4

Properties of Single Loci

Version 22 August 2022

The fact that most principles of quantitative genetics can be expressed without reference to specific genes is precisely why quantitative-genetic analysis is so popular among those who study complex characters. Because this same feature can be cause for suspicion, a primary goal of the next few chapters is to clarify the ways in which quantitative genetics is grounded in fundamental Mendelian concepts. Prior to illustrating the connections between the properties of single genes and the expression and transmission of polygenic traits, we review some very basic and essential vocabulary.

It is well known that the genetic information encoding for characters resides on extremely long strands of deoxyribonucleic acid (DNA) molecules called **chromosomes**. We are still deciphering the function of the vast majority of DNA in organisms, and many believe that a substantial portion of it has no function, at least in multicellular species (reviewed in detail in WL Chapter 8). Historically, the term **gene** referred to a DNA sequence that encodes for a particular product (a protein or RNA), with their chromosomal locations called **loci**. Most organisms have two copies of each of several chromosomes, in which case they are said to be **diploid**. Because DNA replication is an imperfect process, mutations arise, and as a consequence the two "copies" of each gene carried by diploid individuals need not be identical. The various forms of a gene are called **alleles**. This historical definition of a gene, which was largely based on the ability to score functional variants, has been replaced in the genomics era with a much broader definition given that we can now directly score variation in the DNA itself. While it is straightforward to access whether a sequence codes for a protein or a large RNA, we are still in the early stages of detecting all of constellation of sequences that can influence gene regulation (Chapter 21). For example, a single base change in a noncding sequence may influence the expression of a gene on a different chromosome. The former region would be a trait locus in the sense that it can have a functional role and result in trait variation.

Gene loci that exhibit more than one allele are the subject of genetics. Such loci are said to be **polymorphic**, whereas loci at which all gene copies are identical are **monomorphic**. A substantial fraction of the gene pool in many species is polymorphic. The possible reasons for this are the subject of a long-standing debate in population genetics and molecular evolution (Kimura 1983; Gillespie 1991; examined in detail in WL Chapter 8). Many mutant alleles are extremely deleterious and are rapidly eliminated by natural selection, while others have only small or no effects at the phenotypic level and remain in the population until they are fixed or lost by chance (WL Chapter 2). Still others are maintained at intermediate levels by a balance between opposing evolutionary forces (WL Chapter 7).

Not all organisms are diploid. Prokaryotes have only a single copy of each gene and are referred to as **haploid**. Many of the protists and lower plants (algae, mosses, and ferns) also have conspicuous haploid stages in their life cycles, as do the fungi and some animals (males of both rotifers and haplo-diploid insects). Organisms with ploidy levels higher than diploid are known as **polyploids**. A **tetraploid** individual contains four sets of homologous chromosomes, whereas a **hexaploid** contains six. Polyploidy is extremely widespread among plants. It is relatively rare among sexual animals, but common among parthenogenetic species.

Even in diploids, some genes are effectively haploid. Such is the case for genes carried in organelles (mitochondria and chloroplasts). Although there may be hundreds of

copies of organelle genes per cell, they are generally inherited uniparentally and are essentially all the same. Genes residing on the **sex chromosomes** of organisms with a genetic sex-determination mechanism also have a special ploidy status. In mammals, for example, males carry X and Y chromosomes (and are said to the **heterogametic** because of the mismatching sex chromosomes), whereas females are XX, so that X-linked genes occur only in single "copies" in males. In some organisms, such as birds, moths, and butterflies, the heterogametic (WZ) sex is female, while males are ZZ. In order to distinguish sex chromosomes from the remaining pairs, the latter are referred to as **autosomes**. In this book, unless stated otherwise (see especially, Chapter 29), we will be dealing with autosomal loci in diploid populations.

The remainder of this chapter is concerned with the quantification of various properties of single loci. We start by reviewing the concepts of allele and genotype frequencies, showing how the two are connected in an ideal situation that is closely approximated in many natural settings. We next show how the phenotypic effects of different alleles can be described in terms of additive and dominance effects. The genotypic frequencies and effects are then incorporated into expressions for the additive and dominance components of genetic variance at the locus. Finally, we show how the additive effects of an individual's genes define its breeding value for any particular trait. While several of the concepts covered in this chapter may seem rather abstract and far removed from the analysis of multilocus traits, they are becoming increasingly tractable as we move from the use molecular methods localize **quantitative-trait loci** (**QTLs**). These are small chromosomal regions showing variation that influence trait values. Under classic QTL mapping using linkage approaches (Chapters 5, 18, and 19), these regions are on the megabase (10^6 nucleotides) scale, while finer resolution (on the kilobase scale, 10^4 nucleotides) is offered by **genome-wide association studies** (**GWAS**) which exploit population-level linkage disequilibrium (Chapters 5 and 20). Ideally, the goal is to isolate **quantitative-trait nucleotides** (**QTNs**), but this goal has remained elusive (Chapter 21).

ALLELE AND GENOTYPE FREQUENCIES

When denoting the **genotype** at a single locus, we refer to the pair of alleles that a (diploid) individual carries at the locus. Individuals that have two identical alleles are called **homozygotes**, whereas those that have different alleles are **heterozygotes**. If, for example, we denote the alleles at a particular diallelic locus as B_1 and B_2, there are three possible genotypes: B_1B_1 and B_2B_2 homozygotes, and B_1B_2 heterozygotes. There may, of course, be more than two alleles, and hence more than three genotypes, present at a locus in a population. In particular, with k alleles, there are $k(k+1)/2$ diploid genotypes (k homozygotes and $k(k-1)/2$ heterozygous).

Allele frequencies are defined uniquely by genotype frequencies. Suppose that P_{11}, P_{12}, and P_{22} represent the proportions of the population that are, respectively, B_1B_1, B_1B_2, and B_2B_2. If these are the only possible genotypes at the locus, then by definition, $P_{11} + P_{12} + P_{22} = 1$. If there are N individuals in the population, then $P_{11}N$ individuals contain two B_1 alleles and $P_{12}N$ individuals contain a single B_1 allele. Because there are a total of $2N$ genes in the population for each autosomal locus, the frequency of the B_1 allele is simply

$$p_1 = \frac{2\,P_{11}\,N + P_{12}\,N}{2\,N} = P_{11} + \frac{1}{2}\,P_{12} \tag{4.1}$$

Thus, the general rule for a diploid, autosomal locus is that the frequency of an allele is estimated by the observed frequency of homozygotes plus one-half the observed frequency of all heterozygotes containing that allele.

A direct application of this approach involves **single nucleotide polymorphisms** (**SNPs**), which are usually diallelic, but in theory could segregate as many as four alleles (corresponding the nucleotides A, G, T, and C). Through either hybridization techniques (such as DNA chips) or high-throughput DNA sequencing, one can score the exact frequencies of the genotypes associated with a particular SNP. In the absence of sequence data,

it is usually impossible to be certain about the genotypic state of any particular locus for complex morphological and behavioral characters. In some cases, however, the majority of the genetic variation for a character depends on a single locus with large effects, which may offer the possibility that allele and genotype frequencies can be directly estimated. This was the fortuitous case in many of Mendel's classic experiments with peas, and a number of genetic disorders in humans appear to be products of mutant alleles at single loci. Data for a wing-color polymorphism in a British moth are discussed in the following example.

Example 4.1. Fisher and Ford (1947) were able to distinguish three wing-color patterns in the tiger moth *Panaxia dominula,* and through breeding experiments, the polymorphism was found to result from two alleles segregating at a single locus. WL Examples 9.1 and 9.2 review evidence for selection acting on this specific locus. The following table summarizes the distribution of genotype frequencies observed in a population in 1946.

Color Pattern	*dominula*	*medionigra*	*bimacula*	Total
Genotype	B_1B_1	B_1B_2	B_2B_2	
Sample Size (N_{ij})	905	78	3	N = 986
Frequency (P_{ij})	0.918	0.079	0.003	1.000

What are the estimated frequencies of the two alleles? Using Equation 4.1, the frequency of the B_1 allele is found to be

$$p_1 = 0.918 + \frac{0.079}{2} = 0.958$$

and because there are only two alleles, the frequency of B_2 is $p_2 = 1 - p_1 = 0.042$.

THE TRANSMISSION OF GENETIC INFORMATION

The Hardy-Weinberg Principle

From the standpoint of evolutionary analysis, it is crucial to understand how allele and genotype frequencies change from generation to generation (a full treatment of this topic is given by WL). Such changes may result from natural selection, mutation, differential migration, inbreeding, or random drift due to gene sampling in finite populations. All of these forces will be considered in due course, but for now we restrict our attention to a highly idealized situation—an autosomal locus uninfluenced by selection and mutation. By assuming the population to be effectively infinite in size and randomly mating, we also eliminate the possibility of inbreeding and random drift. We further assume that generations are discrete and that the population is closed to immigrants.

Although such an idealized situation is never realized perfectly, in many cases (such as over modest time scales) it is close enough to the truth for practical purposes. Under the ideal model, simple and predictable relationships emerge between allele and genotype frequencies, within and between generations. It is therefore an essential point of departure, much like the ideal gas laws in physics.

In sexual populations, individuals do not necessarily produce offspring whose genotypes match their own. Prior to reproduction, sexual individuals produce haploid gametes by a special form of cell division called **meiosis** (Figure 4.1). Thus, with respect to a single locus, a B_1B_2 heterozygote **segregates** two types of **gametes**—half B_1 and half B_2. The

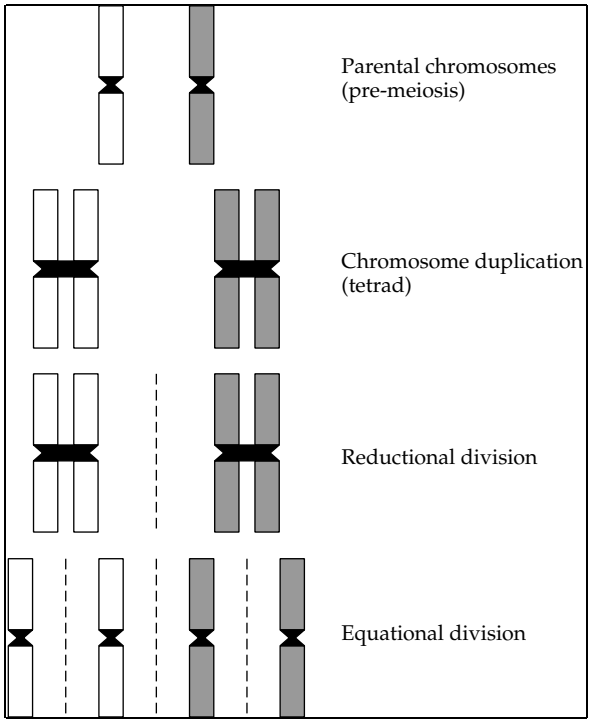

Figure 4.1 Idealized schematic of meiotic production of gametes. Only a single chromosome pair is shown, with the two **homologs** denoted by the different shading. At the onset of meiosis, **sister chromatids** are formed by duplication of each homolog and the homologous pairs come together to form a **tetrad**; although it is not shown, some exchange of material (e.g., **gene conversion** and/or **crossingover**) between homologs may occur at this time. Two meiotic divisions (reductional followed by equational) then produce four haploid products. The maternal and paternal chromosomes migrate to opposite cells during the **reductional division**, and the sister chromatids are isolated into four potential haploid gametes after the **equational division**.

diploid state is restored when gametes from two parents fuse to form a **zygote**. Consequently, at a diallelic locus, a heterozygous parent can potentially produce three types of progeny (B_1B_1, B_1B_2, and B_2B_2), whereas homozygous parents can produce at most two.

Consider a population consisting of separate sexes (**dioecious**) with discrete, nonoverlapping generations. We denote the frequencies of B_1 and B_2 alleles in females in generation 0 by $p_{1f}(0)$ and $p_{2f}(0)$, and those in males by $p_{1m}(0)$ and $p_{2m}(0)$. Under random mating, the expected genotype frequencies in the next generation are obtained from the products of the respective gamete frequencies. For example, because the probability of drawing a B_1 female gamete is $p_{1f}(0)$ and that of drawing a B_1 male gamete is $p_{1m}(0)$, the expected frequency of B_1B_1 zygotes is $p_{1f}(0)\,p_{1m}(0)$. Similarly, the expected frequencies of B_1B_2 and B_2B_2 zygotes are $p_{1f}(0)\,p_{2m}(0) + p_{2f}(0)\,p_{1m}(0)$ and $p_{2f}(0)\,p_{2m}(0)$, respectively. Provided the locus is autosomal, the frequency of the B_1 allele will now be the same in both offspring sexes, because the subpopulations of sons and daughters both acquire half their genes from mothers and half from fathers. Substituting into Equation 4.1, the B_1 allele frequency in generation 1 is

$$\begin{aligned} p_1 &= p_{1f}(0)\,p_{1m}(0) + \frac{p_{1f}(0)\,p_{2m}(0) + p_{1m}(0)\,p_{2f}(0)}{2} \\ &= \frac{p_{1f}(0)\,[\,p_{1m}(0) + p_{2m}(0)\,] + p_{1m}(0)\,[\,p_{1f}(0) + p_{2f}(0)\,]}{2} \\ &= \frac{p_{1f}(0) + p_{1m}(0)}{2} \end{aligned}$$

Likewise, the new frequency for the B_2 allele is $p_2 = 1 - p_1 = [\, p_{2f}(0) + p_{2m}(0)\,]/2$.

Under the conditions of our idealized population, in the next, and in all subsequent, generations, the allele frequencies are both constant and the same in both sexes, with p_1 being the frequency of B_1 and $p_2 = 1 - p_1$ being the frequency of B_2. Further, under random mating (following the logic above), the B_1B_1, B_1B_2, and B_2B_2 genotypes will be found in frequencies p_1^2, $2\,p_1\,p_2$, and p_2^2. Such proportions are known as **Hardy-Weinberg frequencies**, after the two investigators who first pointed out the above relationship (Hardy 1908; Weinberg 1908). The Hardy-Weinberg frequencies can also be obtained directly by multiplying out the terms of the binomial expansion, $(\,p_1 + p_2)^2$. By this means, the Hardy-Weinberg law can be extended to any number of alleles. Suppose, for example, that four alleles (B_1, B_2, B_3, B_4) are present at the locus of interest. The Hardy-Weinberg frequencies for the various genotypes are obtained by squaring the quantity $(\,p_1 + p_2 + p_3 + p_4)$. The expected frequency of a genotype homozygous for the B_i allele is p_i^2, while that for a B_iB_j heterozygote is $2\,p_i\,p_j$. One application of the Hardy-Weinberg (HW) proportions is a quick quality-control (QC) scan for genomic data, as sequencing and genotyping errors can generate departures from HW.

Provided that all of the assumptions of the Hardy-Weinberg model are met, we can summarize its implications as follows. First, it takes no more than a single generation to equilibrate and stabilize the gene (allele) frequencies in the two sexes. Second, only one additional generation is required for the stabilization of the (autosomal) genotype frequencies into the predictable Hardy-Weinberg proportions. These results have obvious implications for the analysis of natural populations. Even if genotype frequencies in a study population are vastly different from Hardy-Weinberg expectations, for example because of natural selection or population subdivision, they can be rendered close to the idealized proportions by imposing an artificial program of random mating for one or two generations.

Sex-Linked Loci

The preceding results do not extend to sex-linked loci. As noted above, when the male is the heterogametic sex, females are diploid for X linked loci, but males are haploid. Thus, for every mating pair, there are three X chromosomes, and the frequency of the B_1 allele in the population is $p_1 = [\, p_{1m}(0) + 2p_{1f}(0)]/3$. In the absence of any forces operating differentially on the alleles, this frequency will be maintained indefinitely. However, the gene frequency will not necessarily be p_1 in both of the sexes. Because males only receive an X chromosome from their mother, the male frequency of the B_1 allele in any generation (t) is necessarily equal to the frequency in females in the previous generation $(t-1)$,

$$p_{1m}(t) = p_{1f}(t-1) \tag{4.2a}$$

On the other hand, fathers and mothers each contribute an X chromosome to their daughters, so the frequency of the B_1 allele in females is equal to the average gene frequency across the two sexes in the previous generation,

$$p_{1f}(t) = \frac{p_{1f}(t-1) + p_{1m}(t-1)}{2} \tag{4.2b}$$

The general solution to these equations is

$$p_{1f}(t) - p_1 = \left[-\frac{1}{2}\right]^t \left[\, p_{1f}(0) - p_1\,\right], \qquad p_{1m}(t) = p_{1f}(t-1) \tag{4.2c}$$

Thus, the approach to the equilibrium allele frequency ($p_{1f} = p_{1m} = p_1$) in the two sexes is gradual and oscillatory if the locus is X linked (Figure 4.2). The deviation of the allele

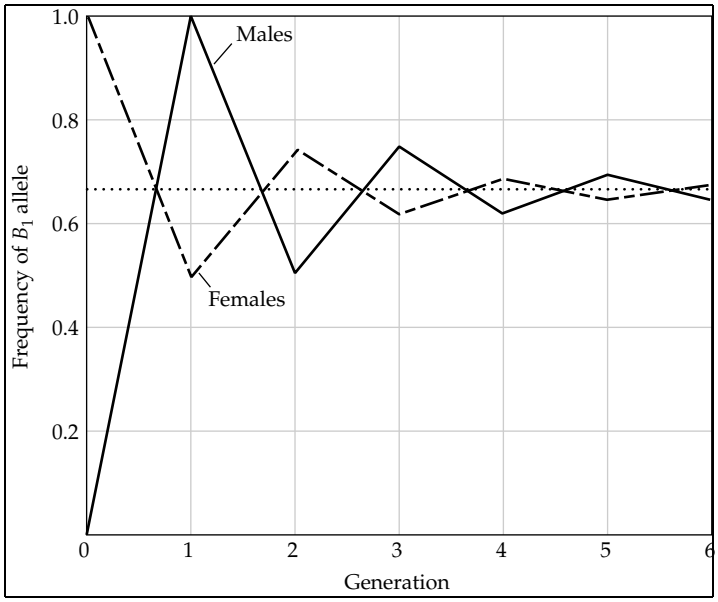

Figure 4.2 The dynamics of gene frequency change for an X-linked gene, B_1, under random mating. An extreme case is illustrated—initially, all females are homozygous for the B_1 gene, $p_{1f}(0) = 1$, while all of the males are haploid for the alternate allele, $p_{1m}(0) = 0$. Consequently, all males contain the B_1 allele in the following generation, while all females are heterozygous. The dotted line represents the population level gene frequency, $p_1 = [\,p_{1m}(0) + 2p_{1f}(0)\,]/3 = 0.67$, towards which both of the sexes converge over time.

frequency from p_1 is halved each generation for both males and females, but the sign changes from generation to generation.

Polyploidy

Another situation in which the Hardy-Weinberg principle is not met exactly arises in polyploid organisms. Because of the high frequency of polyploidy in plants, this case has been examined extensively by Fisher (1947) and Crow (1954) among others. It will only be considered briefly here for a tetraploid species, individuals of which propagate two genes per locus through gametes. The way in which sets of chromosomes assort during meiosis in polyploids depends on the degree of homology between ancestral chromosomes (Marsden et al. 1987). At one extreme are **allopolyploids** that originate by interspecific hybridization. In this case, provided the chromosomes of the parental species are sufficiently different, they will not pair. Meiosis is then identical to that for diploid organisms, except for the doubled number of chromosomes. At the other extreme, **autopolyploids** derive both chromosome sets from the same species. In this latter case, complications for HW can arise because full sets of four, rather than two sets of two, chromosomes can pair during meioses.

For the remainder of our discussion of polyploidy, we will assume that the four sets of chromosomes are sufficiently similar that **tetravalents** (combinations of four homologes), rather than **bivalents**, are formed during meiosis. This condition raises the possibility that some gametes will contain two copies of one of the four genes carried by the parent (i.e., a parent with genotype $B_1B_2B_3B_4$ may produce a B_1B_1 gamete), a result that arises when a **crossover** (a reciprocal exchange of DNA) occurs between replicated arms of two of the four chromosomes during meiosis. The production of such a gamete is referred to as a **double reduction**, and we denote its probability by c. Of the $(1 - c)$ gametes that are not doubly reduced, one-third will contain genes that came from the same parent, and the other two-thirds will contain one paternally derived and one maternally derived gene (Figure 4.3). Here we assume the presence of only two alleles and random assortment of the four homologs. Letting, p_i be the frequency of the B_i allele and $p_{ij}(t)$ be the frequency of B_{ij}

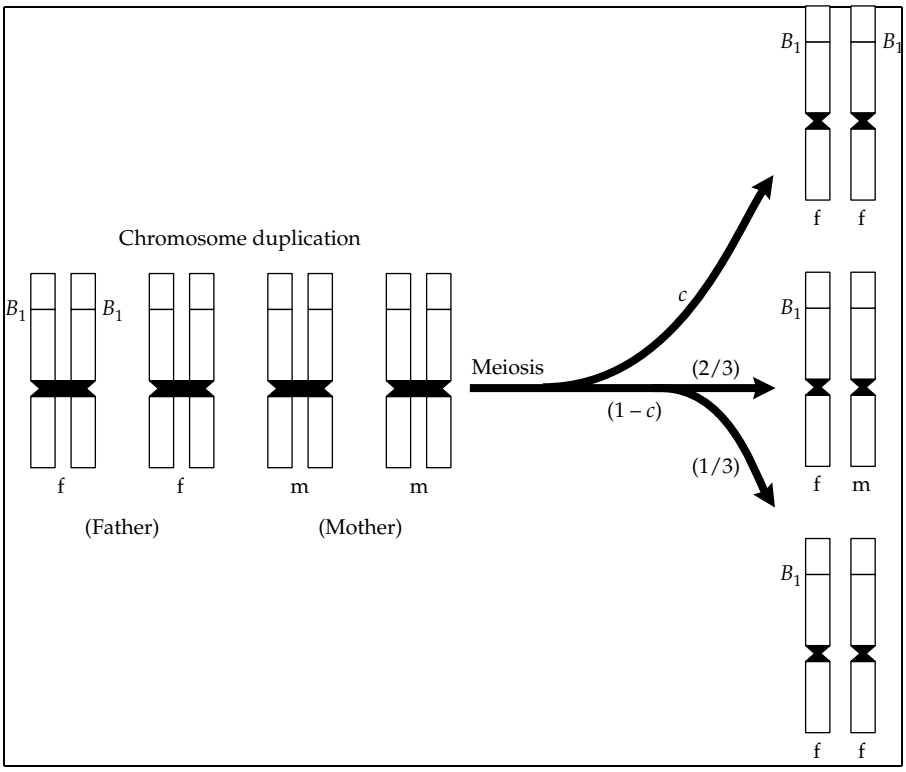

Figure 4.3 The production of three types of (diploid) gametes by a tetraploid individual. This example focuses upon a single paternally derived allele, B_1. The letters f and m refer to chromosomes in an indivdual derived from its father and mother. With four chromosomes (rather than the two of a diploid), the reductional division of meiosis isolates two chromosomes at random into each of the resulting two cells. Subsequent equational division generates the gamete types shown at the right. Here c is the probability that the allele of interest will become associated with itself during gametogenesis as a result of a double reduction. If this does not occur (with probability $1 - c$), there is a $2/3$ chance that chromosome B_1 will be associated in a gamete with a maternally derived chromosome and a $1/3$ chance that it will be associated with the other paternally derived chromosome.

gametes in generation t, the following dynamic equations hold:

$$p_{ii}(t) = c\,p_i + \frac{1-c}{3}\left[\,p_{ii}(t-1) + 2\,p_i^2\,\right] \tag{4.3a}$$

$$p_{ij}(t) = \frac{1-c}{3}\left[\,p_{ij}(t-1) + 2p_ip_j\,\right] \tag{4.3b}$$

(Crow and Kimura 1970, pp. 52–53). The equilibrium solution to these equations is obtained by setting $p_{ii}(t) = p_{ii}(t-1)$ and $p_{ij}(t) = p_{ij}(t-1)$,

$$p_{ii} = (1-f)\,p_i^2 + f\,p_i \tag{4.3c}$$

$$p_{ij} = (1-f)\,p_i\,p_j \tag{4.3d}$$

where $f = 3c/(2+c)$. This equilibrium is approached only gradually. The equilibrium genotype frequencies can be obtained as products of the appropriate equilibrium gametic frequencies.

In the absence of crossing-over between homologous pairs of chromosomes, $c = 0$, $f = 0$, and the equilibrium frequency of gamete types is simply equal to the product of the

respective allele frequencies. However, if $c > 0$, the equilibrium genotype frequencies are not so simple. Consider the extreme case of free recombination (which is approximately the case when a locus is sufficiently far from its centromere). After chromosomal duplication during gametogenesis, eight chromosomes are assorted, two into each of four gametes. Conditional on one of these being transmitted to a gamete, then of the remaining seven possibilities, one will be identical by descent. Thus, for free recombination, $c = 1/7$, $f = 0.2$, and the equilibrium gamete frequencies are $p_{ii} = 0.2p_i(1+4p_i)$ and $p_{ij} = 0.8p_ip_j$. In essence, if there is any crossing-over, polyploidy results in a sort of "internal inbreeding," reducing the frequency of heterozygous gametes. Wricke and Weber (1986), and especially Gallais (2003), provide a very useful coverage of the many complications that polyploidy introduces in quantitative-genetic formulations.

Age Structure

One final complication with respect to the idealized model is age structure. Up to now we have been assuming a population with discrete, nonoverlapping generations, such as an annual plant with no seed carry-over across years or a univoltine insect. In populations composed of several age classes (the majority of land plants and animals), the generations overlap, and this causes the approach of genotype frequencies towards the Hardy-Weinberg expectations to be gradual, even in the case of an autosomal locus. This property arises because the genotypes of new recruits are a function of the allele frequencies specific to the reproductive age classes. Juvenile age classes only influence the change in genotype frequencies through mortality, but as they mature they begin to add copies of their genes to the population. The genotype frequencies become stable only after the allele frequencies become homogenized across age classes and sexes.

Of equal significance is the fact that the allele frequencies themselves can be unstable in an age-structured population even in the absence of genotypic differences in age-specific survival and reproduction. Further complexities are introduced by the scheme of mating between the various age classes. All of these subjects are taken up in detail by Charlesworth (1974, 1994) and Gregorius (1976), while Caswell (2001) examines the modeling of age- and stage-structured populations. The critical point is that when newly founded populations have significant age structure, fluctuations in both gene and genotype frequencies may occur for a substantial period of time even in the absence of selection and drift.

Testing for Hardy-Weinberg Proportions

When data are available on genotype frequencies in a population, it is standard practice to cross-check these with their Hardy-Weinberg expectations. Lack of concordance between the two implies that at least one assumption of the Hardy-Weinberg model is violated, and this often instigates further investigation. Several statistical techniques have been proposed (Weir 2010), the historically most popular of which is the χ^2 (Chi-square) test,

$$X = \sum_{i=1}^{n}\sum_{j\geq i}^{n} \frac{\left(N_{ij} - \widehat{N}_{ij}\right)^2}{\widehat{N}_{ij}} \tag{4.4a}$$

where N_{ij} and $\widehat{N}_{ij}$ are the observed and expected numbers of genotype B_iB_j in a sample.

An alternative, but closely related, approach is the likelihood-ratio based **G-test**,

$$G = -2\sum_{i=1}^{n}\sum_{j\geq i}^{n} N_{ij} \ln\left(\frac{\widehat{N}_{ij}}{N_{ij}}\right) \tag{4.4b}$$

which is becoming more widely used. Such likelihood-based tests have a number of desirable statistical features (Appendix 4). The test statistic G has a sampling distribution very similar to the well-known χ^2 distribution (Appendix 5). That is, if a population in Hardy-Weinberg equilibrium is sampled many different times and G calculated each time, the

frequency distribution of the observed G values will be nearly χ^2 distributed. Thus, the test for Hardy-Weinberg proportions compares the observed statistic G with the cumulative χ^2 distribution. If G exceeds the level at which there is a 5% chance of obtaining a higher χ^2, then the chance that these data from a locus showing Hardy-Weinberg proportions is unlikely.

Regardless of which approach to testing for Hardy-Weinberg frequencies is taken, it should be kept in mind that some of the conditions underlying the Hardy-Weinberg theorem may be violated without causing detectable departures of observations from expectations. For example, if the product of the survivorships of the two homozygotes is equal to the square of the heterozygote survival, the zygotic frequencies after selection will still be in Hardy-Weinberg proportions (Lewontin and Cockerham 1959). Thus, a failure to reject the Hardy-Weinberg model should be interpreted with caution. In particular, even a large amount of allele-frequency change over the previous generation still results (under random mating) in the offspring being in HW proportions (albeit with the new allele frequency). Thus Equations 4.4a and 4.4b simply test *one* feature of Hardy-Weinberg: Can genotypic frequencies be completely described by allele frequencies. This is ensured with random mating when no other evolutionary forces act between the formation of zygotes and the test for HW proportions. The deeper feature of Hardy-Weinberg that is untested is the prediction of no allele-frequency change over time.

Example 4.2. As an example of the application of Equation 4.4b, we return to the data in the table of Example 4.1. The best estimates for the Hardy-Weinberg expectations are obtained from the observed allele frequencies: $\widehat{N}_{11} = p_1^2 N = 904.9$, $\widehat{N}_{12} = 2\,p_1 p_{\,2} N = 79.3$, and $\widehat{N}_{22} = p_2^2 N = 1.7$. Applying these and the observed values (N_{11}, N_{12}, and N_{22}) from the table,

$$G = -2\,[\,905\,\ln(904.9/905) + 78\,\ln(79.3/78) + 3\,\ln(1.7/3)\,] = 1.0293$$

Under the null hypothesis of Hardy-Weinberg frequencies, the sampling distribution of G is a function of the number of degrees of freedom, which in the case of the Hardy-Weinberg test is the number of genotypic classes minus the number of allele frequencies that must be estimated from the data minus one. Here, it was necessary to estimate one parameter (p_1) from the data, so there is $3 - 1 - 1 = 1$ degree of freedom. The 95% critical value for a χ_1^2 is easily obtained in `R` using `qchisq(0.95,1)`, which returns a value of 3.841. The p value for our observed G value is also easily obtained using `1-pchisq(1.0293,1)`, which returns a p value of 0.310. Therefore, the observed data are not significantly different from Hardy-Weinberg expectations. Note that the Chi-square statistic for this data

$$X = \frac{(905 - 904.9)^2}{904.9} + \frac{(78 - 79.3)^2}{79.3} + \frac{(3 - 1.7)^2}{1.7} = 1.0154$$

which has a p value of 0.314, again showing little departure from HW expectations.

Example 4.3. An interesting application of Hardy-Weinberg is **Snyder's ratio**. The human geneticist Laurence Snyder was interested in testing whether the inability to detect the bitter chemical PTC, a *nontaster*, was a Mendelian recessive. Snyder (1933) noted that normal Mendelian procedures, such as a testcross back to a nontaster (*tt*) to determine whether a taster was homozygous or heterozygous (*TT* or *Tt*), is not an option in human genetics. While all the offspring of nontaster x nontaster parents were nontasters, consistent with a recessive, Snyder observed 15% nontaster offspring from taster x taster crosses (marriages) and 37% for taster x nontaster crosses. Were these values consistent with his model? Letting q denote the frequency of *t*, then under random mating, the frequency of taster $\times$ nontaster matings is

$$2 \cdot q^2[(1-q)^2 + 2q(1-q)] = 2q^2(1-q)(1+q)$$

The leading factor of two arises because the mating could be taster father $\times$ nontaster mother, or vise versa. Of these matings, only $Tt \times tt$ produce nontasters, with half of their offspring expected to be nontasters. Hence, the expected fraction of nontaster offspring from taster $\times$ nontaster families is

$$\frac{(1/2)2q^2[2q(1-q)]}{2q^2(1-q)(1+q)]} = \frac{q}{1+q} = S$$

Proceeding in the same fashion, the fraction of nontaster offspring from taster $\times$ taster matings is S^2. More generally, if we let D denote the (putative) dominant phenotype and R the recessive phenotype, then

Mating	D $\times$ D	D $\times$ R	R $\times$ R
Fraction of R offspring	S^2	S^1	$S^0 = 1$

These expected fractions are Snyder's ratios. With an estimate of q in hand (for example, the square root of the frequency of the R phenotype), one can test for goodness of fit to these ratios. When doing so, Snyder found that nontaster was consistent with a Mendelian recessive. Further discussion of this concept, including extensions to incomplete penetrance and epistasis, can be found in Taylor and Prior (1939), Li (1953), Trankell (1956), Furusho (1960), Doolittle (1968), and Otto et al. (1994).

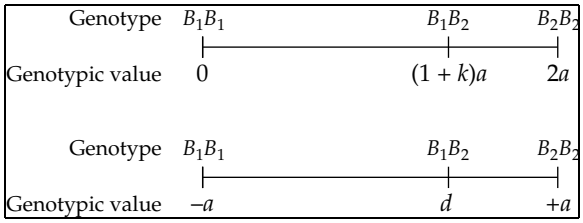

Figure 4.4 Two ways of representing genotypic values for a diallelic locus. A third representation simply adds a constant value, C, to each genotype, so that (for example) the lower representation becomes $C - a$, $C + d$, $C + a$. In all representations, $2a$ is the difference between the two homozygotes and $ka = d$ is the departure of the heterozygote from the average value of the two homozygotes.

CHARACTERIZING THE INFLUENCE OF A LOCUS ON THE PHENOTYPE

In Chapter 3, we encountered the concept of partitioning the phenotype value (z) of an individual into a **genotypic value** (G) and an **environmental deviation** (E),

$$z = G + E$$

where G is the expected phenotype (for a given genotype) resulting from the joint expression of all of the genes underlying the trait. For a multilocus trait, G is a potentially complicated function. For now, however, we are concerned only with the direct contribution of a single autosomal locus, in which case things are quite tractable. We start with the special case in which there are only two alleles. The three genotypic values can then be represented by the scale at the top of Figure 4.4, with $2a$ representing the difference between the mean phenotypes of B_2B_2 and B_1B_1 homozygotes, and k providing a measure of dominance. Alleles B_1 and B_2 behave in a completely additive fashion when $k = 0$, whereas $k = +1$ implies complete dominance of the B_1 allele, and $k = -1$ implies complete dominance of the B_2 allele. If $k > 1$, the phenotypic expression of the heterozygote exceeds that of both homozygotes, and the locus is said to exhibit **overdominance**, whereas $k < -1$ implies **underdominance**.

The fact that we have set the genotypic value of the B_1B_1 homozygote equal to zero may seem troublesome, but it is desirable because it leads to some algebraic simplifications.

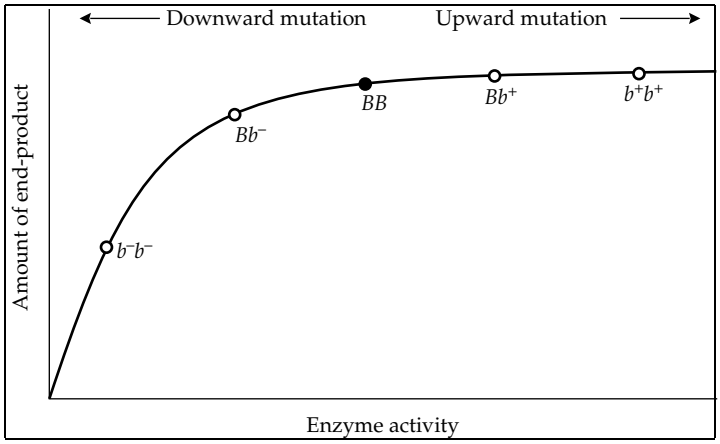

Figure 4.5 The relationship between the activity of a gene product and the **flux** or concentration of an end-product in an enzymatic pathway. *BB* represents the "wild-type" genotype. Upward and downward mutations with the same magnitude of change in enzyme activity are represented as b^+ and b^- alleles. Note that the b^+ allele has only a small effect on the end product, whereas the downwardly acting mutation b^- has a large homozygous effect, and exhibits pronounced recessiveness relative to *B*.

One can always rescale by the simple addition of some constant, such as in the bottom representation in Figure 4.4. The key parameters are not the actual genotypic values, but rather the difference ($2a$) between homozygotes and the departure ($d = ka$) of the value of the heterozygote from the average of the two homozygotes. In what follows, we use both k and d for measures of dominance, as some expressions are simpler when based on k, while others are when based on d.

Example 4.4. The logic underlying the scaling of genotypic values may be clarified by reference to a particular example—the Booroola (*B*) gene that influences fecundity in the Merino sheep of Australia (Piper and Bindon 1988). Litter size in sheep has a polygenic basis, but in this particular breed, it is determined largely by a single polymorphic locus. The mean litter sizes for the *bb, Bb,* and *BB* genotypes based on 685 records are 1.48, 2.17, and 2.66, respectively. Taking these to be estimates of the genotypic values (G_{bb}, G_{Bb}, and G_{BB}), the homozygous effect of the *B* allele is estimated by *a* = (2.66-1.48)/2 = 0.59. The dominance coefficient is estimated by taking the difference between *bb* and *BB* genotypes, *a(1+k)* = 0.69, substituting *a* = 0.59, and rearranging to obtain *k* = 0.17 (equivalently, *d* = *ak* = 0.10). This suggests slight dominance of the Booroola gene, but great confidence cannot be placed on this conclusion. Because the standard errors of the mean genotypic values are approximately 0.09, the midpoint between the two homozygotes, 2.07, is not significantly different from 2.17.

THE BASIS OF DOMINANCE

The presence of dominance complicates many formulations in quantitative genetics, but unfortunately it is a fact of life that cannot be ignored. Since the beginning of the last century, there has been much debate on the genetic and physiological basis of dominance. In the early days, the only genes subject to detailed genetic analysis were those that had a major phenotypic effect. Loci involving such genes are usually characterized by striking levels of dominance. For example, the vast majority of genes with major, deleterious effects on fitness are recessive. Does this then indicate that new mutations are inherently recessive? Fisher (1928a, 1928b, 1929, 1958) argued that because rare alleles are found almost entirely in the

heterozygous state, selection should favor alleles at modifier loci that cause heterozygous carriers of deleterious alleles to resemble the normal homozygote. Implicit in this argument is the assumption that the heterozygote initially encodes for an intermediate phenotype. Using physiological arguments, Wright (1929a, 1929b, 1934a, 1934b) strongly disputed this idea. He also pointed out that although dominance relationships are subject to change, the intensity of selection operating on modifier loci is unlikely to ever be strong enough to be an important evolutionary force. One important caveat to this assertion is that Wright assumed a population in mutation-selection balance, so that the strength of selection was on the order of the mutation rate. There are, however, other settings where the strength of selection on dominance modifiers can be much stronger, such as for an adaptive allele on its way to fixation (Haldane 1956; Mayo and Bürger 1997). The debate between Fisher and Wright was intense and at times bitter, and it scarred their relationship permanently.

Much later, Kacser and Burns (1981) developed a general explanation for dominance based on the biochemical principles of enzyme kinetics. Their model is in good accord with Wright's theory. Most gene products are involved in complex biochemical pathways such that the rate of production of a final end-product (phenotype) is regulated at many steps. Consequently, the relationship between gene-product activity (a function of allelic state) and end-product production is often hyperbolic (Figure 4.5). The assumption here is the **safety factor model**: the concentration of the substrate, not the enzyme, is the rate-limiting step. Under this model, Kacser and Burns showed that the "wild-type" activity normally lies on or near the plateau of this hyperbolic relationship. This leads to three predictions:

1. ***Mutations with large effects at the phenotypic level will be biased in a downward direction***. Even if mutations that increase activity occur as frequently as those that decrease it, the former will usually cause imperceptible changes at the phenotypic level. Thus, if a high production rate or end-product concentration is beneficial, we can expect most individually *discernible* mutations to be detrimental. This is a consequence of the nonlinear (hyperbolic) phenotype map.
2. ***The recessivity of downward mutations is an inevitable consequence of the hyperbolic activity-product relationship***. If we take the heterozygote to be intermediate in activity, the allele producing the homozygote with greater activity will always exhibit dominance on the end-product scale, the degree of dominance diminishing as the heterozygote is placed further out on the plateau.
3. ***The smaller the effect of a mutation, the less pronounced will be the level of dominance***. Such a result is expected simply because the relationship between the BB, Bb, and bb genotypic values tends towards linearity as the deviations among their activities are reduced. In principle, dominance is much more likely to be a complicating factor for characters whose variation is influenced by one or two genes of large effect than for quantitative characters encoded by numerous loci whose individual effects are indiscernible.

Because the exact form of the relationship in Figure 4.5 can change with a shift in the genetic background, the **Kacser-Burns model** does not rule out the possibility of evolutionary changes in dominance relationships. It does, however, eliminate the necessity of ad hoc evolutionary explanations, such as modifier loci, to account for the existence of dominance. A number of authors have noted that the Kascer-Burns framework applies to gene products that are enzymes, but not necessarily to products that are strictly structural or regulatory (Hodgkin 1993; Kondrashov and Koonin 2004). Indeed, dominance of new mutations is not uncommon among structural genes (Wilkie 1994; Phadnis and Fry 2005; Agrawal and Whitlock 2011). One explanation for this observation is the **poison subunit model**, wherein a single faulty component in a multimeric protein assembly can poison the entire complex.

Two important observations are generally in good accord with the predictions Kacser-Burns model. First, in a clever analysis of data on the haploid alga *Chlamydomonas reinhardtii*,

Orr (1991) found that when mutations are observed in artificial diploid constructs, they are almost always recessive. Because the heterozygous state almost never exists in a haploid species (except very transiently when zygotes are made), there can be no opportunity for the selection of dominance modifiers; the mutations must be "recessive" at first appearance. Second, as noted by Charlesworth (1979), lethal alleles in *Drosophila* are almost nearly completely recessive, whereas mildly deleterious alleles, whose individual effects are indiscernible, interact in a nearly additive fashion (Chapter 15). This same pattern of larger-effect genes tending to be more recessive was also seen by Phadnis and Fry (2005), who examined a large number of growth-rate mutations in yeast using the single-gene knock-out panel developed by Steinmetz et al. (2002a). A more detailed analysis of the same data by Agrawal and Whitlock (2011) offered a more nuanced view. While they observed the same general pattern as seen by Phadnis and Fry, they also noted that the distribution of dominance coefficients had a large variance, such that some deleterious alleles showed no dominance. They also noted that mutations in structure genes tended be more dominant that genes in other **gene ontology** (**GO**) classes, an observation also noted by Phadnis and Fry.

Kacser-Burns is not the only pathway-based model with implications for nonadditive genetics effects, see Cornish-Bowden (1987), Savageau (1992), Omholt et al. (2000), Bagheri and Wagner (2004), and Bost and Veitia (2013) and references therein, for details. Additional discussion of the evolution of dominance can be found in the reviews by Mayo and Bürger (1997), Bourguet (1999), Bagheri (2006), and Billiard and Castric (2011).

FISHER'S DECOMPOSITION OF THE GENOTYPIC VALUE

What Part of *G* is Passed Onto an Offspring?

One of the foundational pillars of quantitative genetics, the notion of the **additive genetic variance** of a trait, is also one of the field's more nuanced and challenging concepts. Before building the machinery to formally introduce the additive variance, we first consider Fisher's motivational question behind this concept: *What part of the genotype value G of a parent is passed onto its offspring?* The answer is trivial in two settings: offspring produced by cloning (e.g., vegetative propagation or with certain mating systems) and offspring produced by selfing a fully inbred parent. In both cases, the offspring and parental genotypes are identical, and the offspring has the same G value as its parents.

Conversely, the genotypic values for offspring produced by sexual reproduction are rarely the same as either parent. Indeed, offspring from the same parents are expected to show potentially quite dramatic variation in their genotypic values due to Mendelian segregation at parental heterozygous loci (e.g, a $B_1B_2C_1C_2$ parents generates four different gametes). These results arise because a sexually produced offspring has two parents, each of which (for diploids) passes along ***single alleles at each locus*** to its offspring, rather than whole genotypes. There is also a ***context-specific aspect to transmission***. Consider a diallelic locus. Whether the resulting offspring from B_1B_1 parent are B_1B_2 or B_1B_1 depends on the probability that the other parent (often a random individual from some target population) donates a B_1 or a B_2 allele. Hence, the ability of a parent to transmit part of its G value to an offspring depends, in part, on contributions from the other parent, and hence is ***population-specific***.

Fisher showed that the basic building block of quantitative-trait transmission in sexual populations is the **additive** (or **average**) **effect of an allele**—the average amount that an individual receiving that allele deviates from the population mean. As we will show, this quantity is not simply an intrinsic feature of a given allele, as it also depends on the population allele frequencies. Hence, the additive effect is a **population-specific** measure that can change with allele frequencies. The sum of additive effects over all loci influencing a trait in an individual is called its **additive genetic value**, or, equivalently, **breeding value**, for that trait. As we will see, the breeding value can be estimated from either designed crosses, knowledge of the trait values and familial relationships among a set of individuals, or marker-trait associations (genomic selection/prediction, polygenic risk scores). The

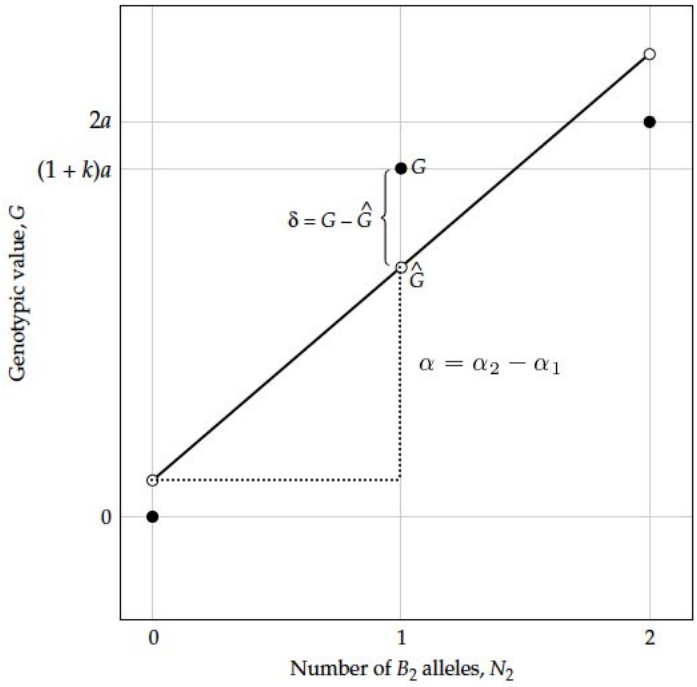

Figure 4.6 Linear least-squares regression (solid line) of the genotypic value G of a single locus on the gene content of allele B_2 (N_2). From left to right, the points represent the B_1B_1, B_1B_2, and B_2B_2 genotypes, corresponding to N_2 values of, respectively, 0, 1, and 2. Solid circles represent the true genotypic values, while open circles are the values, $\widehat{G}$, expected on the basis of average effects alone (Equation 4.7). The deviation between G and $\widehat{G}$ for each genotype is δ, the dominance deviation.

additive genetic variance is simply the ***population variance in breeding values***. If the fraction of trait variance that is additive is small (i.e., the narrow-sense heritability, h^2, is small), then offspring only weakly resemble their parents, which results in (among other things) inefficient selection on a trait (Equation 3.28). To formally develop these concepts, we now consider Fisher's (1918) decomposition of the genetic value.

Fisher's Decomposition

We start by assuming a diallelic locus, whose alleles B_1 and B_2 have frequencies of p_1 and p_2, respectively (where $p_2 = 1 - p_1$). The number of copies of a particular allele (say B_2) in a genotype ($N_2 = 0$, 1, or 2 for diploids) is referred to as the **gene dosage** (or **gene content**). As noted above, unless this allele interacts additively with all other alleles, there will be a nonlinear relationship between the gene content and the genotypic value. It is, nevertheless, useful to consider the best linear approximation to this relationship, as this leads to a partitioning of the genotypic values into their "expected" values based on additivity ($\widehat{G}$) and deviations from those expectations resulting from dominance (δ), see Figure 4.6.

The preceding points can be formalized by least-squares regression of genotypic values on the number of B_1 and B_2 alleles in the genotype (N_1 and $N_2 = 2 - N_1$),

$$G_{ij} = \widehat{G}_{ij} + \delta_{ij} = \mu_G + \alpha_1 N_1 + \alpha_2 N_2 + \delta_{ij} \tag{4.5a}$$

The genotypic value (G_{ij}) of genotype B_iB_j is a function of the mean genotypic value in the population (μ_G), the slopes of the regression (α_1 and α_2), the predictor variables (N_1

and N_2), and the residual error (δ_{ij}). Regression theory (Chapters 3 and 10) provides a powerful interpretation of the regression slopes, α_i. Namely, a unit change in the number of i alleles (N_i) results in an expected change of α_i in the genotypic value, so that α_i is the predicted impact on G from an individual carrying a B_i allele. This partitioning of genotypic values into various components is one of several major advances developed in Fisher's 1918 paper. Many of the innovative ideas in this classic paper are presented in his characteristically cursory (some would even say cryptic) manner, but a useful interpretative guide is provided by Moran and Smith (1966).

Unlike the univariate regression discussed in Chapter 3, Equation 4.5a is a **multiple regression**, the properties of which are discussed in Chapter 10. For the two-allele case, however, we can reduce the model to a standard univariate regression by noting that, for any individual, $N_1 = 2 - N_2$, yielding

$$\begin{aligned} G_{ij} &= \mu_G + \alpha_1 (2 - N_2) + \alpha_2 N_2 + \delta_{ij} \\ &= \iota + (\alpha_2 - \alpha_1) N_2 + \delta_{ij} \end{aligned} \tag{4.5b}$$

where $\iota = \mu_G + 2\alpha_1$ is the intercept. We denote the slope of this regression by

$$\alpha = \alpha_2 - \alpha_1 \tag{4.6}$$

and discuss its meaning shortly. Equation 4.5b implies that the genotypic values predicted by the regression are given by

$$\widehat{G}_{ij} = \mu_G + 2\alpha_1 + \alpha N_2 \tag{4.7a}$$

Note that this form also recovers Equation 4.5a,

$$\widehat{G}_{ij} = \mu_G + \alpha_i + \alpha_j = \begin{cases} \mu_G + 2\,\alpha_1 & \text{for } G_{11} \;\; (N_2 = 0) \\ \mu_G + \alpha_1 + \alpha_2 & \text{for } G_{21} \;\; (N_2 = 1) \\ \mu_G + 2\,\alpha_2 & \text{for } G_{22} \;\; (N_2 = 2) \end{cases} \tag{4.7}$$

We next show that the weighted mean of the coefficients α_1 and α_2 is equal to zero. To accomplish this, return to Equation 4.5a, and taking expectations yields

$$\mu_G = \mu_G + \alpha_1\, E(N_1) + \alpha_2\, E(N_2) + E[\delta_{ij}]$$

The expected value of the residual δ_{ij} is equal to zero by the properties of least-squares regression, and $E(N_1)/2$ and $E(N_2)/2$ are equivalent, respectively, to p_1 and p_2, the frequencies of the B_1 and B_2 alleles. Thus, the previous expression simplifies to

$$p_1\,\alpha_1 + p_2\,\alpha_2 = 0 \tag{4.8}$$

showing that the mean value of α_i over all individuals is indeed zero. Finally, from Equations 4.6 and 4.8 and the fact that $p_1 + p_2 = 1$, we obtain

$$\alpha_2 = p_1\,\alpha \qquad \text{and} \qquad \alpha_1 = -p_2\,\alpha \tag{4.9}$$

Now recall from Chapter 3 that the slope of a univariate regression is simply the covariance between response and predictor variable divided by the variance of the predictor variable. Thus, the slope of the regression in Figure 4.6 is

$$\alpha = \frac{\sigma(G, N_2)}{\sigma^2(N_2)} \tag{4.10a}$$

The terms $\sigma(G, N_2)$ and $\sigma^2(N_2)$ are functions, respectively, of the gene effects (a and k) and allele frequencies (p_1 and p_2). The steps leading up to their computation, under the assumption of random mating, are outlined in Table 4.1. Upon substitution, we obtain

Table 4.1 Properties of a single segregating diallelic locus under random mating. Alternative version of these expressions follow by noting that $p_1 = 1 - p_2$ and $d = ak$.

Genotype	Gene Content (N)	Genotypic Value (G)	Freq.	$G \cdot N$	N^2	Regression Value $(\widehat{G})$	Dominance Deviation $(\delta = G - \widehat{G})$
B_1B_1	0	0	p_1^2	0	0	ι	$-\iota$
B_1B_2	1	$(1+k)a$	$2p_1p_2$	$(1+k)a$	1	$\iota + \alpha$	$(1+k)a - \iota - \alpha$
B_2B_2	2	$2a$	p_2^2	$4a$	4	$\iota + 2\alpha$	$2a - \iota - 2\alpha$

$$
\begin{aligned}
\mu_N &= 2\,p_1p_2(1) + p_2^2(2) \\
&= 2p_2 \\
E(N^2) &= 2p_1p_2(1) + p_2^2(4) \\
&= 2p_2(1+p_2) \\
\mu_G &= 2p_1p_2a(1+k) + 2p_2^2a \\
&= 2p_2a(1+p_1k) \\
E(GN) &= 2p_1p_2a(1+k) + 4p_2^2a \\
&= 2p_2a[\,2p_2 + p_1(1+k)\,] \\
\sigma(G,N) &= E(GN) - \mu_G\mu_N \\
&= 2p_1p_2a[\,1 + k(p_1 - p_2)\,] \\
\sigma^2(N) &= E(N^2) - \mu_N^2 \\
&= 2p_1p_2
\end{aligned}
\qquad
\begin{aligned}
\mu_{\widehat{G}} &= \iota + 2\,p_1p_2\alpha + 2p_2^2\,\alpha \\
&= \iota + 2p_2\alpha \\
\mu_\delta &= -\iota + 2p_1p_2[(1+k)a - \alpha\,] \\
&\quad +2p_2^2(a-\alpha) = 0 \\
E(\widehat{G}^2) &= p_1^2\iota^2 + 2p_1p_2(\iota+\alpha)^2 + p_2^2(\iota+2\alpha)^2 \\
&= \iota^2 + 4p_2\alpha\iota + 2p_2\alpha^2(1+p_2) \\
E(\delta^2) &= p_1^2\iota^2 + 2p_1p_2[\,(1+k)a - \iota - \alpha\,]^2 \\
&\quad +p_2^2(2a - \iota - 2\alpha)^2 = (2p_1p_2ak)^2 \\
\sigma_A^2 &= E(\widehat{G}^2) - \mu_{\widehat{G}}^2 \\
\sigma_D^2 &= E(\delta^2) - \mu_\delta^2
\end{aligned}
$$

$$\alpha = a\,[\,1 + k\,(\,p_1 - p_2\,)\,] = a\,[\,1 + k\,(\,1 - 2p_2\,)\,] \tag{4.10b}$$

Under the assumption of random mating, α is known as the **average effect of allelic substitution**. It represents the average change in genotypic value that results when a B_2 allele is randomly substituted for a B_1 allele. For the purely additive case ($k = 0$), α is simply equal to a. However, for all other cases, α is also a function of k and of the allele frequencies in the population (Figure 4.7). Such behavior results because, with dominance, the phenotypic effect of a gene substitution depends on the status of the unsubstituted allele. If B_2 is a dominant allele ($k > 0$), then α will be inflated relative to the case of additivity if B_2 is rare ($p_1 > p_2$), but diminished if B_2 is common ($p_1 < p_2$). Thus, except in the case of additivity, the average effect of allelic substitution is not simply a function of the inherent physiological properties of the allele. ***It can only be defined in the context of the population***. This shows that Fisher's great insight essentially distills to the follow key conceptual issue: the intrinsic components of the genotypic value (a and k) alone are not sufficient to describe the transmission of the genotypic value from a pair of parents to their offspring. Instead population-weighted metrics of these components (the average effects α_i and dominance deviations δ_{ij}) are required.

PARTITIONING THE GENETIC VARIANCE

Fisher (1918) showed that once the genotypic values have been partitioned in the above manner, it is a relatively simple step to partition the sources of genetic variation at a locus. Recalling the relationship $G = \widehat{G} + \delta$, the total genetic variance may be written as

$$\sigma_G^2 = \sigma^2(\widehat{G} + \delta) = \sigma^2(\widehat{G}) + 2\sigma(\widehat{G}, \delta) + \sigma^2(\delta)$$

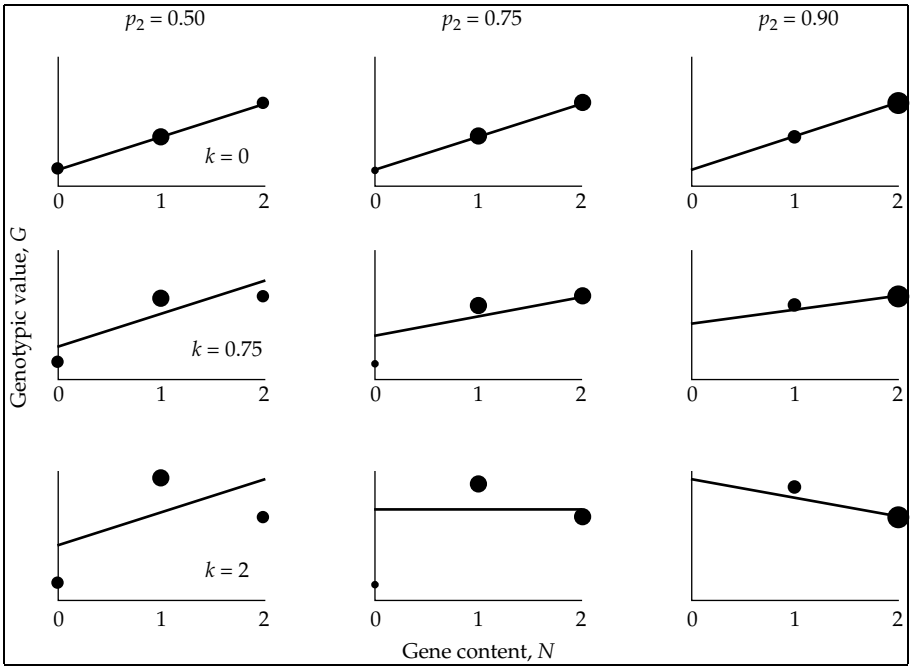

Figure 4.7 The slope α of the linear least-squares regression of genotypic value on gene content as a function of allele frequency, p_2, and degree of dominance, k. The lines denote the regressions, with each of the three points (representing genotypic values) being weighted by their frequency (denoted by the different-sized circles). The columns of graphs give results for different gene frequencies ($p_2 = 0.50$, 0.75, and 0.90), whereas the rows give results for different modes of gene action ($k = 0.00$, additivity; $k = 0.75$, partial dominance; and $k = 2.00$, overdominance). Note that when going across a row (the same model of gene action), the location of the points does not change, but their size (representing their relative abundance) does. Except for the case of complete additivity, the resulting regressions differ with different allele frequencies, as unless all of the point initially fall on a straight line, the different weights of the points change to slope of the optimal linear fit (minimizing of the weighted sum of squared residuals). In the case of overdominance, the slope changes sign as the allele frequency changes; when $p_2 = 0.75$, the slope is zero, i.e., there is no additive genetic variance. When $p_2 = p_1 = 0.5$, the slope $\alpha = a$ regardless of the degree of dominance.

From the property of least-squares regression (Chapter 3), the regression prediction ($\widehat{G}$) is uncorrelated with the residual error (in this case, δ). Thus, the total genetic variance attributable to a locus simplifies to the sum of additive and dominance components. Hereafter, we denote these components as σ_A^2 and σ_D^2, such that

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 \tag{4.11}$$

Statistically speaking, σ_A^2 is the amount of the variance in G that is explained by the regression on N_2 (or equivalently, on N_1), whereas σ_D^2 is the residual variance for the regression. Biologically speaking, σ_A^2 is the genetic variance associated with the average additive effects of alleles (the **additive genetic variance**), and σ_D^2 is the genetic variance associated with dominance effects (the **dominance genetic variance**).

All of the information necessary to compute these two components of genetic variance for a diallelic locus is contained in Table 4.1, and leads to

$$\begin{aligned}\sigma_A^2 &= 2p_1p_2\alpha^2 \\ &= 2a^2p_1(1-p_1)\left[1+k(1-2p_2)\right]^2 \\ &= 2p_1(1-p_1)\left[a+d(1-2p_2)\right]^2\end{aligned} \tag{4.12a}$$

$$\sigma_D^2 = (2p_1p_2ak)^2 = 4d^2p_1^2(1-p_1)^2 \tag{4.12b}$$

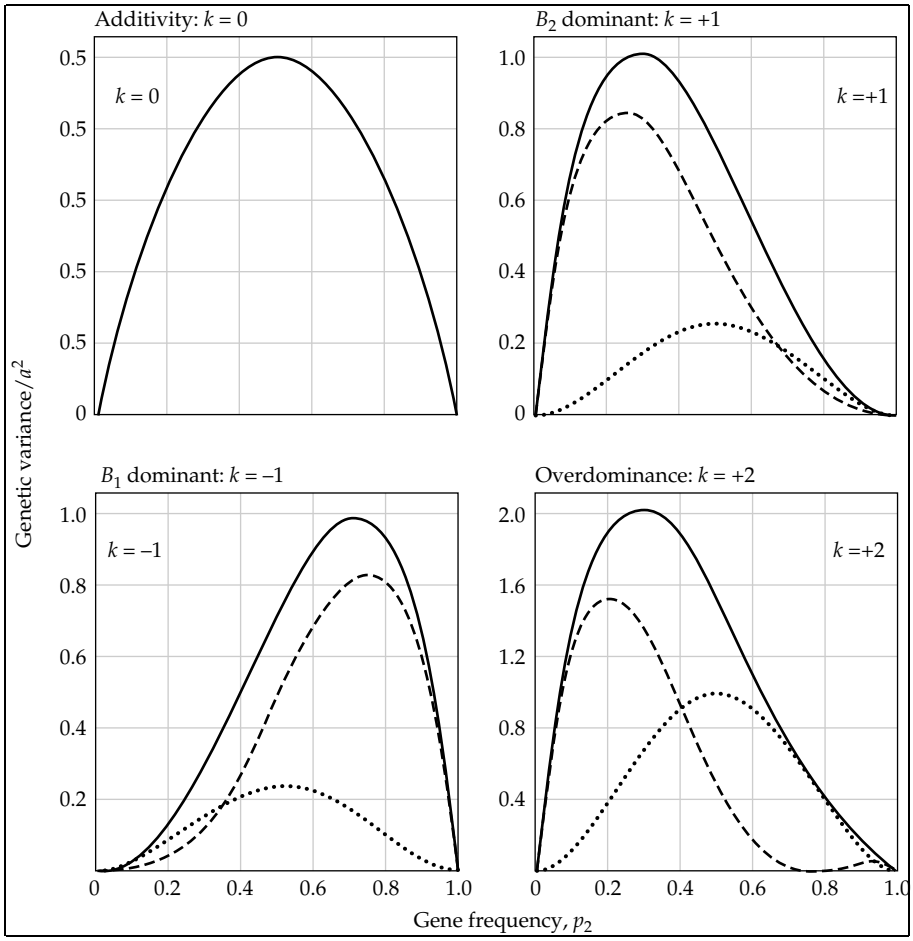

Figure 4.8 The dependence of the components of genetic variance at a locus on the frequency of the B_2 allele. The solid line denotes the total genetic variance, the dashed line the additive genetic variance, and the dotted line the dominance genetic variance. Four cases are illustrated: $k = 0$ (additivity), $k = +1$ (dominance of the B_2 allele), $k = -1$ (dominance of the B_1 allele), and $k = +2$ (a case of overdominance). In the case of additivity, all of the genetic variance is of the additive type. The vertical axes are scaled such that, for any particular case, the actual variances are obtainable by multiplying by a^2, where a is half the difference between homozygous B_1B_1 and B_2B_2 genotypic values.

Finally, letting $H = 2p_1(1 - p_1)$ denote the heterozygosity, we can also express these variances are

$$\sigma_A^2 = \alpha^2 H \qquad \text{and} \qquad \sigma_D^2 = d^2 H^2 \tag{4.12c}$$

Both components of variance depend upon the gene frequencies, the dominance coefficient k, and the homozygous effect a (Figure 4.8). In the case of purely additive allelic effects ($k = 0$ and hence $\alpha = a$), the additive genetic variance reaches a maximum at $p_1 = p_2 = 0.5$, the gene frequency at which heterozygosity H is most pronounced. With dominance, however, the additive genetic variance is maximized at a higher frequency of the recessive allele. This occurs because rare recessive alleles cause little genetic variance, only making a significant impact on trait variance when enough recessive homozygotes appear. Note from Equation 4.12c that σ_D^2 is maximized at $p_1 = p_2 = 1/2$ for all values of k.

A common misconception is that the relative magnitudes of additive and dominance variance provide information on the additivity of gene action. Equations 4.10b and 4.12a show that this generalization does not hold true. Through its influence on α, dominance contributes to the additive genetic variance, and for certain allele frequencies, can cause σ_A^2

to reach much higher levels than in the case of alleles with purely additive effects (Figure 4.8). Even in the case of complete dominance, σ_D^2 is unlikely to greatly exceed σ_A^2, and it is often substantially smaller. In the case of overdominance $(k > 1)$, probably not a common situation (Chapters 11 and 12), there is always an intermediate gene frequency at which σ_A^2 is zero. This occurs when the least-squares regression of G on N_2 has a slope equal to zero (Figure 4.7). At this allele frequency, there is no variance in breeding values, which implies that the expected offspring mean is the same for any parent from this population mated to a random parent.

More formally, from Equation 4.12c, we have

$$\frac{\sigma_A^2}{\sigma_D^2} = \frac{\alpha^2 H}{d^2 H^2} = \frac{a^2\,[\,1 + k\,(\,1 - 2p_2\,)\,]^2}{a^2 k^2 H} = \frac{[\,1 + k\,(\,1 - 2p_2\,)\,]^2}{k^2 H} \tag{4.12d}$$

While this ratio might appear a bit opaque, notice that unless the allele frequencies are intermediate, so that when the heterozygosity $H = 2p_1p_2$ is not small, the additive variance is usually larger than the dominance variance. The exception is for a rare recessive allele (e.g., $k = -1, p_2 \ll 1$, in which case $\sigma_A^2/\sigma_D^2 \simeq 2p_2 \ll 1$). However, in this case even though $\sigma_D^2 > \sigma_A^2$, both variance components are extremely small. Hence, when one allele is somewhat rare, the additive variance (even with large amounts of dominance) is usually larger than the dominance variance (Hill et al. 2008).

Another way to highlight this distinction between dominance and dominant genetic variance was noted by Cheverud and Routman (1995). They stressed the difference between **physiological dominance**, namely the value of d that is a fixed constant (in our single-locus case), and **statistical dominance**, namely the δ_{ij} generated by least-square fit given the allele frequencies. One can have a large value for the former, and yet still have small values of the later, and hence little dominance variance. We will also note this distinction in Chapter 5 for epistasis, the nonadditive interaction of alleles across loci.

ADDITIVE EFFECTS, AVERAGE EXCESSES, AND BREEDING VALUES

Fisher's partitioning of the genotypic value into additive and dominance components is useful because, in randomly mating diploid species, a parent donates only one allele per locus to each of its offspring. The transmitted allele exhibits its additive effect when randomly combined with a gene from other parents. The dominance deviation of a parent, which is a function of the interaction between the two alleles it carries, is eliminated when gametes (which each carry only one of these alleles) are produced. Thus, one can think of $\widehat{G}$ and δ, respectively, as the heritable and nonheritable components of an individual's genotypic value.

Before clarifying this concept further, however, we need some formal definitions. Two different measures of the effect of an allele were proposed by Fisher (1918, 1941): the average ***excess*** α_i^* and the additive (or average) ***effect*** α_i. As will be shown below, these two measures are equivalent in a randomly mating population, the first having a simple biological interpretation, the second being defined as a least-squares regression parameter.

The **average excess** α_2^* of allele B_2 is the difference between the mean genotypic value of individuals carrying at least one copy of B_2 and the mean genotypic value of a random individual from the entire population,

$$\alpha_2^* = \left(\, G_{12}\, P_{12\,|2} + G_{22}\, P_{22\,|2}\,\right) - \mu_G \tag{4.13a}$$

where $P_{ij|i}$ is the conditional probability of a B_iB_j genotype given that one allele is B_i. This is a completely general definition, but initially we will continue to focus on a diallelic locus under random mating, in which case $P_{ij|i} = p_j$, with p_j being the frequency of allele B_j. Under these conditions, Equation 4.13a becomes

$$\alpha_2^* = G_{12}\, p_1 + G_{22}\, p_2 - \mu_G \tag{4.13b}$$

Table 4.2 Conditional mean genotypic values of progeny under random mating, and their deviations from the mean genotypic value in the population, $\mu_G = 2ap_2(1 + p_1\, k)$.

Parental Genotype	Breeding Value	Mean Genotypic Value of Progeny	Deviation of Expected Progeny Mean from μ_G
B_2B_2	$2\alpha_2$	$a[2p_2 + p_1(1 + k)]$	α_2
B_1B_2	$\alpha_1 + \alpha_2$	$a[p_2 + (1 + k)/2]$	$(\alpha_1 + \alpha_2)/2$
B_1B_1	$2\alpha_1$	$ap_2(1 + k)$	α_1

This follows because of all individuals receiving a B_2 allele from one parent, a proportion p_2 (under random mating) receive another B_2 allele from the second parent, while a proportion p_1 receive a B_1 allele from the second parent. Because the genotypic values in these two cases are respectively $2a$ and $a(1 + k)$, subtraction of the population mean μ_G (Table 4.1) from the conditional mean yields

$$\begin{aligned}\alpha_2^* &= \{\, p_1[\, a(1 + k)\,] + p_2(2a)\,\} - 2\, a\, p_2\, (\, 1 + p_1\, k\,) \\ &= p_1\, a\, [\, 1 + k\, (\, p_1 - p_2\,)\,] = p_1\, \alpha \end{aligned} \tag{4.14a}$$

Namely, a function of the average effect of allelic substitution (Equation 4.10b). In the same manner, the average excess of allele B_1 is found to be

$$\alpha_1^* = -p_2\, a\, [\, 1 + k\, (\, p_1 - p_2\,)\,] = -p_2\, \alpha \tag{4.14b}$$

Note that one of the average excesses is positive and the other negative because they are defined as deviations from the population mean genotypic value and hence have expected value zero. In particular,

$$E[\alpha_i^*] = p_1\alpha_1^* + p_2\alpha_2^* = p_1(-p_2\, \alpha) + p_2(p_1\, \alpha) = 0$$

The **additive effects**, α_i, on the other hand, are defined to be the least-squares regression coefficients of genotypic value on gene content. They are obtained by finding the α_1 and α_2 that miminize the mean-squared residual deviation

$$\begin{aligned} M &= E(\, \delta_{ij}^2\,) = E[\, (G_{ij} - \mu_G - \alpha_i - \alpha_j)^2\,] \\ &= (G_{11} - \widehat{G}_{11})^2\, P_{11} + (G_{12} - \widehat{G}_{12})^2\, P_{12} + (G_{22} - \widehat{G}_{22})^2\, P_{22} \end{aligned}$$

where P_{ij} is the frequency of the *ij*th genotype. Again, this is a general definition. For the special case of a randomly mating population (with $P_{11} = p_1^2$, $P_{12} = 2p_1p_2$, and $P_{22} = p_2^2$), setting the partial derivatives of M with respect to α_i equal to zero, and solving gives

$$\alpha_2 = p_1\, a\, [\, 1 + k\, (p_1 - p_2)\,] = p_1\, \alpha \tag{4.15a}$$

$$\alpha_1 = -p_2\, a\, [\, 1 + k\, (p_1 - p_2)\,] = -p_2\, \alpha \tag{4.15b}$$

Comparing these expressions with Equations 4.14a and 4.14b, we find that additive effects are identical to average excesses in randomly mating populations. The α_i are often referred to as **average effects**, but we use the term *additive effects* to discriminate them from average effects of higher-order gene actions (such as dominance).

An individual's **breeding value**, hereafter denoted by A, is the sum of the additive effects of its genes. In other words, the breeding value of a B_1B_1 homozygote is simply $2\alpha_1$, that of a heterozygote is $(\alpha_1 + \alpha_2)$, and that of a B_2B_2 individual is $2\alpha_2$. Because breeding values are defined as deviations from the population mean, ***the expected breeding value of a random individual is zero***.

For random-mating populations, an extremely useful relationship emerges from these definitions for additive effects and breeding values. Consider the expected genotypic values of progeny produced by the parental genotypes. In the case of B_2B_2 parents, a proportion p_2 of the offspring will also be B_2B_2, in which case their genotypic value is $2a$, and a proportion p_1 will be B_1B_2 with genotypic value $a(1+k)$. The average genotypic value of offspring from a B_2B_2 parent is therefore $p_2(2a) + p_1\, a(1+k) = a\,[\,2p_2 + p_1\,(1+k)\,]$. When the population mean, μ_G, is subtracted, we obtain (after some simplification) α_2 (Equation 4.15a). Deviations of expected progeny phenotypes from the population mean are given for the other two parental genotypes in Table 4.2. The results in this table show that when mating is random ***the breeding value of a genotype is equivalent to twice the expected deviation of its offspring mean phenotype from the population mean***. The deviation is multiplied by two because only one of the two parental genes is passed on to each offspring.

Thus, we can estimate the breeding value of an individual by mating it to many randomly chosen individuals from the population and taking twice the deviation of its offspring mean from the population mean. This result is a special case of the more general statement that the ***expected offspring deviation from the population mean is given by the average breeding value of its two parents***. Treating one of those parents as random (and hence, a breeding value of zero), recovers our estimate of twice the offspring deviation for the breeding value of a parent. We thus arrive at the answer to Fisher's question on the transmission of G: ***the fraction of G that is, on average, passed from parent to offspring is A/2***. Chapter 31 discusses the estimation of breeding values under very general settings.

Example 4.5. Consider the consequences of the Booroola gene (described in Example 4.4) in two hypothetical random-mating populations with gene frequencies of 0.5 and 0.1. We assume that the phenotypic means within genotypic classes are known without error, so that they are equivalent to the genotypic values. The additive and dominance genetic variances are, respectively, the mean-squared breeding values and the mean-squared dominance deviations because both types of effects have means equal to zero, so that $\sigma^2(x) = E[x^2]$.

	$p_B = 0.5$			$p_B = 0.1$		
	bb	Bb	BB	bb	Bb	BB
Genotypic Value (G_{ij})	1.48	2.17	2.66	1.48	2.17	2.66
Genotype Frequency (P_{ij})	0.25	0.50	0.25	0.81	0.18	0.01
Mean Genotypic Value						
$\mu_G = P_{bb}G_{bb} + P_{Bb}G_{Bb} + P_{BB}G_{BB}$			2.120		1.616	
Additive Effects						
$\alpha_B = p_B G_{BB} + p_b G_{Bb} - \mu_G$			0.295		0.603	
$\alpha_b = p_b G_{bb} + p_B G_{Bb} - \mu_G$			−0.295		−0.067	
Breeding Values						
$A_{ij} = \alpha_i + \alpha_j$	−0.59	0.00	0.59	−0.134	0.536	1.206
$\overline{A} = P_{bb}A_{bb} + P_{Bb}A_{Bb} + P_{BB}A_{BB}$			0.00		0.00	
Dominance Deviations						
$\delta_{ij} = G_{ij} - (\mu_G + \alpha_i + \alpha_j)$	−0.05	0.05	−0.05	−0.002	0.018	−0.162
$\overline{\delta} = P_{bb}\,\delta_{bb} + P_{Bb}\,\delta_{Bb} + P_{BB}\,\delta_{BB}$			0.00		0.00	
Genetic Variance Components						
$\sigma^2_A = P_{bb}\,A^2_{bb} + P_{Bb}\,A^2_{Bb} + P_{BB}\,A^2_{BB}$			0.1740		0.0808	
$\sigma^2_D = P_{bb}\,\delta^2_{bb} + P_{Bb}\,\delta^2_{Bb} + P_{BB}\,\delta^2_{BB}$			0.0012		0.0003	
$\sigma^2_G = \sigma^2_A + \sigma^2_D$			0.1752		0.0811	

The key features from this example are that the breeding values BV and variance compo-

nents *depend on the population* (here defined by allele frequency), and that the ratio of additive to genetic variance is >0.99 in both populations, despite a large *BB* dominance deviation when $p_B = 0.1$.

Although this example is somewhat artificial in that we employed arbitrary gene frequencies, the basic approach has been exploited in the analysis of human genetic disorders. Biochemical and genomic studies are used to identify **candidate loci** that are potential contributors to the variation of the trait of interest, and the genotypes of random individuals are identified by use of molecular markers or direct sequencing. The average phenotypic values within each genotypic class provide estimates of the genotypic values of candidate genotypes, which can then be used to estimate the fraction of the total phenotypic variance that is associated with the locus. Details on this **measured-genotype approach** are given in Chapter 17.

EXTENSIONS FOR MULTIPLE ALLELES AND NONRANDOM MATING

Although the preceding results were obtained under the assumption of a diallelic locus, they are readily generalized to situations with an arbitrary number of alleles, as well as to nonrandomly mating populations. The algebra necessarily becomes more tedious, but some very useful principles emerge that will be relied upon heavily in subsequent chapters. In addition to presenting a more general treatment, the remainder of the chapter will serve as a review of the concepts introduced earlier in the chapter.

Average Excess

When n alleles are present, the average excess, α_i^*, for any allele B_i is given by

$$\alpha_i^* = \sum_{j=1}^{n} P_{ij|i}\, G_{ij} - \mu_G \tag{4.16a}$$

where $P_{ij|i}$ is the conditional probability of a B_iB_j genotype given that one allele is B_i. Under random mating, $P_{ij|i} = p_j$ and this reduces to

$$\alpha_i^* = \sum_{j=1}^{n} p_j\, G_{ij} - \mu_G \tag{4.16b}$$

where p_j is the frequency of the *j*th allele.

Example 4.6. Here we show how the average excess α_i^* of an allele i can be related to $\sigma(G, N_i)$, the covariance between genotypic value and the number of copies of that allele. This result will be useful in the following sections. To compute $\sigma(G, N_i) = E(G \cdot N_i) - E(N_i) \cdot E(G)$, we start with the fact that $E(G) = \mu_G$, so we merely require expressions for $E(N_i)$ and $E(G \cdot N_i)$. The mean number of alleles of type i at the locus, $E(N_i)$, is straightforward. Because there are two genes at each locus, and the frequency of allele i is p_i, $E(N_i) = 2p_i$. To obtain $E(G \cdot N_i)$, we use **ordered-genotype** notation, with P_{ij} now denotes the probability of getting allele i from the mother and allele j from the father. We assume that $P_{ij} = P_{ji}$, so $P_{ij} = 2P_{ij}$ when $i \neq j$. Because the variable N_i takes on only two nonzero values, two and one, the expected cross-product is

$$E(G \cdot N_i) = (G_{ii} \cdot 2) \cdot P_{ii} + \sum_{j \neq i}^{n} (G_{ij} \cdot 1) \cdot 2P_{ij} = 2p_i \sum_{j=1}^{n} P_{ij|i} G_{ij}$$

where the last step follows from the definition of a conditional genotype probability (Equation 3.3a) as $P_{ij\,|\,i} = P_{ij}/p_i$. Putting the above results together, and recalling Equation 4.16a,

$$\sigma(G, N_i) = 2\,p_i \left[\sum_{j=1}^{n} P_{ij|i} G_{ij} - \mu_G \right] = 2\,p_i\,\alpha_i^* \tag{4.17a}$$

Under the assumption of random mating, average excesses are identical to additive effects, and

$$\sigma(G, N_i) = 2\,p_i\,\alpha_i \tag{4.17b}$$

Additive Effects

As in the diallelic case, with n alleles the additive (or average) effects are defined to be the set of α_i that minimizes $E(\delta_{ij}^2)$, obtained from the least-squares solution for the multiple regression

$$G = \mu_G + \sum_{i=1}^{n} \alpha_i\,N_i + \delta \tag{4.18}$$

This expression is the n-allele extension of Equation 4.5a, with N_i being the number of copies of allele i carried by an individual. For example, for the genotype G_{34}, $\sum \alpha_i\,N_i = \alpha_3 + \alpha_4$, and $\delta_{34} = G_{34} - \mu_G - \alpha_3 - \alpha_4$.

Multivariate regressions are covered in detail in Chapter 10, and here we simply cite the basic result—the regression coefficients (i.e., the α_i) are defined by the set of equations

$$\sigma(G, N_i) = \sum_{j=1}^{n} \alpha_j\,\sigma(N_i, N_j) \qquad \text{for } 1 \le i \le n \tag{4.19}$$

Expressed in this way, the definitions of the average effects are not immediately transparent, and the general solution to these equations is rather involved (Kempthorne 1957). However, under random mating, the solutions for the α_i are simplified greatly and can be expressed in two ways. First, drawing from Example 4.6

$$\alpha_i = \frac{\sigma(G, N_i)}{2p_i} \tag{4.20a}$$

Second, an equivalent, and even more transparent, solution follows from Equation 4.16b,

$$\alpha_i = \sum_{j=1}^{n} p_j\,G_{ij} - \mu_G \tag{4.20b}$$

i.e., under random mating, the average effects are equal to conditional mean deviations from μ_G.

If mating is nonrandom, but genotype frequencies are given by

$$P_{ii} = (1-f)p_i^2 + fp_i \tag{4.21a}$$
$$P_{ij} = 2(1-f)p_ip_j \tag{4.21b}$$

as occurs under regular inbreeding (Chapter 12), then

$$\alpha_i = \frac{\alpha_i^*}{1+f} \tag{4.22}$$

where f, the inbreeding coefficient, is the fractional reduction of heterozygote frequencies relative to those expected under random mating.

Additive Genetic Variance

To obtain the variance associated with the additive effects, we first need a result from regression theory. Consider the regression $y = \mu + \sum \beta_i x_i + e$. Because the total variance of a response variable y equals the variance accounted for by the regression plus the residual variance σ_e^2 (Chapter 10), it follows that the variance accounted for by the predictor variables is $\sum \beta_i \sigma(y, x_i)$. This can be immediately seen by noting

$$\sigma_y^2 = \sigma(y, y) = \sigma\left(y, \mu + \sum \beta_i x_i + e\right) = \sum_{i=1}^{n} \beta_i \sigma(y, x_i) + \sigma_e^2$$

Drawing the analogy with Equation 4.18, where the additive effects arise by considering the genotype G as a response variable ($y = G$) and the gene contents N_i as predictor variables ($x = N_i$), the variance associated with the additive effects becomes

$$\sum_{i=1}^{n} \alpha_i \sigma(G, N_i)$$

Thus, recalling the result from Example 4.6 that $\sigma(G, N_i) = 2 p_i \alpha_i^*$, the additive genetic variance is

$$\sigma_A^2 = 2 \sum_{i=1}^{n} p_i \alpha_i \alpha_i^* \tag{4.23a}$$

as noted by Fisher (1941) and Kempthorne (1957). This general definition for the additive genetic variance holds for both randomly and nonrandomly mating populations. In the former case, $\alpha_i = \alpha_i^*$, and Equation 4.23a reduces to

$$\sigma_A^2 = 2 \sum_{i=1}^{n} p_i \alpha_i^2 \tag{4.23b}$$

which with $n = 2$ (a diallelic locus) reduces further to Equation 4.12a. Thus, under random mating, σ_A^2 for a locus is simply equal to the mean-squared additive effect, multiplied by two to account for diploidy. More generally, because $E[\,\alpha\,] = 0$, $\sigma_A^2 = E[\,(\alpha_i + \alpha_j)^2\,]$, which can involve covariance terms under nonrandom mating (when alleles at a locus are correlated).

From Equation 4.22, it follows that under regular inbreeding,

$$\sigma_A^2 = 2(1 + f) \sum_{i=1}^{n} p_i \alpha_i^2 \tag{4.23c}$$

In general, inbreeding inflates the additive genetic variance by causing correlations among the effects of alleles within the same individuals. However, because the α_i themselves are a function of f, inbreeding does not necessarily simply increase the additive genetic variance by the factor $(1 + f)$. From Kempthorne (1957),

$$\alpha_i = \left(\frac{1 - f}{1 + f}\right) \alpha_{ir} + \left(\frac{f}{1 + f}\right) (G_{ii} - \mu_G) \tag{4.24}$$

where α_{ir} and μ_G, respectively, denote the additive effect of allele *i* (under random mating) and the mean phenotype in the noninbred population. If gene action is additive, then $G_{ii} - \mu_G = 2a$, $\alpha_i = \alpha_{ir} = a$, and the additive genetic variance in an inbred population is, in fact, $(1 + f)$ times greater than that under random mating. However, with any level of

dominance, $\alpha_i \neq \alpha_{ir}$ under inbreeding, and the change in additive genetic variance with f is not likely to be linear. See WL Chapters 11 and 23 for a full treatment.

Finally, we consider the general definition of the breeding value (A_{ij}) under random mating. Parents with genotype B_iB_j transmit alleles i and j with equal frequency, and the expected additive effect of the allele contributed by their mates is equal to zero. Thus, the expected deviation of the mean phenotype of offspring of genotype B_iB_j from the population mean is

$$\left(\mu_G + \frac{\alpha_i + \alpha_j}{2}\right) - \mu_G = \frac{A_{ij}}{2} \tag{4.25}$$

which is half the breeding value of the parental genotype. Returning to Equation 4.18,

$$\begin{aligned} G_{ij} &= \mu_G + \alpha_i + \alpha_j + \delta_{ij} \\ &= \mu_G + A_{ij} + \delta_{ij} \end{aligned} \tag{4.26}$$

Thus, the genotypic value at any locus can be decomposed into four quantities: the mean genotypic value for the population, the additive effects of the two genes (whose sum is the breeding value), and a dominance deviation due to the interaction between the genes. Because μ_G is a constant, and A and δ are (by the properties of least-squares regression) uncorrelated, it follows from Equation 4.26 that the genetic variance can be represented as

$$\sigma^2_G = \sigma^2(\alpha_i + \alpha_j) + \sigma^2(\delta_{ij}) \tag{4.27a}$$

This is a completely general definition, applying even to the case of nonrandom mating (although as noted above, the definitions of the α_i and δ_{ij} change with the degree of inbreeding). For the special case of random mating, α_i and α_j are uncorrelated, and

$$\sigma^2_G = \sigma^2_A + \sigma^2_D \tag{4.27b}$$

Hence, σ^2_A has a very specific and useful meaning. ***Under random mating, the additive genetic variance is equivalent to the variance of breeding values of individuals in the population.*** Note that if $\sigma^2_A = 0$, then all breeding values are zero (the mean value). In this case, the expected offspring mean for *any* parent is simply the population mean. Hence, offspring no more resemble their parents than they do a random individual from the population. Likewise, when the additive variance is nearly zero, there is only a very small spread among the breeding values in the populations, so there is corresponding very little variance in the expected offspring mean for any parent.

SUMMARY

Summing up, the additive effect of an allele, the breeding value of an individual, and the additive-genetic variance of a population are hierarchically related measures of genetic effects (Table 4.3). All of this notation can be quite confusing, particularly when the nonsubscripted α is used to denote the average effect of allelic substitution. We used the latter quantity in our introduction of the one-locus model for historical reasons and because it provides useful insight into the two-allele situation. However, we will not be using it much in the remainder of the book, nor will we be utilizing the concept of average excess (the latter plays a significant role in considerations of selection response, which is covered in WL). Unless otherwise noted, we will be dealing with randomly mating populations, so our reference to the additive effect of an allele will be consistent with the conditional mean deviation definition (Equation 4.20b), as well as formally equivalent to a least-squares regression coefficient. Further commentary on the relationship between average excesses and additive effects can be found in Falconer (1985) and Templeton (1987).

Table 4.3 Summary of quantities used to measure genetic effects. Note that *a, d,* and *k* are physiological parameters, while the remaining terms in the table are statistical parameters (population-specific functions of the physiological parameters).

Homozygous effect, a Dominance coefficient, k, $d = ka$	Intrinsic properties of allelic products. Not functions of allele frequencies, but may vary with genetic background (Example 5.1).
Additive (or average) effect, α_i Average excess, α_i^*	Properties of *alleles* in a particular population. Functions of homozygous effects, dominance coefficients, and genotype frequencies.
Breeding value, A	Property of a particular *individual* in a particular population. Sum of the additive effects of an individual's alleles.
Additive genetic variance, σ_A^2	Property of a particular *population*. Variance of the breeding values of individuals within the population.