12

Inbreeding Depression

Nature thus tells us, in the most emphatic manner, that she abhors perpetual self-fertilization.
Darwin 1876. Version 4 September 2022

Inbred individuals are almost always less fit than progeny of nonrelatives, and, similarly, many traits decline from favorable values under inbreeding. The decline in the mean phenotype with increasing homozygosity within populations, known as **inbreeding depression** (**ID**), is often interpreted as heterosis-in-reverse. However, as will be seen below (as well as in Chapter 13), there are some important distinctions between the genetic mechanisms contributing to inbreeding depression *within* populations and heterosis *between* populations.

The serious study of the impact of inbreeding begins with Darwin (1868, 1876), although published results go back to at least Knight (1799) and numerous cultures have long traditions against marriage of close relatives (Brown 1991). Darwin's interest was no doubt influenced by the fact that he married his first cousin, Emma Wedgwood. Many of Darwin's ten children had significant health issues, and he was concerned that this was, in part, the result of marrying a close relative. Ávarez et al. (2015) examined Darwin's concern by looking at male fertility in the Darwin-Wedgwood pedigree, finding a significant impact of the husband's inbreeding coefficient on both family size and reproductive duration.

The near universal existence of inbreeding depression bears importantly on many basic issues in evolutionary biology as well as on a number of practical issues in human genetics, agriculture, and conservation biology. For example, the deleterious consequences of self-fertilization are likely to be the leading selective forces responsible for the evolution of various aspects of mating systems in plants (Darwin 1876; Lande and Schemske 1985; Schemske and Lande 1985; Charlesworth and Charlesworth 1987; Uyenoyama 1993; Waller 1993; Carr and Dudash 2003; WL Chapter 23), and of behavioral mechanisms for avoiding mating with close relatives in animals (Shields 1982; Thornhill 1993; Pusey and Wolf 1996). The observations of maize breeders that crosses between inbred lines yield substantially more grain than the inbreds themselves (East 1908; Shull 1908; Jones 1918; Sprague 1983) has given rise to a situation in which corn farmers are now almost entirely reliant on seed-producing companies for hybrid seed (Figure 12.1). Finally, the loss of fitness due to the development of inbreeding depression in small populations is a major concern in endangered species management (Templeton and Read 1984; Lacy et al. 1993; Hedrick 1994; Lynch et al. 1995a, 1995b; Lynch 1996; Saccheri et al. 1998; Hedrick and Kalinowski 2000; Richards 2000; O'Grady et al. 2006; Leberg and Firmin 2008; Jamieson and Allendorf 2012; Frankham et al. 2013, 2014; Franklin et al. 2014; Caballero et al. 2017a, 2017b; Pérez-Pereira et al. 2022).

It is widely appreciated that inbreeding depression is an inevitable consequence of dominance. When gene action is entirely additive (no dominance or epistasis), the average phenotypic effects associated with alleles are independent of the genetic background. Hence, inbreeding depression cannot occur for characters with a purely additive genetic basis (as inbreeding changes *genotype*, not *allele*, frequencies). With dominance, however, the average phenotypic effect of an allele changes with a change in genotype frequencies, even in the absence of allele frequency change, because the allelic effect is now a function of the genetic background (Equation 4.10b).

Because there are many types of dominance, this simple explanation for inbreeding depression by no means provides a complete understanding of the process. We start by showing how the two competing hypotheses on the genetic mechanism of inbreeding depression—partially recessive deleterious alleles versus overdominance—lead to some very similar predictions that are in accordance with empirical observations, but also to some major differences that are less easy to resolve empirically (Charlesworth and Charlesworth

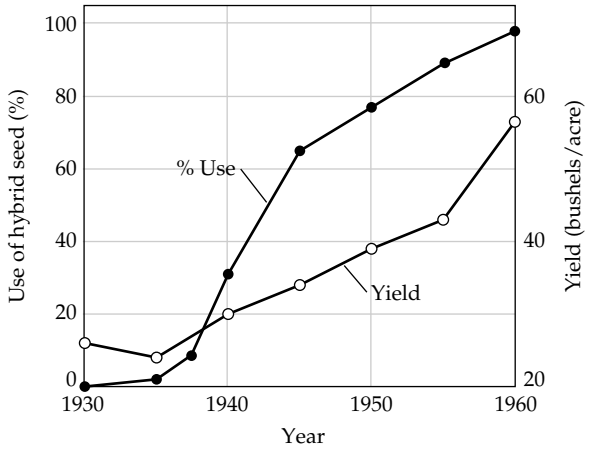

Figure 12.1 Historical change in the reliance on hybrid corn by United States farmers and the increase in mean annual harvest. Duvick (2005) estimated that 50-60% of the yield gain was genetic, with the remainder due to improved agronomic practices. (After Sprague 1983.)

1999; Roff 2002; Carr and Dudash 2003; Kristensen and Sørensen 2005; Charlesworth and Willis 2009; Kristensen et al. 2010). Second, we provide a brief outline of the basic statistical issues that arise in the analysis of inbreeding depression. Third, we review the large body of existing data, showing that inbreeding depression exists, at least to some degree, for essentially all characters in all populations of diploid organsims. We close by reviewing how molecular-marker analysis is starting to refine our understanding of some of these issues.

THE GENETIC BASIS OF INBREEDING DEPRESSION

In the absence of allele-frequency change, inbreeding shifts the genotype frequencies in a population in a very simple way. Let f denote the inbreeding coefficient for the population, i.e., the probability that an individual carries two alleles that are identical by descent at a locus (Chapter 7). At any given *locus*, a partially inbred population has a fraction $(1 - f)$ of noninbred individuals, whose genotype frequencies are in the Hardy-Weinberg proportions. The remaining proportion (f) of the population that is inbred consists entirely of homozygous classes at this locus, each of which has a frequency equal to the respective allele frequency (Table 12.1). With this information in hand, it is straightforward to derive quantitative expressions for the two mechanistic hypotheses for inbreeding depression.

The **dominance hypothesis** (Davenport 1908; Bruce 1910; Keeble and Pellew 1910; Jones 1917) argues that inbreeding depression is caused by the expression of deleterious recessive genes in homozygous individuals. (We retain the use of the term dominance only for historical reasons; **partial dominance** is a more apt term in that the hypothesis focuses explicitly on dominance due to partially to completely recessive genes). Consider a diallelic locus, where the frequency of the deleterious allele is q, and the fitnesses of the three genotypes are denoted as 1, $1 - hs$, and $1 - s$ (Table 12.1). Here, s measures the selection against homozygotes for the deleterious allele, and h is a measure of dominance, with $h = 0.5$ implying additivity and $0 < h < 0.5$ implying that the deleterious allele is partially recessive. Multiplying genotype frequencies (Table 12.1) by their fitnesses, the mean fitness in a population inbred to level f becomes

$$\overline{W}_f = \overline{W}_0 - fpqs(1 - 2h) \tag{12.1a}$$

where

$$\overline{W}_0 = 1 - 2pqsh - q^2 s \tag{12.1b}$$

Table 12.1 Genotypic frequencies and fitnesses under the two dominance-based hypotheses for inbreeding depression. Two alleles (B, b) are assumed to be present, with respective frequencies p and q.

Genotype	Frequency	Fitness		Phenotype for Arbitrary Character
		Partial Dominance	Overdominance	
BB	$p^2(1-f)+pf$	1	$1-t$	$2a$
Bb	$2pq(1-f)$	$1-hs$	1	$a+d$
bb	$q^2(1-f)+qf$	$1-s$	$1-s$	0

is the mean fitness in the random-mating base population. Provided that $h < 0.5$, then $(1-2h)$ is necessarily positive. Thus, with partially recessive deleterious alleles, mean fitness is expected to decline linearly with increasing inbreeding coefficient f.

Unless the mutation rate is very high, deleterious alleles are expected to be maintained at low frequencies by selection (WL Chapter 7), so it can be assumed in Equation 12.1a that $p = 1-q \simeq 1$, showing that the expected decline in fitness due to complete inbreeding ($f = 1$) at a locus is approximately $qs(1-2h)$. For a randomly mating population in selection-mutation balance, if u is the mutation rate from the beneficial to the deleterious allele and $u < h^2 s$, then $q \simeq u/(hs)$ (Haldane 1927; WL Chapter 7). Thus, for large randomly mating populations, the decline in fitness resulting from complete inbreeding at a locus is approximately $[u/(hs)]s(1-2h) = u(1-2h)/h$. This result is ***independent of the intensity of selection*** at the locus (s) because of the inverse relationship between the equilibrium frequency of a deleterious allele and its selection coefficient.

A second potential explanation for inbreeding depression is referred to as the **overdominance hypothesis** (East 1908; Shull 1908; Hull 1946). The idea here is that something special about the heterozygous state causes increased vigor relative to both homozygotes. Letting s and t denote the proportional reduction in fitness of the two homozygotes relative to that of the heterozygote (Table 12.1), then

$$\overline{W}_f = \overline{W}_0 - fpq(s+t) \tag{12.2a}$$

where

$$\overline{W}_0 = 1 - p^2 t - q^2 s \tag{12.2b}$$

As in the case of partial recessives, the overdominance hypothesis leads to the prediction that mean fitness will decline linearly with increasing f. However, contrary to the situation with partial recessives, the loss of fitness *increases* with the strength of selection $(s+t)$ maintaining the polymorphism in the random-mating population. In a large randomly mating population, heterozygote superiority leads to a balanced polymorphism with $\widehat{p} = s/(s+t)$ and $\widehat{q} = 1-\widehat{p} = t/(s+t)$ (Haldane 1927). Hence $f\,[\,\widehat{p}\,\widehat{q}\,(s+t)\,] = f[\,st/(s+t)\,]$. If, for example, $s = t$, the loss of fitness per locus under complete inbreeding is $s/2$.

These alternative hypotheses for inbreeding depression have extremely different evolutionary, and applied, implications. Under the dominance hypothesis, inbreeding depression is an inevitable consequence of recurrent mutation at the genomic level, implying that much of the genetic variation within populations must be associated with the constant influx of deleterious alleles. Although selection *removes* some of these alleles each generation, mutation replaces them (WL Chapters 7 and 26). Conversely, under the overdominance hypothesis, variation is *maintained* by selection favoring the heterozygous state at multiple loci. Here, variation is maintained even in the absence of mutation pressure.

Considerable uncertainty exists as to whether overdominance with respect to fitness is a common phenomenon (Hedrick 2012; Charlesworth 2015; Thurman and Barrett 2016). Only

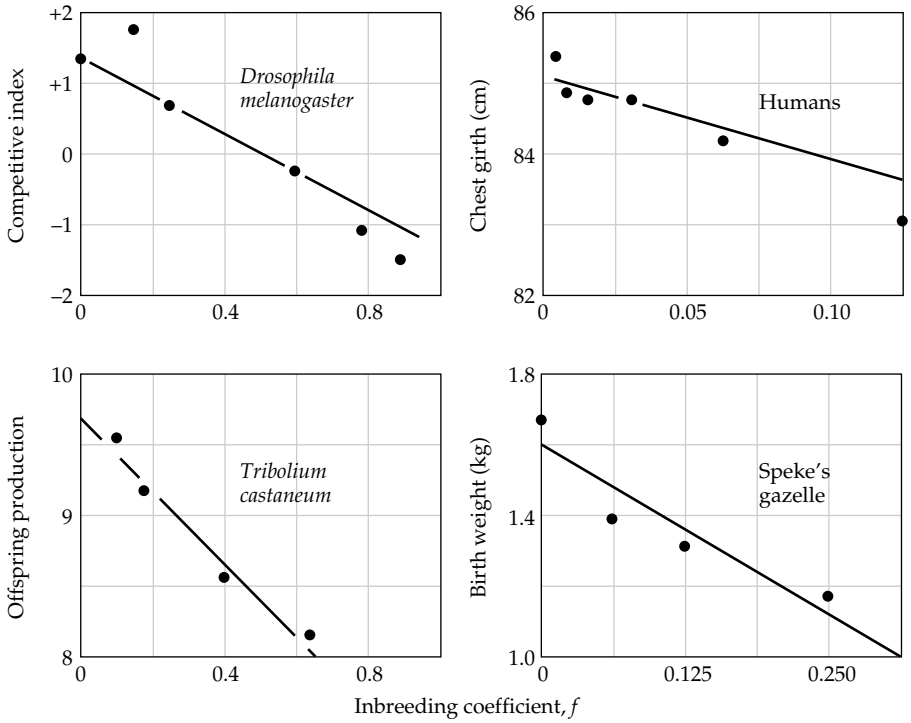

Figure 12.2 Change in mean phenotypes as a function of inbreeding. References: *Drosophila* (Latter and Robertson 1962); humans (Barrai et al. 1964); *Tribolium* (Rich et al. 1984); Speke's gazelle (Templeton and Read 1983).

rarely has it been suggested by studies with molecular markers, and most of those studies are open to alternative interpretations (discussed below). Nevertheless, as cogently pointed out by Crow (1948, 1952), even if overdominance is quite rare, it warrants serious consideration as a contributing factor in inbreeding depression. To see why, consider the expected reduction in fitness under both hypotheses when gene frequencies are in equilibrium. Under the dominance hypothesis, the maximum inbreeding depression per locus, arising with very small h, is approximately u/h. Because u is likely to be on the order of 10^{-5} or smaller for most loci, and the evidence suggests that h is usually greater than 0.1 or so (discussed below), the per-locus inbreeding depression resulting from partial dominance is expected to be quite small, $\leq 10^{-4}$. With overdominance, complete inbreeding leads to the loss of the fittest genotype, so the reduction in fitness is potentially quite large. Thus, even if overdominance is a rare phenomenon, only a few such loci need to exist for its contribution to inbreeding depression to rival, or even surpass, that caused by a much larger number of loci displaying partial dominance.

The linear decline in the means of fitness-related traits with an increase in the inbreeding coefficient, observed in many sets of data (Figure 12.2), is consistent with both the partial dominance and overdominance hypotheses. There is, however, a major distinction between the two hypotheses with respect to the expected distribution of mean phenotypes among inbred lines. If overdominance is the major cause of inbreeding depression, all inbred lines must eventually perform *below* the mean of the randomly mating base population, because a pure line of the best-performing genotype (a heterozygote) cannot be attained. If, on the other hand, partial dominance is the major factor, it should be possible to produce a pure inbred line that performs *at least as well* as the most outstanding member of the base population. If large numbers of loci contribute to the trait of interest, the probability of producing such a line may be quite low. Nevertheless, such lines have been obtained in several studies (Smith 1952; Williams 1959; Wienhues 1968; Busch et al. 1971; Pooni et al. 1994; Uddin et al. 1994), see Figure 12.3.

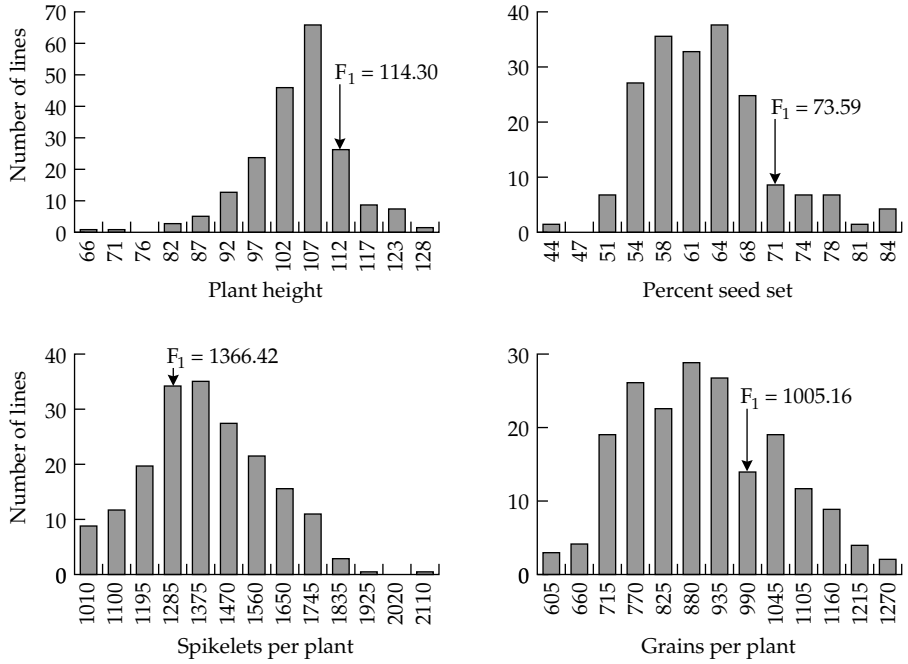

Figure 12.3 Frequency distributions for four characters in recombinant inbred lines of rice, compared to the mean of the F_1 progeny obtained from a cross between two homozygous lines. In all four cases, the F_1 performance exceeds that of both parental lines (i.e., they display heterosis; Chapter 13). More interestingly, some individual inbred lines *exceed* the performance of the parental lines. The recombinant inbred lines were obtained by randomly sampling 194 individuals from the F_2 population, and taking each of them through six rounds of selfing and single-seed descent. (From Xiao et al. 1995.)

A More General Model: Dominance

The preceding paragraphs have focused on the consequences of inbreeding for *fitness*. A more general account of the change in the mean of an arbitrary character (as opposed to the very special trait of fitness) under inbreeding proceeds as follows. Recalling the genotypic frequencies for an inbred population and multiplying them by their respective genotypic values (Table 12.1), a general expression for the mean genotypic value for a single diallelic locus is

$$\mu_f = (1-f)[\,p^2(2a) + 2pq(a+d) + q^2(0)\,] + f[\,p(2a) + q(0)\,] = \mu_0 - (2pqd)f \qquad (12.3a)$$

where $\mu_0 = 2p(a+qd)$ is the mean genotypic value in the randomly mating base population. Summing over all loci, the total inbreeding depression is $\mu_f - \mu_0 = -fI$, where

$$I = 2\sum p_i q_i d_i \qquad (12.3b)$$

Recalling (Equation 4.12b) that the dominance genetic variance in a randomly mating population in gametic-phase equilibrium is $\sigma_D^2 = \sum (2p_i q_i d_i)^2$, it is clear that ***dominance variance is* necessary, *but not* sufficient**, for inbreeding depression to occur. Note that both I and σ_D^2 are maximized a intermediate allele frequencies (p_i = q_i = 0.5, as occurs in the progeny from an inbred line cross). Because the sign of d_i may vary from locus to locus, it is possible for considerable canceling to occur among the effects at different loci, leading to negligible inbreeding depression despite substantial dominance genetic variance. In other words, significant inbreeding depression requires **directional dominance** ($d > 0$), namely that the value of the heterozygote tends to be closer to the value of the larger homozygote. While the values of I and σ_D^2 can be completely uncoupled (if the signs of the d_i are entirely

random), they often tend to be associated. For example, Misztal et al. (1997) examined 14 traits in Holstein cattle, and found that the correlation between values of I and σ_D (taking the square root of the variances to make the comparison on the same scale) was around 80 percent.

The connection between dominance variance and inbreeding depression when a large number of loci underly a trait deserves further comments. Assuming n equivalent loci (all have the same effect and allele frequency), the dominance variance becomes $\sigma^2_D = n\left[2dp(1-p)\right]^2$. Assuming a very large number of underlying loci, then d must scale as $\delta/\sqrt{n}$ in order for the dominance variance to be bounded as n becomes very large (where δ is an arbitrary, but finite, constant). With this scaling,

$$\sigma^2_D = 4np^2(1-p)^2(\delta/\sqrt{n})^2 = 4\delta^2p^2(1-p)^2 \tag{12.3c}$$

Turning to inbreeding depression, Equation 12.3b becomes

$$I = 2np(1-p)d = 2\sqrt{n}p(1-p)\delta \tag{12.3d}$$

Hence, for this special case of very numerous equivalent loci, $I/\sigma_D \simeq \sqrt{n}$. As noted by Robertson and Hill (1983), as the number of loci showing directional dominance increases, I can become increasingly larger for a fixed value of dominance variance. Biologically what this implies is that if there is a strong trend for segregating loci to show directional dominance, then when a large number of loci underlies a trait, one can have significant inbreeding depression, even with an apparently small value of σ^2_D.

Hence, while a trait can display very low amounts of ID even when σ^2_D is large, the converse can also be true, with a trait showing large ID even when σ^2_D is small. We have seem a similar phenomena in Chapter 5, where we noted that a considerable amount of biological interactions (such as dominance) does not necessarily translate into a significant interaction *variance*. Indeed, for a rare recessive ($k = 1, q \simeq 0$), then Equations 4.12c and 4.10 imply that $\sigma^2_A/\sigma^2_D \simeq 2/q$, so that even with complete dominance, essentially all of the genetic variance is additive. Conversely, Equation 12.3b gives the contribution from this locus to I as $\simeq (2qa)f$, where $2a$ is the difference between homozygotes and $d = ak$ (Chapter 4). Hence, if ID is generated by rare recessive alleles, one expects to find genes involved in ID also being associated with significant additive variance. Indeed, Yengo et al. (2021) found that human genomic regions associated with ID are enriched for sites contributing additive genetic variance for the impacted trait.

There are two (non-exclusive) theories for why directional dominance may tend to be pervasive. First, we previously demonstrated that directional dominance may be a fairly generally property of enzymatic systems (Figure 4.5). Second, natural selection provides a filter to increase the segregating frequency of deleterious alleles that are more recessive. When new deleterious mutations arise, those with heterozygous values that are intermediate between the wildtype and mutant allele are rapidly removed from the population. Conversely, alleles that are recessive tend to persist for longer amounts of time, creating an enriched set of segregating alleles that is biased in favor of directional dominance, even if such a bias did not exist in the original collection of new mutations. Hence, traits under directional selection are often *expected* to show directional dominance.

Example 12.1. A large empirical study with the flour beetle *Tribolium castaneum* provides some perspective on the role of dominance in inbreeding depression. López-Fanjul and Jódar (1977) derived 105 lines from a large base population and maintained them by single brother-sister matings for 8 generations (to $f = 0.785$). Despite the large sample sizes, the authors could find no evidence that inbreeding caused a shift in the mean rate of egg laying by virgin females at 33 or 28°C. Independent estimates of the heritabilities for these traits, obtained by full-sib correlation and daughter-mother regression, were 0.34 $\pm$ 0.02 and 0.33 $\pm$ 0.01 at 33°C, and 0.33 $\pm$ 0.02 and 0.26 $\pm$ 0.02 at 28°C. Recalling from Chapter 7 that heritabilities

estimated from full-sib analysis are inflated by dominance genetic variance relative to those obtained by parent-offspring analysis, only for the second temperature is there any evidence of dominance genetic variance, and this is slight (and any additional full-sib variance could be entirely due to maternal effects, rather than dominance; Chapter 23). Thus, the absence of inbreeding depression for rate of egg laying by virgins is not surprising. However, the study population was not immune to the effects of inbreeding, because two other traits, rate of egg laying by fertilized females and egg viability, exhibited substantial declines with inbreeding.

A More General Model: Epistasis

Because inbreeding depression is a consequence of nonlinear interactions between gene effects, it stands to reason that epistasis may complicate matters. However, provided the base population is in gametic-phase equilibrium, only epistasis involving some dominance component (such as AD or DD) contributes to inbreeding depression within populations (Anderson and Kempthorne 1954; Bulmer 1980; Hill 1982a; Lynch 1991). This result arises because although inbreeding causes a change in single-locus *genotypic* frequencies (and hence in terms in Equation 5.6 involving D), it does not (in the absence of selection and gametic-phase disequilibrium) alter the *gametic* frequencies in the population (just as allele frequencies do not change under inbreeding). With this in mind, a general expression for inbreeding depression can be obtained using the composite effects approach (Chapter 11).

Letting $-\delta_1$ be the expected change in the mean caused by single-locus dominance effects (summed over all loci) under complete inbreeding, then $-f\delta_1$ is the expected change at inbreeding level f (δ_1 equals I from Equation 12.3b). The composite additive $\times$ dominance effect may also be altered under inbreeding. It depends only on the inbreeding at single loci and may be represented as $-f(\alpha\delta)$. The composite dominance $\times$ dominance effect depends on whether one ($\delta_2^{[1]}$) or two ($\delta_2^{[2]}$) loci are inbred (have both alleles IBD). Assuming unlinked loci, the probabilities of these two situations are, respectively, $2f(1-f)$ and f^2. Thus, the shift in the mean through the inbreeding-induced alteration of dominance $\times$ dominance effects can be represented by $-2f(1-f)\delta_2^{[1]}$ in the first case and $-f^2\delta_2^{[2]}$ in the second. Summing up terms,

$$\mu_f = \mu_0 - f[\delta_1 + 2\delta_2^{[1]} + (\alpha\delta)] - f^2(\delta_2^{[2]} - 2\delta_2^{[1]}) \tag{12.4a}$$

or more succinctly,

$$\mu_f = \mu_0 - f\Delta_1 - f^2\Delta_2 + \cdots \tag{12.4b}$$

Thus, the expected mean phenotype under inbreeding can be written as a polynomial function of f, with the coefficients Δ_1 and Δ_2 being functions of multiple types of nonadditive gene action. Note that if $\Delta_2 > 0$, the effect in inbreeding depression is *accelerated* as f increases, a condition often called reinforcing, negative, or **synergistic**, **epistasis**. Conversely, if $\Delta_2 < 0$, then the rate of inbreeding depression *decelerates* as f increases, leading to positive, diminishing, or **antagonistic**, **epistasis**.

Equation 12.4b indicates that a net contribution of epistasis to inbreeding depression may sometimes be detected as a nonlinear relationship between the mean phenotype and level of inbreeding. Because the composite inbreeding effects (Δ_i) can be positive or negative, a variety of forms of this relationship is possible. For example, Gulisija et al. (2007) used local regression (LOESS) methods to show nonlinearity in ID with f, finding both synergistic and antagonistic epistasis over their collection of production traits in Jersey cattle. ID-based evidence of synergistic epistasis was found for pollen size in *Mimulus* (Kelly 2005) and for pupal number in *Drosophila* (Domíngues-García et al. 2019).

Despite such success stories, the possibility that the epistatic effects involving different loci may cancel each other implies that a lack of nonlinearity cannot be taken as definitive evidence for the absence of important dominance epistatic interactions between loci. Even in the absence of any canceling effect, large departures from linearity are unlikely unless

epistasis is very pronounced (Δ_2 large) for the simple reason that f^2 is small relative to f, especially with small f. Moreover, when a nonlinear response is observed, care must be taken to ensure that it is not simply due to the selective elimination of lines as inbreeding proceeds. Likewise, if selection is occurring on the trait during inbreeding (as might be expected with fitness-related traits), then purging (see below) of more deleterious alleles can occur. This is especially true for deleterious alleles that interact synergistically, as selection would disproportionately remove these, creating a bias against their detection. Hence, loss of lines and/or selection of alleles among surviving lines will both result in the value of I in some set of more inbred, surviving, lines being potentially different—and perhaps *rather* different—from the base-population starting value (Equation 12.3b). Curik et al. (2001) and Domíngues-García et al. (2019) examined other issues involving the regression of the mean on f when epistasis is present.

Finally, we note that although it is often stated that heterosis (the tendency for F_1 phenotypes to exceed the mean phenotypes of its two parental lines; Chapter 13) and inbreeding depression are different manifestations of a common phenomena (directional dominance), this equivalency is not strictly correct. As we will show in Chapter 13, heterosis is genetically equivalent to $(2\delta_1^c - \alpha_2^c)$, where δ_1^c and α_2^c are the composite dominance and composite additive $\times$ additive effects of genes in the two parental lines. On the other hand, under complete inbreeding within populations, the decline in the mean phenotype is defined by the sum $[\delta_1 + (\alpha\delta) + \delta_2^2]$, assuming the base population is in gametic-phase equilibrium. Thus, dominance is a factor in both heterosis and inbreeding depression, but it is not a necessary condition for heterosis, which can arise entirely as a function of additive $\times$ additive epistasis (α_2^c). In addition, inbreeding depression, but not F_1 heterosis, is a function of additive $\times$ dominance and dominance $\times$ dominance interactions. These differences in the genetic underpinnings of heterosis and inbreeding depression are a consequence of the extreme degree of gametic-phase disequilibrium that exists in the first generation of a line cross. Further complexities arise when the F_1 progeny of a line cross are subsequently selfed (Lynch 1991). As will be examined in Chapter 13, the presence of additive epistasis (terms involving A, but not D) can generate **outbreeding depression**, the opposite of heterosis, wherein the F_1 has *reduced* fitness relative to the parental lines.

Example 12.2. Some insight into the impact of epistasis on inbreeding depression was offered by the two-locus analysis of Curik et al. (2001). They assumed that the two loci were unlinked and in linkage equilibrium. They further assumed *identity equilibrium*: locus one being inbred in an individual does not influence the probability that locus two is also inbred—namely, if f is the probability of inbreeding at the first locus, then f^2 is the probability of inbreeding at both loci (more this concept later in the chapter). They coded the two-locus genotypic values as follows:

	BB	*Bb*	*bb*
CC	$a_1 + a_2 + aa_{12}$	$a_1 + d_2 + ad_{12}$	$a_1 - a_2 - aa_{12}$
Cc	$d_1 + a_2 + da_{12}$	$d_1 + d_2 + dd_{12}$	$d_1 - a_2 - da_{12}$
cc	$-a_1 + a_2 - aa_{12}$	$-a_1 + d_2 - ad_{12}$	$-a_1 - a_2 + aa_{12}$

With this coding, the change in the mean under inbreeding becomes

$$\mu_f - \mu_0 = -f\delta_1 - f(\alpha\delta) - 2f\delta_2 + f^2\delta_2$$

Letting p_1 denote the frequency of allele *B* and p_2 the frequency of allele *C*, the composite terms become

$$\begin{aligned}
\delta_1 &= 2\,[p_1(1-p_1)d_1 + p_2(1-p_2)d_2] \\
(\alpha\delta) &= 2\,[(2p_2-1)p_1(1-p_1) + (2p_1-1)p_2(1-p_2)]\,da_{12} \\
\delta_2 &= 4\,[p_1p_2(1-p_1)(1-p_2)]\,dd_{12}
\end{aligned}$$

As an example of how these were obtained, consider δ_2, a function of dd_{12}, the latter term only appears when the genotype *BbCc* is present. In inbreds, its appearance requires that both loci are noninbred, which occurs with probability $(1-f)(1-f)$. Hence, the decrease in the mean of an inbred due to dd_{12} is $[2p_1(1-p_1)(1-f)]\,[2p_2(1-p_2)(1-f)]$.

An important item to note is that the additive $\times$ dominance term, ($\alpha\delta$), can ***change sign with allele frequencies***, as $(2p-1)$ can change sign (as p_i moves from below 0.5 to above 0.5). Hence, as noted by Curik et al. (2001), one could formally assert that apparent changes in the amount of inbreeding depression over time (purging, see below) may not be due to the purging of deleterious alleles per se, but rather due to allele-frequency change that changes the sign of the composite ($\alpha\delta$) term.

METHODOLOGICAL CONSIDERATIONS: CONTROLLED CROSSES

A number of difficulties arise in attempts to test for inbreeding depression. Some of these are associated with the selective consequences of the inbreeding depression itself. In humans, for example, there is evidence that consanguineous couples, whose early offspring die from the expression of lethal recessives, compensate by reproducing until viable replacements have been born (Schull and Neel 1972). Some plants may behave in a similar manner by selective abortion of embryos (Willson and Burley 1983). Lack of knowledge of such compensation can lead to underestimates of the deleterious consequences of inbreeding. Keeping these difficulties in mind, we will now consider the statistical aspects of two common approaches to quantifying inbreeding depression using controlled crosses. These matters are taken up in more detail in Lynch (1988a). We examine the analysis of inbreeding depression in natural populations (based on either pedigree, or marker-based, estimates of f, and hence not requiring controlled crosses) later in the chapter.

Before proceeding, a brief introduction to the temporal dynamics of the inbreeding coefficient f under regular systems of mating is necessary. The general theory is covered elsewhere (Crow and Kimura 1970; Hartl and Clark 2006; WL Chapters 3 and 23), and because the vast majority of studies on inbreeding depression involve either self-fertilization or full-sib mating, we simply give the results for these special cases. Starting from a random-mating base population at time 0, the average inbreeding coefficient at a locus after t generations of self-fertilization in the absence of selection is

$$f(t) = 1 - \left(\frac{1}{2}\right)^t \tag{12.5a}$$

The quantity $[1-f(t)] = (1/2)^t$ is equivalent to the fraction of the heterozygosity in the base population that is still present after t generations of selfing. Thus, a single generation of selfing reduces the number of heterozygous loci within individuals by 50%. Thereafter, the heterozygosity declines geometrically towards zero, such that only 1.6% of the original heterozygosity remains after six generations of selfing. With full-sib mated lines, the inbreeding coefficient must be computed with the recurrence equation,

$$f(t) = \frac{f(t-1)}{2} + \frac{f(t-2)}{4} + \frac{1}{4} \tag{12.5b}$$

letting $f(-1) = f(0) = 0$. Thus, under full-sib mating, the first generation of inbred progeny has $f(1) = 0.25$, i.e, the heterozygosity within individuals is reduced by 25% relative to that in the random-mating base population. The inbreeding coefficient then progressively approaches one, although more slowly than in the case of selfing. More generally, as developed in Chapter 7, the expected level of inbreeding (f_i) for individual i with parents j and k in a pedigree can be expressed in terms of the coefficients of coancestry, with

$$f_i = \theta_{jk} \quad \text{and} \quad f_i = 2\theta_{ii} - 1 \tag{12.5c}$$

Single-generation Analysis

A common short-term test of inbreeding depression involves the comparison of the mean phenotypes of offspring from random matings with those from a specific class of consanguineous mating. In any such analysis, both types of individuals should be raised simultaneously in a random design to eliminate the possibility that the differences in means are a product of the environment (Appendix 9). Ideally, the offspring of both types of matings should be derived from several mothers to minimize the importance of maternal effects, and all mothers certainly should be derived from the same base population.

With such an experimental design, an approximate t test can be constructed for the null hypothesis of no inbreeding depression. Here we take the null model to be one of purely additive gene action. Consider the situation in which n progeny are assayed from each of L independent families, both in the control and in the inbred population. Under the null hypothesis of no inbreeding depression, the difference between the observed mean phenotypes of noninbred and inbred offspring ($\Delta\bar{z} = \bar{z}_O - \bar{z}_I$) has expectation zero, and the observed difference must be evaluated against its sampling variance, the sum of the variances of $\bar{z}_O$ and $\bar{z}_I$. Several factors contribute to this variance, as can be seen by referring to the definition of the sample mean

$$\bar{z} = \frac{1}{Ln}\sum_{i=1}^{L}\sum_{j=1}^{n}(A_{i\cdot} + a_{ij} + E_{i\cdot} + e_{ij})$$

where $A_{i\cdot}$ is the mean genotypic value associated with the *i*th family, a_{ij} is the deviation from that value for the *j*th member of the family, $E_{i\cdot}$ is the maternal (or, more generally, common-family) effect associated with the *i*th family, and e_{ij} is the residual environmental effect on the *ij*th individual.

We start by considering the control. First, the variance in the control mean caused by environmental effects specific to individuals is $\sigma_e^2/(Ln)$ because all such effects are distributed independently among the Ln individuals. Second, the variance caused by maternal (or general environmental, i.e., common-family) effects is σ_E^2/L; it is only divided by L because L mothers (families) contribute to the control mean. Third, because the segregational variance within families in the random-mating base population is $\sigma^2(a_{ij}) = \sigma_A^2/2$, and the effects of such residual variation are random with respect to individuals, the contribution of the within-family genetic variance to the variance of the mean is $\sigma_A^2/(2Ln)$. Finally, the among-family variance, $\sigma^2(A_{i\cdot})$, is also $\sigma_A^2/2$, and it contributes $\sigma_A^2/(2L)$ to the sampling variance of the control mean. Summing up terms, the variance of the control (outbred) mean phenotype becomes

$$\sigma^2(\bar{z}_O) = \frac{1}{L}\left[\frac{1}{2}\left(1 + \frac{1}{n}\right)\sigma_A^2 + \sigma_E^2 + \frac{\sigma_e^2}{n}\right] \tag{12.6a}$$

Now consider the situation for a sample of progeny derived by selfing L mothers with n progeny sampled per mother. The sampling variance of the mean resulting from general (common-family) and specific environmental effects is exactly the same as in the control. In the first generation of selfing, the *within-family* segregational variance is also identical to that within the control, $\sigma_A^2/2$. However, because of inbreeding, the variance *among families* is σ_A^2, twice that in the control families, where σ_A^2 is still defined as the genetic variance in the base (control) population. Thus, the expected variance of the mean of the sample of selfed progeny is

$$\sigma^2(\bar{z}_S) = \frac{1}{L}\left[\left(1 + \frac{1}{2n}\right)\sigma_A^2 + \sigma_E^2 + \frac{\sigma_e^2}{n}\right] \tag{12.6b}$$

The situation for full-sib mating is a little more complicated, but assuming that all progeny within lines are derived from a single brother-sister mating,

$$\sigma^2(\bar{z}_{FS}) = \frac{1}{L}\left[\left(\frac{7}{8} + \frac{3}{8n}\right)\sigma_A^2 + \sigma_E^2 + \frac{\sigma_e^2}{n}\right] \tag{12.6c}$$

The above expressions depart slightly from those in Lynch (1988a) and appear to be more accurate. WL Chapter 23 gives general expressions for different levels of inbreeding under a number of different family structures.

Although these formulae give exact expectations of the variances of control and inbred line means under the additive model, they are difficult to implement unless one has prior information on the components of variance in the base population. However, the basic structures of the formulae yield a very useful result. Note that for the cases of selfing and sib-mating, $\sigma^2(\Delta\bar{z}_S) = \sigma^2(\bar{z}_o) + \sigma^2(\bar{z}_S)$ and $\sigma^2(\Delta\bar{z}_{FS}) = \sigma^2(\bar{z}_o) + \sigma^2(\bar{z}_{FS})$, respectively. In both of these cases, provided the sample size within families (n) is at least two, then $\sigma^2(\Delta\bar{z}) \leq 2\sigma^2(z_O)/L$, where $\sigma^2(z_O)$ is the phenotypic variance within the control line. Thus, a conservative test for inbreeding depression based on a single generation of consanguineous mating employs the test statistic

$$\tau = \frac{|\Delta\bar{z}|}{\mathrm{SD}(z_O)\sqrt{2/L}} \tag{12.7}$$

where $\mathrm{SD}(z_O)$ is the observed phenotypic standard deviation in the random-mating population. Sampling distributions of means are usually approximately normally distributed, so τ may be treated as t-distributed with $L - 1$ degrees of freedom.

In the case of self-compatible plants that produce multiple flowers, there is a simple way to further increase the power of a test of inbreeding depression. For any pair of parent plants (A and B), both reciprocal outcrosses (A $\times$ B and B $\times$ A) and two inbreds (A $\times$ A and B $\times$ B) can be produced. Because the two parents contribute equal numbers of genes to both inbred and outbred progeny, variance from general (maternal) environmental effects and parent sampling do not contribute to $\sigma^2(\Delta\bar{z})$ in this case, and for any pair of parent individuals, the test statistic

$$\Delta\bar{z}_{A,B} = \frac{(\bar{z}_{AA} + \bar{z}_{BB}) - (\bar{z}_{AB} + \bar{z}_{BA})}{2} \tag{12.8a}$$

has an expected value equal to zero under the null hypothesis of no inbreeding depression. If n replicates are assayed within each of the four groups of progeny, the expected sampling variance of $\Delta\bar{z}_{A,B}$ is

$$\sigma^2(\Delta\bar{z}_{A,B}) = \frac{(\sigma_A^2/2) + \sigma_e^2}{n} \tag{12.8b}$$

Because the numerator of this expression is less than $\sigma^2(z_O)$, a conservative test for inbreeding depression associated with any pair of parents is provided by

$$\tau = \frac{|\Delta\bar{z}_{A,B}|}{\mathrm{SD}(z_O)/\sqrt{n}} \tag{12.9}$$

where $\mathrm{SD}(z_O)$ is again the phenotypic standard deviation of outcrossed individuals.

Multigenerational Analysis

A common long-term approach to quantifying inbreeding depression on a trait is to regress the mean phenotype on the inbreeding coefficient (Figure 12.2). In studies of this sort, the data usually represent progressively inbred generations derived from the same population, a protocol that introduces a series of statistical problems. First, when the different classes of inbreeding are assayed in different generations (the usual case in animals), the possibility arises that any trend in the mean may be caused by a shift in the environment. An important example of this may be improved husbandry as knowledge about a captive species accrues, resulting in a captive environment that becomes more favorable over time (Hedrick and Kalinowski 2000). Similarly, a newly formed population may show adaptation to captivity, with its fitness increasing as generations of captivity proceed.

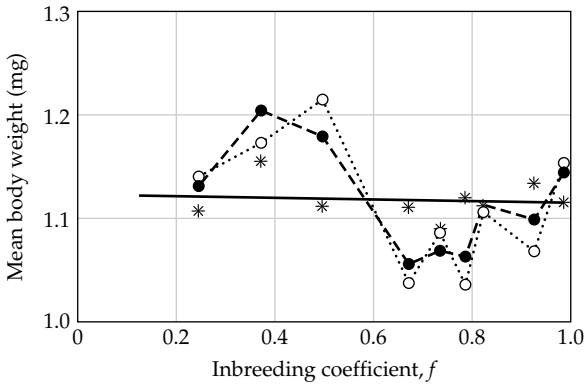

Figure 12.4 The observed trend for mean body weight in inbred (●) and control (○) lines of *Drosophila melanogaster* confounds inbreeding and generational (environmental) trends. The trend from the former can be estimated by use of Equation 12.12, with corrected values given by (∗). Their partial regression is represented by the solid line. (Data from Kidwell and Kidwell 1966.)

Second, the usual assumptions underlying hypothesis testing in (OLS) regression theory are violated in at least two ways: (1) because the means are based upon individuals that are descendants of each other, the data are not independent, and (2) the sampling variance of the means varies with f because of the loss of genetic variance with inbreeding. The problem of nonindependence of data is a particularly serious one, because it diminishes the effective degrees of freedom in an analysis. For example, once a population of selfers is almost completely inbred, the subsequent generations are no longer free to vary genetically (except by mutation). Nonindependence of data can also cause spurious nonlinearities in the apparent response to inbreeding—if one data point lies above the regression line, the preceding and subsequent ones are likely to do so as well. The result is that regression residuals are likely both correlated and heteroscedastic, requiring a GLS, rather than OLS, regression framework (Chapter 10; WL Chapter 18).

A partial resolution of the nonindependence problem is given below, and the complete solution is outlined in Example 12.4. First, however, some attention needs to be given to the correction of data for environmental shifts between generations. Plant breeders have been able to avoid this issue by storing seed from progressive generations of inbreeding and then growing representatives of all generations simultaneously in a randomized design (Russell et al. 1963; Hallauer and Sears 1973; Cornelius and Dudley 1974). Even here, it is assumed implicitly that seed storage time does not influence performance and that general environmental effects experienced by the parents are not transmitted to the progeny. With most animals, embryo storage is either not currently reliable or not (generally) economically feasible, so there is need for a statistical means of correcting the data. The question of interest is whether a trend in the inbred-line means is influenced by a temporal shift in the environment. The issue is not trivial as can be seen from the striking parallel directional trend in control and full-sib mated lines of *Drosophila melanogaster* shown in Figure 12.4.

A solution to the environmental trend problem was proposed by Muir (1986a, 1986b), who suggested the use of a **parallel control** as a means of assaying changes in the environment. There are two important considerations in the choice of a control. First, it is essential that the temporal phenotypic changes in the control are entirely attributable to environmental causes. This condition will essentially hold if clones or highly inbred lines are used. A random-bred base population may also serve as an adequate control, provided the character of interest is not modified by selection during the course of the experiment and that the population is large enough that significant genetic drift is unlikely to occur. Second, given a choice of control lines, the one that provides the strongest signal of the environment, i.e., explains a maximum amount of the variance in the inbred line means, is most desirable.

Finally, the assumption is no (or very little) genotype $\times$ environment interaction, so that the control and inbred lines equally (additively) react to the environment. WL Chapter 18 examines the use of control populations in more detail.

We start by considering the mean phenotypes of the control (C) and inbred (I) lines at generation t to be functions of general environmental effects common to both of them (E), special environmental effects unique to each of them (e_C and e_I), and genetic change confined to the inbred population, $\Delta\mu_G(t)$,

$$\bar{z}_I(t) = \mu_I(0) + E(t) + e_I(t) + \Delta\mu_G(t) \tag{12.10a}$$

$$\bar{z}_C(t) = \mu_C(0) + E(t) + e_C(t) \tag{12.10b}$$

Because the general environmental effects, $E(t)$, are the only common components of the inbred and control line means, a partial regression of the observed inbred line means, $\bar{z}_I(t)$, on the observed control line means, $\bar{z}_C(t)$, and the inbreeding coefficient, $f(t)$, provides a way of factoring out any general trend of the environment,

$$\bar{z}_I(t) = a + b\bar{z}_C(t) + If(t) + e(t) \tag{12.11}$$

where I is the estimated inbreeding depression (i.e., the expected difference in mean phenotypes of noninbred and completely inbred individuals; Equation 12.3b), and $e(t)$ is the deviation of the mean at generation t from its regression prediction. Applying Equation 12.11, after removing any environmental trend, the corrected means for the inbred lines become

$$\bar{z}^*_I(t) = \bar{z}_I(t) - b[\,\bar{z}_C(t) - \bar{z}_C\,] \tag{12.12}$$

where $\bar{z}_C$ is the mean phenotype of the control lines over all generations. The estimated inbreeding depression (I) is equivalent to the regression of the $\bar{z}^*_I(t)$ on $f(t)$. Figure 12.4 shows a rather striking example of how the application of Muir's approach can overcome a trend obscured by environmental factors.

Finally, we return to the problem of hypothesis testing, assuming that the means have been corrected adequately for general environmental trends prior to analysis. Because it ignores the nonindependence of data, ordinary least-squares (OLS) regression of $\bar{z}^*_I(t)$ on $f(t)$ leads to downwardly biased estimates of the standard error of I, often by a factor of three or four (Lynch 1988a; WL Chapter 18). An expression for the sampling variance of I, which fully accounts for the correlational structure of the data, under the null hypothesis of a neutral character with an additive genetic basis is worked out in Lynch (1988a), and the mixed model solution is outlined in Example 12.4. Lynch's solution, portrayed graphically in Figure 12.5, assumes that the regression is performed on a progressive series of inbred lines (e.g., self-fertilization, full-sib mating, or first-cousin mating), starting with $f = 0$ and proceeding for k generations to a final level of inbreeding of $f(k)$. The plotted values are *minimum estimates* of the sampling variance of I, because it is assumed that the variance in the environment makes no contribution to the sampling variance of the means. The sampling variance of I depends primarily on the additive genetic variance in the base population, the number of inbred families, and the level of inbreeding in the final generation. Because the sampling variance of I declines with increasing L (number of lines) and increasing $f(k)$, it is clear that for a fixed amount of resources, the smallest unit of inbreeding (selfing or full-sib mating) should be employed while maximizing the number of lines.

Many of the preceding statistical problems can be avoided when data are available for contemporaneous (and hence presumably experiencing very similar environments) groups of individuals inbred to various degrees. Such is typically the case in the analysis of human and domesticated animal populations where pedigrees are known, and the same can be accomplished in experiments that simultaneously mate various classes of relatives and assay their progeny in a common environment. In both cases, provided the individuals with different levels of f are unrelated, the problem of nonindependent data is eliminated, and provided all individuals are assayed contemporaneously, the need for a temporal control

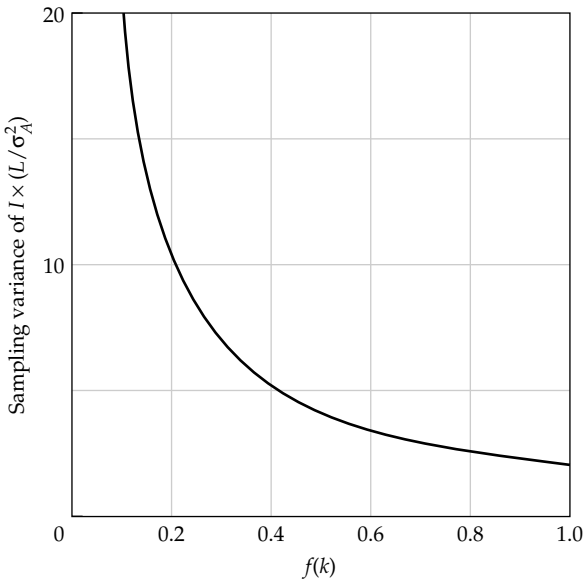

Figure 12.5 The minimum sampling variance of the regression coefficient of consecutive line means on their respective inbreeding coefficients, under the assumption of purely additive gene action and ignoring environmental effects. $f(k)$ is the inbreeding coefficient in the final generation. To obtain the actual sampling variance of I, the points on the ordinate must be multiplied by σ_A^2/L, the ratio of the additive genetic variance in the base population to the number of inbred lines. (From Lynch 1988a.)

is removed. Ordinary least-squares regression of the group mean phenotypes on f then provides a simple approximation of I and its standard error. Such a regression is only an approximation, as it has heteroscedastic residuals because the sample variance changes with the level of inbreeding. The delicate issue in natural settings is estimating the level of inbreeding among individuals in the population (as opposed to a pedigree) sample. We examine marker-based approaches for estimating f values later in the chapter.

Example 12.3. Consider an experimental design in which the means of $L = 10$ full-sib mated lines are assayed from generation 0 with $f = 0$ to generation 9 with $f(k) = 0.859$ (obtained using Equation 12.5b). Reading off Figure 12.5 at $f(k) = 0.859$, we find the point on the ordinate to be 2.5. The expected sampling variance of the slope I under the null model of no inbreeding depression is obtained by multiplying 2.5 by σ_A^2/L, which gives $\sigma_A^2/4$. If an estimate of σ_A^2 is available, then in this case, two standard errors of the slope is estimated by $\sqrt{\text{Var}(A)}$. Because our treatment ignores environmental sources of variance, it is clear that with this design any regression coefficient whose absolute value is less than the square root of the additive genetic variance in the base population must be considered consistent with the null model of no inbreeding depression.

Example 12.4. One can formally accommodate the complicated residual structure associated with inbreeding experiments by using very minor modifications of the methods for GLS and BLUP analyses of selection experiments developed in WL Chapters 18 and 19. We first sketch the simplest analysis under GLS (Chapter 10). The basic model can be written as

$$R_I(t) = \mu_t - \mu_0 = If(t) + e(t) \tag{12.13a}$$

where $R_I(T)$ denotes the cumulative change in the mean from inbreeding depression. Note that this is akin to the cumulative response to selection that was assumed in WL Chapter 18.

To apply a GLS regression framework, the residual error structure is required. This is given by the matrix $\mathbf{V}$, whose elements are given by WL Equation 18.15c, namely

$$V_{ij} = \begin{cases} \left(\frac{1}{M_0} + 2f_i\right) h^2\sigma_z^2 + \sigma_z^2/M_i & \text{for } i = j \\ \left(\frac{1}{M_0} + 2f_i\right) h^2\sigma_z^2 & \text{for } i < j \end{cases} \tag{12.13b}$$

where M_i individuals are scored in generation i and σ_z^2 is the inital phenotypic variance. Let $\mathbf{R}$ denote the observed vector of cumulative changes in the mean and $\mathbf{F}$ denote the vector of cumulative levels of inbreeding, whose ith elements are, respectively, $R_I(i)$ and $f(i)$. The resulting GLS estimator for the amount, I, of inbreeding depression is given by

$$\widehat{I}(GLS) = (\mathbf{F}^T\mathbf{V}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{V}^{-1}\mathbf{R} \tag{12.13c}$$

whose resulting sampling variance is given by

$$\text{Var}\left[\widehat{I}(\text{GLS})\right] = (\mathbf{F}^T\mathbf{V}^{-1}\mathbf{F})^{-1} \tag{12.13d}$$

WL Example 18.3 provides a worked example (for a selection response experiment), and WL Chapter 18 discusses the incorporation of control populations and sample variances for more complicated designs.

One can test for the presence of epistasis by using a quadratic regression

$$R_I(t) = \mu_t - \mu_0 = If(t) + \beta[f(t)]^2 + e(t) \tag{12.13e}$$

with the same error structure as given by Equation 12.13b. The vector of the GLS estimates for I and β follows from Equation 12.13c with $\mathbf{F}$ now being an $n \times 2$ matrix, whose ith row is given by $[f(i), f^2(i)]$. Epistasis is indicated when β is significantly different from zero.

BLUP machinery (Chapters 10 and 31) can also be used to estimate I. This approach requires significantly more information than a simple GLS estimate, as we need the pedigree (or at least relationship estimates) for every *measured individual* in the sample (rather than just the *generation means* that are required for a GLS regression). The only modification required to the approach given in WL Chapter 19 (for the BLUP analysis of selection experiment) is that the observation for individual i now becomes

$$y_i = \mu + A_i + If_i + e_i \tag{12.13f}$$

(WL Equation 23.2c), where A_i denotes the breeding value for individual i, and f_i is its inbreeding level. Hence, one simply adds an additional fixed effect, I, to the model and then estimates and tests its value (like any standard fixed effect). The relatedness and inbreeding among all the measured relatives is accommodated through the relationship matrix, $\mathbf{A}$. Misztal et al. (1997) discusses how to incorporate dominance (beyond I) into the model.

Ritland's Method for Partial Selfers

Most empirical attempts to measure inbreeding depression on *fitness* involve assays of individuals in controlled environments. Because lab conditions often deviate substantially from the situation in nature, one is then left wondering how generalizable the results are to field settings. To eliminate this problem, Ritland (1990a, 1990b) proposed a technique for partially selfing populations of plants that involves essentially no disturbance of individuals in nature and requires no direct estimation of individual fitness. Applying neutral molecular markers (Chapter 8) to progeny arrays, it is possible to estimate the fraction of seed that adults produce by self-fertilization (η), as well as to survey the change in genotype frequencies in a population within and between generations. Genotype frequency change across generations are a function of the degree of selfing in the parents, while the within-generation changes are a function of genotype-specific fitnesses.

Ritland (1990a, 1990b) suggested several ways in which marker information can be exploited to infer indirectly the fitness consequences of inbreeding. Here, we simply point out the simplest situation, which arises when a population has attained an equilibrium state of inbreeding, i.e., a balance between the production of excess homozygosity by selfing and its loss by selection. The inbreeding coefficient (f) of surviving individuals can be computed using the marker-based approaches discussed in Chapter 8 and also later in this chapter. The ratio of fitnesses of selfed to outcrossed individuals is then estimated by

$$w = \frac{2(1-\eta)f}{\eta(1-f)} \tag{12.14}$$

Note that when $\eta > 0$ and $f = 0$ (i.e., the adult population is in Hardy-Weinberg equilibrium), $w = 0$, implying that selfed progeny have zero fitness. More general estimators that allow for generational changes in η and f, which appear to be common (Dole and Ritland 1993), are provided in Ritland (1990a, 1990b).

Applications of Equation 12.14 to partially selfing plant populations have generally yielded estimates of w that are slightly lower than those obtained by direct observations of the performance of selfed and outcrossed progeny in experimental populations (Eckert and Barrett 1994; Kohn and Biardi 1995; Schultz and Ganders 1996). Although violations in the assumptions of Ritland's model (such as an absence of biparental inbreeding and an absence of linkage between marker and fitness loci) can lead to biased estimates of w, the bias does not generally appear to be large. Thus, the empirical results tentatively suggest that the inbreeding depression observed in manipulated populations may generally be lower than that expressed in natural settings. This difference may occur because manipulative studies often fail to fully account for all components of fitness (such as seedling survival) or because the effects of deleterious genes are ameliorated in more benign environments (discussed below).

Epistasis and Inbreeding Depression

An unresolved issue is the extent to which epistasis is involved in inbreeding depression. Recall (Equation 12.4b) that a nonlinear relationship between the mean phenotype and the inbreeding coefficient is an indicator of the presence of epistasis involving dominance effects, which can be tested by using quadratic regressions (Example 12.4). As stressed above, this approach can be biased if lines are lost, or altered, by selection during inbreeding.

A simple way to test for nonlinearity, which avoids the pitfalls of regression, is to compare the change in mean phenotype (per increment in f) between two low levels of f and two high levels of f. Provided the two ranges of f are nonoverlapping, the two observed *changes* are statistically independent, even if the individuals at all four points in time are related. Letting the four observed mean phenotypes, in order of increasing f, be $\bar{z}_1$, $\bar{z}_2$, $\bar{z}_3$, and $\bar{z}_4$, a measure of nonlinearity is then given by

$$\Delta I = \frac{\bar{z}_2 - \bar{z}_1}{\Delta f_L} - \frac{\bar{z}_4 - \bar{z}_3}{\Delta f_H} \tag{12.15a}$$

where $\Delta f_L = f_2 - f_1$, and $\Delta f_H = f_4 - f_3$. Under a linear response, both of the ratios in Equation 12.15a should have the same expected value, but should differ if response is nonlinear. A conservative estimate of the sampling variance of ΔI is given by

$$\text{Var}(\Delta I) = \frac{[\text{SE}(\bar{z}_2)]^2 + [\text{SE}(\bar{z}_1)]^2}{(\Delta f_L)^2} + \frac{[\text{SE}(\bar{z}_4)]^2 + [\text{SE}(\bar{z}_3)]^2}{(\Delta f_H)^2} \tag{12.15b}$$

A test statistic for nonlinearity is then

$$\tau = \frac{|\Delta I|}{\sqrt{\text{Var}(\Delta I)}} \tag{12.15c}$$

which under the null hypothesis of linearity should be *t*-distributed with degrees of freedom equal to the number of inbred lines in the analysis. To ensure that ΔI is not a function of the differential extinction of lines, only the lines surviving to contribute to $\bar{z}_4$ should be used in such an analysis. Domíngues-García et al. (2019) used this approach to show an acceleration of ID with higher f values for pupal number in *Drosophila*. Willis (1993) used a very similar approach to test for epistasis for life-history characters in the monkey flower (*Mimulus guttatus*). Although he did not correct for line loss, he found very little evidence for epistasis. Again, see Curik et al. (2001) and Domíngues-García et al. (2019) for additional discussion of detection of epistasis by regression on f.

Variance in Inbreeding Depression

Evolutionary biologists interested in the origins of diverse mating systems, particularly in plants, have reason to be concerned with the potential for variance in inbreeding depression among members of the same population (Holsinger 1988; Johnston and Schoen 1994; Uyenoyama et al. 1994; Schultz and Willis 1995). Such variation would seem to be necessary to foster the evolution of alternative forms of mating (WL Chapter 23). To assay whether such variation exists, plant population biologists often use ratios of fitness of selfed progeny to outcrossed progeny as a measure of inbreeding depression. This practice raises some statistical problems in that the ratio of expected values is a biased estimate of the expected value of a ratio (Appendix 1). Some of the issues are discussed by Johnston and Schoen (1994), but some of the formulae in their paper are incorrect. Using an improved approximation for the expected value of a ratio (Equation A1.19a), an unbiased estimate of the performance of selfed relative to outcrossed progeny is given by

$$w_i = \frac{\overline{W}_{Si}}{\overline{W}_{Oi}} \left[\frac{1}{1 + [\sigma^2(W_{Oi})/(n_i \overline{W}_{Oi}^2)]} \right] \tag{12.16}$$

where $\overline{W}_{Si}$ and $\overline{W}_{Oi}$ are the observed mean fitnesses of selfed and outcrossed progeny derived from individual i, $\sigma^2(W_{Oi})$ is the variance in fitness of outcrossed progeny, and n_i is the number of outcrossed progeny assayed, all for the *i*th individual. Using the variance in w_i estimates overestimates the variance in inbreeding depression, as the former includes the sampling (error) variance. ANOVA approaches can be used to help separate error variance from the variance in inbreeding depression (Pray and Goodnight 1995), and Moorad and Wade (2005) have suggested specific mating designs to accomplish this goal.

Using measures such as Equation 12.16 to quantify variance in inbreeding depression among individuals raises a number of difficult and unresolved issues. A central problem is that inbreeding depression is not just a property of the *individual*, but of the individual's *prospective mates* as well. It is straightforward enough to estimate an individual's fitness through selfing, but what about the situation with species with separate sexes? An individual's sibs will generally differ with respect to fitness, so the fitness of progeny from full-sib matings will depend on which sibs are employed as mates. The situation is even more extreme when one considers the fitness of individuals produced through outcrossing. Ideally, one would like an estimate of the fitness of outcrossed progeny averaged over all potential mates, but with most species (other than plants), only a small number of matings per individual are possible.

THE EVIDENCE FOR INBREEDING DEPRESSION

Although few of the existing studies of inbreeding depression have fully accounted for all of the difficulties pointed out above, the aggregate of evidence for inbreeding depression is overwhelming. While substantial variation of inbreeding depression exists among species, among traits within species, and even between the sexes (Ebel and Phillips 2016; Clark et al. 2019), almost all organisms exhibit it to some degree. Here we only summarize some of the better-documented cases. An extensive survey of the early literature is available in

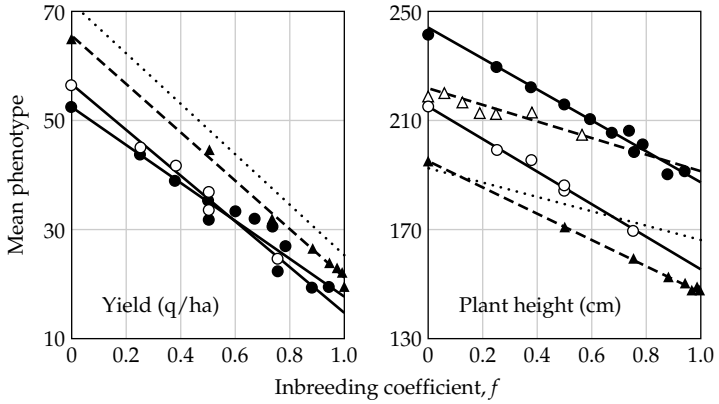

Figure 12.6 The response of mean grain yield and plant height to inbreeding in maize. Data are from: (●, ○) Cornelius and Dudley (1974); (closed triangles) Hallauer and Sears (1973); (Δ) Sing et al. (1967); (· · ·; only the regression line is available) Good and Hallauer (1977). The two studies of Cornelius and Dudley are for the same lines grown in different years. The variation in intercepts is presumably due to differences among base populations as well as among environments in which the experiments were performed.

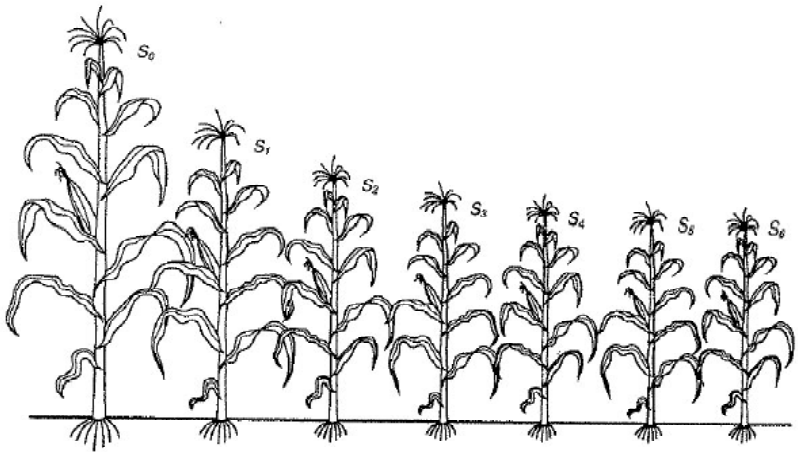

Figure 12.7 Impact of inbreeding on maize height. S_i denotes the number of generations of selfing. S_0 is the hybrid between two inbred lines, and may (depending on the cross) show heterosis, while S_1 is their selfed offspring, in which half the heterotic advantage is removed (Chapter 13). The decline under further generations of selfing reflects inbreeding depression.

Wright (1978), and recent reviews are given by Shields (1982), Charlesworth and Charlesworth (1987), Thornhill (1993), Husband and Schemske (1996), Crnokrak and Roff (1999), Keller and Waller (2002), and Leroy (2014).

More and better data on the phenotypic consequences of inbreeding are available for maize than for any other organism. Hallauer and Miranda (1981) and Hallauer et al. (2010) review the evidence, which was recognized as early as 1876 by Darwin. There have been some very well conceived experiments involving prolonged selfing and full-sib mating in lines derived from a genetically diverse base population (Sing et al. 1967; Hallauer and Sears 1973; Cornelius and Dudley 1974; Good and Hallauer 1977; Lamkey and Smith 1987; Benson and Hallauer 1994). The experiments are very large (involving up to 250 independent lines), and the potential influence of temporal changes in the environment has been minimized by the simultaneous analysis of stored seed. Almost without exception, vegetative, reproductive,

Table 12.2 A survey of the (scaled) inbreeding depression observed in laboratory populations of *Drosophila*. Here, $I_s = (\bar{z}_O - \bar{z}_I)/\bar{z}_O = 1 - (\bar{z}_I/\bar{z}_O)$, where $\bar{z}_O$ and $\bar{z}_I$ are, respectively, the means of the random-mating base and the completely inbred population (obtained by linear extrapolation). This scaled change (reduction as a fraction of the original mean) induced by inbreeding allows us to compare its impact over traits. Results marked with an asterisk were obtained from studies involving only one or two chromosomes; in these cases, extrapolation to the entire genome was done by assuming that each major chromosome arm constitutes 20% of the genome, and that the effects are multiplicative across chromosomes. Negative values imply an increase in character value with inbreeding. All data are for *D. melanogaster,* except for *D. subobscura* (Hollingsworth and Maynard Smith 1955), *D. pseudoobscura* (Dobzhansky and Spassky 1963; Dobzhansky et al. 1963; Marinkovic 1967), and *D. willistoni* (Malogolowkin-Cohen et al. 1964).

Character	I_s	Reference
Competitive ability	0.84	Latter et al. 1995
	0.97	Latter and Sved 1994
Egg-to-adult viability	0.57	Garcia et al. 1994
	0.44	Mackay 1985a
	0.66*	Malogolowkin-Cohen et al. 1964
	0.48*	Dobzhansky et al. 1963
	0.06	Tantaway and Reeve 1956
Female fertility	0.81	Mackay 1985a
	0.18	Tantaway and Reeve 1956
	0.35	Hollingsworth and Maynard Smith 1955
Female rate of reproduction	0.32	Latter et al. 1995
	0.56	Mackay 1985a
	0.96	Hollingsworth and Maynard Smith 1955
	0.57	Marinkovic 1967
Male mating ability	0.52*	Hughes 1995
	0.92	Partridge et al. 1985
	0.76	Sharp 1984
Male longevity	0.18*	Hughes 1995
Male fertility	0.00*	Hughes 1995
	0.22*	Dobzhansky and Spassky 1963
Male weight	0.07*	Hughes 1995
	0.10	Mackay 1985a
Female weight	−0.10	Kidwell and Kidwell 1966
Abdominal bristle number	0.05	Mackay 1985a
	0.06	Kidwell and Kidwell 1966
	0.00	Rasmuson 1952
Sternopleural bristle number	−0.01	Mackay 1985a
	0.00	Rasmuson 1952
Wing length	0.03	Tantaway 1957
	0.01	Tantaway and Reeve 1956
Thorax length	0.02	Tantaway 1957

and physiological characters exhibit significant shifts in the mean phenotype with inbreeding. Cases have arisen in which the regressions of $\bar{z}$ on f appear to be nonlinear (Hallauer and Sears 1973; Good and Hallauer 1977), but in all cases the departure from linearity is small. Two characters that give no evidence of nonlinearity are total grain yield and plant height (Figures 12.6 and 12.7). Starting from a genetically diverse base population, complete inbreeding results in an approximately 65% decline in yield and an approximately 25% decline in plant height.

Several independent investigations of inbreeding depression have been performed

with laboratory stocks of *Drosophila*, a sample of which are listed in Table 12.2. Although substantial variation exists among the results from different studies, a pattern emerges. Primary fitness characters such as viability, fertility, and egg production tend to exhibit very high levels of inbreeding depression (averaging approximately a 50% reduction in the mean), while morphological characters (bristle numbers, body weight and length), which are perhaps more remotely related to fitness, change by only a few percent, if at all. The latter traits are known to exhibit substantial levels of additive genetic variance, while the fitness characters tend to have lower heritabilities (Mousseau and Roff 1987; LW Chapter 6). Thus, in *Drosophila* there appears to be a major difference in the way the genetic variance for morphological and fitness characters is partitioned: mostly additive for the former, and substantially dominant in the latter.

Selection theory helps explain why the additive genetic variance for fitness should be low and why dominance should be directional for fitness-related characters (WL Chapters 5–7, 24–28). Alleles with favorable effects on fitness should move rapidly towards fixation, regardless of their degree of dominance, and dominant alleles with deleterious effects will be eliminated rapidly. However, deleterious recessive alleles will be maintained at low frequencies by mutation pressure. For characters only weakly related to fitness or under stabilizing selection for an intermediate optimum, directional dominance may be less pronounced because mutations that cause a shift in the mean in either direction will be selectively equivalent.

Several large surveys provide firm empirical justification for the incest taboos that exist in many, but not all, populations of humans. Despite this potential for significant inbreeding effects, almost 11% of the global population consists of couples related as at least second cousins ($f \geq 0.0156$) and their progeny (Bittles and Black 2010). Rarely is it possible to obtain data for more extreme situations than first-cousin marriages, but by linear extrapolation the data are sufficient to demonstrate that more extreme inbreeding would lead to substantial depression in body size and IQ (Table 12.3). With the advent of molecular markers, more refined estimates of f can be generated for otherwise random human samples (see below). For example, in a British population of mainly European extraction, 0.03% ($\sim 1/3600$) had **extreme inbreeding** (genomes with 10% or greater of their total size being runs of homozygosity; see below).

The consequences of inbreeding for juvenile mortality and the incidence of congenital effects in human are well known and are examined in the next section from a somewhat different perspective. One precautionary note is in order here. Analyses based on inbreeding depression that rely on natural mating assemblages (as is always true in human studies) run the potential risk that progeny with different levels of f are products of genotypically different groups of parents. If parents that tend to inbreed also tend to be genetically low on the fitness scale, the apparent level of inbreeding depression in the progeny may be substantially exaggerated. As we detail shortly, molecular-marker based estimates of ID for traits weakly correlated with fitness are especially susceptible to **social confounding**, with some confounder influencing both increased levels of inbreeding and the trait of interest, creating an association between them.

Because domesticated animals tend to have higher levels of f (and greater variances), smaller sample sizes are required to detect significant ID effect in livestock relative to humans. Extensive reviews exist on the deleterious consequences of inbreeding in domesticated animals: beef cattle (Dinkel et al. 1968), dairy cattle (Turton 1981; Howard et al. 2017), dogs (Scott and Fuller 1965), horses (Cothran et al. 1986), sheep (Lamberson and Thomas 1984; Wiener et al. 1992a–1992c), swine (Dickerson et al. 1954; Bereskin et al. 1968), and salmonid fishes (Wang et al. 2002), while Leroy (2014) presented a meta-analysis of 57 studies covering seven livestock species. Leroy found that the average scaled change under inbreeding, I_s (Table 12.2; the fraction the reduction in terms of the initial mean) was 0.137, while the average when restricted to production traits was 0.351. Leroy argued that the stronger impact of inbreeding on production traits reflected their past history of selection (akin to life history traits typically showing larger inbreeding effects than other traits).

Table 12.3 The decrease (I) in the mean expected upon complete inbreeding in humans, equal to the mean for noninbred individuals minus the expectation at $f = 1$ (obtained by linear extrapolation). These I values are based on pedigree-based estimates of f. Marker-based estimates of f offer more power (in many settings) and are examined shortly.

Trait	Site	I	Reference
Birth weight (kg)	Japan	5.4*	Morton 1958
	United States	1.7	Slatis and Hoene 1961
Adult height (cm)	Hutterites, U. S.	56	Barrai et al. 1964
	Italy	3	Mange 1964
	Japan	20	Schull 1962
		21	Neel et al. 1970
IQ	Japan	42	Neel et al. 1970
		43	Kudo et al. 1972
		73	Schull and Neel 1965
	United States	42	Slatis and Hoene 1961
Prereproductive survival (%)	Global	70	Bittles and Neel 1994

*This value is obviously too high because it gives a birth weight less than zero.

A nonlinear response of the mean phenotype to the inbreeding coefficient has been seen in many of these studies. However, because few studies incorporate appropriate controls or account for the nonindependence of data in inbred lines, it is difficult to say whether the apparent nonlinearities in the data are simply statistical artifacts, as opposed to real reflections of directional epistasis. In organisms with extensive parental care, still another explanation exists. If *maternal performance* (the ability to raise young) is adversely affected by inbreeding, then an individual's phenotype will be influenced not only by its own level of inbreeding but also by that of its mother. As we will see in the next chapter, the same is also true with heterosis.

An elegant experiment performed with laboratory mice (White 1972) illustrated the impact of maternal inbreeding. When progeny with several different levels of inbreeding were crossfostered with mothers inbred to different degrees, maternal inbreeding was found to have approximately half the impact on juvenile weight as individual inbreeding (Figure 12.8). Similarly, Domíngues-García et al. (2019) also found that material inbreeding accounted for about 50% of the ID effect on *Drosophila* pupal number. Other experiments with mice have verified the effects of maternal inbreeding on progeny performance (Bowman and Falconer 1960; Falconer and Roberts 1960; McCarthy 1967; Nagai et al. 1971), and convincing but less extensive data exist for humans (Schull et al. 1970), sheep (Wiener et al. 1992a–1992c), and birds (Sittmann et al. 1966; van Noordwijk and Scharloo 1981). Recent marker-intensive studies have also found an impact of inbreeding on maternal performance in natural populations of Red deer (Huisman et al. 2016) and Soay sheep (Bérénos et al. 2016; Pemberton et al. 2017).

A brief aside is in order here on inbred mouse lines that are widely used in biomedical research. An important use of such lines is as a control over genetic heterogeneity in studies conducted in different labs. It must be stressed that most lines die out during the rapid inbreeding to form such lines. For example, 95% of all lines from the mouse collaborative cross mapping population (Chapter 18) became extinct, in large part due to male infertility (Shorter et al. 2017). Hence, the combinations of alleles that have survived (together) are a very non-random sample of alleles in the base population, so that the resulting Δ_i values (Equation 12.4) in the surviving inbred lines are likely different from their base-population value at the start of inbreeding.

As mentioned earlier, an important interpretational issue with the impact of inbreeding depression is that it can be environmentally dependent (Hedrick and Kalinowski 2000;

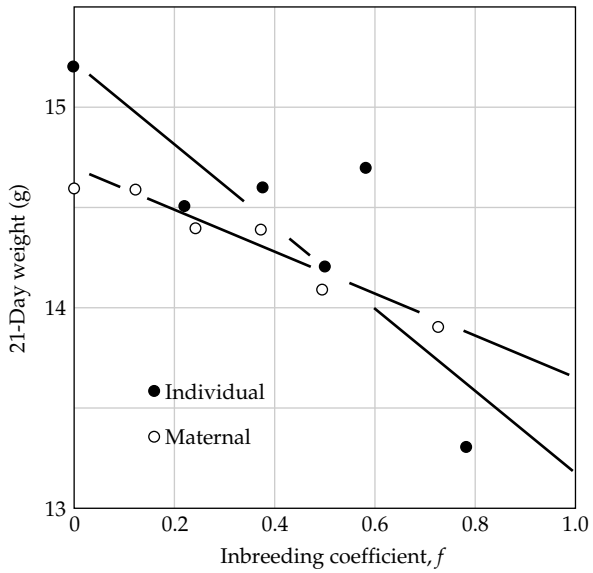

Figure 12.8 The decline in offspring size in laboratory mice as a function of individual and maternal inbreeding. The results of two experiments have been combined after adjusting for mean differences in the control lines.

Keller and Waller 2002; Cheptou and Donohue 2011; Pemberton et al. 2017), often with larger effects in more stressful environments (Armbrster and Reed 2005; Fox and Reed 2011). Hence, results from laboratory settings may undersell the importance of inbreeding depression in nature. Experiments with *Drosophila* support the contention that inbreeding depression is more severe in stressful environments (Hoffmann and Parsons 1991; Miller 1994), as do those with the flour beetle *Tribolium* (Pray et al. 1994) and with mice (Jiménez et al. 1994). On the other hand, in a very large study comparing 38 human populations, Bittles and Neel (1994) found that the effects of inbreeding on survival to age 10 were independent of the mortality rate of noninbred progeny, which ranged from 3 to 40%. Likewise, while numerous studies with plants have documented increased inbreeding depression under extreme conditions (Antonovics 1968; Schemske 1983; Dudash 1990; Schmitt and Ehrhardt 1990; Wolfe 1993), there were also many cases in which the influence of the environment was negligible (Johnston 1992; Heywood 1993; Charlesworth et al. 1994; Ouborg and Van Treuren 1994; Nason and Ellstrand 1995; Norman et al. 1995). While we know of no cases in which inbreeding depression is more intense in benign laboratory or greenhouse environments than in natural settings, Pemberton et al. (2017) suggest that some caution may be in order in assuming that the laboratory estimates tend to err on the conservative side with respect to the situation in nature. They suggested that artificial settings might sometimes be more stressful than natural populations. In particular, organisms have evolved to handle the stresses in their natural settings, but artificial settings (such as recent captive husbandry of endangered species) may impose stresses they have yet to encounter in nature.

Finally, as we detail in Chapter 21, gene expression can also be viewed as a quantitative trait. Hence, it is not surprising that a number of studies have show expression levels change under inbreeding. It is less clear if such changes represent inbreeding depression, as it is unclear how a change in expression maps into fitness. One can imagine settings where increased expression at a particular gene decreases fitness, and settings were the converse is true. However, one very interesting observation on expression and ID was offered by García et al. (2012, 2013) in *Drosophila*. Starting from a common base population, a number of inbred lines were extracted, which showed variation in ID. The three lines showing the least amount of ID in fitness (scored by pupal number) were contrasted with the three

(surviving) lines showing the largest ID in fitness, and with the control (the starting base population), for gene expression. As expected, significant differences in gene expression were seen over lines. Surprisingly, however, the three lines showing the *least* amount of ID had the *largest* differences in gene expression relative to the control (the outbred line). The authors suggest that gene regulation was being used to somewhat buffer the impact of ID.

PURGING INBREEDING DEPRESSION

Concern has arisen that inbreeding in small captive populations of endangered species has led to a loss in fitness, possibly an irretrievable one (Ralls et al. 1979; Ralls and Ballou 1982a, 1982b; Lacy et al. 1993; Hedrick and Kalinowski 2000; Richards 2000; O'Grady et al. 2006; Leberg and Firmin 2008; Jamieson and Allendorf 2012; Frankham et al. 2013, 2014; Franklin et al. 2014; Theodorou and Couvet 2015; Caballero et al. 2017a, 2017b; Pérez-Pereira et al. 2022). Almost every capture animal species for which breeding records have been kept shows an elevated mortality rate in inbred versus outbred progeny. The impact of inbreeding on extinction is also seen in natural populations. Saccheri et al. (1998) and Nonaka et al. (2018) found that the local extinction risk significantly increased with inbreeding level in a Finnish metapopulation of the Glanville fritillary butterfly (*Melitaea cinxia*). Keller (1998) examined the inbreeding levels in survivors following a crash in an British Colombian population of song sparrows (*Melospiza melodia*). The average inbreeding coefficient of survivors was 0.0065, while the average was 0.0312 among the 206 birds that perished, with all birds with inbreeding levels over 0.06 dying. For these reason, special efforts are generally made to restrict matings of endangered species to nonrelatives. However, the options are limited in many cases because essentially all remaining members of the species are descendants of the same pedigree. Similar situations must eventually arise in closed populations of elite breeds of domesticated species.

As a potential strategy for dealing with such extreme situations, Templeton and Read (1983, 1984) advocated intentional inbreeding, combined with rapid population expansion, as a means for exposing deleterious recessive genes, and **purging** them from a remnant population by selection. Implicit in such a management program is the assumption that rare deleterious recessives, not overdominant alleles, are the primary agents of inbreeding depression. The observation of purging some of the inbreeding depression under selfing traces back to at least Darwin (1876). He noted that while most inbred lines of the common morning-glory (*Ipomea purpurea*) tended to suffer in vigor and seed set, one of his selfed lines showed surprising vigor. In honor of its performance, he named this line *Hero*.

Evidence for purging of deleterious genes has been recorded in several laboratory studies with flies (Ehiobu et al. 1989; Bryant et al. 1990; Garcia et al. 1994; Latter et al. 1995; Swindell and Bouzat 2006; Ávila et al. 2010; López-Cortegano et al. 2016; Pérez-Pereira et al. 2021), mice (Bowman and Falconer 1960; Lynch 1977; Connor and Bellucci 1979), the bivalve *Argopecten* (Zheng and Zhang 2012), and the gastropod *Physa* (Noël et al. 2016), a fact that provides further support for the dominance hypothesis. The evidence for natural purging in small, historically isolated, wild animal populations is mixed. Facon et al. (2011) found that invasive populations of ladybird beetles (*Harmonia axyridis*), which presumably passed through bottlenecks, had only a fraction of the inbreeding depression found in native populations. Visscher et al. (2001) suggested that purging occurred in a small feral population of English Chillingham cattle. Robinson et al. (2018) suggested that purging may account for the long-term persistence and lack of inbreeding depression in island foxes (*Urocyon littoralis*). Populations of this species, distinct from the mainland species, have existed on the Channel Islands of California for over 9,000 years. While these populations show extremely low levels of molecular diversity, morphological data showed a lack of congenital defects common in many highly-inbred populations, and some of the island populations have strongly rebounded following recent severe bottlenecks.

Likewise, evidence that a prolonged bottleneck resulted in at least some purging has been seen in some populations of birds, namely Stewart Island robins (*Petroica australis*

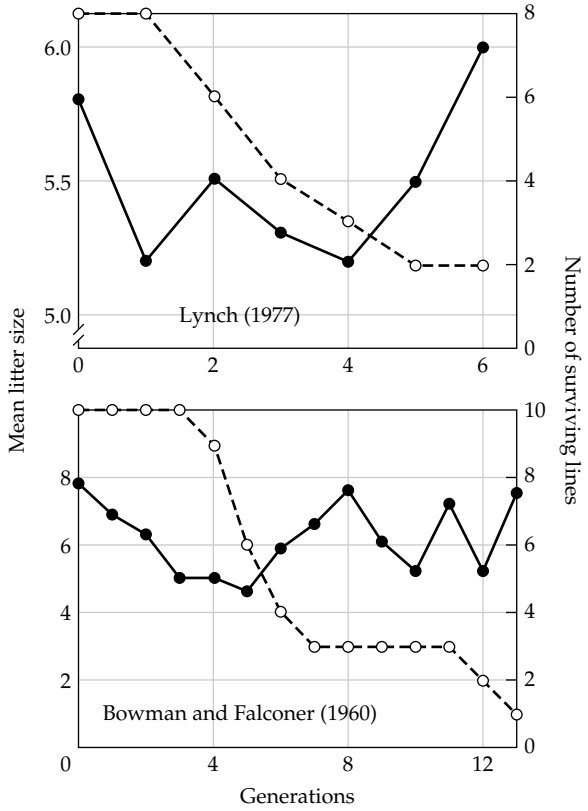

Figure 12.9 Purging of deleterious alleles through inbreeding: the response of mean litter size (●) to inbreeding in the mouse. In both studies, replicate lines were established from the same random-mating base population, and subsequently maintained by single brother-sister matings. Differential extinction of the lines (○) was caused by the failure to produce at least one reproductively competent male and female offspring. In both studies, there is initially a sharp drop in the litter size averaged over all lines, but as the less prolific lines go extinct, the grand mean returned to its initial level. Such results are not expected under the overdominance hypothesis, because with population sizes of only two individuals, most loci are expected to lose all but one allele by random genetic drift within a few generations.

rakiura (Laws and Jamieson 2011), but not for others, such as a flightless rail (the Takahē) *Prophyrio hochstetteri* (Grueber et al. 2010), or the Chatham Island black robin, *Petroica traversi* (Kennedy et al. 2014). Weiser et al. (2016) made an unexpected observation on the last species: while close inbreeding reduced survival when mothers had a modest level of inbreeding, it actually *improved* survival with more inbred mothers. They suggested this occurs because "a highly inbred chick with a highly inbred parents inherits a proven genotype and thus experiences a fitness advantage," a situation potentially akin to Darwin's *Hero* lineage.

However, despite some apparent occurrences, care needs to be taken in extrapolating the idea of purging as a general conservation tool (Hedrick 1994; Lynch et al. 1995a; Lynch 1996; Hedrick and Kalinowski 2000; Hedrick and García-Dorado 2016). Ballou (1997) found, at best, only very modest support for purging in 25 captive mammalian populations. Similarly, Boakes et al. (2007) performed a meta-analysis of 88 species of vertebrates from around 120 zoos. While they inferred some purging, they suggested that its benefits are rarely appreciable. They also suggested (as did Crnokrak and Barrett 2002) that cases where fitness seems to improve during inbreeding may simply reflect adaptation to captivity.

One of the major concerns with purging is that when inbreeding is associated with small population size, it is accompanied by an enhanced probability of fixation of mildly

deleterious genes by random genetic drift, which can permanently reduce mean population fitness. In full-sib mated lines of mice, for example, purging is only accomplished at the expense of extreme selection (extinction) among replicate lines (Figure 12.9). An example of the drawbacks of purging was offered by Leberg and Firmin (2008), who examined the impact of bottlenecks on mosquito fish (*Gambusia affinis*). While the passage through serial bottlenecks do not seem to result in any purging, it did result in an increase in extinction rate and a decline in growth rates. A final point is that because the rate of production of new deleterious alleles by spontaneous mutation is quite high (Chapter 15), any purging of the mutation load from a population is at best a transient phenomenon.

Related to Templeton and Read's (1983, 1984) notion of purging is the idea that obligately self-fertilizing plants should be largely free of inbreeding depression (Wright 1969; Lande and Schemske 1985; Husband and Schemske 1996; reviewed in WL Chapter 23). Here too, some precautionary comments are in order. First, Byers and Waller (1999) found that purging tended to be a rather "inconsistent" force in plants, with the prediction of reduced inbreeding depression in lines with a high rate of selfing being rather uneven. However, Angeloni et al. (2011) found that inbreeding depression significantly increased with population size in plants. Second, the effects of close inbreeding on the evolution of inbreeding depression needs to be considered from *two frames of reference*. By eliminating the genetic variation within an isolated population (or within a lineage of obligate selfers), long-term inbreeding can lead to a situation in which any further inbreeding has essentially no genetic consequences, simply because the progeny genotypes are essentially identical to those of their mothers. Once this situation has been reached, then from the standpoint of the *current population*, there is no inbreeding depression. However, as noted above, this need not be the case from the standpoint of the *ancestral population* if, during the inbreeding process, different deleterious recessive genes have become fixed within different selfing lineages.

The latter point is nicely illustrated by a study of self-fertilized lines of the normally outcrossing aquatic plant *Eichhornia paniculata* (Barrett and Charlesworth 1991). Inbreeding caused an immediate depression in fitness, but after only two generations of selfing, there was no further decline, suggesting that the vast majority of loci affecting fitness had become homozygous within lines. Nevertheless, despite the absence of inbreeding depression *within* the derived lines, crosses *between* lines exhibited a substantial increase in mean fitness, as expected if the different lines had become fixed for different deleterious recessives. Over time, as new deleterious mutations become sequestered in different selfing lineages, the beneficial effects of outcrossing are expected to increase further (Lynch et al. 1995b). Enhanced fitness in outcrossed progeny of normally self-fertilizing plants has been demonstrated in numerous species (Levin 1989; Charlesworth et al. 1990; Holtsford and Ellstrand 1990; Ågren and Schemske 1993; Van Treuren et al. 1993; Latta and Ritland 1994; Johnston and Schoen 1995). Clearly, the absence of local inbreeding depression provides no information about the mutation load harbored by a population.

While mating system can impact the level of inbreeding, so to can the genetic system. Brucker (1978) and Werren (1993) suggested that inbreeding depression should be reduced for haplodiploid species (such as Hymenoptera and Acarian mites), due to the haploid nature of the male being unable to cover up deleterious recessives (at least for those traits expressed in males). A meta-analysis by Henter (2003) found that while inbreeding depression can be substantial, as a group, haplodiploid taxa tended to experience less inbreeding depression than diploid insects.

Theoretical work suggests that slow inbreeding over a prolonged period appears to be the most efficient process for purging (Hedrick 1994; Wang et al. 1999; Glémin 2003; Boakes and Wang 2005; García-Dorado 2012; Caballero et al. 2017a). The success of purging also depends on the underlying genetic architecture, and hence may be successful for some traits, but not others, even in the same population. Willis (1999) examined the role of genes of large versus small effect on inbreeding depression. He started with 1200 lines from a base population of *Mimulus guttatus* (monkey flower), which were selfed by single-seed descent (SSD) for five generations, resulting in 335 surviving lines. The SSD method will select

against genes of very large effect, but the effective population size of an SSD line is too small to selectively influence genes of smaller effects (WL Chapter 7). The resulting SSD lines were then intercrossed to form a large "purged" population. The fitness of outcrossed and selfed progeny were examined using parents from both the purged and ancestral populations. In selfed offspring from the purged population, a limited decline in inbreeding depression and a limited increase in inbred fitness was observed, suggesting that while some purging did occur (through the removal of major genes), the bulk of inbreeding depression was caused by alleles of smaller effect that were not efficiently selected against under SSD.

Any purging strategy represents a delicate balance of inbreeding strength. It must be strong enough to allow selection to efficiently remove deleterious recessives over the time course of the purging yet still be weak enough to avoid fixing mildly deleterious alleles via drift. Fast inbreeding (very small N_e) can quickly purge highly deleterious alleles, either by within-line selection or via among-line selection due to extinction. However, the cost is the fixation of mildly deleterious alleles. Conversely, while slow inbreeding can effectively reduce much of the concealed load from inbreeding depression, it requires much more time, as the weak levels of inbreeding only results in a mild increase in the selection coefficient on deleterious alleles. Hence, failure of many experiments to show significant purging has been attributed to either too strong an amount of inbreeding during the purge phase or too short a duration under slow inbreeding (López-Cortegano et al. 2018).

Slow inbreeding was practiced in both a set of large populations of *Drosophila* and on smaller (N_e of 50 to 100) offshoot lines by López-Cortegano et al. (2016) and Pérez-Pereira et al. (2021). They found that 100–200 generations of mild inbreeding (modest N_e) produced an eight- to ten-fold reduction in I (relative to the base popualtion value) when the lines were subsequently tested under strong inbreeding. They also noted that selection under competitive conditions was much more efficient that selection under more benign conditions in reducing ID.

While conservation biologists tend to be extremely hesitant in applying active purging, Hedrick and García-Dorado (2016) have suggested that commonly applied approaches for maximum avoidance of inbreeding, such as minimum kinship and minimal inbreeding approaches, may be somewhat counterproductive in the long run. Such approaches may tend to minimize any actual selection that is occurring, reducing any potential for some modest background-level purging that may improve the population over time.

A final strategy for conversation biologists to deal with inbreeding depression is **genetic rescue**: introducing a very limited number of individuals from close by, and related, populations. Because this runs the risk of outbreeding depression (hybrids having lower fitness), we examine this strategy in the next chapter. However, an inbred-based cautionary tale is offered by the work of Hedrick et al. (2019) on an isolated population of grey wolves on Isle Royal in Lake Superior. This population formed around 1950 from mainland grey wolves. Around the late 1990s, a single male immigrated to the population, likely via an ice bridge that formed in 1997. He immediately became the alpha male on the island, with his descendants quickly dominating the ancestry of the population. The initial result of this natural genetic rescue was to likely prevent a collapse of the heavily inbred population during the early 2000s. However, his descendants quickly dominated the pedigree, resulting (by 2019) in a population of only two wolves, with an expected offspring inbreeding of almost 0.5. Hence, the apparently successful rescue was too successful, resulting in a dominance of the pedigree, and hence an eventual extinction among his highly inbred descendants.

Example 12.5 Approaches have been suggested to adjust estimates of the base-population I values to account for purging. The first class are based on using known pedigrees (Ballou 1997; Boakes et al. 2007; Gulisija and Crow 2007), while second method, the **inbreeding-purging** (**IP**) model, uses curve-fitting to estimate model parameters (García-Dorado 2012; García-Dorado et al. 2016; López-Cortegano et al. 2016). Simulations by López-Cortegano et al. (2018) found that the IP approach was generally both more powerful and more robust than pedigree-based methods, so we briefly outline the technique here. García-Dorado's (2012) suggestion was to

regress means on a purge-adjusted inbreeding value g (rather than on f), which is given by the recursive relationship

$$g_t \approx \left[\frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right) g_{t-1}\right] (1 - (2d)\, f_{t-1}) \tag{12.17a}$$

Here, f_t is the pedigree inbreeding value, while g is reduced due to removal of deleterious recessives via purging, quantified by the model-fitted pruning coefficient d. Under this model, the expected mean in generation t is given by

$$\mu_t \approx \mu_0 \exp(-\delta\, g_t) \tag{12.17b}$$

where δ, the true rate of inbreeding depression, is also fit by the model. See López-Cortegano et al. (2016) and Pérez-Pereira et al. (2021) for examples.

THE NUMBER OF LETHAL EQUIVALENTS

Morton et al. (1956) developed a simple regression technique for summarizing the deleterious consequences of inbreeding for attributes classified by *incidence* (i.e., binary traits), such as survival, disease expression, and so on. They defined a **lethal equivalent** (**detrimental equivalent** for traits other than survival) as any group of genes "that if dispersed in different individuals . . . would cause on average one death." Thus, a lethal equivalent can consist of a single lethal gene or of a large number of mildly deleterious genes. We note in passing that the lethal equivalent literature either reports **gametic** or **zygotic** (twice the gametic) **equivalents**. Unless otherwise stated here, our use is for the number of lethal equivalents in a gamete.

Assuming that the environment and all loci act independently in determining survivorship, the probability of survival at inbreeding level f can be written as

$$S_f = (1 - P_e) \prod_{i=1}^{n} [1 - P_i(f)] \tag{12.18a}$$

where P_e is the genotype-independent probability of dying from environmental causes, and $P_i(f)$ is the probability of dying as a result of deleterious genes at the ith locus when the inbreeding coefficient is equal to f. From Table 12.1,

$$P_i(f) = f q_i s_i + (1 - f)[\, q_i^2 s_i + 2q_i(1 - q_i)(sh)_i\,] \tag{12.18b}$$

where q_i is the frequency of the deleterious allele, and s_i and $(sh)_i$ are the probabilities of mortality for homozygotes and heterozygotes, respectively. If the probability of dying from any single cause is small, then the approximation $(1 - x) \simeq e^{-x}$ implies $\Pi(1 - x_i) = \exp(-\sum x_i)$, yielding

$$S_f \simeq \exp\left[-P_e - \sum_{i=1}^{n} P_i(f)\right] = \exp[-(A + Bf)] \tag{12.19a}$$

where

$$A = P_e + \sum_{i=1}^{n} q_i[q_i s_i + 2(1 - q_i)(sh)_i] \tag{12.19b}$$

is the sum of probabilities of mortality in the random-mating population, and

$$B = \sum_{i=1}^{n} q_i[s_i - q_i s_i - 2(1 - q_i)(sh)_i] \tag{12.19c}$$

is the excess sum of probabilities of mortality that would exist in a completely inbred population. Logarithmic transformation of Equation 12.19a leads to

$$\ln S_f \simeq -A - Bf \tag{12.20}$$

Thus, the composite measures A and B can be estimated by regressing the natural logarithm of survivorship (or trait incidence) on the inbreeding coefficient.

A slight problem with this approach arises in cases in which the observed survivorship for certain inbreeding classes is zero, because the logarithm of zero is undefined. Templeton and Read (1984) suggested the small sample-size correction

$$S'_f = \frac{1 + N'_f}{2 + N_f} \tag{12.21}$$

where N_f and N'_f are the numbers of total and surviving individuals at inbreeding level f. With no observed survivors, this quantity rapidly approaches $\approx 1/N_f$ as the total sample size increases. Morton et al. (1956) recommended the use of weighted least-squares analysis (Example 10.8), weighting the data by $N_f S_f/[1-S_f]$, the inverse of the sampling variance of $\ln S_f$, and iterating the regression by substituting expected for observed S_f in the weights. (Recall that a similar procedure is used in the estimation of line-cross variances; Example 11.3) Finally, Kalinowski and Hedrick (1998) suggested estimators based on maximum likelihood (Appendix 4), while Armstrong and Cassey (2007) advocate the use of generalized linear mixed models (GLMM; Chapter 14) for the estimation of B. See Nietlisback et al. (2019) for a review of recent development in GLM estimators of B.

Unfortunately, the mean number of lethal equivalents per gamete, $\sum q_i s_i$, cannot be separated cleanly from other terms in the definitions of A and B. However, because $A+B = P_e + \sum q_i s_i$, the number of lethal equivalents per gamete must be between B and $A + B$, assuming $(sh)_i \geq 0$. As will be seen below, estimates of $A + B$ are usually not much greater than B, so the use of B as an approximation of the effective number of lethals is not greatly troubling.

Practical situations often arise in which one only has data for noninbred individuals and a single class of inbred individuals. A regression is not possible in this case, but the parameters can still be estimated by

$$A = -\ln S_0 \tag{12.22a}$$

$$B = -\frac{\ln(S_f/S_0)}{f} \tag{12.22b}$$

where, for example, f is $1/2$ for self-fertilization and $1/4$ for full-sib mating. Using the methods of Appendix 1, the large-sample variance for B in this case is found to be

$$\text{Var}(B) \simeq \frac{1}{f^2}\left(\frac{1-S_f}{S_f N_f} + \frac{1-S_0}{S_0 N_0}\right) \tag{12.23}$$

Results from Vertebrates

The regression method of Morton et al. (1956) has been extensively applied to humans, with several independent studies indicating that the average number of lethal equivalents per gamete, B, for a number of traits is on the order of one to two (Table 12.4). Results from traits in other vertebrate species are, for the most part, very similar, suggesting on the order of 0.5 to 3 lethal equivalents per gamete (Table 12.4). This translates into one to six lethal equivalents per zygote ($2B$ for a diploid), enough to kill the average individual a few times over if fully expressed in the homozyogous state.

Of the existing data, those for the European bison, Holstein cattle, and the Stewart Island robin, which all exhibit no significant lethal load, are the most anomalous. There is

Table 12.4 Estimates of the effective number of lethals per gamete for vertebrates (bounded by B to $A + B$). All estimates were obtained by regression, except those marked by an asterisk (obtained with Equations 12.22a and 12.22b), and by a dagger (†) for maximum-likelihood or GLMM-based estimates.

Species	Trait	A	B	Reference
Humans	Survival to maturity			
	France, 1919–1925	0.16	2.87	Morton et al. 1956
	Chicago, 1936–1956	0.18	1.55	Slatis et al. 1958
	Fukuoka, Japan	0.07	0.67	Yamaguchi et al. 1970
	Nagasaki, Hiroshima, and Hirado, 1948–65	0.10	0.67	Schull and Neel 1972
	Survival to age 10	0.20	0.70	Bittles and Neel 1994*
	Conspicuous abnormalities	0.10	1.16	Slatis et al. 1958
	Mental retardation	0.01	0.80	Morton 1978
	Congenital heart disease	0.01	0.32	Gev et al. 1986
Speke's gazelle	1-year viability	0.42	3.75	Templeton and Read 1984
European bison	2-year viability	0.26	0.13	Slatis 1960
Sheep	Survival, 1.5–5 years	0.09	0.39	Wiener et al. 1992c*
Swine	Embryo survival	0.27	1.01	Pisani and Kerr 1961
Cattle	Survival through calving			
Holstein		0.16	0.02	Pisani and Kerr 1961
Jersey		0.18	1.15	
Hereford		0.19	0.64	MacNeil et al. 1989*
Great tit	Survival to fledging	0.36	0.84	van Noordwijk and Scharloo 1981
Japanese quail	16-week survival	0.60	1.91	Sittmann et al. 1966
Chicken	18-month survival	0.82	2.10	Pisani and Kerr 1961
Takahē (flightless rail)	Lifetime fitness		8.03	Grueber et al. 2011
Collared flycatchers	Juvenile survival		7.47	Kruuk et al. 2002†
Stewart Island robin	Juvenile survival		0.24	Laws and Jamieson 2011†
North Island robin	Juvenile survival		6.71	Armstrong and Cassey 2007†
Large ground finch	Juvenile survival		4.47	Keller et al. 2002†
Cactus finch	Juvenile survival		4.27	Keller et al. 2002†
Song sparrow	Juvenile survival		1.32	Keller 1998
Great tit	Juvenile survival		1.30	Szulkin et al. 2007†
Chatham Island robin	Juvenile survival		3.42	Kennedy et al. 2014†

no obvious explanation for the Holstein data; they are quite inconsistent with those from other breeds. However, it is known that earlier in this century the European bison was reduced to only a dozen individuals. Thus, it is possible that the heterozygosity in this species was largely eliminated by extensive inbreeding during the population bottleneck. As mentioned above, the same may be true for the Stewart Island robin. We also note that a study on congenital birth defects, birth weight, and gestational age for people of the Indian state of Tamil Nada revealed no evidence of inbreeding depression (Rao and Inbaraj 1980). These results, which are quite unusual for humans, may also be related to the decline in the lethal load caused by previous inbreeding. Approximately 40% of the marriages in this population were between second cousins or closer relatives.

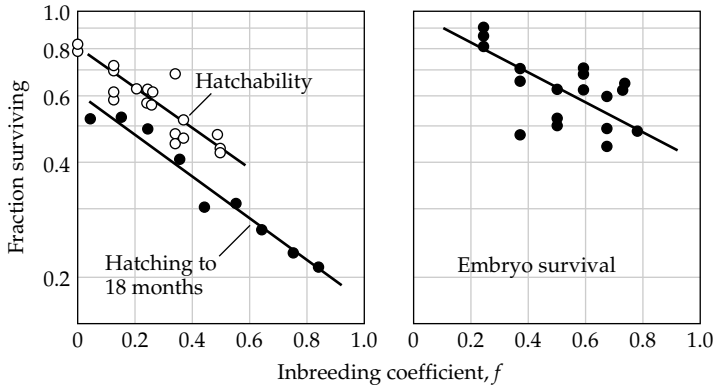

Figure 12.10 Survivorship of white-leghorn chickens (left) and Poland-China pigs (right) as a function of level of inbreeding. (After Pisani and Kerr 1961.)

A critical assumption underlying lethal-equivalent analysis is that the effects of different loci on survivorship are independent. Directional epistatic interactions between pairs of loci will give rise to nonlinearities in plots of log survival versus inbreeding coefficient, but as noted above, these can be detected only if survivorship estimates are available for several levels of inbreeding ideally over a large range of f values. Other than the data in Figure 12.10, which yielded no evidence of nonlinearity, few data sets are extensive enough to evaluate this matter.

Results from *Drosophila*

Through the use of **balancer-chromosome techniques** in *Drosophila,* it is possible to obtain a finer picture of the types of deleterious genes that contribute to inbreeding depression. Recall (Figure 5.8) that this procedure enables one to isolate intact chromosomes from natural populations and to assay their homozygous performance with respect to a control chromosome (the **balancer**). By crossing two lines, each one carrying a different chromosome, it is also possible to assay the relative performance of chromosomal heterozygotes. The ratio of the two relative performances provides a measure of the fitness of chromosomal homozygotes $(f = 1)$ relative to that of heterozygotes $(f = 0)$.

Greenberg and Crow (1960) reasoned that this approach might be exploited to partition the deleterious load in populations into components due to alleles with various magnitudes of effects. The partitioning of fitness classes is arbitrary, but the technique is nevertheless general. The usual procedure has been to classify as **lethals** those chromosomes that, when homozygous, yield less than 10% of the viability observed in random heterozygotes. Chromosomes with relative viabilities greater than 10%, but less than 100%, are referred to as **detrimentals.** Nearly all chromosomes extracted from natural *Drosophila* populations have relative viabilities less than one when in the homozygous state, so this categorization encompasses essentially all chromosomes. Denoting the mean viabilities of chromosomal heterozygotes, detrimental chromosomal homozygotes, and all homozygotes as, respectively, S_0, S_D, and S_T, then from Equation 12.22b,

$$B_T = \ln S_0 - \ln S_T \tag{12.24a}$$

$$B_D = \ln S_0 - \ln S_D \tag{12.24b}$$

$$B_L = \ln S_D - \ln S_T \tag{12.24c}$$

The different components of the deleterious load are additive, because they are the summations of effects of individual viability mutations, i.e., $B_T = B_D + B_L$. In other words, B_D (the **deleterious load**) estimates the total number of lethal equivalents resulting from the cumulative effects of all deleterious genes that are individually nonlethal, whereas B_L (the

Table 12.5 Partitioning of the total number of lethal equivalents (B_T) into the subcomponents resulting from detrimental (B_D) and lethal (B_L) factors. N is the number of studies over which the data, summarized from Simmons and Crow (1977), are averaged.

Species	Chromosome	N	B_T	B_D	B_L
Drosophila melanogaster	II	16	0.483	0.236	0.247
	III	3	0.691	0.284	0.407
Drosophila pseudoobscura	II	3	0.450	0.246	0.204
	III	1	0.578	0.352	0.226
Drosophila willistoni	II	1	0.766	0.380	0.386
	III	1	0.690	0.506	0.184

lethal load) estimates the additional number of lethal equivalents resulting from recessive lethals being present on a subset of the chromosomes.

Simmons and Crow (1977) summarized a large number of studies employing this approach using the second and third chromosomes in several species of *Drosophila.* The data were remarkably consistent (Table 12.5). The total number of (gametic) lethal equivalents associated with each chromosome, in each species, ranges from 0.5 to 0.8. Noting that chromosomes II and III each comprise approximately 40% of the *Drosophila* genome, these observations suggest that the average drosophilid carries approximately three (zygotic) lethal equivalents, similar to the situation in vertebrates. Averaging over all of the data, the detrimental and lethal loads per chromosome are, respectively, 0.33 and 0.28. Thus, about half of the total lethal equivalents are associated with lethal recessives, and about one in three chromosomes carries such a gene.

Finally, the idea that deleterious alleles do not interact epistatically can be tested with the balancer-chromosome technique. In the absence of average interchromosomal epistasis, the number of lethal equivalents expressed in individuals homozygous for both chromosomes II and III should not be significantly different from the sum of the loads obtained for individuals homozygous for just chromosome II and for just chromosome III. Only a few studies of this nature have been undertaken, and the results are somewhat mixed. The overall picture, albeit a weak one, is that if epistasis exists among the genes on the two chromosomes, it is weak and synergistic (positively reinforcing) (Simmons and Crow 1977).

Results from Plants

Equation 12.22b has been used extensively in estimating the number of lethal equivalents in coniferous trees of economic importance. The usual approach has been to compare self-pollinations to outcrosses using a mixture of pollen from several distant trees. Most of the emphasis has been on embryonic mortality, which is easily assayed by counting unfilled seeds. The number of lethal equivalents expressed at this stage is exceptionally high, ranging from one to five per gamete (Table 12.6). Most studies also show a high variance in the number of lethal equivalents per individual ($2B$ for diploids), with few individuals completely free of them and some carrying as many as 30 (Figure 12.11). Longer-term studies (Park and Fowler 1982, 1984; Fowler and Park 1983) indicated that most of the lethal equivalents affecting survival are expressed at the embryonic stage, with approximately one to two additional lethal equivalents per gamete influencing subsequent survivorship. It is conceivable that the extraordinarily high mutation load in conifers is a consequence of their long generation time, which may magnify the mutation rate on a per generation basis (Klekowski 1988). Fundamentally different results arise with short-lived herbaceous plants, where B (Equation 12.22b) for probability of germination is consistently less than one (Table 12.6). Such plants do, however, express additional lethal equivalents in the form of survival to maturity and reproductive performance (see Charlesworth and Charlesworth 1987 for a summary).

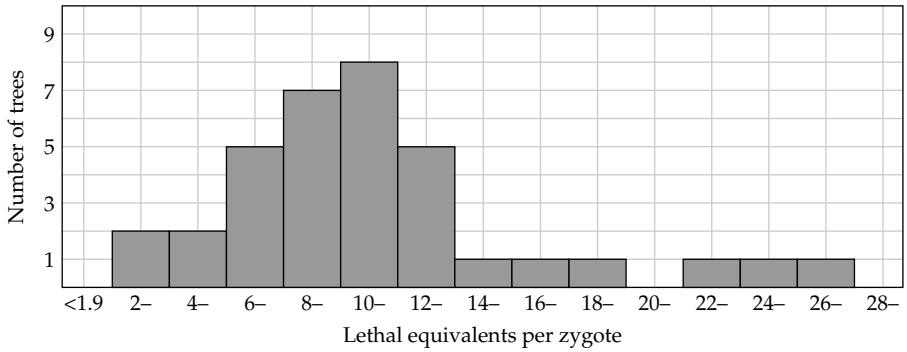

Figure 12.11 Frequency distribution of the mean number of lethal equivalents per zygote ($2B$) for 35 trees in a population of Douglas fir. (After Sorensen 1969.)

Table 12.6 Estimates of lethal equivalents per gamete (B; obtained using Equation 12.22b) affecting early embryonic survival in conifers and herbaceous angiosperms. Estimates for conifers were taken directly from the literature, while those for angiosperms were computed from data on percent germination for outcrossed and selfed seed.

Species	B	Reference
Conifers		
Nobel fir *(Abies procera)*	1.7	Sorensen et al. 1976
Tamarack *(Larix laricina)*	5.4	Park and Fowler 1982
Norway spruce *(Picea abies)*	4.8	Koski 1971
White spruce *(Picea glauca)*	5.0	Fowler and Park 1983
	4.4	Coles and Fowler 1976
Black spruce *(Picea mariana)*	2.4	Park and Fowler 1984
Ponderosa pine *(Pinus ponderosa)*	2.0	Sorensen 1970
Scots pine *(Pinus sylvestris)*	4.4	Koski 1971
	3.6	Savolainen et al. 1992
Loblolly pine *(Pinus taeda)*	4.2	Franklin 1972
	4.8	Bishir and Namkoong 1987
Virginia pine *(Pinus virginiana)*	5.0	Bishir and Namkoong 1987
Douglas fir *(Pseudotsuga menziesii)*	5.0	Sorensen 1969
Short-lived angiosperms		
Begonia hirsuta	0.04	Ågren and Schemske 1993
Begonia semiovata	0.11	Ågren and Schemske 1993
Clarkia tembloriensis	0.07	Holtsford and Ellstrand 1990
Lychnis flos-cuculi	0.39	Hauser and Loeschcke 1994
Mimulus guttatus	0.16	Latta and Ritland 1994
Raphanus sativus	0.01	Nason and Ellstrand 1995
Salvia pratensis	0.67	Ouborg and Van Treuren 1994
Schiedea lydgatei	0.91	Norman et al. 1995

Hedrick (1987b) has reviewed the extensive literature on genetic load in ferns. The majority of studies have been performed by selfing gametophytes and counting the proportion of spores that germinate, ignoring the load expressed following germination. Again, the mean number of lethal equivalents per gamete, which ranges from 0 to 1.3, appears to be substantially lower than that found in conifers. Although the fern data are still limited, they suggest that the species-specific loads are inversely proportional to the frequency of self-fertilization in nature, as expected when inbreeding purges lethals from a population.

Example 12.6. An interesting study on the importance of early selection under inbreeding was given by Hedrick et al. (2016). Myburg et al. (2014) found that 28 selfed offspring from a single *Eucalyptus* parent had a much higher fraction of heterozygosity in surviving offspring relative to the expectation under Mendelian segregation. The selfed parent was heterozygous at 9590 markers. Heterozygosity at these loci in the 28 offspring averaged 65.5%, a highly significant departure from the expected 50% value. Hedrick et al. noted that the locations of the excesses were widely distributed over the genome. Evidence suggested that this was due to several deleterious alleles, rather than single lethals, within each region. The data further suggested that likely over one hundred genes contributed to inbreeding depression. As discussed below, this observation of **distorted Mendelian segregation ratios** has been seen in earlier studies (with far fewer markers). For example, Launey and Hedgecock (2001) using the Pacific oyster (*Crassostrea gigas*), and Bierne et al. (1998) using flat oyster (*Ostrea edulis*), both noted that initially Mendelian ratios in families became increasingly more skewed toward heterozygotes over time. Finally, Hedrick et al. modeled the type of overdominant architecture required to generate such an excess of heterozygosity in the *Eucalyptus* offspring. The found that over 20 overdominant loci, each with a selective advantage of 0.6, would be required. Such genes with large effect did not seem to be located.

PARTIAL RECESSIVES vs. OVERDOMINANCE

The observation of inbred lines that equal or exceed the average performance of individuals in outcrossed populations is a serious challenge to the contention that overdominance is the primary mechanism of inbreeding depression. Nevertheless, there is still a substantial amount of controversy on the subject, fostered to a large degree by a number of puzzling early observations with allozyme loci (to be considered shortly). Here, we summarize results from biometrical analyses that bear on the question of mode of dominance. None of these results supports the idea that overdominance is a common mode of gene action.

The (*A*+*B*)/*A* Ratio

In the previous section, we defined the model for lethal equivalents in terms of partially recessive deleterious alleles. The same logic can be used to redefine the model in terms of overdominant gene action (Table 12.1). This approach again gives rise to Equation 12.20, but with the definitions of A and B altered to

$$A = P_e + \sum_{i=1}^{n} \left(q_i^2 s_i + p_i^2 t_i \right) \tag{12.25a}$$

$$B = \sum_{i=1}^{n} q_i p_i (s_i + t_i) \tag{12.25b}$$

Morton et al. (1956) noticed a useful feature of this model. Recall that for a balanced polymorphism maintained by overdominance, the equilibrium allele frequencies are $q_i = t_i/(s_i + t_i)$ and $p_i = s_i/(s_i + t_i)$. Substituting these into the previous expressions,

$$B = \sum_{i=1}^{n} \frac{s_i t_i}{s_i + t_i} \tag{12.26a}$$

$$A = P_e + B \tag{12.26b}$$

Thus, if inbreeding depression is primarily a consequence of overdominance, then the ratio $(A + B)/A = (P_e + 2B)/(P_e + B)$ is constrained to be less than or equal to two in

populations that are in a state of balanced polymorphism. Strictly speaking, this result applies to a diallelic locus. With k alleles per locus, the constraint is $(A + B)/A \leq k$ (Crow 1958; Lewontin 1974). Returning to Table 12.4, we see that this ratio is usually on the order of 10 or more. Thus, unless a very large number of alleles are maintained in a delicately balanced polymorphism *at each locus* contributing to inbreeding depression (which seems quite unlikely from theory; WL Chapter 5), the results from lethal-equivalent analyses seem generally inconsistent with the overdominance model.

Estimating the Average Degree of Dominance

Accepting that the linear decline in log fitness in a lethal-equivalent analysis is, in fact, a consequence of multiple partially recessive alleles, then some further inference about the mode of gene action can be made. Using Equations 12.19b and 12.19c, it can be shown that

$$\frac{B}{A+B} \leq 1 - \frac{\sum 2q_i s_i h_i}{\sum q_i s_i} \tag{12.27}$$

Recalling (WL Chapter 7) that under selection-mutation balance, $q_i = u_i/(h_i s_i)$, and rearranging, Equation 12.27 implies that for a population in equilibrium,

$$\tilde{h}_1 = \frac{\sum u_i}{\sum (u_i/h_i)} \leq \frac{A}{2(A+B)} \tag{12.28}$$

The expression on the left is equal to the harmonic mean of the dominance coefficients among *newly arising mutations.* An upwardly biased estimate of this quantity is provided by the ratio $A/[2(A + B)]$, the bias approaching zero as the environmental contribution to mortality becomes negligible ($P_e \rightarrow 0$). Application of Equation 12.28 to the data in Table 12.4 shows that, in the vast majority of cases, $\tilde{h}_1$ is in the range of 0.02 to 0.15. It should be kept in mind that $\tilde{h}_1$ will tend to exceed the average dominance coefficient of *segregating* deleterious alleles because mutant alleles with higher degrees of expression are more easily removed by selection. However, because a harmonic mean is always less than the arithmetic mean, these two sources of bias may approximately cancel, leaving $A/[2(A + B)]$ as a reasonable estimator of the arithmetic mean h of segregating alleles. In any event, the data clearly suggest that the majority of deleterious alleles influencing early survival are quite recessive.

When highly inbred lines are available, less biased methods for estimating the average degree of dominance exist, as pointed out by Mukai et al. (1974). Consider a single diallelic locus with the relative fitnesses of the *BB*, *Bb*, and *bb* genotypes being 1, $(1-hs)$, and $(1-s)$. If a randomly mating base population is inbred (through a series of lines) to complete homozygosity, then (from Table 12.1) a fraction q of the lines will have genotype *bb* and fitness $(1 - s)$ at the locus, while the remaining fraction p will have genotype *BB* and fitness 1. Now suppose that the inbred lines are randomly paired and mated. *BB* $\times$ *BB* matings will then occur with frequency p^2, giving rise to *BB* progeny with fitness 1. Similarly, *bb* $\times$ *bb* matings will occur with frequency q^2, giving rise to *bb* progeny with fitness $(1 - s)$, and *BB* $\times$ *bb* matings will occur with frequency $2pq$, giving rise to *Bb* progeny with fitness $(1 - hs)$. Summing over all loci, the genetic variances for log fitness among inbred lines (G_p), among midparent values (G_{mp}), and among F_1 progeny (G_o) are, respectively,

$$\sigma^2(G_p) = \sum p_i q_i s_i^2 \tag{12.29a}$$

$$\sigma^2(G_{mp}) = \sum p_i q_i s_i^2/2 \tag{12.29b}$$

$$\begin{aligned} \sigma^2(G_o) &= \sum 2p_i q_i s_i^2 [\,(1 - 2p_i q_i)h_i^2 - 2q_i^2 h_i + q_i(1 + q_i)/2\,] \\ &\simeq \sum 2p_i q_i s_i^2 h_i^2 \end{aligned} \tag{12.29c}$$

The approximation in Equation 12.29c follows from the reasonable assumption that the frequencies of deleterious alleles are kept low by selection, i.e., $q_i \ll 1$. In addition, the covariance among offspring and midparent genotypic values is

$$\begin{aligned}\sigma(G_o, G_{mp}) &= \sum p_i q_i s_i^2 [\, h_i(1 - 2q_i) + q_i] \\ &\simeq \sum p_i q_i s_i^2 h_i \end{aligned} \tag{12.29d}$$

Thus, half the genetic regression of offspring on midparent values has the expected value

$$\frac{b(o, mp)}{2} = \frac{\sigma(G_o, G_{mp})}{2\sigma^2(G_{mp})} = \frac{\sum p_i q_i s_i^2 h_i}{\sum p_i q_i s_i^2} \tag{12.30a}$$

where the slope follows from Equation 3.14b. Recalling that under selection-mutation balance, $q_i = u_i/(h_i s_i)$ and $p_i \simeq 1$, this expression reduces to

$$\frac{b(o, mp)}{2} = \tilde{h}_2 \simeq \frac{\sum u_i s_i}{\sum u_i s_i / h_i} \tag{12.30b}$$

Like Equation 12.28, this expression is a harmonic mean estimate of the average degree of dominance. In this case, however, each allele is weighted by $(u_i s_i)$, the product of the mutation pressure to the allele and the homozygous mutational effect. If s_i and h_i are uncorrelated, then Equation 12.30b, like 12.28, provides an estimate of the harmonic mean dominance coefficient of new mutations (Watanabe et al. 1976). However, data on newly arisen mutations suggest that such independence of u and s is unlikely (Chapter 15).

Keeping in mind these interpretative limitations, we now consider the situation for viability in *Drosophila.* Averaging over four studies for which the appropriate data are available (Tantaway 1957; Dobzhansky et al. 1963; Malogolowkin-Cohen et al. 1964; Garcia et al. 1994), we estimate $\tilde{h}_1 \leq 0.14 \pm 0.05$. Thus, consistent with our broader interpretation of the data in Table 12.4, most deleterious segregating genes that influence viability appear to be quite recessive. On the other hand, three studies with *D. melanogaster* (Mukai et al. 1972; Mukai and Yamaguchi 1974; Watanabe et al. 1976) obtained estimates of $\tilde{h}_2$ that average 0.30 ± 0.05. Working with the same species, Hughes (1995) obtained estimates of $\tilde{h}_2$ for additional characters in males: 0.08 for body size, 0.17 for mating ability, 0.14 for mortality rate, and 0.30 for longevity. Finally, data from Wills (1966) and Strickberger (1972) yield estimates of $\tilde{h}_2 = 0.27$ and 0.18 for viability in *D. pseudoobscura.* None of these estimates of $\tilde{h}_2$ are strictly comparable to those for $\tilde{h}_1$, not only because different weights are employed in the two definitions, but also because all of the studies employing Mukai's regression method have excluded chromosomal lines with highly deleterious effects. Thus, to the extent that lethal or semilethal alleles have lower dominance coefficients than do mildly deleterious alleles, as the data clearly suggest (Chapter 15), $\tilde{h}_2$ estimates are expected to be higher than those for $\tilde{h}_1$. Summing up the extensive data for *Drosophila,* most segregating deleterious mutations appear to be recessive, with the average h for all deleterious alleles being on the order of perhaps 0.1, and that for mildly deleterious alleles being on the order of 0.15 to 0.3.

There is an unfortunate void on information on the average degree of dominance in other organisms. However, the regression technique of Mukai et al. (1974) is easily extended to certain species, most notably those that reproduce in nature by obligate self-fertilization. Individuals within such populations are as close to being completely homozygous as one can obtain, and provided that forced outcrosses can be implemented, it is possible to use Equation 12.30b to obtain an estimate of the dominance coefficient averaged over the entire genome. Under obligate selfing, most lethal recessives should be purged, leaving most of the deleterious mutation load in the detrimental class. Johnston and Schoen (1995) used the regression method to obtain estimates of the average degree of dominance in four populations of the annual plant *Amsinckia.* Their average estimates of $\tilde{h}_2$ were: 0.30 for flower

Table 12.7 Possible explanations for heterozygosity-fitness correlations (HFCs).

Direct effects
Functional Overdominance Hypothesis
Marker locus under direct (and overdominant) selection.
Indirect effects
Local Effect Hypothesis
Marker in LD with selected loci (overdominance or associative overdominance).
Requires linkage disequilibrium (LD).
LD block size can be assessed with marker data.
General Effect Hypothesis
Inbreeding depression.
Marker heterozygosity predicts genome-wide heterozygosity.
Requires identity disequilibrium (IDE), which in turn requires variance in f.
IDE can be estimated by the g_2 statistic.

number, 0.36 for survivorship from germination to flowering, 0.10 for seed production, and 0.21 for total fitness. These results are close to those for detrimentals in *Drosophila.* Additional methods for estimating the degree of dominance using line-cross analysis are covered in Chapter 25.

CORRELATION BETWEEN MULTILOCUS HETEROZYGOSITY AND FITNESS

There is a long history of studies finding weak correlations between marker heterozygosity and fitness components (such as growth rate), a phenomena usually denoted in the literature as **heterozygosity-fitness correlations** (**HFCs**). Because this observation has a population-genetic context that has changed over time, some background history is in order to provide the proper setting before we examine HFCs in detail. As we will see, possible explanations for HFCs reflect many aspects of the overdominance versus partial dominance debate over the cause of inbreeding depression. As with that discussion, the current consensus is that most HFCs simply reflect inbreeding depression (Szulkin et al. 2010). While the analysis of HFCs marks the first serious marker-based approach for examining the impact of inbreeding, we consider HFCs in some detail before concluding the chapter with more direct marker-based estimates of f and what they tell us about the impact of inbreeding.

The first widespread set of molecular markers were **allozymes** (or **isozymes**), which scored variation in the protein products produced by a given gene (Lewontin and Hubby 1966). Under the method of **electrophoresis**, proteins that vary in total charge run at different speeds on an acrylamide gel under an electric current, resulting in discrete bands corresponding to the different variants (Hubby and Lewontin 1966; Lewontin and Hubby 1966). Allozyme markers were scored over a very wide range of populations and and organisms, with one of the early observations being an association in some studies between the amount of allozyme heterozygosity and fitness, with fitness increasing with total heterozygosity (Nevo 1978).

During this allozyme era from the past century, the field of population genetics was engaged in a raging debate about the role (or lack thereof) of overdominance (heterozygotes having the highest fitness; WL Chapters 5 and 7) in maintaining genetic variation. Many investigators initially assumed that allozyme HFCs were simply a reflection of a heterozygous advantage in fitness for some of the markers—the **functional overdominance hypothesis** (Frydenberg 1963; Koehn and Shumway 1982). Given that allozymes *directly* score gene products, this was not an unreasonable hypothesis (Mitton 1994).

With allozymes, the focus was largely on the direct fitness effects (or lack thereof) of these marker loci. As genomic technology advanced, researchers moved to random DNA markers, most of which were in noncoding regions, and hence were less likely to be directly

under selection. HFCs with such markers are thus generated *indirectly* by either selection at linked sites (marker loci in LD with loci under selection; the **local effect hypothesis**) or by genome-wide effects such as inbreeding, with marker heterozygosity reflecting heterozygosity across the genome of an individual (the **general effect hypothesis**). Table 12.7 summarizes these three competing hypotheses (David et al. 1995; David 1998; Hansson and Westerberg 2002, 2008), which are detailed below.

Under the overdominance hypothesis (direct effects), one would expect individual fitness to increase with the number of heterozygous allozyme loci per individual, either because the loci themselves are heterotic or because they mark heterotic regions of the genome. Because heterozygosity is assumed to be unconditionally advantageous, *this correlation should hold in populations with any level of inbreeding*. In contrast, under the general effect hypothesis, a correlation between **multilocus heterozygosity** (**MLH**) and fitness should arise *only* if there is a correlation between MLH and the level of individual inbreeding, f. Hence, before examining HFCs in more detail, we first need to consider the nature (if any) of the correlation between MLH and f.

Identity Disequilibrium, Variance in f, and the g_2 Statistic

Up to this point, we have largely assumed that all inbred individuals experience the same value of f. This has mainly been for convenience. In reality, the level of inbreeding varies over individuals, and such variance can generate HFCs.The feature required for heterozygosity-f correlations is **identity disequilibrium** (**IDE**). This refers to situations in which there is a correlation across loci for the probability of alleles being identical by descent, which requires ***variance in the level of inbreeding*** in the sampled population (Weir and Cockerham 1973). It is important to stress the differences between LD and IDE. The former refers to nonrandom association of *alleles* in *gametes* (correlation of alleles within a gamete). As we saw in Chapter 5, the extent of linkage disequilibrium can be assessed using markers to quantify LD block size, with the likelihood of the local effect hypothesis increasing with the average block size (as any given marker is impacted by ever-larger regions of the genome). In contrast, identity disequilibrium refers to nonrandom associations of *diploid genotypes* within *zygotes*. Linkage is *not* required for IDE, which reflects the genome-wide level of f within an individual. Indeed, for a model with partial selfing, no selection and no drift, Weir and Cockerham (1973) found that the IDE between two completely linked loci is only twice that for unlinked loci.

As was the case with LD, one can use markers to infer the amount of IDE within a population. A simple test for its presence was suggested by Balloux et al. (2004). Their **heterozygosity-heterozygosity correlation test** randomly partitions marker loci into two groups, and then computes the correlation between the number of heterozygous loci in group one vs. group two over individuals.

A more formal IDE metric, g_2, was introduced by David et al. (2007). This parameter follows by considering the expectation that *pairs* of loci are jointly heterozygous within an individual, which can be written as

$$E[H_i H_j] = E[H_i]\, E[H_j]\, (1 + g_2) \tag{12.37a}$$

where H_i is a zero-one indicator variable of whether locus i is a heterozygote. If heterozygosity is uncorrelated over loci, then $E[H_i H_j] = E[H_i]\, E[H_j]$ and $g_2 = 0$. If the observation that one locus is heterozygous in an individual inflates the probability that other loci in that individual are also heterozygous, then $g_2 > 0$. David et al. proposed an approximate estimator of g_2, with an improved version offered by Stoffel et al. (2016). This metric is called g_2 because k-way disequilibrium metrics (g_k) were also proposed, with the expectation of joint heterozygosity at k loci within an individual being written as

$$E\left[\Pi_{i=1}^{k} H_i\right] = \left(\Pi_{i=1}^{k} E\left[H_i\right]\right)(1 + g_k) \tag{12.37b}$$

The connection between g_2 and the variance in inbreeding levels among individuals within in the population, $\sigma^2(f)$, was shown by David et al. (2007) to be

$$g_2 = \frac{\sigma^2(f)}{(1-\overline{f})^2} \tag{12.37c}$$

where $\overline{f}$ is the average level of inbreeding in the population. In the absence of variance in the level of inbreeding, $g_2 = 0$ and there is no IDE. Thus even a highly inbred population can *fail* to show IDE if all individuals are *inbred to the same level*.

A meta-analysis by Miller and Coltman (2014) of estimated g_2 values in vertebrates found that only 26 of 129 estimates were significant. Among all studies, the average value of g_2 was 0.007, while among the studies where g_2 was significant, its average value was 0.025. This implies that, given one locus is heterozygous, the probability that a second is a heterozygote is inflated by 2.5% relative to its Hardy-Weinberg expectation. An earlier analysis of 19 vertebrate populations by Grueber et al. (2011) found a mean value for $\sigma^2(f)$ of 0.0044 and a mean value for $\overline{f}$ of 0.053, suggesting a rough value for g_2 of $0.004/(1-0.052)^2 = 0.005$, consistent with the larger summary of Miller and Coltman. An important observation regarding $\sigma^2(f)$ was made by Huisman et al. (2016) in their study on red deer (*Cervus elaphus*), who found that this variance typically *decreases* with age, presumably reflecting selection against inbred individuals.

Finally, it is important to note that an estimate of g_2 simply provides information on the amount of IDE within a sample, and this estimate is *trait-independent*. As noted by a number of investigators, one can have a nonsignificant value of g_2 and yet still have a significant HFC, because the former may be estimated with a small set of markers, whereas the latter (when IDE is present) is the cumulative effect over an entire genome (Szulkin et al. 2010; Kardos et al. 2014; Miller and Coltman 2014). Hence, the joint observation of a nonsignificant g_2 and a significant HFC does not rule out the general effects hypothesis.

Example 12.7. Forstmeier et al. (2012) examined a five-generation pedigree of zebra finches, finding the following distribution of pedigree-expected f values (f_{ped}) among 1210 sampled individuals

Number	Frequency	f_{ped} value
1100	0.909	0
24	0.020	0.01563
6	0.005	0.03125
29	0.024	0.0625
4	0.003	0.125
47	0.039	0.25

For this data, it quickly follows that $\overline{f}$ = 0.01209 and $\overline{f^2}$ = 0.00258. Hence, from Equation 2.5,

$$\text{Var}(f) = \frac{n\big(\overline{f^2} - [\overline{f}]^2\big)}{n-1} = \frac{1210\big(0.00258 - 0.01209^2\big)}{1209} = 0.00244$$

As we discuss below, this is a biased lower underestimate of $\sigma^2(f)$, as it simply uses the *expected value* of f for various sets of relatives (their predigree values), and ignores the variation in their *realizations* of any particular f about this expected value (Chapters 7 and 8). Substituting into Equation 12.37c gives an estimated g_2 value of

$$\widehat{g_2} = \frac{\text{Var}(f)}{(1-\overline{f})^2} = \frac{0.00244}{(1-0.01209)^2} = 0.00250$$

Generating HFCs Under Indirect Effects: Identity Disequilibrium (IDE)

Two (non-exclusive) mechanisms can indirectly generate HFCs for marker loci not themselves under selection, both of which require correlations between marker loci and loci under selection. The first is linkage disequilibrium (the local effect hypotheses; Table 12.7), which we will examine shortly. The second is identity disequilibrium, where the number of heterozygous marker loci that an individual carries provides some information on its level of inbreeding (the general effect hypothesis; Table 12.7). As Equation 12.37c shows, IDE requires not just inbreeding, but more specifically, *variance* in the level of inbreeding over individuals in the population, $\sigma^2(f) > 0$.

Under inbreeding depression, we generally expect a negative correlation between fitness and f, namely, $\rho(w, f) < 0$. How does this translate into a HFC? Szulkin et al. (2010) noted that the expected correlation between fitness and an estimator, $\widehat{f}$, of f can be expressed as

$$\rho(w, \widehat{f}\,) = \rho(\widehat{f}, f)\, \rho(w, f) \tag{12.38a}$$

Hence, $\rho(w, \widehat{f}\,) < \rho(w, f)$, with the fraction of any true correlation between inbreeding and fitness being captured by HFCs depending on how accurately MLH predicts f. Similarly, the regression slope of fitness on our estimator of f can be written as

$$\beta(W, \widehat{f}\,) = \beta(W, f)\, \beta(f, \widehat{f}\,) \tag{12.38b}$$

Most HFC studies use either H, the total number of heterozygous marker loci in an individual, or the **standardized heterozygosity** (Coltman et al. 1999),

$$H_s = \frac{H}{\overline{H}} \tag{12.39a}$$

namely the number of heterozygous loci standardized by the average number of heterozygous loci in a random individual from the sampled population. Szulkin et al. (2010) showed that the correlations between H and H_s with f are

$$\rho^2(f, H) = g_2 \left(\frac{\overline{H}^{\,2}}{\sigma^2(H)} \right) = \sigma^2(f) \left(\frac{\overline{H}^{\,2}}{\sigma^2(H)(1 - \overline{f})^2} \right) \tag{12.39b}$$

and

$$\rho^2(f, H_s) = \frac{g_2}{\sigma^2(H_s)} = \frac{\sigma^2(f)}{\sigma^2(H_s)(1 - \overline{f})^2} \tag{12.39c}$$

where the last steps follows from Equation 12.37c, with $\sigma^2(H)$ and $\sigma^2(H_s)$ being, respectively, the variance among H and H_s values among the individuals in the population (or our sample).

An important question in the design of HFC studies is the required number, and type, of markers needed to detect an HFC generated under the general effects hypothesis (the presence of *both* IDE in the genome *and* inbreeding depression in the focal trait used as the fitness proxy). This was nicely addressed by Miller et al. (2014), who showed that for m markers, each with an average heterozygosity of h, then

$$\rho^2(f, H_s) = \frac{m g_2 h}{1 - h + (m - 1) g_2 h} \tag{12.40a}$$

For example, using the estimated g_2 value ($\simeq 0.003$) from Example 12.7, the expected correlation between H_s and f for 1000 SNPs, each with $h = 0.2$, is

$$\rho^2(f, H_s) = \frac{m g_2 h}{1 - h + (m - 1) g_2 h} = \frac{1000 \cdot 0.003 \cdot 0.2}{1 - 0.2 + 999 \cdot 0.003 \cdot 0.2} = 0.429$$

Hence, this design would pick up just over 40% of the variance in fitness explained by the level of inbreeding. Note that Equation 12.40a can be rearranged (Miller and Coltman 2014) to yield the required number of markers, m_r, to obtain a expected correlation of r between H_s and f,

$$m_r = \left(\frac{r^2}{1-r^2}\right)\left(\frac{1-g_2(1+h)}{g_2 h}\right) \tag{12.40b}$$

Again using the g_2 value from Example 12.7, how many SNPs (with $h = 0.2$) are required to have H_s account for 90% ($r^2 = 0.9$) of the variance in f? Equation 12.40b yields

$$m_r = \left(\frac{0.9}{1-0.9}\right)\left(\frac{1-0.003(1+0.2)}{0.003 \cdot 0.2}\right) = 14,946$$

An important caveat for Equations 12.40a and 12.40b is the assumption that markers are *independent*, namely, they are not in LD. The effective population size and linkage map set an upper limit on the number of independent markers (Chapters 5 and 20). Hence, the more proper terminology for Equation 12.40b is not the required number of markers, but rather the **equivalent number of independent markers**, which may be far less than the actual number of markers that were scored.

Equation 12.40a makes the important point that the product of number of markers and their average heterozygosity determines the strength of the correlation. Early DNA-based HFC studies typically used no more than a few dozen microsatellites (Chapman et al. 2009). The advantage of this class of markers is that they are highly polymorphic. Roughly speaking, many alleles (length variants) at a given microsatellite are at nearly equal frequencies, so that a microsatellite locus with k alleles has an approximate h value of $(1 - 1/k)$, as there are k different homozygotes, each with approximate frequency $1/k^2$. SNPs, on the other hand, have lower values of h (as they are typically bialleic with unequal frequencies, so that typically $h \ll 0.5$), and hence more are required to have the same power as microsatellites. As an example, in a study of two bighorn sheep populations by Miller et al. (2014), microsatellites had h values of around 0.6, while SNPs had values of around 0.2. Hence, for these values, three times as many SNPs were required to have the same power as microsatellites. Genomic technology has now evolved to the point where it is relatively trivial, and inexpensive, to score thousands (or more) of SNPs quickly. As a result, HFC studies are shifting entirely to SNPs.

Miller and Coltman (2014) noted from Equation 12.40b that *most HFC studies are significantly underpowered* in terms of the numbers of scored markers. In particular, observe that the correlation between H_s and f increases with the parameter mg_2h. An investigator can increase this value by scoring more markers (increasing m) or using more polymorphic markers (h). Also, the appropriate sampling design can increase the sample value of g_2 (Miller et al. 2014). This might be accomplished by sampling more individuals (so that the sample includes more inbred individuals), or by a more hierarchical design that attempts to capture a subset of individuals that are more inbred than random.

In large, randomly mating populations, the association between MLH and f is expected to be negligible because essentially all individuals trace through pedigrees with similar (and very low) levels of inbreeding, rendering the variance in f among individuals, and hence g_2, nearly insignificant. Thus, if the general effects hypothesis is correct, such populations should not exhibit HFCs. However, even large populations can contain considerable structure, with localized inbreeding, so care must be taken in declaring the general effect hypothesis to be invalid simply because a HFC is observed in a large population. Further, even in a large stable population, migration can generate IDE, with migrants generally being more outbred relative to residents, generating $\sigma^2(f) > 0$.

Bierne et al. (2000) suggested that one optimal setting to generate HFCs in a large population is during the immediate aftermath of a bottleneck (and subsequent expansion to yield the currently observed large size). A large pre-bottleneck population that is randomly mating is expected to have the potential of considerable inbreeding depression, while a small population might have much of its inbreeding depression purged (see above). Hence,

small populations, while potentially having significant IDE, may have little inbreeding depression, while a historically large population has high depression potential but little IDE. A bottleneck generates IDE while keeping much of the variation needed for inbreeding depression.

Example 12.8. Szulkin et al. (2010) examined a HFC study by Pujol et al. (2005) on an important crop in developing countries, cassava (*Manihot esculenta*). Using eight microsatellites, the mean number of heterozygous loci per individual was $\overline{H}$ = 3.731, the sample variance in H was 6.104, and an estimate of g_2 based on these markers was 0.057. Finally, the observed correlation between H and growth rate (taken as a surrogate for fitness) was $\rho^2(w, H) = 0.031$. Equation 12.39b gives an estimate of the correlation between f and H as

$$\rho^2(f, H) = g_2 \left(\frac{\overline{H}^2}{\sigma^2(H)} \right) = 0.057 \left(\frac{3.731^2}{6.104} \right) = 0.13$$

Hence, H accounts for only 13% of the actual variance in f among individuals. Note that Equation 12.37c implies that $\sigma^2(f) < g_2 = 0.057$. Using the Pujol data, Szulkin et al. (2010) estimated that $\overline{f} = 0.071$, which from Equation 12.37c yields

$$\sigma^2(f) = (1 - \overline{f})^2 \, g_2 = (1 - 0.071)^2 \cdot 0.057 = 0.049$$

Finally, rearranging Equation 12.38a allows one to estimate the underlying fitness-inbreeding correlation,

$$\rho^2(w, f) = \frac{\rho^2(w, H)}{\rho^2(f, H)} = \frac{0.031}{0.13} = 0.238$$

Thus, while almost 24% of the variation in growth rate was due to differences in inbreeding level, only 13% of this effect was captured by using H, resulting in H accounting for only 3% of the variance in growth rate.

Now suppose that this study was carried out with 300 SNPs, each with an average heterozygosity of 0.2. From Equation 12.40a, the expected correlation between H_s and f becomes

$$\rho^2(f, H_s) = \frac{m g_2 h}{1 - h + (m-1) g_2 h} = \frac{300 \cdot 0.057 \cdot 0.2}{1 - 0.2 + 299 \cdot 0.057 \cdot 0.2} = 0.81$$

From Equation 12.40b, the expected number of SNPs (with $h = 0.2$) in order to have H_s explain 95% of the variation in f is

$$m_r = \left(\frac{r^2}{1 - r^2} \right) \left(\frac{1 - g_2(1+h)}{g_2 h} \right) = \left(\frac{0.95}{0.05} \right) \left(\frac{1 - 0.057(1 + 0.2)}{0.057 \cdot 0.2} \right) = 1553$$

Generating HFCs Under Indirect Effects: Linkage Disequilibrium (LD)

The second mechanism for indirectly generating HFCs is linkage disequilibrium (the local effect hypothesis; Table 12.7). While overdominance at these linked selected loci can generate HFCs, so too can partial dominance. To see this latter point, consider the situation for a marker locus with two alleles, M_1 and M_2, neither of which has any direct influence on fitness, and assume that each marker allele is tightly linked (and in complete linkage disequilibrium) with a deleterious allele at a different locus. Letting lowercase letters denote deleterious alleles, then the gametic states associated with the marker locus are AM_1b and aM_2B. This is an extreme case of repulsion disequilibrium (Chapter 5), as there are only three genotypes in the population associated with the marker: AM_1b/AM_1b, AM_1b/aM_2B, and aM_2B/aM_2B. Now suppose that each deleterious allele (a, b) reduces fitness by the

fraction s in the homozygous state, by hs in the heterozygous state, and that the effects of the two loci are independent. The resulting fitnesses associated with the marker locus become $(1 - s)$ for the two homozygous classes and $(1 - hs)^2$ for the heterozygous class. Under this scenario, the heterozygous marker class will exhibit greater fitness than the homozygous classes provided $h < (1 + s)/2$.

This apparent heterozygote superiority, solely an artifact of deleterious alleles being in repulsion disequilibrium, is known as **associative overdominance** (Frydenberg 1963; Ohta 1971). Notice that when fitness is analyzed on the direct scale of measurement, associative overdominance can arise even with additivity $(h = 0.5)$ or with slight dominance of the deleterious allele. If $h > (1 + s)/2$, **associative underdominance** occurs, wherein the heterozygotes exhibit reduced fitness. These latter two peculiarities disappear if fitness is measured on a logarithmic scale. Associative overdominance or underdominance can still arise in this case, but assuming small s, it depends more simply upon whether h is less than or greater than 0.5, i.e., on whether the deleterious alleles are partially recessive or partially dominant. We highlight this scaling property because many HFC analyses are performed on non-log transformed data. The important point is that even for marker loci in Hardy-Weinberg equilibrium (i.e., with no evidence of inbreeding), with no direct effects on fitness, and with no functional overdominance elsewhere in the genome, associative overdominance arises if loci carrying partially recessive deleterious alleles are linked to the marker and in repulsion disequilibrium.

Now consider the opposite situation: coupling disequilibrium, such that the two marked stretches of DNA are aM_1b and AM_2B. The fitnesses associated with the M_1M_1, M_1M_2, and M_2M_2 genotypes are then $(1 - s)^2$, $(1 - hs)^2$, and 1. Most HFC studies simply consider whether the fitness of heterozygotes exceeds the *average* of the homozygous classes (without respect to homozygous genotype). Thus, the relevant observation is that associative overdominance with respect to the mean logarithm of fitness for the two homozygous classes will arise if $h < 0.5$, the same conclusion that we arrived at with repulsion disequilibrium. For non-log transformed fitness, the requirement is $h < (2 + s)/4$.

Thus, the general conclusion is that linked loci carrying partially recessive deleterious alleles in linkage disequilibrium, whether in repulsion or in coupling, will always lead to the ***appearance of overdominance***, a phenomena known as **pseudo-overdominance** (Birchler et al. 2006). The fact that situations fully in accord with the dominance hypothesis can lead to observations fully compatible with the predictions of the overdominance hypothesis is an obvious problem. Because essentially all populations have some degree of structure (either variance in inbreeding and/or gametic-phase disequilibria), it can be difficult to draw rigorous conclusions about the mechanism of inbreeding depression (or about the advantages of heterozygosity at individual loci) from descriptive surveys of the relationship between MLH and fitness-related characters. Fortunately, with the advent of dense marker maps, we can gain some insight into the extend of LD with a genome and of IDE within a sample, both of which can help inform our understanding of HFCs.

We expect that a large randomly mating population will have little LD, as the time back to the most recent common ancestor will be long, allowing recombination to remove LD for all but the most tightly-linked loci (Chapter 5). When LD extends over only very short distances, the chance that a relatively small random set of markers will just happen to capture linked selected loci is expected to be rather unlikely. Hence, as was the case for IDE, we (at first blush) expect that the local effect hypothesis will be unable to generate any substantial HFCs in a large populations. However, again as with IDE, caution is in order. A currently large population may have experienced bottlenecks and expansions in the past, which can generate significant LD. The use of dense markers can estimate the average LD block size in a population sample, allowing one to probe the likelihood of LD effects.

In summary, both of the indirect hypotheses (local and general effects; Table 12.7) require population structure and demographic features that generate LD and/or IDE. If both types of disequilibria are present, then for local effects to dominate general effects, the cumulative fitness effects of nearby selected loci in LD with a given marker must exceed

the cumulative effects from IDE with selected loci in the rest of the genome. As stressed by Szulkin et al. (2010), this is unlikely unless the amount of IDE is extremely small or the LD reach of the scored markers spans a significant fraction of the genome.

HFCs: The Data

Given the above discussion, the most informative MLH surveys come from observations on historically large, random-mating populations, for which gametic-phase and identity disequilibria are likely to be minimized (subject to the caveats mentioned above). Studies of this nature with adequate statistical power raise serious questions about the generality of the overdominance hypothesis when they do not yield a positive correlation between MLH and fitness. What do the data tell us? For most organisms that have been studied, predominantly small, but significant, correlations are often found between heterozygosity and fitness-related characters (Mitton and Grant 1984; Zouros and Foltz 1987), typically accounting for around 1% of the variance (Chapman et al. 2009). There are, however, numerous exceptions in which no such pattern is found.

A substantial body of data for marine bivalves suggests that most populations exhibit a positive correlation between growth rate and allozyme MLH (Singh and Zouros 1978; Zouros et al. 1980, 1988; Koehn et al. 1988; Gaffney 1990; Gaffney et al. 1990; David et al. 1995). It should be stressed that these studies often used key enzymes in important metabolic pathways, enhancing the chance that the markers were under direct selection. Interestingly, samples from these same populations almost always exhibit a deficiency of heterozygotes, at least in early-age cohorts. In addition, there appears to be substantial multilocus disequilibrium—an excess of highly homozygous and highly heterozygous individuals. Many studies have documented the elimination of the heterozygote deficiency as cohorts age, which suggests that MLH is positively correlated with survival as well as with growth rate. The presence of disequilibria in new recruits clearly indicates the potential for these results to be a simple consequence of associative overdominance, rather than an intrinsic advantage to allozyme heterozygosity, a conclusion that is bolstered by the observation that increases in heterozygosity with cohort aging only appear at loci that initially have heterozygote deficiencies. Given that most marine bivalves are broadcast spawners, it is difficult to envision how a high degree of genetic structure (generating IDE) could arise within populations via restricted mating. However, as Bierne et al. (2000) noted, bottlenecks can generate HFCs via indirect effects by generating both LD and IDE, and they suggested that this could easily apply to bivalve populations.

An alternative explanation for both the heterozygote deficiencies and the multilocus disequilibria at some loci is the presence of either null alleles or of aneuploidy (missing chromosomes). If either situation is common, as appears to be the case (see Gaffney et al. 1990 for a summary), then a fraction of the individuals that are scored electrophoretically as homozygotes will actually be either active/null heterozygotes or chromosomal haploids. If such individuals have reduced activity for important metabolic and/or developmental functions, as seems likely, their undetected presence will cause a downward bias in the estimated fitnesses of homozygous classes (Foltz 1986). Thus, in addition to yielding apparent heterozygote deficiencies, null alleles and aneuploidy will promote the appearance of heterozygote advantage.

Pogson and Zouros (1994) have argued that the failure of random bivalve microsatellite (SSR) markers to show the fitness correlations seen with allozymes is a point in favor of functional overdominance of allozymes. The same observation was also made for rainbow trout (*Oncorhynchus mykiss*) by Thelen and Allendorf (2001). However, the presence of null alleles for allozymes, but not for DNA-based markers, could easily account for this difference. Allozymes score the final protein product, which can fail to be produced by a given gene copy for any number of reasons, resulting in a haploid product from an individual being scored as a homozygote. DNA-based markers generally do not suffer this problem. Hence, DNA markers will not count heterozygotes involving a nonfunctional alleles as homozygotes (and likely with reduced fitness). Support for this idea comes from the ob-

servation that Pogson and Zouros found that their microsatellites markers, unlike some of their allozymes, showed only small heterozygote deficiencies.

Apparent heterozygote advantages have also been recorded for growth rate in some species of trees (Mitton et al. 1981; Ledig et al. 1983; Strauss 1986; Strauss and Libby 1987). However, other extensive surveys have failed to find any such relationship (Bush and Smouse 1991; Savolainen and Hedrick 1995). In *Pinus radiata,* stands with greater allozyme heterozygote deficiencies exhibit higher correlations between individual heterozygosity and growth rate (Strauss and Libby 1987), results that are reminiscent of those obtained in marine bivalves. If not an artifact of null alleles or aneuploidy, such results may be a simple consequence of multilocus homozygosity acting as a proxy for the variation in degree of inbreeding within individual stands of trees (IDE; the general effect hypothesis).

Attempts have been made to find an association between MLH and fitness in many other organisms, some with success, others not (reviewed, among others, by Britten 1996; Reed and Frankham 2003; Chapman et al. 2009; Miller and Coltman 2014). Most older (pre-2000) studies were based on small numbers of individuals and a modest number of allozyme loci, and those that have found positive correlations are subject to the challenge that the results are an artifact of population structure (generating $\sigma^2(f) > 0$), rather than a consequence of true functional overdominance (Houle 1989a). Three of the largest studies that were performed in this era, *D. melanogaster* by Houle (1989a), fungus beetles by Whitlock (1993), and brook trout by Hutchings and Ferguson (1992), all failed to find an allozyme MLH-fitness association. Conversely, Hoffman et al. (2014) noted that an initially weak association between MLH (based on ~30 microsatellites) and parasite load seen in harbor seals (*Phoca vitulina*) saw the strength of this association increase almost five-fold when MLH was based on ~15,000 SNPs. Equation 12.40a predicts such an increase when such a correlation is real.

A number of attempts have been made to develop statistical approaches that could definitively resolve the associative overdominance issue for allozyme data. For example, Smouse (1986) reasoned that a deeper understanding of the mechanisms of inbreeding depression could be obtained by examining the fitnesses of the alternative homozygous classes within loci. Assuming two alleles per locus, under the overdominance hypothesis, the rarer of the two homozygous classes should have the lowest fitness (WL Chapter 5). Smouse's adaptive distance model transiently attracted some followers, until Houle (1994) showed that the model fits cases of associative overdominance as well as cases of functional overdominance.

One proposed experimental approach for determining the importance of local versus general effects in generating HFCs is to examine heterozygosity-fitness correlation in sibs, such as the offspring of selfed individuals (Leary et al. 1987; Fu and Ritland 1994; Bierne et al. 1998; Hansson et al. 2001). The proposed logic was that all members in a sibship have the same value of f, controlling for differences in inbreeding levels. The presence of correlations between MLH and fitness within such families is then taken as evidence *against* the general effects hypothesis. However, this approach is potentially a bit misleading, as while the *expected* value of f is the same in all of the sibs, the *realized* values show variation about this expectation, variation that potentially can be partly captured by variation in heterozygosity among sibs.

Finally, a more promising approach for distinguishing between the alternative hypotheses listed in Table 12.7 is to use dense molecular maps to localize QTLs (Chapters 17–21) involved in inbreeding depression (and heterosis). This is examined in detail in the next chapter in the search for loci underlying heterosis. The basic conclusion is that, for those cases where a QTL involved in heterosis initially appears to show overdominance, finer mapping almost always reveals that it was generated by pseudo-overdominance, with linked deleterious recessive alleles acting as a single gene in the initial linkage block.

Summary: What Generates Most HFCs?

Early allozyme-based HFCs studies were motivated by the search for direct selection via

overdominance, although this was only one of three potential explanations (Table 12.7). With the advent of random DNA markers, the focus of HFCs studies has somewhat shifted. In contrast with allozymes, which were often *specifically chosen* because they scored key protein products in important metabolic pathways, *direct* selection on random DNA markers was viewed as unlikely and HFCs involving these markers are due to *indirect effects*. These arise via correlations between scored markers and unscored loci under selection, through either LD with nearby selected sites (the local effect hypothesis) or IDE with selected sites over the entire genome (the general effect hypothesis). Of these, a stronger impact from general effects is much more likely, as the total fraction of the genome in LD with a modest set of scored markers is likely to be rather small, so that genome-wide effects via IDE would likely swamp any locally-based signal. The conclusion is that most observed HFCs are simply manifestations of inbreeding depression (Szulkin et al. 2010), arising because inbred individuals are less fit, and MLH predicts (or at least ranks) the inbreeding levels of sampled individuals (when IDE is present). Finally, population structure and demography dictate the likelihood of HFCs. Indirect effects are not expected in large, stable, and closed populations, but can be generated by localized mating, bottlenecks, or migration, all of which can be difficult to rule out for any particular population.

MOLECULAR MARKER-BASED ANALYSIS OF INBREEDING DEPRESSION

As detailed above, most information on the impact of inbreeding has historically come from either controlled crosses (regressing trait means on f) or from data on an observed (or inferred) single generation of inbreeding in nature. One could also examine inbreeding depression using pedigree data, with inbreeding estimates, f_{ped}, following from the pedigree coefficients of coancestry (Chapter 7; Equation 12.5c). Unfortunately, these estimates have potential problems beyond the obvious difficulty in obtaining pedigrees for most natural populations. First, they depend on both the accuracy and the depth of the pedigree, as the assumption is that the base of the pedigree (the founders) consists of unrelated and outbred individuals, resulting in f_{ped} tending to underestimate the true value of f. Fortunately, a pedigree going back only 3–5 generations tends to capture most of the recent inbreeding, so that in many settings, this underestimation is often small (Balloux et al. 2004; Keller et al. 2011; but see Ferenčaković et al. 2017). Second, there is a Mendelian sampling variance associated with f (and other relationships in general; Chapters 7 and 8), such that f_{ped} represents an *expected* value, about which there can be considerable variation for any particular *realization* (Hill and Weir 2011, 2012). Hence, two individuals with the same predicted value of f_{ped} can have realized values that are rather different, a fact that can be exploited to control for non-genetic factors, such as using within-sib comparisons (e.g., Clark et al. 2019).

Runs of Homozygosity (ROHs)

As developed in Chapter 8, marker-based estimates of f attempt to address these issues: they can estimate the realized f values with or without a pedigree, and can be applied to essentially any population. Thus, in theory, one could examine the impact of inbreeding depression from a random collection of individuals from a target population, *provided* that there is *sufficient variation* in the level of inbreeding in the sample.

Recall that marker-based estimates of f are either based on genome-wide averaging over single marker (point) estimates (e.g., f_{HOM}, f_Y; Equations 8.19–8.22) or by using very dense markers to detect local runs of homozygosity, ROHs (Broman and Weber 1999; Chapman and Thompson 2003; McQuillan et al. 2008; Keller et al. 2011; Peripolli et al. 2016; Druet and Gautier 2017; Ceballos et al. 2018; Yengo et al. 2021). A ROH is defined as a continuous DNA segment that is completely homozygous, leading to an estimate of the fraction of the genome that is autozygous (inbred) as

$$\widehat{f}_{ROH} = \frac{\text{total length of ROHs}}{\text{genome size}} \tag{12.41}$$

Table 12.8 ROH-based estimates of the scaled strength of inbreeding, I_s (as defined in Table 12.2) for selected traits in cattle and pigs. Here $*$ and $**$ denote,, respectively, significance at the 5% and 1% levels.

Character	I_s	Reference
Holsteins		
205-day milk yield	0.236^{**}	Bjelland et al. 2013
Daily milk yield	0.250^{**}	
Average fat %	0.028	
Average protein %	-0.033	
Iberian pigs: Number of piglets at birth		
ROH $\geq$ 0.5 MB	0.313^{*}	Saura et al. 2015
ROH 0.5 to 5 MB	-0.760	
ROH $\geq$ 5 MB	0.254^{*}	
Iberian pigs: Number of piglets born alive		
ROH $\geq$ 0.5 MB	0.325^{**}	Saura et al. 2015
ROH 0.5 to 5 MB	-0.482	
ROH $\geq$ 5 MB	0.256^{*}	

As discussed in Chapter 8, typically a region is declared to be an ROH when it exceeds some size threshold, based on either physical distance (base pairs) or genetic distance (expected number of recombinants). Long ROHs are consistent with recent inbreeding, while short blocks of ROH are generated by distant inbreeding, although sufficiently short ROHs could also be allozygous (outbred) regions, due to the inheritance of a set of markers as a single linkage block. Threshold size represents a tradeoff between the control of false positives (longer ROHs have lower false-positive rates of autozygosity) versus reduced power (ignoring shorter runs that, cumulatively, may be biologically important). It is also a tradeoff between a few recent (and potential rather large) events (long ROHs) versus a deeper and more complete sampling of pedigree ancestors to capture more ancient events (short ROHs). This lack of a clearly defined threshold creates some amount of arbitrariness for ROH-based estimates of f, as the smaller the threshold ROH value, the more ancient the assumed base population (Chapter 8). Finally, it is often assumed that the genomic distribution of ROH lengths (under inbreeding) should be roughly exponential (Chapter 8), but Yengo et al. (2021) found that the distribution in humans was better fit by a mixture of two exponentials, likely reflecting genomic variation in recombination rate.

Because of the higher frequency of homozygotes in inbred individuals, offspring from the mating of close relatives have higher risks for presenting recessive Mendelian (single major gene) diseases (Vogel and Motulsky 1996). Similarly, several complex diseases show correlations with ROHs, with a higher fraction of the genome showing ROHs in disease cases as opposed to controls (non-diseased), an observation often called the **homozygosity burden** (reviewed by Ku et al. 2011; Ceballos et al. 2018; Pemberton and Szpiech 2018). For example, Keller et al. (2012) examined roughly 22,000 individuals ($\sim$9,400 cases, $\sim$12,500 controls) and found that the odds of schizophrenia increased by about 17% for each 1% increase in the fraction of the genome that showed ROHs. This implies a 2.7-fold increase for first-cousin inbreeding, and a 1.29-fold increase for second-cousin inbreeding. Similarly, Howrigan et al. (2016) found an association between ROH and lower general cognitive ability, with most of the signal from long, rare ROH (suggesting an impact from very recent inbreeding). However, Johnson et al. (2016), using a sample size almost twice as large as Keller et al. ($\simeq 40,000$), were unable to replicate their correlation between ROH and schizophrenia risk. This is a strong cautionary tale in that the case-control ascertainment corrections suitable for GWAS (essentially, control for between-sample allele-frequency differences; Chapter 20) may not be sensitive enough to control for more subtle population structures (such as very small differences in recent inbreeding) required when using ROH.

A single generation difference in inbreeding rates between cases and control that would not generate an allele-frequency difference (and hence would not needed to be adjusted by a GWAS) can create a false ROH-trait association.

As with controlled cross data, one can regress individual trait values on their estimated f_{ROH} values, with (as before) the slope of the regression being the expected inbreeding depression in a fully inbred individual (assuming no epistasis). Table 12.8 presents a few such estimates from domesticated animals. The results of Saura et al. (2015) are especially interesting, as they compared whether most of the inbreeding depression signal was coming from short ROHs (distant ancestors) or long ROHs (more recent ancestors). For the two traits they examined, most of the signal came from longer ROHs. However, the effective population size of their pig population was very small, so this could simply be a manifestation of the observation of Keller et al. (2011) that power for short ROHs decreases with N_e. Conversely, for semen quality traits in cattle, Ferenčaković et al. (2017) found an ROH threshold size of 2MB gave a better model fit than a threshold of 4MB, so that much of the ID signal was coming form short ROH regions.

Just as the ROH length associated with ID provides information about its underlying genetic architecture, so too does which f statistic is associated with the bulk of the ID. As noted by Clark et al. (2019), point-based estimates (such as genome-wide averages of excessive marker homozygosity) have low power to detect ID effects from nearby rare alleles, as LD is very weak between a common (such as most SNP markers) and rare allele (Example 20.1). As a result, only a small fraction of excess marker homozygosity is likely to translate into excess causal site homozygosity. Conversely, ROH can detect signals from both rare and common alleles when they are made homozygous by being contained within a ROH. Both Clark et al. (2019) and Yengo et al. (2021) used this approach to probe the genetic basis of ID in humans. Clark et al. examined a 100 traits, 21 of which had sample sizes over 500,000. They found that the signal from the 32 traits with significant ID (involving reproductive success, risky behaviors, cognitive ability, body size, and health) was mainly due to ROH ($>$ 1.5 MB), but not to homozygosity tagged by common variants (point-based estimators of f). Hence, the ID was mainly from rare, rather than common, alleles (a topic returned to in Chapters 20 and 21). They noted that this observation was consistent with the partial dominance model (with deleterious recessives kept at low frequency by selection), but not the overdominant model (where causal alleles frequencies are predicted to be much higher, and thus accessible to being tagged by common SNPs). Conversely, Yengo et al. found a more trait-specific pattern. Some of their 11 traits (from a UK Biobank sample of 350,000) had the majority of their ID accounted for by SNP-based point estimates, while others had ROH being more important.

Homozygosity and ROH Mapping

In 1953, C. A. B. Smith noted that inbred offspring could be used to map recessive disease genes by looking for an excess of homozygotes in a candidate gene region in cases versus controls. However, he quickly dismissed this idea as being wholly impractical given the scarcity of scorable human markers. Bolstered by the ever-increasing availability of such markers, Lander and Botstein (1987) independently proposed this idea and named it **homozygosity mapping**. Morton (1991) rightly noted that this is better called **autozygosity mapping**, as the idea is to search for regions of excess autozygosity in cases versus controls.

A major stumbling block for linkage-based mapping (Chapters 5 and 19) of Mendelian recessive disorders (single recessive genes that have a major impact on disease risk) is that cases are rare in natural populations (on the order of q^2, the squared frequency of the disease allele). As a result, it may be difficult to obtain a sufficient number of the multiple-case families that are needed for traditional linkage analysis (Wong et al. 1986). Most cases are usually found as singletons, and hence contain little traditional linkage information (Chapters 5, 17, and 19; Ott 1991). For example, Lander and Botstein (1987) noted (at the time of their publication) that there are around 100 living cases of Bloom's syndrome, but only nine known families with two or more cases (eight with two and one with three),

providing very little traditional linkage information.

However, under homozygosity mapping, even singletons contain useful information. As noted by Lander and Botstein, the fraction of all cases from autozygosity in inbred individuals is $fq/[fq + (1-f)q^2]$, which is $\simeq 1$ when $q \ll f$. Conversely, the chance that a random location is autozygous is f, for an approximate odds ratio of $1/f$ in support of linkage for each singleton. For example, a singleton case from a first cousin marriage ($f = 1/16$) contains the same amount of linkage information as a family with three affected offspring (Lander and Botstein 1987). In the above-mentioned case of Bloom's syndrome, there are at least 24 cases from marriages of cousins, offering far more information that from the limited number of multiple-case families. Thus even a modest number (10 to 40) of inbred singletons quickly gives significant power to test a particular candidate region (Lander and Botstein 1987). Extensions, and discussions on limitations, of this basic approach are provided by Mueller and Bishop (1993), Houwen et al. (1994), Kruglyak et al. (1995), Miano et al. (2000), Botstein and Risch (2003), Laurier et al. (2006), Leutenegger et al. (2006), and Hildebrandt et al. (2009). In particular, autozygosity can be assessed using ROHs (e.g., Lencz et al. 2007).

VanRaden et al. (2011) and Sahana et al. (2013) extended this idea to map regions in cattle that contain lethals (or synthetic lethals). These would appear as haplotypes that are relatively common, but which never appear in ROHs. Similar logic can be applied to crudely localize QTLs for complex traits by using **ROH mapping**. For a binary trait, one searches for chromosomal regions that are enriched for ROHs in cases relative to controls. For continuous traits, one regresses trait value on the present/absence of ROHs in specific regions or around specific SNPs (Lencz et al. 2007; Kardos et al. 2016; Ferenčaković et al. 2017). The basic idea follows from standard GWAS methods (Chapters 20 and 21). For example, Ferenčaković et al. (2017) modified the basic GWAS linear model (for the effect of SNP j) as follows

$$z_i = \mu + I\, f_i + \beta_{1j}\, x_{ij} + \beta_{2j}\, \mathrm{ROH}(ij) \tag{12.42}$$

Here z_i is the trait value of individual i, f_i its level of genome-wide inbreeding, x_{ij} is number of copy of the minor allele at SNP j in individual i ($x_{ij} = 0, 1, 2$), and $\mathrm{ROH}(ij)$ is an indicator variable, equaling one when SNP j in individual i is covered by a ROH. Using this approach, they found 53 SNPs associated with inbreeding depression (β_{2j} significantly less than zero) for cattle semen quality. Similar ROH mapping approaches were used by Pryce et al. (2014) to localize specific regions of ROHs that resulted in inbreeding depression in days for calving interval and milk yield, and by Saura et al. (2015) for porcine regions influencing fertility (number of piglets).

It is, however, important to note the difference between the architecture for ID and the architecture for total trait variation. Chapters 17–21 detail approaches for probing the latter, while ROH mapping focuses on only that gene subset of the full architecture generating an ID signal (i.e., segregating sites displaying directional dominance). If rare recessives are a key part of ID, then there is likely more of a stochastic component to ID gene signals relative to the full architecture. For example, Howard et al. (2015) found that different ROH regions impacted ID for Jersey production traits in US vs. Australian cattle populations, with the majority of regions at low frequency (as might be expected from sporadic, large-effect, and rare, variants). Zhang et al. (2019) also observed in potatoes that large-effect regions involved in ID also tended to be line-specific. They also noted that they tend to be found in genomic regions of high recombination. In humans, Yengo et al. (2021) also found that genomic regions involved in ID are 20-fold enriched in regions of higher recombination. They also found that ID-impacting regions were similarly enriched for conserved sites (those presumably under strong functional constraints).

Which Marker-based Estimate Should One Use to Detect ID?

As discussed above, there are several options for using markers to look at the impact of inbreeding on trait values. The most basic is the MLH score, looking for correlations between heterozygosity values and trait value (fitness being the classic example). More sophisticated

marker-based estimates, which require specification of base-population properties, were introduced in Chapter 8 (Equations 8.19–8.22), and finally we can detect ROHs as the number of markers becomes sufficiently dense. What is the best marker-based approach to detect, and quantify, ID in a trait? While the answer initially depends on the number of available markers, the optimal choice becomes a bit more nuanced as the number of markers becomes very large.

The least assumption-burdened approach is MLH score, the number of marker loci which are heterozygous in an individual (H), or its standardized value, H_s (Equation 12.39a). Note that an MLH score does not require any knowledge of base-population allele frequencies. As a result, these heterozygosity measures *do not* directly estimate f, but rather attempt to *rank individuals in the sample* from lowest to highest f values, a procedure that is only appropriate if there is some variance in inbreeding levels, namely that $\sigma^2(f)$, and therefore g_2 (Equation 12.37), are nonzero. As might be expected, this is a crude measure of the amount of inbreeding, as MLH and f, while correlated when $g_2 > 0$, do not measure the same quantity. A good example of this is seen in the work of Slate et al. (2004) on Coopworth sheep in New Zealand. They examined ~600 sheep with a 7–10 generation pedigree that were also scored for 138 SSRs markers (roughly equally spaced over the 26 autosomes), finding that MLH was only weakly correlated with pedigree f ($r^2 \simeq 0.03$). Further, pedigree f, but not MLH score, detected evidence of inbreeding depression on several traits.

While methods based on genome-wide averages of single-point estimates of f (Equations 8.19–8.22) are clearly an improvement over MLH approaches—as they directly return estimates of f for a given individual—these have low accuracy when the number of markers is less than a few hundred SSRs (or a few thousand SNPs), even when inbreeding is frequent and severe (Balloux et al. 2004; Slate et al. 2004; Pemberton 2004; Carothers et al. 2006). The situation is even worse for populations experiencing less frequent, and less severe, inbreeding, where $\sigma^2(f)$ and g_2 are expected to be very small. Equations 12.40a and 12.40b provide some insight into the accuracy of single-point methods, given some initial estimate of g_2. Although these equations examine the correlation between f and H_s, they still provide a good benchmark for single-point approaches.

Given results like these, it was not surprising that pedigree-based estimates, f_{ped}, were regarded until recently as the gold standard for estimates of f (e.g., Pemberton 2004, 2008). Thus, when one has access to only a very modest number of markers, but has a fairly accurate pedigree in hand, use pedigree-based estimates of f, potentially using the markers to check for any obvious pedigree errors (Chapter 8). As the number of markers further increases—say at least 10,000 SNPs are scored—in most settings single-point estimators become quite accurate, accounting for upwards of 90% of the true f values, and become preferred over pedigree-based estimates (Kardos et al. 2015).

As the marker density continues to increase to the point where accurate estimates of ROHs become feasible, the natural questions becomes which approach, single-point averages versus ROHs, is more powerful? There is actually considerable subtlety in the answer. In part, one tradeoff between these approaches depends on the reference population, as single-point methods require base-population allele frequencies, while ROH does not. However, the chosen size threshold for ROHs is, itself, an assumption about the base population, but rather than on its allele frequencies, on the time scale that we regard as setting the base. It might seem that ROH-based approaches might be preferred because they appear to directly measure f, rather than infer it by the fraction of excessive homozygotes or excessive correlation among alleles at a locus, and they are rather insensitive to allele frequency estimates. Indeed, several authors have suggested that ROH is favored over single-point methods (Keller et al. 2011; Gazal et al. 2014; Solé et al. 2017; Nietlisback et al. 2019).

However, simulations by Yengo et al. (2017) found that modifications of single-point-based estimates of ID can be less biased than ROH-based estimates, especially when a very dense SNP set is available (such as the millions used by human geneticists). They first noted a subtle problem with single-SNP based methods: the allele frequency and LD structure

of marker SNPs might not reflect that of the causative SNPs, which can introduce bias (Chapters 20 and 21). They resolved this concern by binning (or **stratifying**) marker SNPs into a small number, ℓ, of classes, based on both their minor allele frequencies and **LD score**. The latter is given by the sum of the pairwise correlations between a focal SNP and all scored SNPs within a defined distance (Equation 20.13b), and provides a measure of the LD structure around a given SNP (Yang et al. 2010; Bulik-Sullivan et al. 2015). They found that such a stratification largely removed any bias for the correlation-based estimator (f_Y; Equation 8.22a), but not the homozygosity-based estimator (f_{HOM}, Equation 8.19c). They then used Equation 8.22a to estimate f within each of the LD-minor allele frequency classes, an approach they refer to as **LD and minor allele frequency stratified inference** (**LDMS**). A multiple regression was then performed, with the trait value of individual i modeled by

$$z_i = \mu + \sum_{j=1}^{\ell} b_j f_{i,j} + e, \quad \text{with} \quad I = \sum_{j=1}^{\ell} b_j \tag{12.43}$$

where $f_{j,i}$ is the estimate of f for individual i using SNPs from LDMS category j. An overall estimate of the impact, I, from inbreeding depression (the change in mean under complete inbreeding) is obtained by the sum of the category-specific b_j estimates. They found that ROH estimates of ID were consistently larger (upwardly biased) relative to those based on LDMS-corrected SNP estimators, but that the latter produced more significant examples of ID (due to f_Y having a lower variance that F_{ROH}; Yengo et al. 2018).

Simulations by Caballero et al. (2021) partly resolved this apparent discrepancy between point-based and ROH estimates. They found that in low N_e settings (such as used by Nietlisback et al. 2019), f_{ROH} gave less biased estimates of ID than SNP-based methods. Conversely, with larger N_e (such as used in the simulations of Yengo et al. 2018), SNP-based methods were better behaved. Further complicating matters, Villanueva et al. (2021) noted that point-based estimators can sometimes given inconsistent values. In the case of these current uncertainties, our suggestion is to compute and then report, ID estimates using a number of different approaches.

Example 12.8. Bérénos et al. (2016) examined data from a long-term study of an isolated population of Soay sheep on the Scottish island of St. Kilda, followed since 1985. Information was available on roughly 6700 individuals, most of which could be placed into a pedigree. Pedigree- and genomic-based estimators were compared using ~37,000 SNPs. The table below presents the correlations among various estimators of f: f_{ped} is the pedigree-based estimator (Equation 12.5c); f_Y is Yang's correlation estimator (Equation 8.22a); f_{HOM} is a homozygosity estimator (Equation 8.19c); and f_{ROH} is a ROH estimator (Equation 12.41) using a 5 MB threshold. Values above the diagonal are correlations based on individuals with four known grandparents, while the values below the diagonal are based on a larger set of individuals where only both parents and one maternal grandparent were known.

	f_{ped}	f_Y	f_{HOM}	f_{ROH}
f_{ped}		0.72	0.67	0.77
f_Y	0.65		0.92	0.87
f_{HOM}	0.60	0.91		0.86
f_{ROH}	0.71	0.86	0.83	

Notice that the ROH estimator had the highest correlation with the pedigree-based estimator (although both point-wise estimators were also highly correlated with f_{ped}), and that all three genomic-based estimators were more correlated with each other than with pedigree-based estimate. Using the SNP marker data, the estimated value of g_2 was 0.0014 ± 0.0002. Assuming an average heterozygosity of 0.2, Equation 12.40a suggests that the correlation between one of the point-wise estimates (f_H) and the true f was approximately

$$\rho^2(f, f_H) = \frac{37000 \cdot 0.0014 \cdot 0.2}{1 - 0.2 + 37000 \cdot 0.0014 \cdot 0.2} \simeq 0.93$$

Using these different f estimators, the impact of inbreeding depression was scored on three fitness-related traits (birthweight, hindleg length, and weight at 4 months). The table below presents estimates of $I_s = (\bar{z}_O - \bar{z}_I)/\bar{z}_O$, the fractional reduction in the mean under complete inbreeding (Table 12.2) and its standard error. These values were based on whether the level of inbreeding in the individual or its mother was the regression variable, and also partitioned over the method used to estimate f. Significant results at the 5% and 1% levels are, respectively, denoted by $*$ and $**$.

Trait	Focal	f_{ped}	f_{HOM}	f_Y	f_{ROH}
Birthweight	Individual	−0.42 (0.16)	−0.26 (0.29)	−0.06 (0.11)	−0.07 (0.14)
	Maternal	0.01 (0.34)	−1.49 (0.49)**	−0.48 (0.20)**	−0.90 (0.26)**
Hindleg	Individual	−0.18 (0.06)**	−0.30 (0.10)**	−0.09 (0.04)**	−0.12 (0.05)**
	Maternal	−0.07 (0.09)	−0.24 (0.15)	−0.09 (0.06)	−0.11 (0.08)
4 m. weight	Individual	−0.63 (0.17)**	−0.83 (0.26)**	−0.31 (0.11)**	−0.42 (0.14)**
	Maternal	−0.60 (0.27)*	−1.20 (0.42)**	−0.43 (0.17)**	−0.64 (0.22)**

Note that f_{HOM} estimators usually gave the largest values for the amount of inbreeding depression (but also had the largest standard errors), while f_Y estimators gave both the smallest standard errors and smallest estimated values among the genomic-based methods. The genomic-based estimators (f_{HOM}, f_Y, f_{ROH}) were more powerful than the pedigree estimate (f_{ped}) in detecting maternal effects. As mentioned by the authors, part of the increased power under genome-based methods is due to larger sample size, as pedigree-based estimators included only those individuals with pedigree information.

Marker-based Estimates of Inbreeding Depression in Natural Populations

In population surveys on the impact of inbreeding, the amount of variation in f in the sample is critical. If individuals in the sample show too little variation in f, there is very little power (g_2 is too small; Equation 12.37c), requiring a modified sampling design—intentionally enriched for inbreds—to have any power. Examples of the latter approach would be to sample a series of smaller, isolated populations, or from populations known to have higher rates of consanguineous matings.

Keller et al. (2011) found that single-point estimators (Equations 8.19–8.22) showed the highest population variation in f, while pedigree-based estimates showed the lowest (in large part because f_{ped} is an *expectation*, not a *realization* with its inherent variance about its mean). The population variance in f based on ROHs showed intermediate values. Using ROHs, Keller et al. found that very large samples (10,000 to over 60,000) from a randomly mating human population are needed for sufficient variation in f to have power to detect the impact of inbreeding depression. In a study of 16 health-related traits using roughly 350,000 individuals from 102 cohorts, Joshi et al. (2015) found correlations between height, lung volume, cognitive ability, and educational attainment with fraction of the genome in ROHs, implying the presence of directional dominance in these traits. A number of other metabolic traits did not show any such associations. When the expected f value among inbreds is suspected to be small, the ROH threshold should be set lower to capture the effects of distant inbreeding contained in shorter blocks of ROH, although this approach may have poor power if N_e is small (Keller et al. 2011).

Another concern, especially in human populations, is the role of **confounding effects**. Socioeconomic and cultural factors can influence *both* a trait of interest *and* the level of inbreeding within a group (the propensity for mating among relatives). One example of such **social confounding** was seen in the Dutch study by Abdellaoui et al. (2013) that initially found a *negative* correlation between ROH and major depressive disorder (MDD). However, when the authors did a more detailed examination of their data, they found that much of the signal was coming from religiosity (R), with individuals scoring high

on this trait having increased ROH (due to a higher chance of pairing with like-minded individuals), but also a lower MDD risk. Hence, $MDD \leftarrow R \rightarrow ROH$. In a second study from the same group, Abdellaoui et al. (2015) examined the level of educational attainment. While there was a barely significant ($p = 0.045$) negative association between the genomic fraction of ROHs and the educational attainment of a focal individual, the offspring of more educated parents had fewer ROHs. Abdellaoui et al. were able to show that this was entirely a function of the physical distance between parental birthplaces, with individuals achieving a higher level of education tending to travel further, and, therefore, were less likely to mate with (distant) relatives, than individuals who did not migrate. Clark et al. (2019) suggested that one design to control for social confounding is to contrast ROH levels *within sibships*, finding that ID effects estimated from such a design averaged around 20% smaller than estimates based on an uncontrolled design.

A final issue is that most studies of the impact of inbreeding on fitness measure a single episode of selection, which can give a misleading picture (WL Chapter 29). An example of this is the work of Grueber et al. (2010) on a New Zealand population of the flightless rail, the Takahē (*Prophyrio hochstetteri*). While inbreeding depression was small at most of the individual life-history stages, its cumulative effect was quite substantial. Hence, estimates of the impact of inbreeding depression based on a single fitness component (typically early survival) can significantly underestimate the total impact of inbreeding depression (O'Grady et al. 2006).