individual  $h^I$  and maternal heterosis  $h^M$  in sheep (estimation of  $h^I$  and  $h^M$  is discussed later in this chapter). 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. From Nitter (1978).

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	≈ 0		

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  $F_1$  seed, and this seed is generated anew each generation, with the  $F_1$  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  $B \times C$  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.

Winters (1952) suggested a hybridization scheme, **rotational crossbreeding**, that 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. We examine this scheme in detail at the end of the chapter.

### **Crossbreeding in Animals: Heterosis in *Bos indicus* $\times$ *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*  $\times$  *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  $\times$  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  $\times$  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  $\times$  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*  $\times$  *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*  $\times$  *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*  $\times$  *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., Hariana 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).

Hybrid populations are typically outside of both Hardy-Weinberg and gametic-phase equilibrium. Indeed, this is often the purpose for generating a hybrid — to create an excess of certain genotypes (e.g., heterozygotes) over what would be expected in a population under Hardy-Weinberg. Here we explore how the genetic variances are redistributed within and between hybrids. As always, some reference population is required for the genetic variances, in this case it would be the hypothetical population generated by randomly mating the parental lines until equilibrium is reached. For unlinked loci, both Hardy-Weinberg and gametic-phase equilibrium is reached in the  $F_2$ . With linked loci, additional generations are required to reach linkage equilibrium.

### 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

$$\mu_A = \mu + g_A^I + g_A^M + g_A^{M'} \quad (12.13a)$$

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

$$\mu_{AB} = \mu + \frac{g_A^I + g_B^I}{2} + g_B^M + g_B^{M'} + h_{AB}^I \quad (12.13b)$$

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

$$\mu_{BA} = \mu + \frac{g_A^I + g_B^I}{2} + g_A^M + g_A^{M'} + h_{AB}^I \quad (12.13c)$$

From Equation 12.13a-c, it follows that an estimate of the individual (direct) heterotic effect is the obvious one (Nitter 1978),

$$\frac{\mu_{AB} + \mu_{BA}}{2} - \frac{\mu_{AA} + \mu_{BB}}{2} = h_{AB}^I \quad (12.13d)$$

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

$$\mu_{BA} - \mu_{AB} = (g_A^M + g_A^{M'}) - (g_B^M + g_B^{M'}) \quad (12.13e)$$

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, then 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,

$$\mu_{C \cdot AB} = \frac{2g_C^I + g_A^I + g_B^I}{4} + \frac{h_{CA}^I + h_{CB}^I}{2} + \frac{g_A^M + g_B^M}{2} + h_{AB}^M + g_B^{M'} + \frac{r_{ab}^I}{2} \quad (12.14a)$$

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 12.4 summarizes these coefficients for a variety of crosses. As Equation 12.med indicates, estimates of the various effects can be obtained by suitably weighted combination of line means. Combining 12.13b and 12.14a gives (Nitter 1978),

$$\mu_{C \cdot AB} - \frac{\mu_{CA} + \mu_{CB}}{2} = h_{AB}^M + \frac{r_{ab}^I}{2} \quad (12.14b)$$

Hence, if recombination effects are small (these would be absence in the absence of epistasis), Equation 12.14b allows for a direct estimate of maternal heterotic effects. Table 12.5 summarizes estimates of for individual and maternal heterotic effects for a variety of traits in sheep using this approach.

**Table 12.5.** 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	$g^I$	$h^I$	$r^I$	$g^M$	$h^M$	$r^M$	$g^{M'}$	$h^{M'}$	$r^{M'}$
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	

### ROTATIONAL CROSSBREEDING

The motivation for rotational crossbreeding is to maintain as much heterosis as possible. With a rotational scheme, the relative contributions from the different lines change each generation, eventually reaching limiting values, as the following example shows.

**Example 12.2.** Consider a three-bred rotational crossbreeding scheme where dams from the previous generation are crossed pure-bred sires in a rotating sequence (line  $A$  in one generation,  $B$  in the next,  $C$  in the third and so on). Under this scheme, what fraction of lines from each of the lines are present in any particular generation? The logic is straightforward: in each generation, half of the contribution from the previous dam is passed on, as is half the genes from the sire line. Thus,

Generation	Cross	Percentage of lines:		
		A	B	C
1	$A \times B$	50.0	50.0	0.0
2	$C \times$ gen 1 dam	25.0	25.0	50.0
3	$A \times$ gen 2 dam	62.5	12.5	25.0
4	$B \times$ gen 3 dam	31.3	56.3	12.5
5	$C \times$ gen 4 dam	15.6	28.1	56.3
6	$A \times$ gen 5 dam	57.8	14.1	28.1
7	$B \times$ gen 6 dam	28.9	57.0	14.1
8	$C \times$ gen 7 dam	14.5	28.5	57.0

The asymptotical contributions reached are 57.1% for the sire line, 28.6% for the sire line used in the previous generation, and 14.3% for the sire line used two generations previous.

In a similar fashion, for a four breed ( $A, B, C, D$ ) rotational scheme,

Generation	Cross	Percentage of lines:			
		A	B	C	D
1	$A \times B$	50.0	50.0	0.0	0.0
2	$C \times$ gen 1 dam	25.0	25.0	50.0	0.0
3	$D \times$ gen 2 dam	12.5	12.5	25.0	50.0
4	$A \times$ gen 3 dam	56.3	6.3	12.5	25.0
5	$B \times$ gen 4 dam	28.1	53.1	6.3	12.5
6	$C \times$ gen 5 dam	14.1	26.6	53.1	6.3
7	$D \times$ gen 6 dam	7.0	13.3	26.6	53.1
8	$A \times$ gen 7 dam	53.5	6.6	13.3	26.6

At equilibrium the line contributions cycle among 53.3%, 26.7%, 13.3%, and 6.7%.

Following the logic in Example 12.2, the expected fraction of genetic contributions at equilibrium from each of the  $k$  lines where crossbred dams are kept and crossed in rotation to purebred sires are of the form  $(1/2)*I, (1/2)^2*I, \dots, (1/2)^k*I$  where

$$I = \sum_{i=0}^{\infty} \left(\frac{1}{2^k}\right)^i = 1 + \frac{1}{2^k - 1} \quad (12.15)$$

with the largest fraction  $(I/2)$  from the line last used as a sire and the smallest fraction  $(I/2^k)$  for the line to be next used as a sire.

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

$$\hat{R}_2 = \bar{z}_{AB} - \frac{\bar{z}_{AB} - \bar{P}_2}{3}, \quad \text{where } \bar{P}_2 = \frac{\bar{z}_A + \bar{z}_B}{2} \quad (12.16)$$

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

$$\hat{R}_3 = \overline{SC}_3 - \frac{\bar{z}_{AB} - \bar{P}_3}{7}, \quad \text{where } \overline{SC}_3 = \frac{\bar{z}_{AB} + \bar{z}_{AC} + \bar{z}_{BC}}{3} \quad (12.16b)$$

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. At equilibrium, contribution of lines is 57.14, 28.57, and 14.29 percent

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

$$\hat{R}_4^{(A,B,C,D)} = \overline{SC}_4 - \frac{\overline{SC}_{na} - \bar{P}_4}{15}, \quad \text{where } \overline{SC}_{na} = \frac{\bar{z}_{AC} + \bar{z}_{BD}}{2} \quad (12.17)$$

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-adjacents lines in the rotations.

**Example 12.3.** 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)

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

How well does Equation 12.16 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.

**Example 12.4.** As illustrated in Example 12.3, it would be nice to have standard errors for the predicted values. Here we illustrate how this is done for the two-bred predictor, but the basic approaches easily extends to most of the other predictors developed in this chapter. Rearranging to collect common terms,

$$\hat{R}_2 = \bar{z}_{AB} - \frac{\bar{z}_{AB} - \bar{P}_2}{3} = \left(1 - \frac{1}{3}\right) \bar{z}_{AB} + \left(\frac{1}{6}\right) (\bar{z}_A + \bar{z}_B)$$

Since the estimates are independent and recalling that  $\sigma^2(ax) = a^2\sigma^2(x)$  for a constant  $a$ , it immediately follows that

$$\sigma^2(\hat{R}_2) = \left(1 - \frac{1}{3}\right)^2 \sigma^2(\bar{z}_{AB}) + \left(\frac{1}{6}\right)^2 \left(\sigma^2(\bar{z}_A) + \sigma^2(\bar{z}_B)\right)$$