2

Epistasis and Plant Breeding

*James B. Holland**
U.S. Department of Agriculture—Agriculture Research Service, Plant Science Research Unit, Department of Crop Science, North Carolina State Univ., Box 7620, Raleigh, North Carolina 27695-7620.

*I thank Drs. Edwin Bingham, Charles Brummer, James Crow, Arnel Hallauer, Wyman Nyquist, and David Remington for their helpful comments on and corrections to a previous draft of this article.

Plant Breeding Reviews, Volume 21, Edited by Jules Janick
ISBN 0-471-41847-1

I. INTRODUCTION

Epistasis is the interaction of alleles at different loci. The value of an allele or genotype at one locus depends on the genotype at other epistatically interacting loci, complicating the picture of gene action. A seemingly "favorable" allele at one locus may be an "unfavorable" allele in a different genetic background. There are well-defined cases of interactions occurring at the molecular level between gene products, but the relationship between molecular interactions and complex phenotypes is often not clear. Classical quantitative genetics methods relate observable phenotypic measures to the aggregate statistical effects of alleles and allelic combinations in specific populations. Genetic components of variance are population-dependent, often poorly estimated, and do not necessarily reflect the relative importance of different modes of gene action. DNA markers have simplified the direct estimation of gene action effects, and recent QTL and population genetics studies have revealed that epistatic gene action is more important for plant yield and fitness than was previously evident. Implications of strong epistasis for plant breeding include: (1) epistatic variance can shift to additive variance under drift or inbreeding; (2) epistatic variance contributes to "temporary" response to selection in outcrossing populations that can be captured as a form of heterosis using appropriate breeding procedures but may be otherwise squandered; and (3) fitness or yield is not a simple function of allele frequencies, resulting in rugged adaptive landscapes filled with local fitness optima on which breeding populations can become stranded. If epistasis is important, then genomics tools can be used to identify the nature and components of interacting genic systems and marker-assisted selection schemes can be designed to exploit epistasis.

The literature on theory of and empirical evidence for epistasis in crops is reviewed here. Explicit formulations for additive, dominance, and epistatic genetic effects and variances; inbreeding depression; heterosis; and response to selection are presented to unify the discussion

of the various aspects and implications of epistasis. These formula are given in terms of a two-locus, two-allele model for diploids and are not found as such elsewhere in the literature to my knowledge. Verbal explanations for the implications of these formulas are given, and the reader not interested in mathematical detail can skip the equations with little loss of understanding.

II. GENE ACTION AND STATISTICAL EFFECTS

Epistasis is the dependence of allelic effects at one locus on the genotype at a second locus. Epistasis, then, is an interaction of alleles at different loci, and so is a form of non-additive gene action. The terms *additive* and *non-additive* can be confusing because their meaning depends on whether the scope of inference is a single locus or multiple loci (Falconer and Mackay 1996, p. 119). Additive gene action in reference to a single locus implies the lack of dominant gene action. Additive gene action in reference to two or more loci refers to the lack of epistasis (Table 2.1). Furthermore, confusion arises because epistasis to classical geneticists may refer strictly to the masking effect of dominant alleles at one locus on recessive homozygotes at another locus, while epistasis generally is taken by statistical geneticists to mean any form of non-allelic interaction (Phillips 1998).

In the absence of epistasis, the total genetic value for an individual is simply the sum of the individual locus genotype values because the loci are independent. For example, there can be additive gene action *within* a locus *A*, additive and dominance effects *within* a locus *B*, and additive gene action *between* the two loci (Table 2.1). The additivity between loci in the example in Table 2.1 is demonstrated by the consistency of differences among genotypes within one locus across genotypic classes

Table 2.1. An example of two-locus genotypic values that do not exhibit epistasis. The total genotypic value is the sum of the individual locus genotypic values.

Genotype at locus *A*	Genotype at locus *B*			
	B_1B_1	B_1B_2	B_2B_2	Unweighted marginal mean
A_1A_1	20	20	10	16.7
A_1A_2	15	15	5	11.7
A_2A_2	10	10	0	6.7
Unweighted marginal mean	15	15	5	

at the other locus. For example, A_1A_1 differs from A_1A_2 by a value of 5, whether the genotype at locus B is B_1B_1, B_1B_2, or B_2B_2 (Table 2.1).

Quantitative genetic models can be developed that incorporate the gene action effects of each locus plus the effects of interactions between loci affecting the trait. Such models are useful because they provide a means to define parameters that can be estimated from phenotypic data and that can be related to important concepts such as heritability and response to selection. Explicitly defining the components of the model allows us to understand the relationship between gene action effects and statistical genetic parameters such as additive genetic variance.

The simplest quantitative genetic model that includes epistasis is a two-locus model in which each locus has two alleles (Table 2.2). The additive and dominant gene action effects at each locus can be defined in a manner analogous to the classical definitions given by Falconer and Mackay (1996, Chap. 7), ignoring epistasis. For a non-epistatic model, the additive gene action effect at locus A (a^A) is half the difference between the values of the A_1A_1 and A_2A_2 genotypes. In the two-locus epistatic model, a^A is defined as half the average difference between these genotypic classes measured in either a B_1B_1 or B_2B_2 genotypic background (Tables 2.2 and 2.3). The dominance gene action effect of locus A (d^A) is the difference between the heterozygote A_1A_2 and the average of the two homozygous genotypes at the A locus, if epistasis is ignored. To include epistasis in the model, d^A is defined as the average

Table 2.2. Genotypic values and unweighted marginal means of two-locus, two-allele model of gene action.

	B-locus genotype			
A-locus genotype	B_1B_1	B_1B_2	B_2B_2	Marginal mean
A_1A_1	G_{1111} = m+a^A+a^B+aa	G_{1112} = m+a^A+d^B+ad	G_{1122} = m+a^A−a^B−aa	$G_{11..}$= m+a^A+(1/3)d^B+(1/3)ad
A_1A_2	G_{1211} = m+d^A+a^B+da	G_{1212} = m+d^A+d^B+dd	G_{1222} = m+d^A−a^B−da	$G_{12..}$= m+d^A+(1/3)d^B+(1/3)dd
A_2A_2	G_{2211} = m−a^A+a^B−aa	G_{2212} = m−a^A+d^B−ad	G_{2222} = m−a^A−a^B+aa	$G_{22..}$= m−a^A+(1/3)d^B−(1/3)ad
marginal mean	$G_{..11}$ = m+(1/3)d^A+a^B +(1/3)da	$G_{..12}$ = m+(1/3)d^A+d^B +(1/3)dd	$G_{..22}$ = m+(1/3)d^A−a^B −(1/3)da	$G_{....}$ = m+(1/3)d^A+(1/3)d^B +(1/9)dd

difference between A_1A_2 and the mean of the two homozygous genotypes at the A locus measured in either a B_1B_1 or B_2B_2 genotypic background (Tables 2.2 and 2.3). One can also define four epistatic gene actions in this system (Table 2.3). Additive-by-additive gene action (*aa*) refers to the difference between additive gene action at locus A in B_1B_1 homozygotes and B_2B_2 homozygotes. Equivalently, it refers to the difference between additive gene action at locus B in A_1A_1 homozygotes and A_2A_2 homozygotes. Additive-by-dominant gene action (*ad*) refers to the difference between the additive effect at locus A in B_1B_2 heterozygotes and in B_1B_1 and B_2B_2 homozygotes on average. Dominant-by-additive gene action (*da*) refers to the difference between the additive effect at locus B in A_1A_2 heterozygotes and in A_1A_1 and A_2A_2 homozygotes on average. Dominant-by-dominant gene action (*dd*) refers to the difference between the dominant effect at the A locus in B_1B_2 heterozygotes and in B_1B_1 and B_2B_2 homozygotes on average. This general model can be used to quantify any digenic, two-allele interaction (see, for example, Mather and Jinks 1977, Chap. 5).

Multiplicative gene interaction is a special case of epistasis in which the net genotypic value is the product of, rather than the sum of, effects at different loci. Opinion varies as to whether multiplicative interaction should be considered epistasis because multiplicative effects can be made additive simply by the use of a logarithmic transformation of phenotypic values. For example, experiments designed to test for epistasis

Table 2.3. Gene action parameters of the two-locus, two-allele model.

Midparent value = $m = [G_{1111} + G_{1122} + G_{2211} + G_{2222}]/4$
Additive effect at locus A = $a^A = [G_{1111} - G_{2211} + G_{1122} - G_{2222}]/4$
Dominant effect at locus A = $d^A = [G_{1211} - (\frac{1}{2})(G_{1111} + G_{2211}) + G_{1222} - (\frac{1}{2})(G_{1122} + G_{2222})]/2$
Additive effect at locus B = $a^B = [G_{1111} - G_{1122} + G_{2211} - G_{2222}]/4$
Dominant effect at locus B = $d^B = [G_{1112} - (\frac{1}{2})(G_{1111} + G_{1122}) + G_{2212} - (\frac{1}{2})(G_{2211} + G_{2222})]/2$
Additive-by-additive gene action effect = $aa = [G_{1111} - G_{2211} - G_{1122} + G_{2222}]/4$
Additive-by-dominant gene action effect = $ad = \{G_{1112} - G_{2212} - (1/2)[G_{1111} - G_{2211} + G_{1122} - G_{2222}]\}/2$
Dominant-by-additive gene action effect = $da = \{G_{1211} - G_{1222} - (1/2)[G_{1111} - G_{1122} + G_{2211} - G_{2222}]\}/2$
Dominant-by-dominant gene action effect = $dd = G_{1212} - (\frac{1}{2})(G_{1112} + G_{2212}) - (1/2)[G_{1211} - (1/2)(G_{1111} + G_{2211}) + G_{1222} - (1/2)(G_{1122} + G_{2222})]$

among viability genes consider additivity on a logarithmic scale (equal to multiplicative interactions on the original scale) to be a lack of epistasis (de Visser et al. 1997; Elena and Lenski 1997; Fu and Ritland 1996; Remington and O'Malley 2000). If homozygosity at one gene affects viability by causing lethality in a proportion of the population equal to x_1 and homozygosity at a second viability gene causes lethality in a proportion of the population equal to x_2, then, assuming independence, the combination of homozygosity at both genes is expected to cause lethality in a proportion of the population equal to the product of the two gene effects, $y = x_1x_2$. For viability, therefore, multiplicative interactions are considered independent gene action and deviations from multiplicative interactions are considered epistasis. Multiplicative gene effects on crop heterosis observed without transformations, however, are generally considered evidence of epistasis because they are deviations from a linear additive model (Schnell and Cockerham 1992). The difference between these points of view seems to be what form of gene interaction is expected, i.e., is it "normal" for effects at different loci to be accumulated in an additive or multiplicative manner? There is no clear-cut answer to this question, and it may differ among traits. In this review, epistasis will be considered to be deviations from additive gene action among loci, unless stated otherwise.

Statistical genetic parameters such as genetic components of variance can now be defined in terms of the gene action model given in Tables 2.2 and 2.3. Often, additive, dominance, and epistatic genetic components of variance are defined as functions of the *statistical* effects of alleles and allelic interactions, which are *not* the same as the *gene action* of those alleles and interactions. The relationships among genetic components of variance, statistical genetic effects, and gene action effects must be clarified in order to understand the impact of epistatic gene action on genetic variance components.

The difference between statistical genetic effects and gene action effects can be demonstrated most simply using a single-locus model. For example, assume that a trait is controlled by a single locus with two alleles at which the gene action values of a and d both equal 5 (complete dominance, Table 2.4). The additive statistical effect of an allele i, α_i, is the weighted average effect of allele i in a specified random-mating population. The weighted average effect of an allele is the expected deviation of genotypes from the population mean conditional on their having at least one copy of the allele. In the example in Table 2.4, when allele frequencies are equal ($p_1 = 0.5$), half of the time an A_1 allele is united with another A_1 allele to produce an A_1A_1 genotype with value 10, and half the time allele A_1 is united with allele A_2 to produce a heterozygote,

also with value 10. Thus, the mean of all genotypes conditional on them having one A_1 allele is 10, and the deviation of this conditional mean from the population mean is +2.5. Therefore, the average statistical effect of allele A_1 in this population is $\alpha_1 = +2.5$ (Table 2.2).

If the frequency of allele A_1 is different (e.g., $p_1 = 0.1$, Table 2.4), then the population mean changes ($\mu = 1.9$) and the average effect of the allele changes. The genotype mean conditional on having at least one A_1 allele is still 10, but now the deviation of this conditional mean from the population mean is larger than before, $\alpha_1 = 8.1$ (Table 2.4).

The additive genetic variance (σ^2_A) is the variance of allelic average statistical effects (multiplied by two because there are two alleles per locus per genotype), and therefore σ^2_A also changes when the allele frequencies change (Table 2.4). The dominance deviations of the statistical model also change, and consequently the dominance genetic variance changes (Table 2.4). So, even with the same underlying gene action, different populations may have different statistical genetic parameters, such as heritability and additive and dominance genetic variances, because their allele frequencies differ.

Table 2.4. Statistical genetic model parameters are functions of gene action effects and allele frequencies. Genotype frequencies, population mean, average allelic effects (α_i), dominance deviations (δ_{ij}), additive genetic variance (σ^2_A), and dominance genetic variance (σ^2_D) for a trait affected by a single locus with complete dominance when frequency of the favorable allele, p_1, is 0.5 or 0.1, and the population is in Hardy-Weinberg equilibrium.

		p1 = 0.5	p1 = 0.1
Genotype	Value	Genotype frequency	Genotype frequency
A_1A_1	10	0.25	0.01
A_1A_2	10	0.50	0.18
A_2A_2	0	0.25	0.81
Parameters of the statistical genetic model:			
Population Mean		7.5	1.9
α_1		$10 - 7.5 = 2.5$	$10 - 1.9 = 8.1$
α_2		$5 - 7.5 = -2.5$	$1 - 1.9 = -0.9$
δ_{11}		$10 - 7.5 - 2.5 - 2.5 = -2.5$	$10 - 1.9 - 8.1 - 8.1 = -8.1$
δ_{12}		$10 - 7.5 - 2.5 + 2.5 = 2.5$	$10 - 1.9 - 8.1 + 0.9 = 0.9$
δ_{22}		$0 - 7.5 + 2.5 + 2.5 = -2.5$	$0 - 1.9 + 0.9 + 0.9 = -0.1$
σ^2_A		12.5	7.29
σ^2_D		6.25	0.81

Furthermore, there is no direct relationship between statistical genetic parameters and gene action parameters except in special cases. Even in the special case of no epistasis, additive genetic variance is a function of not only additive (a) gene action effects but also of allele frequency and dominant (d) gene action effects. This may not be obvious from the example in Table 2.4; but consider the changes in the average allelic effects caused by changing d from 5 to zero. The population mean changes, as does the mean of genotypes conditional on their having an A_1 allele. The average effect of A_1, the difference between these two means, would change in the case of $p_1 = 0.1$, and therefore the additive and dominance genetic variances would change as well. Thus, the additive genetic variance is influenced by dominant as well as additive gene action effects.

Critical to the discussion in this paper, epistatic gene action effects influence the average effects of alleles and dominance deviations, and, consequently, the additive and dominance genetic variances. Cheverud and Routman (1995) demonstrated that strong "physiological" epistasis can exist in populations in which most of the epistatic effects contribute to additive genetic variance and in which the epistatic variance component itself is small. To illustrate this, consider the genetic effects for a trait influenced by two unlinked loci, each with two alleles (Table 2.2). To define the genetic components of variance for this genetic system, one must work with the genotypic values in terms of the gene action model given in Table 2.2, and also consider the genotypic frequencies (Table 2.5). Assuming a random-mating population in both Hardy-Weinberg and gametic phase equilibria, the statistical genetic effects can be defined following the approach of least-squares estimation of factorial effects (Cockerham 1954; Kempthorne 1954). Under these assumptions, the statistical genetic effects are defined explicitly in terms of gene action effects and allele frequencies in Table 2.6. Even this simplest of two-locus systems produces complicated definitions of statistical genetic effects! For example, the additive statistical effects of alleles are functions of allele frequencies and additive, dominance, and epistatic gene action effects (Table 2.6). The dominance statistical effects are functions of allele frequencies and dominant gene action and epistatic effects (Table 2.6). The epistatic statistical effects, however, are functions of allele frequencies and solely epistatic gene action effects. Since the additive genetic variance is a function of squared additive statistical effects, then the additive genetic variance is affected by epistatic gene action effects, whereas the epistatic variance (a function of squared epistatic gene action effects) is not affected by additive or dominant effects.

Table 2.5. Genotypic values, genotypic frequencies, and weighted marginal means of two-locus, two-allele model of gene action, assuming Hardy-Weinberg and gametic phase equilibria.

		B-locus genotype			
		B_1B_1	B_1B_2	B_2B_2	
		$f(B_kB_l)$, Single-locus genotype frequency			
		q_1^2	$2q_1q_2$	q_2^2	
A-locus genotype	$f(A_iA_j)$	Two-locus genotypic values (and frequencies)			Weighted marginal mean†
A_1A_1	p_1^2	$m+a^A+a^B+aa$ $(p_1^2q_1^2)$	$m+a^A+d^B+ad$ $(2p_1^2q_1q_2)$	$m+a^A-a^B-aa$ $(p_1^2q_2^2)$	$\mu_{11..} = m+a^A+(q_1-q_2)a^B+2q_1q_2d^B+(q_1-q_2)aa+2q_1q_2ad$
A_1A_2	$2p_1p_2$	$m+d^A+a^B+da$ $(2p_1p_2q_1^2)$	$m+d^A+d^B+dd$ $(4p_1p_2q_1q_2)$	$m+d^A-a^B-da$ $(2p_1p_2q_2^2)$	$\mu_{12..} = m+d^A+(q_1-q_2)a^B+2q_1q_2d^B+(q_1-q_2)da+2q_1q_2dd$
A_2A_2	p_2^2	$m-a^A+a^B-aa$ $(p_2^2q_1^2)$	$m-a^A+d^B-ad$ $(2p_2^2q_1q_2)$	$m-a^A-a^B+aa$ $(p_2^2q_2^2)$	$\mu_{22..} = m-a^A+(q_1-q_2)a^B+2q_1q_2d^B-(q_1-q_2)aa-2q_1q_2ad$
Weighted marginal mean†		$\mu_{..11} = m+(p_1-p_2)a^A+2p_1p_2d^A+a^B+(p_1-p_2)aa+2p_1p_2da$	$\mu_{..12} = m+(p_1-p_2)a^A+2p_1p_2d^A+d^B+(p_1-p_2)ad+2p_1p_2dd$	$\mu_{..22} = m+(p_1-p_2)a^A+2p_1p_2d^A-a^B-(p_1-p_2)aa-2p_1p_2da$	$\mu_{....} = m+(p_1-p_2)a^A+(q_1-q_2)a^B+2p_1p_2d^A+2q_1q_2d^B+(p_1-p_2)(q_1-q_2)aa+2p_1p_2(q_1-q_2)da+2(p_1-p_2)q_1q_2ad+4p_1p_2q_1q_2dd$

†Note that $(p_1^2-p_2^2) = p_1-p_2$ and that $(q_1^2-q_2^2) = q_1-q_2$.

Table 2.6. Statistical genetic parameters of the two-locus, two-allele model, assuming Hardy-Weinberg and gametic phase equilibria.

Model for genotypic value: $G_{ijkl} = \mu_{....} + {}^{A}_{i} + {}^{A}_{j} + {}^{B}_{k} + {}^{B}_{l} + {}^{A}_{ij} + {}^{B}_{kl} + {}_{ik} + {}_{il} + {}_{jk} + {}_{jl} + {}_{ikl} + {}_{jkl} + {}_{ijk} + {}_{ijl} + {}_{ijkl}$

Restrictions of genotypic model:

$$\sum_{i=1}^{2} p_i \alpha_i^A = 0, \sum_{k=1}^{2} q_k \alpha_k^B = 0$$

$$\sum_{i=1}^{2}\sum_{j=1}^{2} p_i p_j \delta_{ij}^A = 0, \sum_{k=1}^{2}\sum_{l=1}^{2} q_k q_l \delta_{kl}^B = 0, \sum_{i=1}^{2} p_i \; {}^{A}_{ij} = 0, \sum_{k=1}^{2} q_k \; {}^{B}_{kl} = 0$$

$$\sum_{i=1}^{2}\sum_{k=1}^{2} p_i q_k \; {}_{ik} = 0, \sum_{k=1}^{2} q_k \; {}_{ik} = 0, \sum_{i=1}^{2} p_i \alpha \alpha_{ik} = 0$$

$$\sum_{i=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} p_i q_k q_l \alpha \delta_{ikl} = 0, \sum_{k=1}^{2}\sum_{l=1}^{2} q_k q_l \; {}_{ikl} = 0, \sum_{i=1}^{2} p_i \; {}_{ikl} = 0$$

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} p_i p_j q_k \; {}_{ijk} = 0, \sum_{i=1}^{2}\sum_{j=1}^{2} p_i p_j \; {}_{ijk} = 0, \sum_{k=1}^{2} q_k \delta \alpha_{ijk} = 0$$

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} p_i p_j q_k q_l \delta \delta_{ijkl} = 0, \sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} p_j q_k q_l \; {}_{ijkl} = 0, \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} p_i p_j q_k \; {}_{ijkl} = 0,$$

$$\sum_{i=1}^{2}\sum_{i=1}^{2} p_i p_j \; {}_{ijkl} = 0, \sum_{k=1}^{2}\sum_{l=1}^{2} q_k q_l \delta \delta_{ijkl} = 0, \sum_{i=1}^{2}\sum_{k=1}^{2} p_j q_k \; {}_{ijkl} = 0, \sum_{i-1}^{2} p_i \; {}_{ijkl} = 0, \sum_{k=1}^{2} q_k \; {}_{ijkl} = 0$$

Overall genotypic mean = $\mu_{....} = m + (p_1 - p_2)a^A + (q_1 - q_2)a^B + 2p_1p_2d^A + 2q_1q_2d^B + (p_1 - p_2)(q_1 - q_2)aa + 2(p_1 - p_2)q_1q_2ad + 2p_1p_2(q_1 - q_2)da + 4p_1p_2q_1q_2dd$

Mean of genotypes, conditional on having one A_1 allele = $p_1\mu_{11..} + p_2\mu_{12..}$

Mean of genotypes, conditional on having one A_2 allele = $p_1\mu_{12..} + p_2\mu_{22..}$

Mean of genotypes, conditional on having one B_1 allele = $q_1\mu_{..11} + q_2\mu_{..12}$

Mean of genotypes, conditional on having one B_2 allele = $q_1\mu_{..12} + q_2\mu_{..22}$

Additive statistical effects:

$\alpha_1^A = p_1\mu_{11..} + p_2\mu_{12..} - \mu_{....} = p_2[a^A + (p_2 - p_1)d^A] + p_1p_2[(q_1 - q_2)(aa - da) + 2q_1q_2(ad - dd)] + p_2^2[(q_1 - q_2)(aa + da) + 2q_1q_2(ad + dd)]$

$\alpha_2^A = p_1\mu_{12..} + p_2\mu_{22..} - \mu_{....} = -p_1[a^A + (p_2 - p_1)d^A] - p_1p_2[(q_1 - q_2)(aa + da) + 2q_1q_2(ad + dd)] - p_1^2[(q_1 - q_2)(aa - da) + 2q_1q_2(ad - dd)]$

$\alpha_1^B = q_1\mu_{..11} + q_2\mu_{..12} - \mu_{....} = q_2[a^B + (q_2 - q_1)d^B] + q_1q_2[(p_1 - p_2)(aa - ad) + 2p_1p_2(da - dd)] + q_2^2[(p_1 - p_2)(aa + ad) + 2p_1p_2(da + dd)]$

$\alpha_2^B = q_1\mu_{..12} + q_2\mu_{..22} - \mu_{....} = -q_1[a^B + (q_2 - q_1)d^B] - q_1q_2[(p_1 - p_2)(aa + ad) + 2p_1p_2(da + dd)] - q_1^2[(p_1 - p_2)(aa - ad) + 2p_1p_2(da - dd)]$

Dominant statistical effects:

$\delta_{11}^A = \mu_{11..} - \alpha_1^A - \alpha_1^A - \mu_{....} = -2p_2^2[d^A + (q_1 - q_2)da + 2q_1q_2dd]$

$\delta_{12}^A = \mu_{12..} - \alpha_1^A - \alpha_2^A - \mu_{....} = 2p_1p_2[d^A + (q_1 - q_2)da + 2q_1q_2dd]$

$\delta_{22}^A = \mu_{22..} - \alpha_2^A - \alpha_2^A - \mu_{....} = -2p_1^2[d^A + (q_1 - q_2)da + 2q_1q_2dd]$

$\delta_{11}^B = \mu_{..11} - \alpha_1^B - \alpha_1^B - \mu_{....} = -2q_2^2[d^B + (p_1 - p_2)ad + 2p_1p_2dd]$

$\delta_{12}^B = \mu_{..12} - \alpha_1^B - \alpha_2^B - \mu_{....} = 2q_1q_2[d^B + (p_1 - p_2)ad + 2p_1p_2dd]$

$\delta_{22}^B = \mu_{..22} - \alpha_2^B - \alpha_2^B - \mu_{....} = -2q_1^2[d^B + (p_1 - p_2)ad + 2p_1p_2dd]$

Epistatic effect contrasts and sums involved in statistical epistatic effects:

$c_{11} = aa - ad - da + dd$

$c_{12} = aa + ad - da - dd$

$c_{21} = aa - ad + da - dd$

$c_{22} = aa + ad + da + dd$

Additive-by-additive statistical epistatic effects:

$\alpha\alpha_{11} = \mu_{1.1.} - \alpha_1^A - \alpha_1^B - \mu_{....}$

$\alpha\alpha_{11} = p_2q_2[p_1q_1(aa - ad - da + dd) + p_1q_2(ad + aa - dd - da) + p_2q_1(da - dd + aa - ad) + p_2q_2(dd + da + ad + aa)]$

$\alpha\alpha_{11} = p_2q_2[p_1q_1c_{11} + p_1q_2c_{12} + p_2q_1c_{21} + p_2q_2c_{22}]$

$\alpha\alpha_{12} = -p_2q_1[p_1q_1c_{11} + p_1q_2c_{12} + p_2q_1c_{21} + p_2q_2c_{22}]$

$\alpha\alpha_{21} = -p_1q_2[p_1q_1c_{11} + p_1q_2c_{12} + p_2q_1c_{21} + p_2q_2c_{22}]$

$\alpha\alpha_{22} = p_1q_1[p_1q_1c_{11} + p_1q_2c_{12} + p_2q_1c_{21} + p_2q_2c_{22}]$

Table 2.6. (*Continued*)

Additive-by-dominant statistical epistatic effects:

$\alpha\delta_{111} = \mu_{1.11} - \alpha_1{}^{A} - 2\alpha_1{}^{B} - \delta_{11}{}^{B} - 2\alpha\alpha_{11} - \mu_{...}$

$\alpha\delta_{111} = p_2q_2{}^2[p_1(c_{11} - c_{12}) + p_2(c_{21} - c_{22})] = 2p_2q_2{}^2\,[p_1(dd - ad) - p_2(dd + ad)]$

$\alpha\delta_{112} = -2p_2q_1q_2[p_1(dd - ad) - p_2(dd + ad)]$

$\alpha\delta_{122} = 2p_2q_1{}^2[p_1(dd - ad) - p_2(dd + ad)]$

$\alpha\delta_{211} = -2p_1q_2{}^2[p_1(dd - ad) - p_2(dd + ad)]$

$\alpha\delta_{212} = 2p_1q_1q_2[p_1(dd - ad) - p_2(dd + ad)]$

$\alpha\delta_{222} = -2p_1q_1{}^2[p_1(dd - ad) - p_2(dd + ad)]$

Dominant-by-additive statistical epistatic effects:

$\delta\alpha_{111} = \mu_{111.} - 2\alpha_1{}^{A} - \alpha_1{}^{B} - \delta_{11}{}^{A} - 2\alpha\alpha_{11} - \mu_{...}$

$\delta\alpha_{111} = p_2{}^2q_2[q_1(c_{11} - c_{21}) + q_2(c_{12} - c_{22})] = 2p_2{}^2q_2[q_1(dd - da) - q_2(dd + da)]$

$\delta\alpha_{121} = -2p_1p_2q_2[q_1(dd - da) - q_2(dd + da)]$

$\delta\alpha_{221} = 2p_1{}^2q_2[q_1(dd - da) - q_2(dd + da)]$

$\delta\alpha_{112} = -2p_2{}^2q_1[q_1(dd - da) - q_2(dd + da)]$

$\delta\alpha_{122} = 2p_1p_2q_1[q_1(dd - da) - q_2(dd + da)]$

$\delta\alpha_{222} = -2p_1{}^2q_1[q_1(dd - da) - q_2(dd + da)]$

Dominant-by-dominant statistical epistatic effects:

$\delta\delta_{1111} = \mu_{1111} - 2\alpha_1{}^{A} - 2\alpha_1{}^{B} - \delta_{11}{}^{A} - \delta_{11}{}^{B} - 4\alpha\alpha_{11} - \mu_{...}$

$\delta\delta_{1111} = p_2{}^2q_2{}^2[c_{11} - c_{12} - c_{21} + c_{22}] = 4p_2{}^2q_2{}^2dd$

$\delta\delta_{1112} = -4p_2{}^2q_1q_2dd$

$\delta\delta_{1122} = 4p_2{}^2q_1{}^2dd$

$\delta\delta_{1211} = -4p_1p_2q_2{}^2dd$

$\delta\delta_{1212} = 4p_1p_2q_1q_2dd$

$\delta\delta_{1222} = -4p_1p_2q_1{}^2dd$

$\delta\delta_{2211} = 4p_1{}^2q_2{}^2dd$

$\delta\delta_{2212} = -4p_1{}^2q_1q_2dd$

$\delta\delta_{2222} = 4p_1{}^2q_1{}^2dd$

In conclusion, the parameters of the statistical genetic model are not necessarily good guides to the underlying gene action. Furthermore, the relationship between gene action effects and statistical genetic effects becomes less direct as more interacting loci contribute to the character in question. Given that even under greatly simplifying conditions classic quantitative genetic theory leads to estimable parameters that do not necessarily indicate the nature of gene action, one might question the usefulness of this theory. In defense of classical quantitative genetics, the statistical genetic models were developed to understand the inheritance of phenotypes that were controlled by many genes, the individual effects of which were unobservable. In order to get around the problem of not being able to classify individuals according to genotype (as you can with discrete, Mendelian characters), quantitative genetics was developed to relate the *combined effects* rather than *individual effects* of alleles and their interactions to observable phenotypic quantities. Therefore, instead of relating genetic variances directly to gene action effects, genetic variances were related to aggregate statistical effects of genes in specific populations. The parameters of the statistical genetic model, such as genetic variances and heritability, have two advantages over the gene action model parameters: they are estimable from phenotypic data alone, with no knowledge of underlying genotypes; and they can often accurately predict the response to different forms of selection. Therefore, quantitative genetics investigations into epistasis would seem to hold some promise for elucidating the ways in which epistasis might influence selection response.

III. EPISTASIS AND MOLECULAR INTERACTIONS

Before reviewing the literature on biometrical investigations of epistasis, however, let us first consider the current understanding of the biology of gene interactions. What sort of genic interactions should we expect to underly phenotypes?

A. Molecular Interactions Lead to Epistasis

Epistasic gene action can be due to duplicate gene interaction (Fig. 2.1), complementary gene interaction (Fig. 2.2), or more complex forms of non-allelic interaction (Fig. 2.3). Epistasis can be mediated by interactions among enzymes encoded by different genes (Fig. 2.1–2.3); or by the interactions at any level between gene and phenotype, from gene transcription (e.g., transcription factors and their targets) to morphology

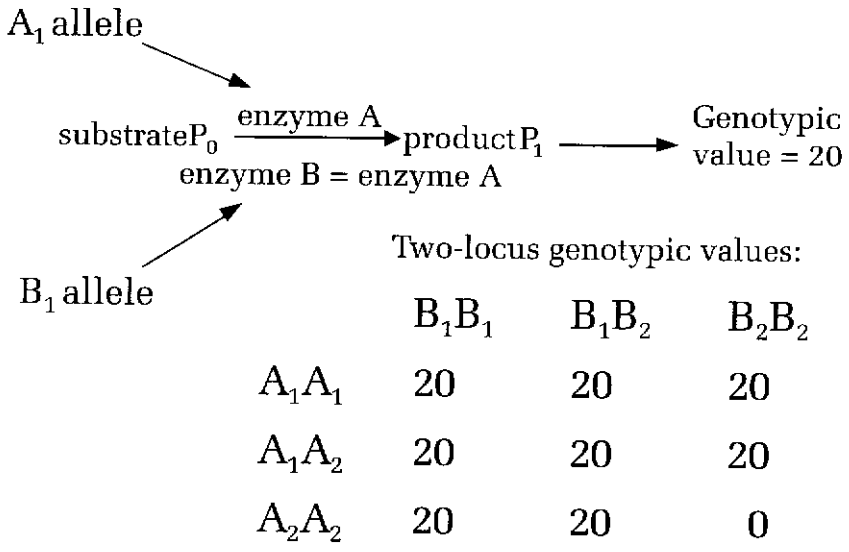
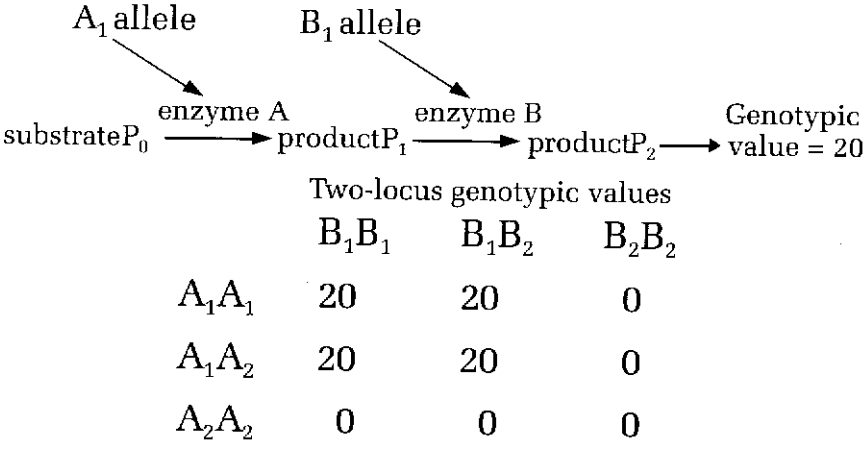


Fig. 2.1. Duplicate gene interaction form of epistasis. A functional gene product from either of two duplicated loci A or B is required to produce the optimum genotype. Only the doubly homozygous recessive genotype $A_2A_2B_2B_2$ confers the inferior genotypic value.

(e.g., features that are adaptive in one morphological background and deleterious in another). Here, I present examples of how genes or gene products interact biochemically or molecularly and describe the resulting phenotypic consequences of different forms of gene interactions on simply-inherited traits.

B. Duplicate Gene Interactions and Genetic Redundancy

Duplicate gene interactions can occur when two or more loci serve the same function, for example, by producing an identical enzyme that results in an observable phenotype (Fig. 2.1). If either locus can produce a functional gene product, then the enzymatic pathway functions. Only when both genes are "knocked out" or homozygous for a recessive, nonfunctioning allele is the biochemical process interrupted, producing a different phenotype. In a diploid F_2 population segregating for recessive mutants at both loci, this results in a 15:1 phenotypic ratio. Examples include antibiosis to an insect pest in maize (*Zea mays*) (Byrne et al. 1998), and gene-for-gene pathogen resistances in which an incompatible



Two-locus genotypic values

	B_1B_1	B_1B_2	B_2B_2
A_1A_1	20	20	0
A_1A_2	20	20	0
A_2A_2	0	0	0

Fig. 2.2. Complementary gene interaction form of epistasis. Functional gene products from both loci A and B are required to produce the optimum genotype. Genotypes homozygous recessive at either of the two complementary loci have inferior genotypic value.

reaction due to one host resistance gene is sufficient to confer resistance (Flor 1956). Polyploidy, genome duplication and genetic redundancy are common in plants (Pickett and Meeks-Wagner 1995), and this suggests that genes with duplicated functions may also be common. As an example, two *stamenoid petal* genes in *Brassica napus* that map to homoeologous chromosomes must both be homozygous for recessive mutant alleles in order to confer the mutant morphology (Fray et al. 1997).

C. Complementary Gene Interactions in Biochemical and Molecular Pathways

Complementary gene interaction can result when two or more genes code for enzymes that function at different points in the same pathway, so that functioning products from all genes in the set are needed to produce the final product (Fig. 2.2). If any gene or gene product in the pathway is non-functioning, then the final product of the pathway is not produced. In an F_2 population segregating at two complementary loci, a 9:7 phenotypic ratio is expected. Biochemical examples of this form of epistasis include the requirement for variant alleles at two genes to produce a unique terpenoid in cotton (*Gossypium hirsutum*) (Kohel and Bell 1999) and the production of maysin in maize (McMullen et al.

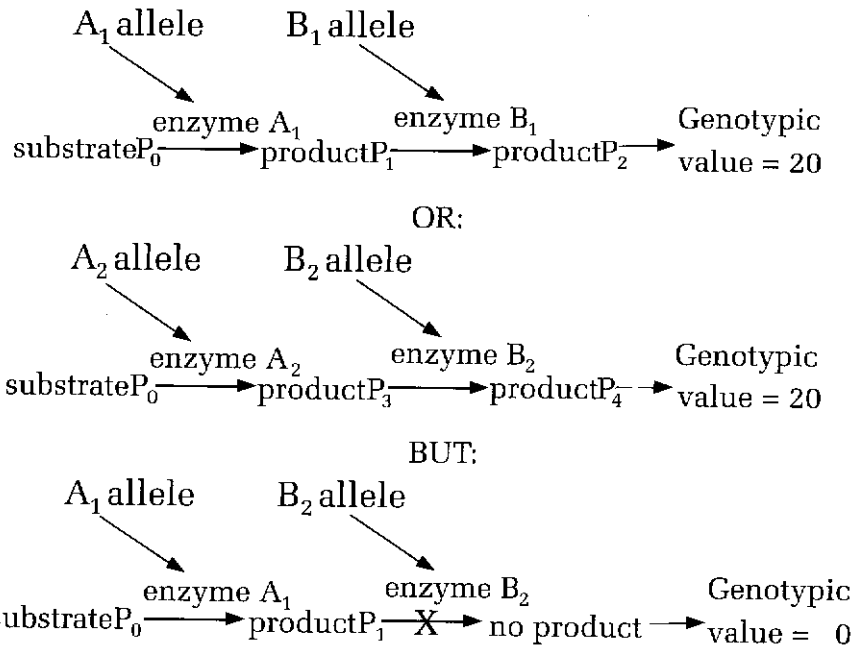


Fig. 2.3. Complicated epistasis. Gene products from allele pair A_1/B_1 function well together and produce an optimum genotype. Similarly, gene products from allele pair A_2/B_2 functional well together and produce an optimum genotype. Products of allele pair A_1/B_2 are not functionally compatible and alone will result in an inferior genotypic value.

1998). At the level of morphology and developmental responses, complementary gene interactions have been observed for photoperiod-sensitivity in sorghum (*Sorghum bicolor*) (Rooney and Aydin 1999), and floral morphology in maize (Doebley et al. 1995).

Another form of epistasis occurs when the genotype at an early step in a pathway can mask genetic variation at genes acting later in the pathway, but the masking effect is not reciprocated, as it is in the complementary gene interaction shown in Fig. 2.2. This is the "classical" definition of epistatic gene action originally given by Bateson (1909, Chap. 4) in reference to the interaction of mouse coat color genes. These types of gene interactions are found commonly in developmental regulatory pathways studied in model systems, including those affecting sex-determination in animals, the cell cycle in yeast, embryonic pattern formation in *Drosophila* (Avery and Wasserman 1992), flowering time

and vernalization response in both barley (*Hordeum vulgare*) and diploid wheat (*Triticum monococcum*) (Tranquilli and Dubcovsky 2000), and floral development in *Arabidopsis* (Parcy et al. 1998). For example, mutations at the early-acting floral development regulator gene, *LFY*, eliminate or radically transform flowers, and this effect masks mutations at the later-acting *AP1*, *AP3*, *AG*, and *PI* regulator genes, which normally act to specify the different floral organs (Parcy et al. 1998). Similar gene interactions are expected to occur in other types of molecular pathways, including disease resistance response pathways (Crute 1998), and signal transduction pathways (e.g., the phytochrome system, Millar et al. 1994). Direct evidence for protein-protein interactions in the phytochrome system exists: the *phyA* protein phosphorylates the *cry1* and *cry2* cryptochrome proteins in vitro (Ahmad et al. 1998), although the biological significance of this particular interaction is not clear (Casal 2000).

Epistasis cannot always be easily classified into duplicate or complementary forms; a rich diversity of genic interactions is possible. For example, mutations in the *Arabidopsis* gene *CAL*, which is homologous to the transcription factor *AP1*, in an otherwise normal genetic background have no phenotypic effect (Martienssen 1999). The sequence similarity and masking effect of wild-type *AP1* alleles on mutant *CAL* alleles suggests that *CAL* and *AP1* interact as duplicate genes. *AP1* mutations in wild-type *CAL* backgrounds, however, have a mutant phenotype, suggesting that *AP1* is epistatic in the classical sense to *CAL*. To further complicate the picture of allelic interactions between these loci, *CAL-AP1* double mutants have a novel phenotype that could not have been predicted from either single-mutant phenotype alone (Martienssen 1999). Martienssen (1999) suggested that these floral identity genes have overlapping, but non-identical functions in common developmental pathways. Thus, interactions among these genes cannot be considered as simply complementary or duplicated, or upstream and downstream in a common pathway, but rather as a more subtle combination of the two types of interactions. Combinations of mutations in these genes will generally result in unpredictable phenotypes.

Production of anthocyanin in maize aleurone tissue is a well-studied example that demonstrates both duplicate and complementary gene interactions mediated by both biochemical activity and transcriptional regulation (Jayaram and Peterson 1990). At least three structural enzymes, encoded by genes *C2* or *Whp*, *A1*, and *A2*, are required to produce anthocyanin and a visible pigment (Coe 1994). Genotypes that are homozygous for a non-functional allele at the *A1* gene, which codes for dihydroflavonol reductase, the second step in the pathway, will lack

aleurone pigment, regardless of the *A2* genotype. Similarly, if *A2* is homozygous for a nonfunctioning allele, no anthocyanin will be produced, no matter what the genotypes are at upstream genes. These complementary gene interactions are mediated at the biochemical level because gene transcription at the *A1* gene is not affected by the genotype at the *A2* gene. A duplicate gene interaction occurs at the biochemical level between the *C2* and *Whp* genes because both genes code for the same enzyme, chalcone synthase (Cone 1994), and both genes must be homozygous for nonfunctional alleles in order to cause a mutant phenotype. Complementary gene interaction is also observed between this set of genes and the regulatory genes *B*, *R*, *C1*, and *Pl*, but this interaction is mediated through transcriptional regulation. A complex transcription factor composed of subunits coded for by either *B* or *R* and *C1* or *Pl* is required to activate transcription of the anthocyanin pathway structural genes (Goff et al. 1992). If the functioning transcription factor complex is not present, the structural genes will not be transcribed. A complementary biochemical interaction occurs between the *B/R* gene pair and the *C1/Pl* gene pair: if either pair cannot produce a functional gene product, then the other pair will still be transcribed but will not be able to activate transcription of the structural genes. Finally, there are duplicate gene interactions between the regulatory genes themselves: *B* and *R* code for functionally duplicate gene products, and *C1* and *Pl* are duplicate genes.

D. Synergistic and Antagonistic Gene Interactions

Other possible forms of genic interaction include synergistic or antagonistic interactions. There can be confusion because the terms synergistic and antagonistic interactions have opposite meanings depending on whether inference is made either to the phenotypic effect of mutations of the genes or to the molecular interaction of the gene products (Casal 2000). Strong synergistic interactions among enzymes that seem to occur at the level of biochemical activity have been reported in yeast (Niederberger et al. 1992). Casal (2000) reviewed the interactions among phytochromes and cryptochromes in plants, and observed that phytochromes *phyA* and *phyB* function synergistically under some light conditions and antagonistically under others. The precise nature of the interactions is not yet known, but almost certainly does not involve direct physical interaction of the *phyA* and *phyB* molecules. The interaction is more likely an outcome of the shared regulatory networks that are functionally "downstream" of both molecules. The two molecules are partially redundant members of a common gene family, and therefore

likely affect similar regulatory pathways. The molecules respond to different light spectra, and there is evidence that *phyA* regulates different signal cascades depending on the light regime. These small variations between otherwise seemingly-redundant molecules generate a complex set of interactions observed between mutations in the genes, and between these mutations and the environment.

Transgene silencing in plants represents an unnatural but clearly defined example of antagonistic gene interactions (Nap et al. 1997). Often, transgene constructs that are expressed at higher levels when present individually in different genotypes have altered expression patterns when combined with other, at least partially homologous, transgenes in a single genotype. The silencing can be non-reciprocal, in which one transgene is expressed but represses expression of the other gene (Matzke et al. 1993), or it can be reciprocal "co-suppression," in which the expression of both transgenes is suppressed (Neuhuber et al. 1994). In addition, transgenes can inhibit expression of endogenous genes with which they share sequence homology (Jorgensen 1995). The suppression of expression caused by the interactions of homologous genes can occur at the level of transcription or post-transcriptionally. The mechanisms by which these interactions are mediated are not completely understood, but there is evidence for DNA-DNA interactions and methylation mediating transcriptional inactivation, and for the involvement of RNA-RNA interactions and RNA degradation mechanisms in post-transcriptional gene silencing (Matzke and Matzke 1995). Further research on silencing mechanisms may lead to a better understanding of normal gene regulation in plants. Many plant genes exist as members of gene families, as duplicate or redundant genes (paralogs), and as members of homoeologous groups in polyploids. Perhaps the mechanisms of gene silencing are also involved in the natural, coordinated regulation of gene expression and interactions among homologous and partially homologous sets of genes in plants. Indeed, by varying the dosage of individual chromosomal segments, Guo et al. (1996) demonstrated that normal maize gene expression is determined by the interaction of regulator and structural genes on different chromosomes.

E. Compensatory Gene Interactions

There are unlimited possibilities for different forms of epistasis, and in the most extreme cases, the description of an allele as "favorable" or "unfavorable" may depend upon the genotype at other loci. Clear examples of this type of interaction occurring among enzymes is lacking, but there is some evidence in plants for this type of "compensatory" gene

interaction occurring at the level of DNA transcription regulation. Research on the development of *Arabidopsis thaliana* has revealed examples of this type of genic interaction.

Development in *Arabidopsis* is controlled in part by the action of floral meristem identity genes, such as *LFY* or *AP1*, and the antagonistic action of the vegetative shoot identity gene, *TFL1*. Plants homozygous for null alleles at the *LFY* or *AP1* loci lose their meristem identity, at least partially, and form flowers with shoot-like characteristics (Ratcliffe et al. 1999). Plants homozygous for null alleles at the *TFL1* gene have shoots that become transformed into flowers. *Arabidopsis* plants transgenically modified to over-express *LFY* or *AP1* look similar to *tfl1tfl1* mutants, while those modified to over-express *TFL1* have a much longer vegetative phase. Thus, over-expression of *LFY* mimics lack of expression of *TFL1*, and vice-versa. When plants are modified to overexpress both *LFY* and *TFL1*, the resulting phenotype is nearly normal (Ratcliffe et al. 1999). Modified expression of *TFL1* leads to a mutant phenotype in an otherwise wild-type background, but in a background in which the expression of *LFY* is similarly modified, it leads to a normal phenotype. The phenotypic effect of over-expression of *TFL1* leads to either a mutant or to a normal phenotype, depending on the genotype at other genes with which it interacts because of the compensatory gene interaction.

Compensatory mutations have been studied primarily within the same gene or in *cis* configurations (Cleghon et al. 1996; Kirby et al. 1995), because it is easier to identify *cis*-acting than *trans*-acting compensatory mutations. Nevertheless, RNA-RNA pairing among small nuclear RNAs (snRNAs) that compose the spliceosome (involved in intron splicing) and RNA-RNA pairing between snRNAs and messenger RNAs (mRNAs) present a clear example of *trans*-acting allele-specific compensatory mutations that occur at the molecular level (Madhani and Guthrie 1994). Base changes at a normal intron splice site in the DNA encoding a "target" gene will prevent pairing between the resulting mRNA and a snRNA that normally binds to the site, and this will prevent proper splicing of the intron, as demonstrated in model systems such as human cell cultures (Sun and Manley 1995) and *Drosophila* (Lo et al. 1994). Similarly, mutations in the DNA encoding the specific snRNA that binds the target site at the site involved in pairing with mRNA prevent proper intron splicing of the normal target gene mRNA (Sun and Manley 1995). If the mutations in the target gene and the sRNA gene are complementary, however, the mutant snRNA can promote proper splicing of the mutant target gene (Lo et al. 1994; Sun and Manley 1995), resulting in a normal phenotype.

Although biochemical and molecular examples of epistatic interactions that cause the value of a genotype at one locus to change from pos-

itive to negative depending on the genetic background are rare, these types of interactions have been observed at the phenotypic level. Examples include flowering time QTLs in *Brassica* (Camargo and Osborn 1996) and oat (*Avena sativa*) (Holland et al. 1997) and interactions between specific P-element insertions on their effects on metabolic phenotypes in *Drosophila* (Clark and Wang 1997).

Examples of *trans*-acting compensatory mutations at the molecular level have not been reported more frequently perhaps because molecular developmental biology research is biased against working with such mutations. Tests of epistasis are a powerful tool that developmental biologists use to determine the order of developmental triggers in gene regulation pathways, but such tests are most effective when alleles conferring complete loss-of-function at the genes under study are used (Avery and Wasserman 1992). Avery and Wasserman (1992) advise "partial loss-of-function mutations should be avoided" in these types of experiments. Some allelic variants at regulatory genes confer quantitative differences in function and perhaps even different kinds of functions, and these "weaker" variants may confer unique epistatic effects that are not observed with alleles that cause genes to be either "on" or "off" (e.g., Ang and Deng 1994). Developmental biologists attempt to understand the basic flow of regulatory gene cascades rather than test for quantitative forms of epistasis in such regulatory mechanisms, but overlooking weaker mutations at regulatory genes may result in a biased view of the variation in these pathways in natural populations.

The result of compensatory mutations can be described as additive-by-additive epistasis and phenotypic values like those in Fig. 2.4 result

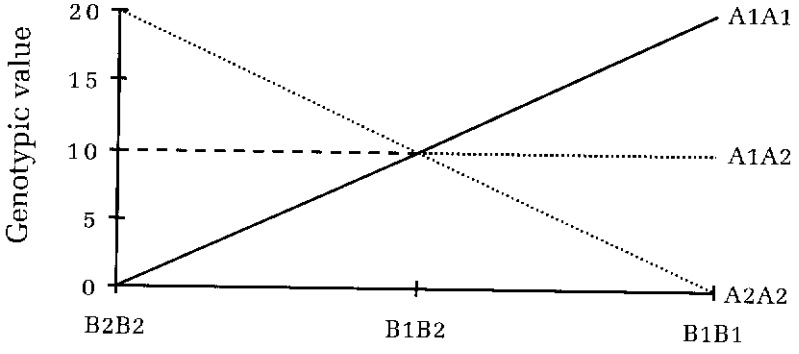


Fig. 2.4. Graphical representation of additive-by-additive form of epistasis that could result from the relationships shown in Figure 3. The optimum genotypic value is conferred by genotypes $A_1A_1B_1B_1$ and $A_2A_2B_2B_2$. Which is the favorable allele at locus A? It depends on the genotype at locus B.

if there is no dominance at either locus involved. This is a very strong, and very extreme form of epistasis, but it illustrates the general principle that in such cases the question, "which is the favorable allele?" becomes rather like a Zen koan. The answer is that the advantage of one allele versus another at one locus depends completely on the genotype at the second locus. There is no favorable allele, there are only favorable and unfavorable combinations of alleles.

IV. COMPLEX MOLECULAR INTERACTIONS UNDERLIE QUANTITATIVE PHENOTYPES (SOMETIMES)

The examples provided above were culled from only a limited number in which gene interactions are observable directly at the phenotypic level and also have well-characterized molecular basis of interaction. Given that many agronomically important traits are inherited multigenically and are likely affected by numerous biochemical pathways and developmental regulatory networks, a better understanding is needed of the effects of multitudinous gene interactions mediated through complex regulatory networks and pathways on quantitative phenotypes. If molecular interactions among genes and gene products are common, as we suspect them to be, how does this affect our current notions of quantitative inheritance? Can we interpret statistical genetic parameters, such as additive genetic variance, epistatic genetic variance, heritability, and heterosis, in terms of molecular interactions? Under what conditions are gene interactions at the molecular level likely to have an important effect on the phenotype? Can interactions occur at higher levels of organization (such as the phenotype) in the absence of molecular interactions? Classical quantitative genetic theory alone will not provide answers to these questions. When more realistic quantitative genetics models are considered, the mathematics quickly become at best unwieldy, such as when multiple alleles and inbreeding are introduced (e.g., Weir and Cockerham 1977) and often intractable, as when multiple loci, linkage, and selection are considered simultaneously (e.g., Kempthorne 1988). In the face of mathematical complexity, therefore, computer-aided simulation studies, metabolic control theory (which interprets gene action in terms of physical chemistry of enzymes), and network theory have been used to address some of these questions. The results may help us to better understand the empirical data to be considered later.

A. Models of Interactions in Metabolic Pathways

Keightley (1989) used metabolic control theory to investigate the nature of interactions between enzymes in common or competing branches of a biochemical pathway. An example of enzymes that compete for a common substrate are those of the citric acid cycle, fatty acid synthesis, cholesterol synthesis, and ketone body synthesis pathways, all of which require acetyl-CoA (Stryer 1988, p. 634). Keightley showed that, in most cases, additive genetic variance is expected to be much larger than epistatic interaction variance for metabolic pathway flux, despite the fact that the metabolic pathways are inherently interactive. Furthermore, most interactions between enzymes are expected to be antagonistic for flux (Keightley 1996a; Szathmáry 1993). In this case, response to selection for increasing fitness is expected to be non- linear, with an accelerating rate of increase (Keightley 1996b). In contrast, interactions between enzymes are most often expected to be synergistic if fitness is related to optimal flux (a non-linear function of maximal flux) or metabolic pool size, rather than maximal flux (Szathmáry 1993). Although Szathmáry (1993) did not discuss the magnitude of epistatic variance for fitness under these conditions, the contrast between his results and Keightley's (1989) indicates that caution should be used when drawing general inferences from these modeling studies, because the results can depend greatly on the model assumptions. The models are limited by the assumptions of linear pathways, nonsaturable enzymes, and lack of feedback regulation (Szathmáry 1993). Nevertheless, Keightley's (1989) results suggest that strong biochemical interactions between enzymes do not necessarily result in significant epistatic variances at the phenotypic level.

B. A Model of Gene Interactions in Developmental Regulation Pathways

The limitations of metabolic control theory prompted the investigation of gene action and interaction models based on the regulation of gene expression, rather than on metabolic flux (Omholt et al. 2000). Gibson (1996) developed a thermodynamic model of transcriptional regulation based on the known properties of the threshold-dependent interaction between a gene involved in embryogenesis in *Drosophila* and its transcriptional activator. The system modeled may be similar to signaling cascades important in plant development and response to the environment. Key conclusions from Gibson's (1996) investigation of the properties of

this model are that epistasis is a natural consequence of the physical interaction of the gene promoter and the activator protein. Furthermore, such regulatory networks likely favor the prevalance of compensatory epistatic interactions in which "the same genetic output can be readily produced by numerous different genotypic combinations" (Gibson 1996). These are precisely the types of epistatic interactions that have the most profound effect on concepts of favorable alleles and selection.

How are molecular interactions related to phenotypes? Moreno (1994) suggested that gene product dosage often has a nonlinear relationship with phenotype, as has been frequently observed for highly redundant developmental regulator genes in *Drosophila*. In this case, allelic changes at any one of a set of redundant loci results in a change in gene activity at the locus, but only a small change in phenotype, because of the non-linear relationship between total gene activity of the set of redundant genes and phenotype. If allelic variants occur simultaneously at more than one locus of the interacting set, then the resulting phenotypic change is expected to be much larger than the sum of the single-locus changes.

In essence, this is an extension to multiple loci of the argument made by Kacser and Burns (1981) that dominant gene action is common because of the often non-linear relationship between gene dosage and biochemical activity. Moreno (1994) predicted that allelic variants at such loci could reside at relatively high frequencies in natural populations because their effects are minimized by the action of redundant loci. If population bottlenecks, selection, or inbreeding shift allele frequencies to the point where variant alleles at more than one locus occur quite frequently, then the epistasis will become apparent at the phenotypic level.

C. Boolean Regulatory Networks as Models for Genetic Regulatory Pathways

Kauffman's (1993) NK model of boolean regulatory networks is a general framework in which multigenic interactions can be modeled. An NK model is a network of N nodes (analogous to genes that act as regulatory switches), each of which has K inputs (analogous to transcriptional activators, enhancers, or repressors). N models the total number of genes in the network and K models the magnitude of interactivity or epistasis. The model is boolean because, at any one moment, each node is in one of two states, on or off (or 0 or 1).

Frank (1999) developed boolean networks with varying values for N and K that expressed a "phenotype" that is a mathematical function of

the state of a subset of the nodes. Populations of networks were created and subjected to many generations of truncation selection for the phenotype followed by "sexual reproduction." Then the quantitative genetics of the network were investigated by measuring the heritability of nodal changes and by measuring the "metabolic additivity" of nodal changes. The heritability is a measure of the average phenotypic effect of changing a node averaged over the whole population of network backgrounds, while the metabolic additivity is the average phenotypic effect of changing one nodal state within a constant network background. The results of this simulation were that highly connected networks (high K) had greater metabolic additivity, but lower heritability. This implies that, if the NK networks are representative of gene regulatory networks, then epistatic genetic variance, as well as epistatic genetic effects, are likely to be important at the phenotypic level. Furthermore, in highly-connected networks, "particular alleles will often be advantageous in one genetic background and disadvantageous in another" (Frank 1999), implicating the strongest form of epistasis, compensatory epistasis, as a natural component of regulatory networks. These results contrast with those of Keightley's (1989), but are similar to those of Gibson (1996), perhaps illustrating a fundamental difference between gene regulatory networks and biochemical pathways.

The question remains, however: how similar are boolean networks to real gene regulatory networks? There is growing evidence that they are surprisingly similar. For example, Yuh et al. (1998) demonstrated that a gene, *Endo16*, that is activated specifically in the midgut of late embryo and larval life stages of the sea urchin, is regulated by a boolean network encoded in its promoter and upstream regulatory regions. At least 15 different proteins bind with high specificity to target sites within this region, and the binding of these regulator proteins acts to turn the different nodes of the network on or off. One of the upstream regulatory regions synthesizes the information encoded in the network and acts as a switch to turn gene expression on or off. The operation of this *cis*-regulatory network can be encoded in the same manner as an analog computer (Yuh et al. 1998). Recently, simple boolean regulatory networks of transcriptional regulators that mimic some natural gene regulation networks have been constructed artificially with recombinant DNA techniques in bacteria (Elowitz and Leibler 2000; Gardner et al. 2000). Biological signaling pathways, however, cannot be divorced from the reality of cells and cell structures in which they occur, and the interactions between signaling pathways and the effects of spatial organization of the cell greatly add to the complexity of gene regulation (Weng et al. 1999).

D. Multiplicative Interactions: Phenotypic Epistasis as an Emergent Property of Additive Physiological Components

Multiplicative interactions among component characters of a complex trait are a form of synergistic epistasis that can lead to heterosis, as proposed originally by Richey (1942). For example, grain yield per plant in a cereal crop is the product of the number of tillers per plant, the number of seeds per tiller, and the weight of seeds. A linear increase in the mean weight of seeds, if all else remains constant, will result in a multiplicative increase in grain yield in this system. There is limited empirical evidence for this form of epistasis (Melchinger et al. 1994; Schnell and Cockerham 1992), but it demonstrates that epistatic interactions can occur at the phenotypic level without an underlying basis of molecular interactions. In principle, epistatic interactions among genes can be an emergent property at the level of a complex character that is not reflected by the gene action among the same genes at the level of component characters. Thus, in theory, epistasis can occur without an underlying molecular interaction!

V. BIOMETRICAL EVIDENCE FOR EPISTASIS

Empirical quantitative genetics investigations into epistasis have focused on estimating epistatic variance components or testing for epistasis based on the statistical significance of deviations from simpler additive-dominant models. Digenic epistatic variances were included in the first statistical genetic model developed. Fisher (1918) developed the model that is still used to partition the total genetic variance in a population into components due to additive variance, dominance variance, and epistatic variance. Cockerham (1954) and Kempthorne (1954) extended this by partitioning the total digenic epistatic variance into additive by additive ($A \times A$), additive-by-dominant ($A \times D$), and dominant-by-dominant ($D \times D$) variance components, and by showing how these epistatic components of variance are involved in the covariance between relatives. Understanding the contribution of the epistatic variances to the covariances among relatives created the opportunity to estimate epistatic components of variances using mating designs and phenotypic observations in the same manner that additive and dominance variance components had been estimated successfully (Cockerham 1963).

A. Outcrossing Plant Species

Hallauer and Miranda (1988, Chap. 5) reviewed four studies that attempted to estimate epistatic components of variance in maize using mating designs. They concluded that epistatic variance is not an important contributor to the genetic variance for yield in maize:

> It seems that epistasis for a complex trait, such as yield, must exist . . . but realistic estimates of additive by additive epistasis have not been obtainable. Hence either the genetic models used are inadequate or epistatic variance is small relative to total genetic variance of maize populations.

One of the major problems with estimation of epistatic variance components is that the coefficients of the additive and additive-by-additive variance components in the covariances of relatives are often quite similar, and the correlation between these coefficients prevents accurate simultaneous estimation of both components of variance. The additive and epistatic variance components become very difficult to distinguish (Brim and Cockerham 1961). Since the first-order parameters (e.g., the additive effects) are fit in the model first, this makes it difficult to estimate the higher order (epistatic) parameters.

To illustrate further the difficulty of estimating epistatic components of variance, the genetic components of variance for the simple two-locus, two-allele system presented in Tables 2.2 and 2.3 and Tables 2.5 and 2.6 were derived for the case where all allele frequencies equal one-half and the population is in Hardy-Weinberg and gametic phase equilibria (Table 2.7). These allele frequencies were chosen for illustration because the epistatic variance components are simple functions of the epistatic gene action parameters and because the epistatic variance components are maximum relative to the additive variance when the allele frequencies are intermediate. Thus, Table 2.7 illustrates the best possible population in which to detect epistatic variance components. If the only form of epistatic variance is additive-by-additive, then in this case, the ratio of additive-by-additive to additive genetic variance is:

$$\frac{\sigma^2_{AA}}{\sigma^2_A} = \frac{\frac{1}{4}(aa)^2}{\frac{1}{2}\left(\left(a^A\right)^2 + \left(a^B\right)^2\right)}$$

Table 2.7. Statistical genetic effects and variances for two-locus, two-allele model, when $p_1 = p_2 = q_1 = q_2 = 0.5$, assuming Hardy-Weinberg and gametic phase equilibria.

Statistical Genetic Effects	**Genetic Variance Components**
Additive effects: $\alpha_1^A = -\alpha_2^A = (\frac{1}{2})a^A + (1/4)ad$ $\alpha_1^B = -\alpha_2^B = (\frac{1}{2})a^B + (1/4)da$	Additive variance: $\sigma^2_A = \sigma^2_{A[A]} + \sigma^2_{A(B)} = 2\Sigma p_i(\alpha^A_i)^2 + 2\Sigma q_k(\alpha^B_k)^2 = (\frac{1}{2})[(a^A)^2 + (a^B)^2] + (1/8)[(ad)^2 + (da)^2] + (\frac{1}{2})[(a^A)ad + (a^B)da]$
Dominant effects: $\delta_{11}^A = \delta_{22}^A = -\delta_{12}^A = -(\frac{1}{2})[d^A + (\frac{1}{2})dd]$ $\delta_{11}^B = \delta_{22}^B = -\delta_{12}^B = -(\frac{1}{2})[d^B + (\frac{1}{2})dd]$	Dominance variance: $\sigma^2_D = \sigma^2_{D[A]} + \sigma^2_{D(B)} = \Sigma p_i p_j(\delta^A_{ij})^2 + \Sigma q_k q_l(\delta^B_{kl})^2 = (1/4)[(d^A)^2 + (d^B)^2] + (1/8)(dd)^2 + (1/4)[(d^A + d^B)dd]$
Additive–by–additive effects: $\alpha\alpha_{11} = -\alpha\alpha_{12} = -\alpha\alpha_{21} = \alpha\alpha_{22} = (1/4)aa$	Additive–by–additive variance: $\sigma^2_{AA} = 4\Sigma p_i q_k(\alpha\alpha_{ik})^2 = (1/4)(aa)^2$
Additive–by–dominant effects: $\alpha\delta_{111} = -\alpha\delta_{112} = \alpha\delta_{122} = -\alpha\delta_{211} = \alpha\delta_{212} = -\alpha\delta_{222} = -(1/4)ad$	Additive–by–dominant variance: $\sigma^2_{AD} = 2\Sigma p_i q_k q_l(\alpha\delta_{ikl})^2 = (1/8)(ad)^2$
Dominant–by–additive effects: $\delta\alpha_{111} = -\delta\alpha_{112} = \delta\alpha_{122} = -\delta\alpha_{211} = \delta\alpha_{212} = -\delta\alpha_{222} = -(1/4)da$	Dominant–by–additive variance: $\sigma^2_{DA} = 2\Sigma p_i p_j q_k(\delta\alpha_{ijk})^2 = (1/8)(da)^2$
Dominant–by–dominant effects: $\delta\delta_{1111} = -\delta\delta_{1112} = \delta\delta_{1122} = -\delta\delta_{1211} = \delta\delta_{1212} = -\delta\delta_{1222} = \delta\delta_{2211} = -\delta\delta_{2212} = \delta\delta_{2222} = (1/4)dd$	Dominant–by–dominant variance: $\sigma^2_{DD} = \Sigma p_i p_j q_k q_l(\delta\delta_{ijkl})^2 = (1/16)(dd)^2$

Even in the best possible case, the square of the additive-by-additive gene action effects must be at least four times the mean of the squared additive gene action effects in order to have a variance component of comparable size. If there are dominant forms of epistasis, the additive variance would become even larger relative to the epistatic component. Epistatic gene action effects contribute to additive genetic variance, and strong epistasis and biochemical interactions can exist in populations that do not have large epistatic variances (Cheverud and Routman 1995; Keightley 1989).

Due to this difficulty, other biometrical approaches have been developed to detect epistasis for quantitative characters. It is probably not a coincidence that biometrical methods that use mean comparisons rather than variance component estimation (for example, generation means analysis and the triple test cross design) have regularly indicated that epistatic effects are important for yield in maize. The triple test cross experimental design proposed by Kearsey and Jinks (1968) (similar to a

design used earlier by Bauman 1959) tests the null hypothesis of no epistatic effects using a comparison of means. Individuals, families, or lines from a population developed from two inbred lines are testcrossed to the two original inbred parents and also to the F_1 of the cross of the two parents. For each member of the population, the contrast $L_1 + L_2 - 2L_3$ is performed, where L_1 is the yield of the testcross to the first inbred parent, L_2 the yield of the testcross to the second inbred parent, and L_3 the yield of the testcross to the F_1. Additive and dominant gene effects sum to zero in this contrast, and only epistatic effects remain. Thus, if the contrast is significantly greater than zero, there is evidence for epistatic effects. The value of the contrasts for each member of the population is squared and then summed, leading to a direct F-test for the null hypothesis of no epistatic effects. This test can be performed before attempting to estimate additive and dominance effects, rather than simultaneously with them, as is required when estimating variance components from mating designs. This is why the triple test cross is a much more powerful test for detecting epistatic effects, although this is still a *conservative* test for epistasis, because across the whole genome, positive and negative epistatic effects can sum to zero in a comparison of phenotypic means.

Using triple test cross designs, both Eta-Ndu and Openshaw (1999) and Wolf and Hallauer (1997) detected significant epistatic effects for grain yield in maize. Similarly, Melchinger et al. (1986) detected significant epistatic effects for grain yield in maize using both generation means analysis and a modified diallel analysis, although epistatic variance was small relative to additive and dominance variances. Using more complex designs, but still based on mean comparisons or regression models using means, Stuber and Moll (1971), Moreno-Gonzalez and Dudley (1981), and Lamkey et al. (1995) reported significant epistasis for grain yield in maize. We face a conundrum: we can detect significant epistatic effects but not significant epistatic variance for grain yield in maize. A major reason for this surely is that effects (first-order statistics) are easier to estimate precisely than variances (second-order statistics), but we are still unsure whether epistasis is important enough to warrant changes in breeding methods. Stuber et al. (1973) suggested that although epistatic effects were evident, their magnitude would not substantially hinder testcross prediction based on models ignoring epistasis. Conversely, Wolf, and Hallauer (1997) argued that favorable epistatic combinations unique to the B73 × Mo17 hybrid partially explained why it was so exceptional and widely grown. Perhaps epistatic effects create the differences between the very good and the outstanding genotypes when most additive effects are already fixed among elite lines.

B. Self-Pollinating Species

In autogamous species, similar results have been found: with experimental designs of weak power to detect epistasis, epistasis was not found, and where more powerful experimental designs were used, epistasis was generally found to be significant. Hanson and Weber (1961) and Brim and Cockerham (1961) employed nested mating designs and covariances of relatives to estimate genetic components of variance in soybean (*Glycine max* L.) populations, and found "only a suggestion that epistatic variability could be important for yield" (Hanson and Weber 1961) and "the additive X additive component is small relative to the additive component of genetic variance" (Brim and Cockerham 1961). Again, the difficulties in estimating epistatic variance components and the lack of a simple relationship between epistatic gene action effects and epistatic variance may have been important reasons for the inability to detect large epistatic variances.

Different results generally have been obtained with the use of experimental designs that have better power to detect epistatic effects. Diallel or NC Design II mating schemes can be used to estimate general and specific combining abilities (GCA and SCA) with a modification of the maize design (Hallauer and Miranda 1988, Chap. 4) that, instead of testing F_1 progenies of the crosses, highly inbred bulks or random sets of lines derived from several generations of selfing from each cross are tested. This modification is necessary for practical purposes, because producing sufficient F_1 seed for field evaluations is difficult in most autogamous species, whereas producing larger quantities of seed of inbred progeny by selfing the hybrid progeny for several generations is relatively easy. This modification leads to a good test for epistasis because the dominance effects that contribute to SCA in F_1 hybrids have been "removed" by selfing to near homozygosity. The GCA component estimated from inbred progenies reflects additive effects, while the SCA component is composed primarily of additive-by-additive epistatic effects. Significant SCA effects, therefore, indicate the presence of epistasis in these types of experiments. While more powerful than variance component-based tests for epistasis, these are still conservative tests because positive and negative epistatic effects may cancel each other out when summed across the whole genome.

Epistasis for yield has been detected in numerous autogamous species using these types of designs, including bread wheat (*Triticum aestivum*) (Busch et al. 1974; Cregan and Busch 1978), soybean (Hanson et al. 1967), oat (Pixley and Frey 1991; Stuthman and Stucker 1975), and rice

(*Oryza sativa*) (Gravois 1994). Upadhayaya and Nigam (1998) used the triple test cross in peanut (*Arachis hypogaea*) and Humphrey et al. (1969) used generation means analysis in tobacco (*Nicotiana tabacum*) and detected significant epistasis. In general, these authors recognized an important implication of epistasis for breeding methodologies: epistasis reduces the correlation between early and later selfing-generation yields, meaning that early generation testing and selection is expected to be less effective than delaying selection until later generations, when most of the additive-by-additive effects have been fixed within lines. This is one of the justifications for the single seed descent breeding method commonly used in soybean and oat.

C. Why Is There More Evidence for Epistasis in Selfing than in Outcrossing Species?

Although epistasis is not always found to be important for yield in autogamous species (e.g., Bitzer et al. 1982), epistasis seems to have been observed more commonly in autogamous species than in maize. Reasons for this may be either better experimental designs to detect epistasis in selfing species or the proposed greater importance of nonallelic interactions in autogamous (particularly disomic polyploid) species. Better experimental designs have contributed to the detection of epistasis in selfing species by avoiding the difficulty of having to estimate dominance as well as epistatic effects. With the dominance form of genic interaction "out of the way," any deviations from predictions based on the additive model can be attributed to epistasis. The difficulty in detecting epistatic variances simultaneously with additive and dominance variances in maize is reflected in the more frequent detection of epistasis in tests based on mean comparisons than on variance components. Thus, epistasis may be easier to detect in selfing than in outcrossing species but of similar importance in both groups of species.

An alternative explanation is that since additive-by-additive forms of epistasis, unlike dominance interactions, are a form of genic interaction that can be "fixed" and exploited in homozygous genotypes, selfing species will tend to exhibit strong epistatic interactions, while outcrossing species will exhibit strong dominance interactions. Mac Key (1970) presented a thorough review of the literature and developed a general theory of genic interactions that explains the seemingly disparate breeding and genetic behavior of both disomic and polysomic polyploids. Mac Key (1970) recognized that "polysomic polyploidy has been found only in connection with allogamy," and suggested that

outcrossing served the purpose of maintaining heterozygosity and intralocus allelic interactions. Allogamous diploid species, such as maize, exhibit inbreeding depression and the phenomenon is exaggerated in the allogamous polyploids, such as alfalfa (*Medicago sativa*). Mac Key (1970) suggested that almost all disomic polyploids are autogamous species that do not exhibit significant inbreeding depression because they compensate for the loss of intralocus allelic interactions with interlocus non-allelic epistatic interactions between homoeologous loci. Ehlke and Hill (1988) demonstrated that if a tetrasomic polyploid were converted to a disomic polyploid, the higher-order intralocus (three- and four-allele) interactions of the tetrasomic form would be transformed into epistatic interactions in the disomic form. This provides autogamous disomic polyploids the ability to fix heterosis in inbred genotypes, for which there is some molecular evidence in wheat (Gomez et al. 1989). Bingham (1979) summarized the situation:

> Polyploidy appears dependant on heterozygosity! The largest group of polyploids, the allopolyploids (disomic polyploids) have fixed heterozygosity in the two or more divergent genomes they possess . . . The autopolyploids (polysomic polyploids) [ensure] their heterozygosity through cross-pollination . . . We can find no example in crop plants of a successful polysomic polyploid species which is self-pollinated.

This implies that epistasis is observed more commonly in autogamous species than in maize because epistatic effects really are more important in selfing species, not because they are hidden in maize as an artifact of experimentation and statistics. A corollary to this implication is that epistasis is likely less important in polysomic polyploid species such as alfalfa than in disomic autogamous polyploid species like wheat and oat. While Mac Key's (1970) theory proposes the greater importance of epistasis among homoeologous loci in disomic versus polysomic polyploids, it says nothing about differences in importance of epistasis among non-homoeologous loci in the different polyploid forms, however. The epistatic model in Fig. 2.1 is an example of homoeologous epistasis (interactions among duplicated genes), while those in Figures 2.2 and 2.3 illustrate non-homoeologous forms of epistasis. Holland et al. (1997) found no evidence for homoeologous epistasis for flowering time in oat, although homoeology is not well-defined in oat (Kianian et al. 1997). Among non-homoeologous loci, autopolyploids can exhibit more *forms* of epistasis in the same manner that they can have more forms of intralocus allelic interactions, such as trigenic and quadrigenic interactions, and higher-order epistasis would be promoted by outcrossing.

VI. EVIDENCE FOR EPISTASIS FROM PLANT EVOLUTION STUDIES

Allard (1988, 1996) summarized the results of numerous isozyme marker studies of the evolution of plant populations, including wild oat (*Avena* spp.), cultivated barley, and maize:

> The single most important genetic mechanism in all three species groups was the assembly of favorable epistatic combinations of alleles of different loci by means of recurring cycles of selection, intercrossing superior selects and inbreeding to near homozygosity leading to stable superior multilocus genotype adapted to specific habitats.

In support of this idea, Rieseberg et al. (1996) found that, following hybridizations of two sunflower species (*Helianthus* spp.) and backcrossing to one parent, a subset of linkage blocks from the donor parent was repeatedly fixed. Similar genome composition was observed in an ancient hybrid sunflower species. They suggested that this was the result of selection for combinations of alleles with favorable epistatic effects, rather than selection for independently acting alleles (which would have resulted in less consistent maintenance of specific combinations of donor alleles). Similar results were reported in interspecific backcross populations of cotton (Jiang et al. 2000).

VII. MOLECULAR MARKER INVESTIGATIONS OF EPISTASIS

Fasoulas and Allard (1962) were perhaps the first to use genetic markers to measure epistasis in plants. They developed four near-isogenic lines (NILs) of barley differing only for their genotypes at two unlinked loci. Each locus conferred discrete phenotypic effects, orange lemmas (*oo*) or smooth awns (*rr*), and the four NIL genotypes were *OORR*, *OOrr*, *ooRR*, and *oorr*. They crossed the NILs to develop a population containing all nine possible two-locus genotypes, and evaluated individual plants phenotypically for numerous traits, including yield components. This permitted a two-factor analysis of variance, as proposed by Fisher (1918), the factors being the two loci (representing the additive and dominant effects of each locus separately), and the interaction between the two loci (representing the epistatic interactions between the two loci). A key to this type of experiment is that it allows an orthogonal partitioning of the additive and epistatic effects; with adequate sampling

sizes, there is little confounding of or correlation between the effects, as occurs in the covariances of relatives in mating designs. Fasoulas and Allard (1962) observed that epistatic interactions were significant for all traits studied and the magnitude of epistatic effects was on average half of the additive effects. Russell and Eberhart (1970) used three morphological marker loci and a similar design to detect significant additive-by-additive and dominant-by-dominant forms of epistasis for grain yield in maize.

Molecular markers have made this type of analysis possible on a genome-wide basis, although the principle remains the same (Holland 1998). With genetic maps containing marker loci about every 20 cM throughout a genome, epistasis between pairs of quantitative trait loci (QTLs) that might reside anywhere in the genome can be detected. DNA markers have been used to document the existence, magnitude, and nature of epistatic interactions among QTLs in crop species (Li 1998).

Early efforts using molecular markers generally did not provide evidence for important epistatic interactions (Tanksley 1993). For example, Edwards et al. (1987) detected numerous QTLs for yield in maize, but found no strong evidence for interactions among them. Their study had very large samples from two populations and good power to detect epistasis, and was limited only by incomplete genome coverage (17 to 20 isozyme loci, with some chromosomes unmarked). Subsequent QTL-mapping studies for yield in maize have either ignored epistasis (Austin and Lee 1996; Veldboom and Lee 1994) or found epistatic interactions among QTLs to be of minor importance or unrepeatable over environments (Lubberstedt et al. 1997; Melchinger et al. 1998). Openshaw and Frascaroli (1997), in contrast, detected numerous epistatic interaction effects of similar magnitude to QTL main effects in a very large sample of maize progeny (976 F_5 testcrosses). Inclusion of epistatic effects along with the main effects in a multiple regression model increased the explanatory power of the model by only a small amount, however.

Cockerham and Zeng's (1996) reanalysis of Stuber et al.'s (1992) data from a NC Design III experiment involving a population derived from the cross of the two most prominent maize inbred lines, B73 and Mo17, and augmented by molecular marker data demonstrated strong evidence for epistasis for grain yield, however. Their analysis was unique in that it did not test for interactions between different marker loci; rather it exploited a unique feature of the Design III, the testcross of each progeny to both parent lines. This provided a test for epistasis based on contrasts between means of marker genotypes *at a single marker locus* testcrossed to different parents, similar to a triple test cross analysis, but on a locus-by-locus basis. What they tested for, and found in abun-

dance, were epistatic interactions between two or more QTLs linked to the same marker locus. They did not attempt to test for epistasis between different marker loci. In addition, important epistatic effects have been found in maize between QTLs affecting biochemical phenotypes (Damerval et al. 1994), floral morphology (Doebley et al. 1995), and insect resistance (Byrne et al. 1996). Interactions between QTLs affecting wood yield components were detected in the outcrossing tree species *Eucalyptus grandis* and *Pinus radiata* (Grattapaglia et al. 1996; Kao et al. 1999).

Studies in self-pollinated species have more regularly given evidence for important epistasis among yield QTLs. Yu et al. (1997) found significant epistasis for yield QTLs in rice and attributed heterosis in an elite rice hybrid to epistasis. Li et al. (1997) also reported major epistatic effects for rice yield QTLs. These results contrast with those of Xiao et al. (1996, 1995) who found no evidence for epistasis in their rice QTL-mapping experiments. Epistasis for yield was also detected in mapping experiments in barley (Thomas et al. 1995), soybean (Orf et al. 1999), and tomato (*Lycopersicon esculentum*) (Eshed and Zamir 1996; but see deVicente and Tanksley 1993). Epistasis among QTLs has been found more frequently for simpler traits, such as flowering time in oat (Holland et al. 1997), rice (Yano et al. 1997), and *Brassica* (Camargo and Osborn 1996); oil and protein contents and height in soybean (Lark et al. 1994, 1995); and disease resistance in rice (Pressoir et al. 1998). Eshed and Zamir (1996) suggested that the different conclusions about the importance of epistasis among QTL-mapping studies may be attributed in large part to the experimental materials used. Epistasis is often not found when mapping is performed in F_2 or recombinant inbred line populations, in which segregation occurs throughout the genome. Tanksley (1993) suggested that these types of mapping populations were not ideal for investigating QTL interactions, whereas near-isogenic lines (NILs), like those used by Fasoulas and Allard (1962), would be more useful for that purpose because they stabilize the genetic background such that the main effects and interactions of the regions of interest are not confounded with genetic effects in other regions of the genome. Doebley et al. (1995) reported a good example of this very phenomenon: epistatic effects between two QTLs affecting plant and floral morphology of maize were not detected in an F_2 population, but were found to be very important in a study using NILs.

Marker-assisted selection studies have sometimes unwittingly turned up evidence for epistasis. For example, the effect of introgression of chromosomal regions from wild into cultivated tomato genotypes was found to vary depending on the recurrent parent genotype (Tanksley and

Hewitt 1988). In some cases, a supposedly favorable allele had unfavorable effects in a specific genetic background (Tanksley and Hewitt 1988). Similar results were reported by Stuber (1994), who mapped QTLs affecting grain yield in maize, ignoring the possibility of epistasis (Stuber et al. 1992), and used marker-assisted selection to transfer favorable QTL alleles from donor parents into elite backgrounds. There seemed to be no advantage to introgressing more than two to four of six favorable donor chromosome segments, and there may be some disadvantage to it (Stuber 1998). The likely explanation is that there are unfavorable epistatic effects between some of the otherwise favorable donor QTL alleles or there are favorable epistatic combinations in the elite background that are disrupted by multiple introgressions (Stuber 1998).

Statistical methods for detecting epistasis in QTL experiments are improving. Maximum-likelihood estimation procedures for detecting and estimating QTL epistatic effects were developed by Wang et al. (1999) and Kao et al. (1999). Wang et al.'s (1999) method searches for epistatic interactions throughout the genome, while other methods allow only for testing for interactions between QTLs with significant main effects. The advantage of being able to test for epistasis among all possible pairs of genomic regions is that gene pairs with additive-by-additive interactions do not necessarily have additive effects, and can therefore be missed by mapping algorithms that test for epistasis only between QTLs with significant main effects. The drawback is greater computational complexity and greater difficulty in determining significance thresholds because of the numerous statistical tests conducted.

VIII. WHY IS THERE MORE EVIDENCE FOR EPISTASIS FROM QTL EXPERIMENTS THAN FROM BIOMETRICAL STUDIES?

QTL-mapping studies seem to provide more evidence for epistasis for yield and other important agronomic traits than classical biometric studies (Li 1998). With molecular markers, the gene action effects of specific chromosomal regions can be estimated, providing two great improvements over classical biometric methods of quantitative genetic analysis. First, estimation of effects in specific chromosomal regions provides considerable power over biometric methods that test for the average or total gene effects of the entire genome (in which case positive and negative interactions among different pairs of loci may sum to near zero). Second, QTL studies can estimate gene action effects (e.g., a and d) rather than

just statistical genetic effects such as average allelic and genotypic effects (e.g., α and δ) in populations of specific allele frequencies. A major limitation of quantitative genetics has been that statistical genetic effects are generally not reliable indicators of the underlying gene action; QTL-mapping methods provide a more direct route to understanding gene action.

A caveat should be added that, while genetic marker methods have improved our ability to resolve complex quantitative traits to their underlying genetic factors, biases and limitations are inherent to QTL-mapping experiments. A difficulty in interpreting QTL-mapping experiment results is that simultaneous detection and estimation of the effects of numerous genetic factors using typical population sizes (75 to 200 progenies or lines) results in multicolinearity, or correlations among the regressor variables, a problem commonly encountered in multiple regression model selection and estimation (Rawlings 1988). In the case of QTL-mapping experiments, linked marker loci (the regressor variables) are obviously correlated. The effects of QTLs linked within 20 cM of each other cannot be reliably separated in typical QTL-mapping experiments (Kearsey and Farquhar 1998). What is less obvious, but still very problematic for selecting those loci that best account for the observed phenotypic variation and estimating their effects, is that unlinked loci generally are also correlated to some extent simply by sampling a finite population. QTL effects estimated at one marker locus, therefore, may be partially confounded with the effects of QTLs on other chromosomes. The result is that QTLs with small effects tend to remain undetected, QTLs with intermediate effects are detected in some samples and not others, and the effects of those QTLs that are detected are overestimated (Beavis 1994; Openshaw and Frascaroli 1998; Utz et al. 2000). Independent progeny samples drawn from the same population may give quite different results (Melchinger et al. 1998). The problem is serious for estimating QTL main effects and becomes worse for estimates of epistasis among QTLs (Utz et al. 2000). Furthermore, making multiple tests for QTLs complicates the determination of a proper statistical threshold for declaring the presence of a QTL (Churchill and Doerge 1994), and this problem also becomes more serious when performing $m(m-1)/2$ tests for epistasis among all pairs of m marker loci in the genome (Holland 1998; Holland et al. 1997). Real certainty in the existence and magnitude of epistatic interactions among QTLs will require isolation of the different allelic combinations into a homogeneous background using NILs and similar approaches (Doebley et al. 1995), where these statistical difficulties can be minimized.

IX. IMPLICATIONS OF EPISTASIS FOR PLANT BREEDING

Does it matter to plant breeders whether epistasis is an important component of gene action for yield? Will the presence of significant epistasis affect breeding methodology? The summary answer, to be defended below, is that *if* epistasis is important, then, current plant breeding methods and paradigms are adequate in the short-term, but will almost surely limit long-term progress.

What are the implications for breeding methodology if epistasis is of general importance? Epistasis is treated only in passing in some standard plant breeding textbooks (Fehr 1987; but see Allard 1999, Chap. 11), and deciding among alternative breeding strategies if epistasis is important is difficult. Baker (1984) suggested that epistasis most likely will be of importance in self-pollinating species because of the larger coefficient of additive-by-additive epistatic variance of covariance of relatives in inbred generations, but he doubted that epistasis would have much impact on breeding cross-pollinated crops. Thus, the implications of epistasis for plant breeding methodology will depend on the breeding system of the crop and the current methods for breeding and seed production because these determine the types of epistatic effects that can be propagated reliably. Specifically, breeding methods for self-pollinating crops, cross-pollinated hybrid crops, and cross-pollinated non-hybrid crops will be affected in different ways. Selfing species allow fixation of additive-by-additive epistatic effects in cultivars, while all types of epistatic effects can be fixed in F_1 hybrid cultivars. The frequencies of various epistatic effects in cultivars that are derived from multiple generations of cross-pollination (approximating random mating), in contrast, may fluctuate over generations.

Below I discuss the influence of epistasis on both inbreeding depression and heterosis and illustrate three effects of epistasis on response to selection: (1) epistatic variance can shift to additive variance under drift or inbreeding; (2) epistatic variance contributes to "temporary" response to selection in outcrossing populations which can be captured as a form of heterosis using appropriate breeding procedures but may be otherwise squandered; and (3) fitness or yield is not a simple function of allele frequencies, resulting in rugged adaptive landscapes filled with local fitness optima on which breeding populations can become stranded. If epistasis is important, then marker-assisted selection schemes should be designed to exploit it.

A. Inbreeding Depression and Heterosis

Inbreeding depression is the difference between the mean of a population inbred to a level measured by the inbreeding coefficient, F, and the random-mated population from which it was derived. Midparent heterosis is the difference between the mean of an F_1 population derived from the cross of two populations and the average of the means of the two parental random-mated populations.

1. Inbreeding Depression. The mean of a population inbred to a degree F, in the absence of gametic disequilibrium, and allowing for two-locus interactions (Kempthorne 1957, Chap. 20; Weir and Cockerham 1977) is:

$$\mu_F = \mu_0 + F\sum_{l=1}^{n}\sum_{i} p_i\delta_{ii}^{l} + F^2\sum_{l=1}^{n-1}\sum_{l'=l+1}^{n}\sum_{i}\sum_{k} p_i q_k \delta\delta_{iikk}^{ll'}$$

where μ_0 is the mean of the random-mated population, n is the number of loci, and the summations are over loci and alleles for the dominance term and over locus pairs and allele pairs for the dominance-by-dominance term. Thus, only dominance and dominance-by-dominance statistical effects contribute to inbreeding depression. In the case of the two-locus, two-allele model, this relates to the gene action parameters as follows:

$$\begin{aligned}\mu_F &= \mu_0 - 2F[(p_1p_2)(d^A + (q_1 - q_2)da + 2q_1q_2dd) + (q_1q_2)(d^B + (p_1 - p_2)ad \\ &\quad + 2p_1p_2dd)] + 4F^2(p_1q_1p_2q_2)dd \\ &= \mu_0 - 2F[(p_1p_2)(d^A + (q_1 - q_2)da) + (q_1q_2)(d^B + (p_1 - p_2)ad)] \\ &\quad + 4F(F-2)(p_1q_1p_2q_2)dd\end{aligned}$$

If allele frequencies are equal ($p_1 = p_2 = q_1 = q_2 = 0.5$), then ad and da gene action effects do not contribute to inbreeding depression and only dominance and dominance-by-dominance epistatic gene action effects contribute. At other allele frequencies, however, ad and da effects can affect inbreeding depression.

Epistatic effects on inbreeding depression are sometimes considered to be deviations from a multiplicative interaction rather than additive model, as discussed in section II of this review. Fu and Ritland (1996) presented evidence that inbreeding depression for fecundity conforms to a model of multiplicative interactions, whereas inbreeding depression for viability deviates from a multiplicative model in *Mimulus guttatus*.

Remington and O'Malley (2000) also found that inbreeding depression for viability in loblolly pine (*Pinus taeda*) was best explained by a multiplicative model.

2. Heterosis. The treatment of heterosis is more complicated because the statistical genetic effects in different populations cannot be related simply to each other due to the differing allele frequencies (Stuber and Cockerham 1966). To demonstrate heterosis in terms of gene action parameters for the special case where the two parental populations have the same two alleles at both loci, but may have them in different frequencies, the allele frequencies at the A locus in one population are denoted p_1 and p_2, and the allele frequencies for the A locus are $p_1 - y_A$ and $p_2 + y_A$, where y_A equals the difference in the frequency of allele 1 between the two populations. Similarly, the allele frequencies at the B locus are written as q_1 and q_2 in one population and $q_1 - y_B$ and $q_2 + y_B$ in the second population. Assuming that the parental populations are in Hardy-Weinberg and gametic phase equilibria, the midparent mean (the average of the two parental population means) can be written as:

$$\begin{aligned}\mu_{MP} = {} & m + (p_1 - p_2 - y_A)a^A + (q_1 - q_2 - y_B)a^B \\ & + [2p_1p_2 + (p_1 - p_2 - y_A)y_A]d^A + [2q_1q_2 + (q_1 - q_2 - y_B)y_B]d^B \\ & + [(p_1 - p_2)(q_1 - q_2) - (q_1 - q_2)y_A - (p_1 - p_2)y_B + 2y_Ay_B]aa \\ & + [2(p_1 - p_2)q_1q_2 - 2q_1q_2y_A + (p_1 - p_2)(q_1 - q_2)y_B - 2(q_1 - q_2)y_Ay_B \\ & + 2y_Ay_B^2 - (p_1 - p_2)y_B^2]ad + [2p_1p_2(q_1 - q_2) + (p_1 - p_2)(q_1 - q_2)y_A \\ & - 2p_1p_2y_B - 2(p_1 - p_2)y_Ay_B + 2y_A^2y_B - (q_1 - q_2)y_A^2]da + [4p_1p_2q_1q_2 \\ & + 2(p_1 - p_2)q_1q_2y_A + 2p_1p_2(q_1 - q_2)y_B + 2(p_1 - p_2)(q_1 - q_2)y_Ay_B \\ & - 2q_1q_2y_A^2 - 2p_1p_2y_B^2 - 2(q_1 - q_2)y_A^2y_B - 2(p_1 - p_2)y_Ay_B^2 \\ & + 2\,y_A^2y_B^2]dd\end{aligned}$$

The mean of the F_1 population derived from the inter-population cross (which is in neither Hardy-Weinberg nor gametic phase equilibria) is:

$$\begin{aligned}\mu_{F1} = {} & m + (p_1 - p_2 - y_A)a^A + (q_1 - q_2 - y_B)a^B + [2p_1p_2 + (p_1 - p_2)y_A]d^A \\ & + [2q_1q_2 + (q_1 - q_2)y_B]d^B + [(p_1 - p_2)(q_1 - q_2) - (q_1 - q_2)y_A - (p_1 \\ & - p_2)y_B + y_Ay_B]aa + [2(p_1 - p_2)q_1q_2 - 2q_1q_2y_A + (p_1 - p_2)(q_1 - q_2)y_B \\ & - (q_1 - q_2)y_Ay_B]ad + [2p_1p_2(q_1 - q_2) + (p_1 - p_2)(q_1 - q_2)y_A - 2p_1p_2y_B \\ & - (p_1 - p_2)y_Ay_B]da + [4p_1p_2q_1q_2 + 2(p_1 - p_2)q_1q_2y_A + 2p_1p_2(q_1 \\ & - q_2)y_B + (p_1 - p_2)(q_1 - q_2)y_Ay_B]dd\end{aligned}$$

Mid-parent heterosis is the difference between the F_1 and mid-parent means, and in this case equals:

$$\text{Heterosis} = \mu_{F1} - \mu_{MP} = y_A^2 d^A + y_B^2 d^B - y_A y_B aa + [(q_1 - q_2)y_A y_B - 2y_A y_B^2 + (p_1 - p_2)y_B^2]ad + [(p_1 - p_2)y_A y_B - 2y_A^2 y_B + (q_1 - q_2)y_A^2]da + [-(p_1 - p_2)(q_1 - q_2)y_A y_B + 2q_1 q_2 y_A^2 + 2p_1 p_2 y_B^2 + 2(q_1 - q_2)y_A^2 y_B + 2(p_1 - p_2)y_A y_B^2 - 2y_A^2 y_B^2]dd$$

Comparing this result to the result for inbreeding depression shows that although heterosis can be considered "simply inbreeding depression in reverse" (Falconer and Mackay 1996) if there is no epistasis, this is not strictly true when there is epistasis.

By defining the effects in reference to the equilibrium population derived from the F_1 by sufficient random mating, heterosis can also be expressed in terms of statistical genetic effects as:

$$\text{Heterosis} = \mu_{F1} - \mu_{MP} = y_A^2 D^A + y_B^2 D^B - y_A y_B AA$$

where $D^A = \delta_{12}^A - (1/2)(\delta_{11}^A + \delta_{22}^A)$, $D^B = \delta_{12}^B - (1/2)(\delta_{11}^B + \delta_{22}^B)$, and $AA = \alpha\alpha_{11} - \alpha\alpha_{12} - \alpha\alpha_{21} + \alpha\alpha_{22}$ (Willham and Pollak 1985). Similar expressions, with varying notation, were given by Hill (1982) and Lynch (1991). The interpretation of these statistical genetic expressions is that heterosis depends on dominance of favorable alleles isolated in the parental populations or by favorable additive-by-additive interactions between alleles in different parental populations, or both (Lynch 1991). The purpose of expressing heterosis in terms of gene action in this review is to emphasize that although heterosis is a function of dominance and additive-by-additive epistasis *statistical* effects (Hill 1982; Lynch 1991; Willham and Pollak 1985), it is also a function of dominance and additive-by-additive, additive-by-dominant, and dominant-by-dominant *gene action* effects. This is true because the statistical dominance and additive-by-additive effects incorporate additive-by-dominant and dominant-by-dominant gene action effects (Table 2.6). The practical result of this is that hybrid cultivars and population-cross cultivars can exploit all forms of epistasis because they contribute to heterosis.

3. Multiplicative Epistasis and Heterosis. Multiplicative gene action is a special case of the general gene action formulae given in Tables 2.2 and 2.3, wherein the epistatic gene action effects are the products of the single-locus gene action effects, scaled to the midparent value:

$$aa = (a^A \times a^B)/m;\ ad = (a^A \times d^B)/m;\ da = (d^A \times a^B)/m;\ dd = (d^A \times d^B)/m.$$

As mentioned in section IV D, one plausible manner in which multiplicative gene action can occur is when a complex character is the

Table 2.8. Example of multiplicative gene action for a complex trait, W, which is the product of traits X and Y, controlled independently by loci *A* and *B*, respectively.

Two-locus genotypic values									
	B-locus genotype								
A-locus genotype	B_1B_1			B_1B_2			B_2B_2		
	X	W	Y	X	W	Y	X	W	Y
A_1A_1	5	15	3	5	15	3	5	5	1
A_1A_2	4	12	3	4	12	3	4	4	1
A_2A_2	1	3	3	1	3	3	1	1	1

Gene action parameters	Trait		
	X	W	Y
m	3	6	2
a^A	2	3	0
a^B	0	4	1
d^A	1	3	0
d^B	0	2	1
aa	0	2	0
ad	0	1	0
da	0	2	0
dd	0	1	0

product of two or more subcomponents; e.g., grain yield is the product of seed number per plant and mean seed weight.

A numerical example of this type of multiplicative gene action is shown in Table 2.8, where a trait, W, is the product of traits X and Y. The genes affecting X and Y operate independently; the *A* locus affects trait X but not Y, while the *B* locus affects trait Y but not X. The two loci exhibit no epistasis for either traits X or Y, but they exhibit multiplicative epistasis for trait W due to the multiplication of components X and Y (Table 2.8). In this situation, heterosis in character W can arise from two sources (see review by Schnell and Cockerham 1992). One source is the multiplication of heterosis observed for character X (due to dominance at locus A) with heterosis observed for character Y (due to dominance at locus B). The second source of heterosis is from multiplicative interactions of the subcomponents directly. This can be expressed as

$$H_W = \mu_{F1(W)} - \mu_{MP(W)} = (\mu_{MP(X)}\mu_{MP(Y)})(h_X h_Y - 1) - (\mu_{P1(X)} - \mu_{P2(X)})(\mu_{P1(Y)} - \mu_{P2(Y)})/4$$

where the first part of the right hand side of the equation is due to the product of heterosis (measured as the ratio of the F_1 mean to the midparent mean) expressed for traits X and Y individually ($h_X = \mu_{F1(X)}/\mu_{MP(X)}$ and $h_Y = \mu_{F1(Y)}/\mu_{MP(Y)}$, respectively), and the second part of the right hand side is due to multiplicative interactions of the traits (Schnell and Cockerham 1992).

As an example of heterosis in a complex trait, assume that there are two populations sharing the same two alleles and the same gene action at both loci as in the example in Table 2.8, but differing for allele frequencies at the two loci. If the first parental population has allele frequencies of $p_1 = 0.2$ and $q_1 = 0.8$, its mean for the sub-component and complex traits will be: $\mu_{P1(X)} = 2.12$; $\mu_{P1(Y)} = 2.92$; and $\mu_{P1(W)} = 6.19$. If the second parental population has allele frequencies of $p_1 = 0.8$ and $q_1 = 0.2$, its mean for the sub-component and complex traits will be: $\mu_{P2(X)} = 4.52$; $\mu_{P2(Y)} = 1.72$; and $\mu_{P2(W)} = 7.77$. Therefore, the midparent trait means are: $\mu_{MP(X)} = 3.32$; $\mu_{MP(Y)} = 2.32$; and $\mu_{MP(W)} = 6.98$. The trait means of the interpopulation cross F_1 will be: $\mu_{F1(X)} = 3.68$; $\mu_{F1(Y)} = 2.68$; and $\mu_{F1(W)} = 9.86$; and the resulting heterosis for the three traits will be: $H_X = 0.36$; $H_Y = 0.36$; and $H_W = 2.88$. Heterosis measured as the ratio of F_1 and of the midparent means for the three traits is: $h_X = 1.108$; $h_Y = 1.155$; and $h_W = 1.413$.

Following Schnell and Cockerham's (1992) formulation, part of the heterosis in W ($H_W = 2.88$) is due to the multiplication of heterosis within components X and Y: $(\mu_{MP(X)}\mu_{MP(Y)})(h_Xh_Y - 1) = (7.7)(1.28 - 1) = 2.16$. In genetical terms, this fraction arises from the multiplication of dominance statistical effects within loci (Schnell and Cockerham 1992). The remaining portion of heterosis is due to the multiplicative effects: $-(\mu_{P1(X)} - \mu_{P2(X)})(\mu_{P1(Y)} - \mu_{P2(Y)})/4 = -(-2.4)(1.2)/4 = 0.72$. In genetical terms, this portion arises from the additive-by-additive statistical epistatic interactions between loci (Schnell and Cockerham 1992).

Schnell and Cockerham (1992) suggested that, with increasing number of pairs of loci involved in epistatic interactions, the first part of heterosis (due to dominance statistical effects) would be the more substantial component of heterosis in the complex trait, and the second part (due to additive-by-additive statistical effects) the lesser. The reason for this is that heterosis at each additional locus involved in the multiplicative interaction is multiplied together, so with more loci affecting heterosis, the first portion is the product of more component heteroses, and so will become larger. In contrast, with increasing numbers of multiplicatively-acting loci affecting the trait, the additive-by-additive statistical effects will tend to contribute less to heterosis because the allele pairs contributing to a positive multiplicative interaction must

be dispersed in the parental populations if they are to contribute positively to heterosis. For example, if the A_1B_1 and A_2B_2 allele pairs are more favorable, the greatest heterotic effect from this epistatic combination will occur when one parent population has a higher frequency of A_1 and a lower frequency of B_1 than the other parent. As more loci contribute to the multiplicative interaction, all of the favorable allele pairs cannot be dispersed between the parental populations. For example, if three loci are involved and the favorable allele pairs are A_1B_1 and A_2B_2, A_1C_1 and A_2C_2, and B_1C_1 and B_2C_2, and A_1 and B_1 are dispersed and A_1 and C_1 are dispersed, then B_1 and C_1 cannot be dispersed simultaneously. B_1 and C_1 would both have to be at higher frequencies in one of the parental populations than in the other, and their epistatic effect would contribute negatively to heterosis. This reasoning also applies to the contributions of additive-by-additive effects to heterosis in the more general case. With many loci contributing to heterosis, the additive-by-additive effects of locus pairs contributing to heterosis will tend to be positive as often as negative, and in net may contribute little to heterosis.

The multiplicative model illustrates how epistatic interactions can occur between two or more sub-components of a complex phenotypic trait in the absence of molecular interactions. How realistic is the multiplicative model? Melchinger et al. (1994) reviewed the literature and found several reports of heterosis arising from multiplicative interactions in crop plants. In most cases in which appropriate data were available to make the determination, heterosis was due primarily to multiplication of sub-component heterosis, rather than to multiplicative epistasis, in accordance with the prediction of Schnell and Cockerham (1992). Melchinger et al.'s (1994) own data on heterosis in crosses between large- and small-seeded *Vicia faba* cultivars, however, demonstrated that multiplicative epistasis can make an important contribution to heterosis. They suggested that the parental cultivars represented different types of cultivars from distinct germplasm pools that were selected for different yield components. A limitation to significant multiplicative epistasis occurring generally for yield is the likely existence of negative genetic correlations among the yield sub-components. For multiplicative interactions to occur, the sub-components must be uncorrelated (Melchinger et al. 1986), but in reality, many pairs of sub-components will compete for a common pool of energy and nutrient resources, resulting in their being negatively correlated in general.

Finally, while there is reason to believe that epistasis generally may not contribute greatly to heterosis, relative to dominance, data on the subject are limited. Lynch (1991) interpreted the results of Moll et al.'s (1965) classic study of the relationship between genetic divergence and

heterosis in maize as evidence that additive-by-additive effects were important positive contributors to heterosis. As mentioned previously, for this to occur requires that favorable allele pairs be dispersed among populations, contrary to the general expectation that coadapted gene complexes are more likely to exist within populations than between populations (Lynch 1991).

B. Epistatic Variance Can Be Transformed into Additive Variance after Bottlenecks

A surprising result of theoretical studies that have investigated the effect of finite population sizes on genetic variances is that, following a population bottleneck, the genetic variance within a sub-population or breeding line may increase, rather than decrease, as expected under a purely additive model (Falconer and Mackay 1996), and as taught more or less as dogma in plant breeding courses (Fehr 1987). If epistasis is important and allele frequencies are initially intermediate, population bottlenecks can result in increased additive genetic variance (Cheverud and Routman 1996; Goodnight 1988; Whitlock et al. 1993). This is counter-intuitive, but consider the simple model of fitness resulting from the epistatic pair of loci described in Figs. 2.4 and 2.5. When the population has intermediate allele frequencies, it is on the saddlepoint in the middle of Fig. 2.5. In this case, additive genetic variance is zero, because within each locus, neither allele is more favorable than the other. Fig. 2.5 assumes a Hardy-Weinberg population, implying a very large population size and stable allele frequencies without selection. If population size is restricted, random drift takes effect, and if, for example, allele A_1 drifts by random chance to less than 50% frequency, while allele B_1 by chance remains close to 50%, then on average the B_1 allele will be most often associated with the A_2 allele, resulting in an unfavorable epistatic combination (Fig. 2.5). The B_2 allele will be commonly associated with the A_2 allele, resulting in a favorable genotype. Now the alleles will have average effects different from zero, causing additive genetic variance. The population will respond to selection: A_2B_2 genotypes will be the most commonly selected, and A_2 and B_2 allele frequencies will increase.

Cheverud and Routman (1996) consider epistasis as a source of increased additive variance following bottlenecks, but another perspective is that epistasis can suppress additive genetic variance in large, random-mating populations with intermediate allele frequencies. Indeed, modifier genes epistatic to other genes can act as "capacitors" for genetic variance, masking the potential genetic variance until allele

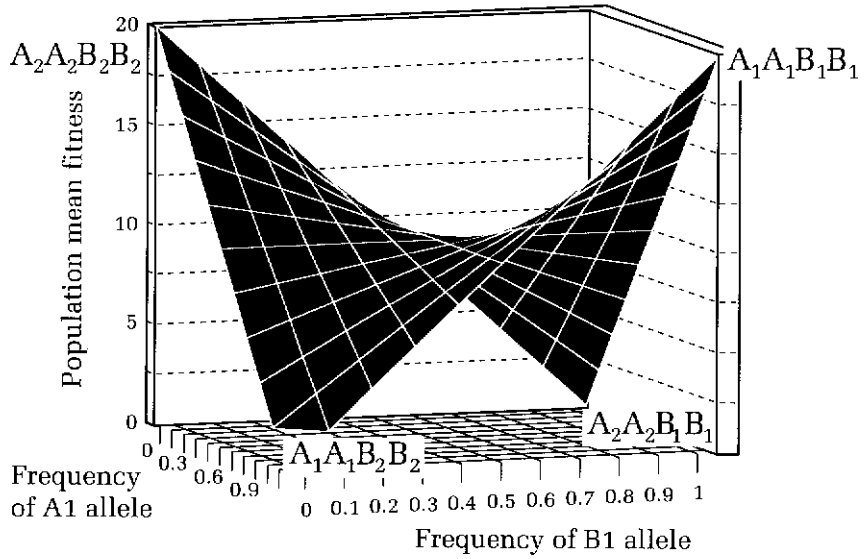


Fig. 2.5. Graphical representation of a 3-dimensional adaptive landscape resulting from the additive-by-additive epistasis shown in Fig. 2.4. Mean population fitness is a function of allele frequencies at loci A and B and is maximum when both A1 and B1 alleles are fixed or when both A2 and B2 alleles are fixed. These two points are adaptive peaks on the fitness landscape. For a population to shift from near one peak to near the other requires an intermediate reduction in population mean fitness.

frequencies at the modifiers shift to allow expression of differences at other loci (Rutherford and Lindquist 1998). This is simply the reverse of the phenomenon described by Moreno (1994) and discussed in section IV B, in which epistatically-acting mutants at low frequency can act additively, but when shifted to higher frequency act more obviously in an epistatic fashion. López-Fanjul et al. (1999) argued that increases in additive variance following bottlenecks were limited to situations in which allele frequencies were intermediate at both loci or at extreme frequencies at one or both loci before the bottleneck and that such conditions are unrealistic in natural populations. Nevertheless, they may be common in plant breeding populations developed by mating unrelated genotypes.

Deliberate inbreeding by sib-mating or selfing is expected to result in increased variation among lines and decreased variation within lines, and this is exploited by plant breeders who wish to enhance gain from among-line selection by maximizing differences among lines and min-

imize differences within lines (Fehr 1987). If epistasis is strong enough, however, there can be more heritable variance within lines than expected. Unfortunately, experiments to empirically determine that epistasis causes increases in genetic variances under inbreeding must be exceedingly rigorous because genetic variances within some subpopulations can increase by chance even under a completely additive model, so large numbers of subpopulations and individuals within subpopulations need to be sampled to have adequate statistical power to reject the additive model (Lynch 1988). Furthermore, dominance effects can also cause increases in genetic variances under drift and inbreeding (Robertson 1951; Weir and Cockerham 1977), so the experimental design must be able to discriminate the effects of epistasis from those of dominance. As a result, there is a paucity of good data on the subject. Recent reports based on extremely large-scale and rigorous experiments have provided differing conclusions. Whitlock and Fowler (1999) found that additive variance within *Drosophila* subpopulations decreased as expected under a completely additive model, while Cheverud et al. (1999) demonstrated that an increase in genetic variance under inbreeding was caused by epistasis in mice. Rasmusson and Phillips (1997) attributed part of the surprisingly high gains from selection within barley populations derived from closely related lines to enhanced genetic variance from epistasis, but this is only anecdotal evidence.

If epistasis really can cause substantial changes in additive genetic variance under inbreeding, predictions of response to selection practiced among partially inbred lines based on genetic variance components estimated from outbred populations may be quite wrong. Within-line selection will result in permanent response from selection due not only to additive genetic variance, but also to additive-by-additive epistatic variance (Cockerham and Tachida 1988), in contrast to the outbred situation, in which epistatic variance only contributes to temporary response to selection (see section IX C). The permanency of this response, however, depends upon not intermating different lines after selection.

Wricke and Weber (1986) concluded that the presence of epistatic variance has little effect on the optimal ratio of among- to within-line selection intensity, but their conclusion did not account for the possibility of additive variance increasing within lines due to drift effects. Perhaps more attention should be given to within-line selection than is commonly given in pedigree breeding programs for self-pollinated and hybrid crops. Similarly, perhaps this is another reason to employ some form of inbreeding in cross-pollinated population improvement schemes in addition to the enhanced ability to purge deleterious recessive alleles under inbreeding.

Gathering the empirical results necessary to choose appropriate levels of among- and within-line selection under inbreeding for traits affected by epistasis will be difficult. Furthermore, direct derivation of predicted responses to selection for a variety of complex gene action models and breeding method combinations would be difficult and tedious. A more fruitful approach in the near term may be computer simulations of heritable variances and selection responses covering a wide range of variables, including different forms and magnitudes of epistasis and different selection schemes. Podlich and Cooper (1998) recently developed software to implement such simulations. Their program allows the user to specify genetic models, including any form of epistasis among any number of genes, heritabilities, and genotype-by-environment interaction effects. This may permit robust theoretical investigations of the effects of epistasis on genetic variances and responses to selection under inbreeding.

C. Temporary Response to Selection

A general formula for response to selection is based on the regression of offspring values on parental values (Falconer and Mackay 1996). The numerator of the response, therefore, is based on the covariance between parent and offspring phenotypes. A simple way to determine the covariance between parent and offspring is to write the statistical genetic models for parent and offspring, and to compute the cross-product of the parent and offspring models. For example, a two-locus diploid statistical genetic model for a parent, X, is:

$$G_{(X)ijkl} = \mu + \alpha_i^A + \alpha_j^A + \delta_{ij}^A + \alpha_k^B + \alpha_l^B + \delta_{kl}^B + \alpha\alpha_{ik} + \alpha\alpha_{il} + \alpha\alpha_{jk} + \alpha\alpha_{jl} + \alpha\delta_{ikl} + \alpha\delta_{jkl} + \alpha\delta_{ijk} + \alpha\delta_{ijl} + \delta\delta_{ijkl},$$

where terms are as defined in Table 2.6. We assume that the population is in Hardy-Weinberg and gametic phase equilibria.

The total genetic variance is obtained as the expectation of the squared difference between the value for G_X and the population mean:

$$\sigma_G^2 = E[(G_X - \mu)^2] = E[G_X^2] - \mu^2 \text{ (Lynch and Walsh 1997, Chap. 2).}$$

Inserting the value for $G_{(X)ijkl}$ into this equation gives:

$$\sigma_G^2 = E[(\mu + \alpha_i^A + \alpha_j^A + \delta_{ij}^A + \alpha_k^B + \alpha_l^B + \delta_{kl}^B + \alpha\alpha_{ik} + \alpha\alpha_{il} + \alpha\alpha_{jk} + \alpha\alpha_{jl} + \alpha\delta_{ikl} + \alpha\delta_{jkl} + \alpha\delta_{ijk} + \alpha\delta_{ijl} + \delta\delta_{ijkl})^2] - \mu^2.$$

The assumptions of Hardy-Weinberg and gametic phase equilibria imply that all of the genetic terms in the above equation are uncorrelated, simplifying the equation to:

$$\sigma_G^2 = E[\mu^2] + E[(\alpha_i^A)^2] + E[(\alpha_j^A)^2] + E[(\delta_{ij}^A)^2] + E[(\alpha_k^B)^2] + E[(\alpha_l^B)^2 + E[(\delta_{kl}^B)^2] + E[(\alpha\alpha_{ik})^2] + E[(\alpha\alpha_{il})^2] + E[(\alpha\alpha_{jk})^2] + E[(\alpha\alpha_{jl})^2] + E[(\alpha\delta_{ikl})^2] + E[(\alpha\delta_{jkl})^2] + E[(\alpha\delta_{ijk})^2] + E[(\alpha\delta_{ijl})^2] + E[(\delta\delta_{ijkl})^2] - \mu^2.$$

We then define the genetic components of variance to be:

$$\begin{aligned}
\sigma_A^2 &= E[(\alpha_i^A)^2] + E[(\alpha_j^A)^2] + E[(\alpha_k^B)^2] + E[(\alpha_l^B)^2] \\
\sigma_D^2 &= E[(\delta_{ij}^A)^2] + E[(\delta_{kl}^B)^2] \\
\sigma_{AA}^2 &= E[(\alpha\alpha_{ik})^2] + E[(\alpha\alpha_{il})^2] + E[(\alpha\alpha_{jk})^2] + E[(\alpha\alpha_{jl})^2] \\
\alpha_{AD}^2 &= E[(\alpha\delta_{ikl})^2] + E[(\alpha\delta_{jkl})^2] + E[(\alpha\delta_{ijk})^2] + E[(\alpha\delta_{ijl})^2] \\
\sigma_{DD}^2 &= E[(\delta\delta_{ijkl})^2],
\end{aligned}$$

so that:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2.$$

The phenotypic value of parent X includes, in addition to the genotypic value of X, a classifiable macro-environmental effect, an unclassifiable micro-environmental effect, effects due to interactions between the genotype and these two environmental effects, plus a measurement error effect (Nyquist 1991). We will always assume no correlation between the non-genetic effects of parents and offspring, however, so only the genetic value, G_X, needs to be considered in the parent-offspring covariance.

The genetic value of an offspring, Y, the progeny of a mating between parent X and some other random member of the population, Z, depends upon the alleles inherited from both parents. Let us assume that Y receives alleles *i* and *k* from parent X and *i'* and *K'* from parent Z (we will consider the possibility that an offspring of X inherits alleles other than *i* and *k* from parent X later). The genetic value of $Y_{ii'kk'}$ is then:

$$G_Y = \mu + \alpha_i^A + \alpha_{i'}^A + \delta_{ii'}^A + \alpha_k^B + \alpha_{k'}^B + \delta_{kk'}^B + \alpha\alpha_{ik} + \alpha\alpha_{ik'} + \alpha\alpha_{i'k} + \alpha\alpha_{i'k'} + \alpha\delta_{ikk'} + \alpha\delta_{i'kk'} + \alpha\delta_{ii'k} + \alpha\delta_{ii'k'} + \delta\delta_{ii'kk'}.$$

The covariance of G_X and G_Y includes the expectations of cross products involving the same terms (which are variances): in this case only

α_i^A, α_k^B, and $\alpha\alpha_{ik}$ are in common between X and Y, so the following quadratic terms contribute to the covariance of parent and offspring:

$$\mathrm{Cov}(G_X, G_Y) = E[(\alpha_i^A)^2 + (\alpha_k^B)^2 + (\alpha\alpha_{ik})^2] = E[(\alpha_i^A)^2 + (\alpha_k^B)^2] + E[(\alpha\alpha_{ik})^2]$$

Additive genetic variance was defined as $\sigma_A^2 = E[(\alpha_i^A)^2] + E[(\alpha_j^A)^2] + E[(\alpha_k^B)^2] + E[(\alpha_l^B)^2]$. $E[(\alpha_i^A)^2]$ equals $E[(\alpha_j^A)^2]$ and $E[(\alpha_k^B)^2]$ equals $E[(\alpha_l^B)^2]$, so $E[(\alpha_i^A)^2 + (\alpha_k^B)^2] = (1/2)\sigma^2{}_A$. Additive-by-additive genetic variance was defined as $\sigma_{AA}^2 = E[(\alpha\alpha_{ik})^2] + E[(\alpha\alpha_{il})^2] + E[(\alpha\alpha_{jk})^2] + E[(\alpha\alpha_{jl})^2]$. The expectations of all squared $\alpha\alpha$ terms are the same, so $E[(\alpha\alpha_{ik})^2]$ equals $(1/4)\sigma_{AA}^2$. Summarizing, the parent-offspring covariance consists of half of the additive genetic variance plus one-fourth of the additive-by-additive epistatic variance:

$$\mathrm{Cov}(G_X, G_Y) = (1/2)\sigma_A^2 + (1/4)\sigma_{AA}^2$$

Returning to consider the possibility that Y inherits alleles other than *i* and *k* from X, we note that offspring of X can inherit one of four possible combinations of alleles at loci *A* and *B*, with the following probabilities:

Prob. (Y inherits *i* and *k* from X) = 1/4
Prob. (Y inherits *i* and *l* from X) = 1/4
Prob. (Y inherits *j* and *k* from X) = 1/4
Prob. (Y inherits *j* and *l* from X) = 1/4.

The covariance between Y and X can be derived for each possible case and weighted by the probability of its occurrence to compute the expected covariance between X and Y. In this case, the covariance between X and Y is the same in all cases, $(1/2)\sigma_A^2 + (1/4)\sigma_{AA}^2$. Therefore, the regression of offspring on parent phenotypes is this covariance, divided by the phenotypic variance among the parents:

$$b = [(1/2)\sigma_A^2 + (1/4)\sigma_{AA}^2] / \sigma_P^2$$

If selection is practiced on both males and females, the covariance between parent and offspring generations is the sum of the covariances between offspring and mother and offspring and father. Therefore, this regression coefficient is generally doubled and equated to narrow-sense heritability, $h^2 = 2b = \sigma_A^2/\sigma_P^2$. Including epistasis in the model illustrates that this is not strictly correct. But if the parent-offspring relationship is the basis of response to selection, then a portion of the additive-by-

additive epistatic variance should be included in the numerator of heritability (Nyquist 1991). If there is a significant amount of additive-by-additive epistatic variance, then this will increase the parent-offspring regression, and the expected response from selection will also be increased. So, additive-by-additive epistatic variance is a part of the heritable genetic variance.

Consider the case, however, when selection is performed in some generation 0, and individuals in generation 1 are mated at random to form generation 2. The response to selection that carries over into generation 2 is given by the sum of the expected covariances of a grandchild in generation 2 and its four grandparents in generation 0. All of the additive effects inherited by a grandchild descend from a selected parent in generation 0, so the additive portion of the parent-grandchild covariance remains unchanged, but only half of the additive-by-additive epistatic effects from child X in generation 1 are transmitted to a grandchild in generation 2, assuming free recombination between the two loci. Thus, response from selection in generation 1 caused by the additive-by-additive variance is reduced by half in generation 2, and this contribution will continue to be halved in each succeeding random-mating generation. The covariance between ancestors and randomly-mated descendants many generations later is σ_A^2. Griffing (1960) showed that the expected response (R) to selection (with selection differential S) observed t generations of random mating after selection is:

$$R = S\left(\frac{\sigma_A^2 + (1-r)^t \sigma_{AA}^2}{\sigma_P^2}\right),$$

where r is the recombination frequency between the two loci. Thus, tight linkage will cause the epistatic portion of the response to selection to be reduced more slowly.

Therefore, we can speak of the additive-by-additive genetic variance as temporarily heritable, and as contributing to a temporary response to selection. Recall that this variance can contribute to permanent response to selection when selection is practiced within partially inbred lines, however (Cockerham and Tachida 1988).

The increase in temporary response to selection caused by additive-by-additive genetic variance can be nearly nullified in some special situations, however, as demonstrated by Kimura (1965). Selection generates gametic disequilibrium (Bulmer 1985, Chap 9; Falconer and Mackay 1996, Chap. 11), and gametic disequilibrium affects the genetic variance (Lynch and Walsh 1997, Chap. 5). Under loose linkage and relatively

small epistasis, and after several generations of selection, populations are not in gametic phase equilibrium, but attain a state of "quasi-linkage equilibrium", in which the ratio $R = f(A_1B_1)f(A_2B_2)/f(A_2B_1)f(A_1B_2)$, where $f(A_iB_k)$ is the frequency of gamete A_iB_k, attains a constant value over generations. When populations are in quasi-linkage equilibrium, the gametic disequilibrium reduces the genetic variance by an amount nearly equal to the value of $(1/4)\sigma^2_{AA}$, and the response to selection is almost exactly what would be predicted by considering the additive genetic variance alone in the parent-offspring covariance (Kimura 1965). For this reason, Crow and Kimura (1970, Chap. 5) suggest that the contribution of epistatic variance to temporary selection response will often be nil, unless epistasis is strong, linkage is tight, or the response to selection in the initial generations, before quasi-linkage equilibrium has been attained, is being considered. Since few breeding programs are conducted as truly long-term recurrent selection programs, however, it is not certain that many plant breeding populations are at or near quasi-linkage equilibrium.

Temporary response to selection may be even greater in polysomic polypoids. For example, consider the response to selection in a tetrasomic tetraploid population, ignoring the possibility of double reduction. Expanding Kempthorne's (1955) model for autotetraploid genotypic values for a single-locus model to include two loci for individual X with alleles *i*, *j*, *k*, and *l* at locus A and alleles *m*, *n*, *o*, and *p* at locus B requires 256 terms. The different types of terms can be summarized as:

$$G_X = \mu + 8\ \alpha_i \text{ effects} + 12\ \beta_{ij} \text{ effects} + 8\ \gamma_{ijk} \text{ effects} + 2\ \delta_{ijkl} \text{ effects} + 16\ \alpha\alpha_{im} \text{ effects} + 48\ \alpha\beta_{imn} \text{ effects} + 32\ \alpha\gamma_{imno} \text{ effects} + 8\ \alpha\delta_{imnop} \text{ effects} + 36\ \beta\beta_{ijmn} \text{ effects} + 48\ \beta\gamma_{ijmno} \text{ effects} + 12\ \beta\delta_{ijmnop} \text{ effects} + 16\ \gamma\gamma_{ijkmno} \text{ effects} + 8\ \gamma\delta_{ijkmnop} \text{ effects} + 1\ \delta\delta_{ijklmnop} \text{ effect},$$

where α_i effects refer to additive effects, β_{ij} effects refer to di-allelic interactions within a locus, γ_{ijk} effects refer to tri-allelic interactions within a locus, δ_{ijkl} effects refer to quadri-allelic effects within a locus, and the other terms are interactions between effects at different loci.

When X is mated at random to another individual, the offspring of X (referred to as Y) will inherit the following number of effects in common with X (obtained by arbitrarily saying that Y inherits alleles *i* and *j* at locus A and *m* and *n* at locus B from X):

$$G_Y \text{ inherits from } G_X\text{: } 4\ \alpha_i \text{ effects} + 2\ \beta_{ij} \text{ effects} + 0\ \gamma_{ijk} \text{ effects} + 0\ \delta_{ijkl} \text{ effects} + 4\ \alpha\alpha_{im} \text{ effects} + 4\ \alpha\beta_{imn} \text{ effects} + 0\ \alpha\gamma_{imno} \text{ effects} + 0\ \alpha\delta_{imnop} \text{ effects} + 1\ \beta\beta_{ijmn} \text{ effect} + 0\ \beta\gamma_{ijmno} \text{ effects} + 0\ \beta\delta_{ijmnop} \text{ effects} + 0\ \gamma\gamma_{ijkmno} \text{ effects} + 0\ \gamma\delta_{ijkmnop} \text{ effects} + 0\ \delta\delta_{ijklmnop} \text{ effects}$$

Therefore, the covariance between tetrasomic parent and offspring is:

$$\mathrm{Cov}(X, Y) = (1/2)\sigma^2_A + (1/6)\sigma^2_D + (1/4)\sigma^2_{AA} + (1/12)\sigma^2_{AD} + (1/36)\sigma^2_{DD},$$

where σ^2_A refers to additive, σ^2_D to digenic dominance, σ^2_{AA} to additive-by-additive epistatic, σ^2_{AD} to additive-by-digenic dominant epistatic, and σ^2_{DD} to digenic-by-digenic epistatic variance components. Levings and Dudley (1963) doubled this value to obtain the numerator of the parent-offspring regression estimator of narrow sense heritability in alfalfa.

This result shows that in tetrasomic species, even more non-additive terms contribute to temporary response to selection than in diploids. Since tetrasomic species have 2*x* gametes, a portion of the digenic within-locus allelic interactions can be transmitted to the progeny. Therefore, additive-by-digenic and digenic-by-digenic epistatic allelic combinations can also be inherited, which is why Dudley et al. (1969) referred to σ^2_{AA} and σ^2_{AD} as "heritable" epistatic components.

Again, the contributions to selection response from non-additive variances decline after the initial generation is random-mated. The contributions including dominance decrease by a factor of 1/3 each generation (Walsh and Lynch 2000, Chap. 4) and become zero when the population achieves single-locus random mating equilibrium (RME, the natural extension of Hardy-Weinberg single-locus equilibrium to tetrasomic species). The portions involving epistasis decline by a factor of (1-*r*) for each generation of random-mating, due to the approach of gametic phase equilibrium. Hill and Haag (1974) computed the expected gains from selection in an autotetraploid (ignoring epistasis) and observed that the digenic dominance variance component did not contribute to the numerator of the response, in contrast to the formula given by Levings and Dudley (1963). The reason for this is that the method used by Hill and Haag (1974) implicitly assumes that the response to selection is computed based on the comparison of a RME population resulting from selection to the original RME population. Thus, Hill and Haag (1974) were computing the permanent response to selection.

Since genotype frequencies can be fixed in inbred or hybrid cultivars, these results are really only of practical importance for cultivars that are populations derived from cross-pollination. Many forage crop cultivars are synthetic or open-pollinated populations, and many of these species are polysomic polyploids (Busbice et al. 1972; Casler et al. 1996; Rumbaugh et al. 1988; Vogel and Pedersen 1993). Thus the enhanced contribution of epistasis to temporary selection response in polysomic polyploids could be exploited in these crops. Although temporary

response to selection declines with each generation of random mating after selection, this does not imply that we should ignore this component of the selection response. This idea contradicts some current thinking about polyomic polyploid breeding methods. For example, Rowe and Hill (1984) observed that there are situations in which relative rankings of autotetraploid population crosses differ in the Syn-1 and RME generations because the Syn-1 is not in equilibrium. They suggested that breeders of autotetraploid crops should allow synthetics to random mate for several generations to approach RME before evaluating them. This is a good suggestion if resulting cultivars are expected to be in RME, but if Syn-1 or other early generation synthetics could be developed as cultivars, then evaluation of Syn-1 generations is appropriate. Rather than considering these temporary response effects to be a nuisance, methods to exploit them should be developed. Breeding schemes that minimize the number of generations from selection to farmers' fields should have an advantage if any of the epistatic variance components contributing to selection response are important (Wricke and Weber 1986).

Hybrid cultivar development has been notoriously difficult in forage species, but recently Brummer (1999) proposed a method to develop semihybrid alfalfa cultivars that will at least capture a good portion of the heterosis in cultivars, and is feasible for current forage crop seed production technologies. Since most forage crops are currently sold as synthetic populations in the Syn-3 or Syn-4 or later generations, neither heterosis nor temporary responses to selection are capitalized upon in farmer's fields. Brummer's (1999) semihybrid method can exploit heterosis and temporary responses due both to dominance and epistasis in forage crops and can be used in the absence of true hybrid seed production methods.

D. Adaptive Landscapes

Wright (1982) developed the concept of adaptive landscapes as a way to illustrate the relationship between allele frequencies and mean fitness of a random-mating population where epistatic gene action is important for fitness. Adaptive landscapes are easy to describe if epistasis and overdominance are not important for fitness. In this case, the mean population fitness is a monotonic function of allele frequency at each locus, i.e., as the frequency of the favorable allele at a locus increases, population mean fitness will continue to increase. Natural or artificial selection for higher fitness will push the population uphill, always resulting in higher fitness. Eventually, the population will become fixed for the favorable allele at all loci affecting fitness, and the population will remain stable

at this maximum fitness. This is often termed a "Fisherian" model of selection and response (Coyne et al. 1997). In terms of quantitative genetics, additive genetic variance would be most important, epistatic variance would be nil, and response to selection would be a function of selection intensity and heritability. The key point is that under this model of genetics and selection, the population will eventually, but unfailingly, attain maximum fitness.

A simple example of a rugged adaptive landscape involving two loci (based on the epistatic gene action in Fig. 2.4) is shown in Fig. 2.5. Assuming the population is in Hardy-Weinberg and gametic equilibria, mean population fitness is a function of allele frequencies at the two loci. The resulting three-dimensional fitness curve has two peaks, when the population is simultaneously fixed for the A_1 and B_1 alleles or for the A_2 and B_2 alleles. If a population starts with lower frequencies of both A_1 and B_1 alleles, selection will decrease those allele frequencies until the A_2 and B_2 alleles are fixed. If a population starts with lower frequency of A_1 and higher frequency of B_1, however, selection will initially push both allele frequencies toward 0.5, at which point the population is at the "saddle point" in the middle Fig. 2.5. This saddle point is an unstable equilibrium point, and from there, the population could eventually progress to either of the fitness peaks. Most interesting is the possibility that one of the two fitness peaks is more fit than the other, a possibility that does not appear in Fig. 2.5, but is easily imaginable. In such a case, the maximum fitness point is considered a "global" fitness peak, and the lower peak is a "local" fitness peak. Depending on the initial allele frequencies in the population, the population can easily end up on the local fitness peak, where it will be stranded, because the only way to get from the local peak to the global peak in this case is by first becoming *less* fit, by going *against* the pressure of selection. Wright (1982) generalized this situation to many loci, resulting in complex, multidimensional adaptive landscape filled with local peaks onto which populations are likely to become stranded if selection is the only evolutionary force. Wright developed his "shifting balance theory" of evolution to suggest a manner in which populations might be able to move from lower to higher peaks on adaptive landscapes, with genetic drift resulting from population subdivision being the force that would allow populations to "go downhill" against selection pressures and cross valleys in the adaptive landscape.

The shifting balance theory of evolution involves many processes besides epistasis and is still hotly debated (Coyne et al. 1997, 2000; Goodnight and Wade 2000; Wade and Goodnight 1998). The debate should not obscure the fact that if epistasis is important (whether or not

the entire shifting balance theory is correct), then rugged adaptive landscapes must exist, and if selection is never counterbalanced, populations will likely become stranded on local fitness peaks and be unable to reach global peaks. Furthermore, Fisherian forms of selection and response are a part of the overall process; within a restricted range (the scope of a single local peak), selection will operate on allele frequencies in a regular, predictable manner. The profound implication of this idea, however, is that long-term selection alone will not guarantee that a population will reach its maximum potential fitness.

Is epistasis of enough importance to result in rugged adaptive landscapes that would force us to rethink concepts of selection and plant breeding? Most likely, the answer depends on one's perspective. For most breeders, progress toward a local fitness peak could be a lifetime's work. With many loci involved and complex epistatic patterns, there are more genotypic possibilities requiring evaluation to understand the adaptive landscapes than a typical breeding program can handle. In addition, the importance of genotype-by-environment interactions implies that the adaptive landscapes will be constantly shifting across environments, and the concept of global fitness maximum may be environment-dependent, anyway.

Nevertheless, in addition to the direct evidence cited previously supporting the importance of epistasis for yield in most crops, epistatic-like interactions and rugged adaptive landscapes seem to be inherent properties of all complex systems (Kauffman and Levin 1987; Lenski et al. 1999). The more parts comprising a system, the more rugged the resulting adaptive landscape and the less likely that any simple selection function will bring the system to a global maximum (Kauffman and Levin 1987). Plant genetic systems are very complicated, molecular and biochemical interactions are known to be extensive (section III), and it is almost certain that fitness and yield are not simple linear functions of allele frequencies. In the long term it is surely worthwhile considering breeding plans that can both maximize immediate gains from selection and ensure the probability of maximizing fitness in a global sense.

We will never be able to fully understand or measure the fitness landscapes of breeding populations and germplasm pools; there are simply more genotypic combinations than can possibly be evaluated. Nevertheless, this is no justification to ignore the good possibility that fitness or yield may not be a linear function of allele frequencies. DNA markers can be used to not only map crop genomes but to also map the topology of the adaptive landscapes of breeding populations. This will require large population sizes, extensive phenotypic evaluations, and perhaps special genetic stocks such as NILs (rather than typical F_2-derived map-

ping populations) in order to obtain good estimates of the fitness of many different multilocus genotypic combinations. The tools of genomics are rapidly developing to the point where interactions at the molecular level can be identified on a large scale (Ito et al. 2000). This may allow the simultaneous evaluation of molecular interactions among gene products along whole biochemical or regulatory network pathways, providing some clues as to what sorts of metabolic and regulatory interactions are really important in plants. We may find that the effects of most QTLs depend primarily on gene regulation rather than protein function, as suggested by a recent QTL cloning study (Frary et al. 2000). In this case, the ability to discriminate genic interactions that occur at the levels of gene transcription, translation, protein stability, or protein-protein interactions may be helpful in selecting and combining alleles that will interact to produce desirable genotypes. To achieve some understanding of the adaptive landscapes and to identify those phenotypic and molecular interactions that are parts of the same process, however, we will need to develop a science of *phenomics* which will relate allelic diversity, allelic and non-allelic gene interactions, and allele-by-environment interactions to the complex phenotypes that have agronomic importance. This will not be possible in a laboratory alone, but will require coordinated field and laboratory investigations to both identify and confirm candidate genes and candidate gene interactions. Again, not all multilocus genotypes can be evaluated due to practical limitations, but a good sample of different genotypes will allow a rough mapping of the adaptive topology. If Kauffman (1993, Chap. 2) is correct that many local peaks cluster near the global optimum, then a preliminary outline of the topography can guide efforts to identify optimal genotypes. Finally, much can be gained by better grounding quantitative genetics theory in biological reality, and at the same time expanding the implications of molecular biology to the phenotypic and population levels: "a conceptual and methodological marriage between mathematical statistics and nonlinear systems dynamics may become quite instrumental if it is cultivated within a molecular genetic framework," (Omholt et al. 2000).

I have presented evidence that DNA marker technologies have provided the best method for measuring epistatic effects on quantitative traits, and DNA markers or other genomic technologies may be required to better exploit epistasis for crop improvement. Currently, marker-assisted selection methods treat QTLs as building blocks that maintain their effects in isolation or in groups. Epistatic effects are considered, at most, a nuisance. If we can reliably identify epistatic effects with markers, we should be able to use them to select multi-locus genotypes rather

than specific QTL alleles in an additive fashion. Software (Charmet et al. 1999) has already been developed to implement these ideas, but this is limited to selection within simplified populations. A challenge for the future will be to more comprehensively measure epistasis and to develop methods to best exploit additive, dominant, and epistatic effects by selection both within and across populations and pedigrees. Ultimately, such methods may allow breeders to bridge the fitness valleys that exist between elite, adapted germplasm pools, and the rich genetic resources that exist for many crops but remain unused because of poor adaptation.

LITERATURE CITED

Ahmad, M., J. A. Jarrillo, O. Smirnova, and A. R. Cashmore. 1998. The CRY1 blue light photoreceptor of Arabidopsis interacts with phytochrome A in vitro. Mol. Cell 1: 939–948.

Allard, R. W. 1988. Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenies. J. Hered. 79:225–238.

Allard, R. W. 1996. Genetic basis of the evolution of adaptedness in plants. Euphytica 92:1–11.

Allard, R. W. 1999. Principles of plant breeding. 2nd Ed. John Wiley and Sons, New York.

Ang, L. H., and X. W. Deng. 1994. Regulatory hierarchy of photomorphogenic loci: allele-specific and light-dependent interaction between the HY5 and COP1 loci. Plant Cell 6:613–628.

Austin, D. F., and M. Lee. 1996. Comparative mapping in F2:3 and F6:7 generations of quantitative trait loci for grain yield and yield components in maize. Theor. Appl. Genet. 92:817–826.

Avery, L., and S. Wasserman. 1992. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. Trends Genet. 8:312–316.

Baker, R. J. 1984. Quantitative genetic principles in plant breeding. p. 147–176. In: J. P. Gustafson (ed.), Gene manipulation in plant improvement. Plenum Press, New York.

Bateson, W. 1909. Mendel's principles of heredity. Cambridge Univ. Press, Cambridge.

Bauman, L. F. 1959. Evidence of non-allelic gene interaction in determining yield, ear height, and kernel row number in corn. Agron. J. 51:531–534.

Beavis, W. D. 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. p. 252–268. Proc. 49th Ann. Corn and Sorghum Res. Conf., Chicago, IL. Amer. Seed Trade Assoc.

Bingham, E. T. 1979. Maximizing heterozygosity in autopolyploids. p. 471–487. In: W. H. Lewis (ed.), Proc. International Conference on Polyploidy: Biological Relevance, Washington University, St. Louis, MO. 24–27 May 1979. Plenum Press, New York.

Bitzer, M. J., F. L. Patterson, and W. E. Nyquist. 1982. Hybrid vigor and combining ability in a high-low yielding eight-parent diallel cross of soft red winter wheat. Crop Sci. 22:1126–1129.

Brim, C. A., and C. C. Cockerham. 1961. Inheritance of quantitative characters in soybeans. Crop Sci. 1:187–190.

Brummer, E. C. 1999. Capturing heterosis in forage crop cultivar development. Crop Sci. 39:943–954.

Bulmer, M. G. 1985. The mathematical theory of quantitative genetics. Oxford Univ. Press, Oxford, U.K.

Busbice, T. H., R. R. Hill, Jr., and H. L. Carnahan. 1972. Genetics and breeding procedures. p. 283–318. In: C. H. Hanson (ed.), Alfalfa science and technology. Agronomy Monograph 15. Am. Soc. Agron., Crop. Soc. Sci. Am., Soil. Soc. Sci. Am., Madison, WI.

Busch, R. H., J. C. Janke, and R. C. Frohberg. 1974. Evaluation of crosses among high and low yielding parents of spring wheat (*Triticum aestivum* L.) and bulk prediction of line performance. Crop Sci. 14:47–50.

Byrne, P. F., M. D. McMullen, M. E. Snook, T. A. Musket, J. M. Theuri, N. W. Widstrom, B. R. Wiseman, and E. H. Coe. 1996. Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. Proc. Nat. Acad. Sci. (USA) 93:8820–8825.

Byrne, P. F., M. D. McMullen, B. R. Wiseman, M. E. Snook, T. A. Musket, J. M. Theuri, N. W. Widstrom, and E. H. Coe. 1998. Maize silk maysin concentration and corn earworm antibiosis: QTLs and genetic mechanisms. Crop Sci. 38:461–471.

Camargo, L. E. A., and T. C. Osborn. 1996. Mapping loci controlling flowering time in *Brassica oleracea*. Theor. Appl. Genet. 92:610–616.

Casal, J. J. 2000. Phytochromes, cryptochromes, phototropin: Photoreceptor interactions in plants. Photochem. Photobiol. 71:1–11.

Casler, M. D., J. F. Pedersen, G. C. Eizenga, and S. C. Stratton. 1996. Germplasm and cultivar development. p. 413–469. In: L. E. Moser, D. E. Buxton, and M. D. Casler (ed.), Cool-season forage grasses. Agrononmy Monograph 34. Am. Soc. Agron., Crop. Soc. Sci. Am., Soil. Soc. Sci. Am., Madison, WI.

Charmet, G., N. Robert, M. R. Perretant, G. Gay, P. Sourdille, C. Groos, S. Bernard, and M. Bernard. 1999. Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. Theor. Appl. Genet. 99:1143–1148.

Cheverud, J., and E. Routman. 1996. Epistasis as a source of increased additive genetic variance at population bottlenecks. Evolution 50:1042–1051.

Cheverud, J. M., and E. J. Routman. 1995. Epistasis and its contribution to genetic variance components. Genetics 139:1455–1461.

Cheverud, J. M., T. T. Vaughn, S. Pletscher, K. King-Ellison, J. Bailiff, E. Adams, C. Erickson, and A. Bonislawski. 1999. Epistasis and the evolution of additive genetic variance in populations that pass through a bottleneck. Evolution 53:1009–1018.

Churchill, G. A., and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Clark, A. G., and L. Wang. 1997. Epistasis in measured genotypes: Drosophila *P*-element insertions. Genetics 147:157–163.

Cleghon, V., U. Gayko, T. D. Copeland, L. A. Perkins, N. Perrimon, and D. K. Morrison. 1996. Drosophila terminal structure development is regulated by the compensatory activities of positive and negative phosphotyrosine signaling sites on the Torso RTK. Genes Dev. 10:566–577.

Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariance among relatives when epistasis is present. Genetics 39:859–882.

Cockerham, C. C. 1963. Estimation of genetic variances. p. 53–94. In: W. D. Hanson and H. F. Robinson (ed.), Statistical genetics and plant breeding. National Academy of Sciences—National Research Council, Washington, DC.

Cockerham, C. C., and H. Tachida. 1988. Permanency of response to selection for quantitative characters in finite populations. Proc. Nat. Acad. Sci. (USA) 85:1563–1565.

Cockerham, C. C., and Z.-B. Zeng. 1996. Design III with marker loci. Genetics 143: 1437–1456.
Coe, E. H. 1994. Anthocyanin genetics. p. 279–281. In: M. Freeling and V. Walbot (ed.), The maize handbook. Springer-Verlag, New York.
Cone, K. 1994. Cloned anthocyanin genes and their regulation. p. 282–285. In: M. Freeling and V. Walbot (ed.), The maize handbook. Springer-Verlag, New York.
Coyne, J. A., N. H. Barton, and M. Turelli. 1997. Perspective: a critique of Sewall Wright's shifting balance theory of evolution. Evolution 51:643–671.
Coyne, J. A., N. H. Barton, and M. Turelli. 2000. Is Wright's shifting balance process important in evolution? Evolution 54:306–317.
Cregan, P. B., and R. H. Busch. 1978. Heterosis, inbreeding, and line performance in crosses of adapted spring wheats. Crop Sci. 18:247–251.
Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory. Harper and Row, New York.
Crute, I. R. 1998. The elucidation and exploitation of gene-for-gene recognition. Plant Pathology 47:107–113.
Damerval, C., A. Maurice, J. M. Josse, and D. de Vienne. 1994. Quantitative trait loci underlying gene product variation: A novel perspective for analyzing regulation of genome expression. Genetics 137:289–301.
deVicente, M. C., and S. D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585–596.
de Visser, J. A. G. M., R. F. Hoekstra, and H. van den Ende. 1997. Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. Evolution 51:1499–1505.
Doebley, J., A. Stec, and C. Gustus. 1995. *tesosinte branched* and the origin of maize: Evidence for epistasis and the evolution of dominance. Genetics 141:333–346.
Dudley, J. W., T. H. Busbice, and C. S. I. Levings. 1969. Estimates of genetic variance in 'Cherokee' alfalfa (*Medicago sativa* L.). Crop Sci. 9:228–231.
Edwards, M. D., C. W. Stuber, and J. F. Wendel. 1987. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution, and types of gene action. Genetics 116:113–125.
Ehlke, N. J., and R. R. Hill, Jr. 1988. Quantitative genetics of allotetraploid and autotetraploid populations. Genome 30:63–69.
Elena, S. F., and R. E. Lenski. 1997. Test of synergistic interactions among deleterious mutations in bacteria. Nature 307:395–398.
Elowitz, M. B., and S. Leibler. 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338.
Eshed, Y., and D. Zamir. 1996. Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics 143:1807–1817.
Eta-Ndu, J. T., and S. J. Openshaw. 1999. Epistasis for grain yield in two F2 populations of maize. Crop Sci. 39:346–352.
Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics, 4th ed. Longman Technical, Essex, U.K.
Fasoulas, A. C., and R. W. Allard. 1962. Nonallelic gene interactions in the inheritance of quantitative characters in barley. Genetics 47:899–907.
Fehr, W. R. 1987. Principles of cultivar development. Volume 1: Theory and technique. Macmillan, New York.
Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Royal Soc., Edinburgh 52:399–433.
Flor, H. H. 1956. The complementary genic systems in flax and flax rust. Adv. Genetics 8:29–54.

Frank, S. A. 1999. Population and quantitative genetics of regulatory networks. J. Theor. Biol. 197:281–294.

Frary, A., T. C. Nesbitt, A. Frary, S. Grandillo, E. van der Knapp, B. Cong, J. Liu, J. Meller, R. Elber, K. B. Alpert, and S. D. Tanksley. 2000. *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–87.

Fray, M., P. Puangsomlee, J. Goodrich, G. Coupland, E. Evans, A. Arthur, and D. Lydiate. 1997. The genetics of stamenoid petal production in oilseed rape (*Brassica napus*) and equivalent variation in *Arabidopsis thaliana*. Theor. Appl. Genet. 94:731–736.

Fu, Y. B., and K. Ritland. 1996. Marker-based inferences about epistasis for genes influencing inbreeding depression. Genetics 144:339–348.

Gardner, T. S., C. R. Cantor, and J. J. Collins. 2000. Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342.

Gibson, G. 1996. Epistasis and pleiotropy as natural properties of transcriptional regulation. Theor. Pop. Biol. 49:58–89.

Goff, S. A., K. C. Cone, and V. L. Chandler. 1992. Functional analysis of the transcriptional activator encoded by the maize B gene: evidence for a direct functional interaction between two classes of regulatory genes. Genes Dev. 6:864–875.

Gomez, L., R. Sanchez-Monge, F. Garcia-Olmedo, and G. Salcedo. 1989. Wheat tetrameric inhibitors of insect α-amylases: Alloploid heterosis at the molecular level. Proc. Nat. Acad. Sci. (USA) 86:3242–3246.

Goodnight, C. J. 1988. Epistasis and the effect of founder events on the additive genetic variance. Evolution 42:441–454.

Goodnight, C. J., and M. J. Wade. 2000. The ongoing synthesis: A reply to Coyne, Barton, and Turelli. Evolution 54:317–324.

Grattapaglia, D., F. L. G. Bertolucci, R. Penchel, and R. R. Sederoff. 1996. Genetic mapping of quantitative trait loci controlling growth and wood quality traits in Eucalyptus grandis using a maternal half-sib family and RAPD markers. Genetics 144:1205–1214.

Gravois, K. A. 1994. Diallel analysis of head rice percentage, total milled rice percentage, and rough rice yield. Crop Sci. 34:42–45.

Griffing, B. 1960. Theoretical consequences of truncation selection based on the individual phenotype. Aust. J. Biol. Sci. 13:307–343.

Guo, M., D. Davis, and J. A. Birchler. 1996. Dosage effects on gene expression in a maize ploidy series. Genetics 142:1349–1355.

Hallauer, A. R., and J. B. Miranda, Fo. 1988. Quantitative genetics in maize breeding. 2nd Edition. Iowa State Univ. Press, Ames.

Hanson, W. D., A. H. Probst, and B. E. Caldwell. 1967. Evaluation of a population of soybean genotypes with implications for improving self-pollinated crops. Crop Sci. 7:99–103.

Hanson, W. D., and C. R. Weber. 1961. Resolution of genetic variability in self-pollinated species with an application to the soybean. Genetics 46:1425–1434.

Hill, R. R., Jr., and W. L. Haag. 1974. Comparison of selection methods for autotetraploids. I. Theoretical. Crop Sci. 14:587–590.

Hill, W. G. 1982. Dominance and epistasis as components of heterosis. Z. Tierz. Zuchtgsbiol. 99:161–168.

Holland, J. B. 1998. EPISTACY: A SAS program for detecting two-locus epistatic interactions using genetic marker information. J. Hered. 89:374–375.

Holland, J. B., H. S. Moser, L. S. O'Donoughue, and M. Lee. 1997. QTLs and epistasis associated with vernalization responses in oat. Crop Sci. 37:1306–1316.

Humphrey, A. B., D. F. Matzinger, and C. C. Cockerham. 1969. Effects of random intercrossing in a naturally self-fertilizing species, *Nicotiana tabacum* L. Crop Sci. 9: 495–497.

Ito, T., K. Tashuro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, and Y. Sakaki. 2000. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interaction in all possible combinations between the yeast proteins. Proc. Nat. Acad. Sci. (USA) 97:1143–1147.

Jayaram, C., and P. A. Peterson. 1990. Anthocyanin pigmentation and transposable elements in maize aleurone. Plant Breed. Rev. 8:91–137.

Jiang, C.-X., W. C. Peng, X. Draye, P. L. Morrell, C. W. Smith, and A. H. Paterson. 2000. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). Evolution 54:798–814.

Jorgensen, R. A. 1995. Cosuppression, flower color patterns, and metastable gene expression states. Science 268:686–691.

Kacser, H., and J. A. Burns. 1981. The molecular basis of dominance. Genetics 97:639–666.

Kao, C.-H., Z.-B. Zeng, and R. D. Teasdale. 1999. Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216.

Kauffman, S., and S. Levin. 1987. Towards a general theory of adaptive walks on rugged landscapes. J. Theor. Biol. 128:11–45.

Kauffman, S. A. 1993. The origins of order. Oxford Univ. Press, Oxford, U.K.

Kearsey, M. J., and A. G. L. Farquhar. 1998. QTL analysis in plants; Where are we now? Heredity 80:137–142.

Kearsey, M. J., and J. L. Jinks. 1968. A general method of detecting additive, dominance, and epistatic variance for metrical traits. Heredity 23:403–409.

Keightley, P. D. 1989. Models of quantitative variation of flux in metabolic pathways. Genetics 121:869–876.

Keightley, P. D. 1996a. A metabolic basis for dominance and recessivity. Genetics 143:621–625.

Keightley, P. D. 1996b. Metabolic models of selection response. J. Theor. Biol. 182:311–316.

Kempthorne, O. 1954. The theoretical values of correlations between relatives in random mating populations. Genetics 40:153–167.

Kempthorne, O. 1955. The correlation between relatives in a simple autotetraploid population. Genetics 40:168–174.

Kempthorne, O. 1957. An introduction to genetic statistics. Wiley, New York.

Kempthorne, O. 1988. An overview of the field of quantitative genetics. p. 47–56. In: B. S. Weir, M. M. Goodman, E. J. Eisen, and G. Namkoong (ed.), Proc. Second International Conference on Quantitative Genetics. Sinauer, Sunderland, MA.

Kianian, S. F., B.-C. Wu, S. L. Fox, H. W. Rines, and R. L. Phillips. 1997. Aneuploid marker assignment in hexaploid oat with the C genome as a reference for determining remnant homoeology. Genome 40:386–396.

Kimura, M. 1965. Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. Genetics 52:875–890.

Kirby, D. A., S. V. Muse, and W. Stephan. 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Nat. Acad. Sci. (USA) 92:9047–9051.

Kohel, R. J., and A. A. Bell. 1999. Genetic analysis of two terpenoid variants in cotton (*Gossypium hirsutum* L.). J. Hered. 90:249–251.

Lamkey, K. R., B. J. Schnicker, and A. E. Melchinger. 1995. Epistasis in an elite maize hybrid and choice of generation for inbred line development. Crop Sci. 35:1272–1281.

Lark, K. G., K. Chase, F. Adler, L. M. Mansur, and J. H. Orf. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc. Nat. Acad. Sci. (USA) 92:4656–4660.

Lark, K. G., J. Orf, and L. M. Mansur. 1994. Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. Theor. Appl. Genet. 88:486–489.

Lenski, R. E., C. Ofria, T. C. Collier, and C. Adami. 1999. Genome complexity, robustness and genetic interactions in digital organisms. Nature 400:661–664.

Levings, C. S., and J. W. Dudley. 1963. Evaluation of certain mating designs for estimation of genetic variance in autotetraploid alfalfa. Crop Sci. 3:532–535.

Li, Z. 1998. Molecular analysis of epistasis affecting complex traits. p. 119–130. In: A. H. Paterson (ed.), Molecular dissection of complex traits. CRC Press, Boca Raton, FL.

Li, Z., S. R. M. Pinson, W. D. Park, A. H. Paterson, and J. W. Stansel. 1997. Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 145:453–465.

Lo, P. C. H., D. Roy, and S. M. Mount. 1994. Suppressor U1 snRNAs in Drosophila. Genetics 138:365–378.

Lopez-Fanjul, C., A. Fernandez, and M. A. Toro. 1999. The role of epistasis in the additive genetic variance after population bottlenecks. Gen. Res. Cambridge 73:45–59.

Lubberstedt, T., A. E. Melchinger, C. C. Schon, H. Utz, and D. Klein. 1997. QTL mapping in testcrosses of European flint lines of maize. I. Comparison of different testers for forage yield traits. Crop Sci. 37:921–931.

Lynch, M. 1988. Design and analysis of experiments on random drift and inbreeding depression. Genetics 120:791–807.

Lynch, M. 1991. The genetic interpretation of inbreeding depression and outbreeding depression. Evolution 45:622–629.

Lynch, M., and B. Walsh. 1997. Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland, MA.

Mac Key, J. 1970. Significance of mating systems for chromosomes and gametes in polyploids. Hereditas 66:165–176.

Madhani, H. D., and C. Guthrie. 1994. Dynamic RNA-RNA interactions in the spliceosome. Ann. Rev. Genet. 28:1–26.

Martienssen, R. 1999. Copying out our ABCs: the role of gene redundancy in interpreting genetic hierarchies. Trends Genet. 15:435–437.

Mather, K., and J. L. Jinks. 1977. Introduction to biometrical genetics. Cornell Univ. Press, Ithaca, NY.

Matzke, M. A., and A. J. M. Matzke. 1995. How and why do plants inactivate homologous (trans)genes? Plant Physiol. 107:679–685.

Matzke, M. A., F. Neuhuber, and A. J. Matzke. 1993. A variety of epistatic interactions can occur between partially homologous transgene loci brought together by sexual crossing. Mol. Gen. Genet. 236:379–386.

McMullen, M. D., P. F. Byrne, M. E. Snook, B. R. Wiseman, E. A. Lee, N. W. Widstrom, and E. H. Coe. 1998. Quantitative trait loci and metabolic pathways. Proc. Nat. Acad. Sci. [USA] 95:1996–2000.

Melchinger, A. E., H. H. Geiger, and F. W. Schnell. 1986. Epistasis in maize (Zea mays L.). 2. Genetic effects in crosses among early flint and dent inbred lines determined by three methods. Theor. Appl. Genet. 72:231–239.

Melchinger, A. E., M. Singh, W. Link, H. F. Utz, and E. von Kittlitz. 1994. Heterosis and gene effects of multiplicative characters: theoretical relationships and experimental results from *Vicia faba* L. Theor. Appl. Genet. 88:343–348.

Melchinger, A. E., H. F. Utz, and C. C. Schon. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveal low power of QTL detection and large bias in estimates of QTL effects. Genetics 149: 383–403.

Millar, A. J., R. B. McGrath, and N.-H. Chua. 1994. Phytochrome phototransduction pathways. Ann. Rev. Genet. 28:325–349.

Moll, R. H., J. H. Lonnquist, J. Velez Fortuno, and E. C. Johnson. 1965. The relationship of heterosis and genetic divergence in maize. Genetics 52:139–144.

Moreno, G. 1994. Genetic architecture, genetic behaviour, and character evolution. Ann. Rev. Ecol. Syst. 25:31–44.

Moreno-Gonzalez, J., and J. Dudley. 1981. Epistasis in related and unrelated maize hybrids determined by three methods. Crop Sci. 21:644–651.

Nap, J. P., A. J. Conner, L. Mlynarova, W. J. Stiekema, and R. C. Jansen. 1997. Dissection of a synthesized quantitative trait to characterize transgene interactions. Genetics 147:315–320.

Neuhuber, F., Y.-D. Park, A. J. M. Matzke, and M. A. Matzke. 1994. Susceptibility of transgene loci to homology-dependant gene silencing. Mol. Gen. Genet. 244:230–241.

Niederberger, P., R. Prasad, G. Miozzari, and H. Kacser. 1992. A strategy for increasing in vivo flux by genetic manipulations. The tryptophan system of yeast. Biochem. J. 287:473–479.

Nyquist, W. E. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10:235–322.

Omholt, S. W., E. Plahte, L. Oyehaug, and K. Xiang. 2000. Gene regulatory networks generating the phenomena of additivity, dominance, and epistasis. Genetics 155:969–980.

Openshaw, S., and E. Frascaroli. 1997. QTL detection and marker-assisted selection for complex traits in maize. p. 44–53. Proc. 52nd Ann. Corn and Sorghum Res. Conf., Chicago, IL. Amer. Seed Trade Assoc.

Openshaw, S., and E. Frascaroli. 1998. QTL detection in large and small samples using real data. p. 127. Proc. Plant and Animal Genome VI, San Diego, CA. http://www.intlpag.org/pag/6/abstracts/141.html.

Orf, J. H., K. Chase, F. R. Adler, L. M. Mansur, and K. G. Lark. 1999. Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Sci. 39:1652–1657.

Parcy, F., O. Nilsson, M. A. Busch, I. Lee, and D. Weigel. 1998. A genetic framework for floral patterning. Nature 395:561–566.

Phillips, P. C. 1998. The language of gene interactions. Genetics 149:1167–1171.

Pickett, F. B., and D. R. Meeks-Wagner. 1995. Seeing double: Appreciating genetic redundancy. Plant Cell 7:1347–1356.

Pixley, K. V., and K. J. Frey. 1991. Combining ability for test weight and agronomic traits of oat. Crop Sci. 31:1448–1451.

Podlich, D. W., and M. Cooper. 1998. QU-GENE: a simulation platform for quantitative analysis of genetic models. Bioinformatics 14:632–653.

Pressoir, G., L. Albar, N. Ahmadi, I. Rimbault, M. Lorieux, D. Fargette, and A. Ghesquiere. 1998. Genetic basis and mapping of the resistance to rice yellow mottle virus. II. Evidence of a complementary epistasis between two QTLs. Theor. Appl. Genet. 97:1155–1161.

Rasmusson, D. C., and R. L. Phillips. 1997. Plant breeding programs and genetic diversity from de novo variation and elevated epistasis. Crop Sci. 37:303–310.

Ratcliffe, O. J., D. J. Bradley, and E. S. Coen. 1999. Separation of root and floral identity in *Arabidopsis*. Development 26:1109–1120.

Rawlings, J. O. 1988. Applied regression analysis: a research tool. Wadsworth & Brooks, Pacific Grove, CA.

Remington, D. L., and D. A. O'Malley. 2000. Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family. Genetics 155:337–348.

Richey, F. D. 1942. Mock-dominance and hybrid vigor. Science 96:280–281.

Rieseberg, L. H., B. Sinervo, C. R. Linder, M. C. Ungerer, and D. M. Arias. 1996. Role of gene interactions in hybrid speciation: Evidence from ancient and experimental hybrids. Science 272:741–745.

Robertson, A. 1951. The effect of inbreeding on the variation due to recessive genes. Genetics 37:189–207.

Rooney, W. L., and S. Aydin. 1999. Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. Crop Sci. 39:397–400.

Rowe, D. D., and R. R. Hill, Jr. 1984. Effect of gametic disequilibrium on means and on genetic variances of autotetraploid synthetic varieties. Theor. Appl. Genet. 68:69–74.

Rumbaugh, M. D., J. L. Caddel, and D. E. Rowe. 1988. Breeding and quantitative genetics. p. 777–808. In: A. A. Hanson, D. K. Barnes, and R. R. Hill, Jr. (ed.), Alfalfa and alfalfa improvement. Agronomy Monograph 29. Am. Soc. Agron., Crop. Soc. Sci. Am., Soil. Soc. Sci. Am., Madison, WI.

Russell, W. A., and S. A. Eberhart. 1970. Effects of three gene loci in the inheritance of quantitative characters in maize. Crop Sci. 10:165–169.

Rutherford, S. L., and S. Lindquist. 1998. Hsp90 as a capacitor for morphological evolution. Nature 396:336–342.

Schnell, F. W., and C. C. Cockerham. 1992. Multiplicative vs. arbitrary gene action in heterosis. Genetics 131:461–469.

Stryer, L. 1988. Biochemistry. 3rd Ed. W. H. Freeman, New York.

Stuber, C. W. 1994. Success in the use of molecular markers for yield enhancement in corn. p. 232–238. Proc. 49th Ann. Corn and Sorghum Res. Conf., Chicago, IL. Am. Seed Trade Assoc.

Stuber, C. W. 1998. Case history in crop improvement: Yield heterosis in maize. p. 197–206. In: A. H. Paterson (ed.), Molecular Dissection of Complex Traits. CRC Press, Boca Raton, FL.

Stuber, C. W., and C. C. Cockerham. 1966. Gene effects and variances in hybrid populations. Genetics 54:1279–1286.

Stuber, C. W., S. E. Lincoln, D. W. Wolff, T. Helentjaris, and E. S. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839.

Stuber, C. W., and R. H. Moll. 1971. Epistasis in maize (*Zea mays* L.). II. Comparison of selected with unselected populations. Genetics 67:137–149.

Stuber, C. W., W. P. Williams, and R. H. Moll. 1973. Epistasis in maize (*Zea mays* L.): III. Significance in predictions of hybrid performances. Crop Sci. 13:195–200.

Stuthman, D. D., and R. E. Stucker. 1975. Combining ability analysis of near-homozygous lines derived from a 12-parent diallel cross in oats. Crop Sci. 25:800–803.

Sun, J. S., and J. L. Manley. 1995. A novel U2-U6 snRNA structure is necessary for mammalian messenger-RNA splicing. Genes Dev. 9:843–854.

Szathmary, E. 1993. Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. Genetics 133:127–132.

Tanksley, S. D. 1993. Mapping polygenes. Ann. Rev. Genet. 27:205–233.

Tanksley, S. D., and J. Hewitt. 1988. Use of molecular markers in breeding for soluble solids contents in tomato—a re-examination. Theor. Appl. Genet. 75:811–823.

Thomas, W. T. B., W. Powell, R. Waugh, K. J. Chalmers, U. M. Barua, P. Jack, V. Lea, B. P. Forster, J. S. Swanston, and R. P. Ellis. 1995. Detection of quantitative trait loci for agronomic, yield, grain and disease characters in spring barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 91:1037–1047.

Tranquilli, G., and J. Dubcovsky. 2000. Epistatic interaction between vernalization genes Vrn-Am1 and Vrn-Am2 in diploid wheat. J. Hered. 91:304–306.

Upadhyaya, H. D., and S. N. Nigam. 1998. Epistasis for vegetative and reproductive traits in peanut. Crop Sci. 38:44–49.

Utz, H. F., A. E. Melchinger, and C. C. Schon. 2000. Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839–1849.

Veldboom, L. R., and M. Lee. 1994. Molecular-marker-facilitated studies of morphological traits in maize. II: Determination of QTLs for grain yield and yield components. Theor. Appl. Genet. 89:451–458.

Vogel, K. P., and J. F. Pedersen. 1993. Breeding systems for cross-pollinated perennial grasses. Plant Breed. Rev. 11:251–274.

Wade, M. J., and C. J. Goodnight. 1998. Perspective: the theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. Evolution 52:1537–1553.

Walsh, B., and M. Lynch. 2000. Selection and evolution of quantitative traits. http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html.

Wang, D. L., J. Zhu, Z. K. Li, and A. H. Paterson. 1999. Mapping QTLs with epistatic effects and QTLxEnvironment interactions by mixed linear model approaches. Theor. Appl. Genet. 99:1255–1264.

Weir, B. S., and C. C. Cockerham. 1977. Two-locus theory in quantitative genetics. p. 247–269. In: E. Pollack, O. Kempthorne, and T. B. Bailey, Jr. (ed.), Proc. First International Conference on Quantitative Genetics, Ames, IA. August 16–21, 1976. Iowa State Univ. Press, Ames.

Weng, G., U. S. Bhalla, and R. Iyengar. 1999. Complexity in biological signaling systems. Science 284:92–96.

Whitlock, M. C., and K. Fowler. 1999. The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster*. Genetics 152:345–353.

Whitlock, M. C., P. C. Phillips, and M. J. Wade. 1993. Gene interaction affects the additive genetic variance in subdivided populations with migration and extinction. Evolution 47:1758–1769.

Willham, R. L., and E. Pollak. 1985. Theory of heterosis. J. Dairy Sci. 68:2411–2417.

Wolf, D. P., and A. R. Hallauer. 1997. Triple testcross analysis to detect epistasis in maize. Crop Sci. 37:763–770.

Wricke, G., and W. E. Weber. 1986. Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin.

Wright, S. 1982. Character changes, speciation, and the higher taxa. Evolution 36:427–443.

Xiao, J., J. Li, L. Yuan, and S. D. Tanksley. 1996. Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific rice cross. Theor. Appl. Genet. 92:230–244.

Xiao, J. H., J. Li, L. P. Yuan, and S. D. Tanksley. 1995. Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745–754.

Yano, M., Y. Harushima, Y. Nagamura, N. Kurata, Y. Minobe, and T. Sasaki. 1997. Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. Theor. Appl. Genet. 95:1025–1032.

Yu, S. B., J. X. Li, C. G. Xu, Y. F. Tan, Y. J. Gao, X. H. Li, Q. Zhang, and M. A. S. Maroof. 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc. Nat. Acad. Sci. (USA) 94:9226–9231.

Yuh, C.-H., H. Bolouri, and E. H. Davidson. 1998. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. Science 279:1896–1902.

Errata for Table 2.6 from Holland, J. B. 2001. Epistatsis and Plant Breeding, in **Plant Breeding Reviews**, *Volume 21, edited by Jules Janick. John Wiley & Sons, New York.*

Table 2.6. Statistical genetic parameters of the two-locus, two-allele model, assuming Hardy-Weinberg and gametic phase equilibria.

Model for genotypic value: $G_{ijkl} = \mu_{....} + \alpha_i^A + \alpha_j^A + \alpha_k^B + \alpha_l^B + \delta_{ij}^A + \delta_{kl}^B + \alpha\alpha_{ik} + \alpha\alpha_{il} + \alpha\alpha_{jk} + \alpha\alpha_{jl} + \alpha\delta_{ikl} + \alpha\delta_{jkl} + \delta\alpha_{ijk} + \delta\alpha_{ijl} + \delta\delta_{ijkl}$

Restrictions of genotypic model:

$$\sum_{i=1}^{2} p_i\alpha_i^A = 0, \sum_{k=1}^{2} q_k\alpha_k^B = 0$$

$$\sum_{i=1}^{2}\sum_{j=1}^{2} p_ip_j\delta_{ij}^A = 0, \sum_{k=1}^{2}\sum_{l-1}^{2} q_kq_l\delta_{kl}^B = 0, \sum_{i=1}^{2} p_i\delta_{ij}^A = 0, \sum_{k=1}^{2} q_k\delta_{kl}^B = 0$$

$$\sum_{i-1}^{2}\sum_{k=1}^{2} p_iq_k\alpha\alpha_{ik} = 0, \sum_{k=1}^{2} q_k\alpha\alpha_{ik} = 0, \sum_{i=1}^{2} p_i\alpha\alpha_{ik} = 0$$

$$\sum_{i=1}^{2}\sum_{k-1}^{2}\sum_{l=1}^{2} p_iq_kq_l\alpha\delta_{ikl} = 0, \sum_{k=1}^{2}\sum_{l=1}^{2} q_kq_l\alpha\delta_{ikl} = 0, \sum_{i=1}^{2} p_i\alpha\delta_{ikl} = 0$$

$$\sum_{i=1}^{2}\sum_{j-1}^{2}\sum_{k-1}^{2} p_ip_jq_k\delta\alpha_{ijk} = 0, \sum_{i=1}^{2}\sum_{j=1}^{2} p_ip_j\delta\alpha_{ijk} = 0, \sum_{k=1}^{2} q_k\delta\alpha_{ijk} = 0$$

$$\sum_{i-1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} p_ip_jq_kq_l\delta\delta_{ijkl} = 0, \sum_{j-1}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} p_jq_kq_l\delta\delta_{ijkl} = 0, \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k-1}^{2} p_ip_jq_k\delta\delta_{ijkl} = 0,$$

$$\sum_{i-1}^{2}\sum_{j=1}^{2} p_ip_j\delta\delta_{ijkl} = 0, \sum_{k=1}^{2}\sum_{l=1}^{2} q_kq_l\delta\delta_{ijkl} = 0, \sum_{i=1}^{2}\sum_{k-1}^{2} p_iq_k\delta\delta_{ijkl} = 0, \sum_{i-1}^{2} p_i\delta\delta_{ijkl} = 0, \sum_{k-1}^{2} q_k\delta\delta_{ijkl} = 0$$