Gene Effects

By William Beavis, Kendall Lamkey, Katherine Espinosa (ISU)

Introduction

In 1918, RA Fisher provided the first major contribution to the modern synthesis by proposing a model that reconciled inheritance of discrete characteristics (Mendel) and continuous, or quantitative, characteristics (Darwin) in breeding populations. Herein, the same theoretical foundations are introduced.

For the beginning student, it will seem that the primary purpose of theory and modeling is to provide interpretations of observational and experimental results. Without this theoretical foundation, there would be no genetic understanding of the results from plant breeding experiments.

However, there is a more important practical justification for learning theoretical models: Theory provides predictions. Predictions are the basis for generating testable hypotheses. Also, with a theoretical model it is possible to simulate many different breeding strategies. These can be compared and the most promising can be used to design and implement the most effective and efficient breeding strategies. Thus, theory provides a rational basis for designing plant breeding programs.

Objectives

- Model Genotypic and Phenotypic Values of individuals in crop breeding populations.
- Integrate genotypic effect models with allele frequency models at single and multiple loci.
- Distinguish and estimate genetic effects, effects of allele substitutions and Breeding Values at single and multiple loci.
- Distinguish and estimate dominance and epistatic deviations from additive effect models.
- Integrate concepts to applied breeding programs with data sets consisting of genotypic (marker) information with phenotypic information for QTL analyses.

Linear Models for Phenotypic Values

Single Locus

The **phenotypic value** of an individual, or group of individuals, is observed when a character or trait is measured. For example, if a corn plant was measured and found to be 275 cm tall, then that would be its phenotypic value for height.

To draw inferences about the genetic properties of a trait, we model phenotypic values using linear components. The most common model consists of a part due to genetics and a part due to non-genetic effects such as the environment. This is usually written as:

$$P = G + E$$

where P is the **phenotypic value**, G is the **genotypic value**, and E represents non-genetic factors.

If we assume that ΣE = 0, then

$$P = G)(\bar{P} = \bar{G}$$

Population Mean

The mean phenotypic value of a population is equal to the mean genotypic value when the non-genetic (environmental) deviations sum to zero.

To calculate the expected genotypic properties of a population for a single locus, we assign arbitrary genotypic values to each locus.

Consider a single locus with two alleles = (A, a)

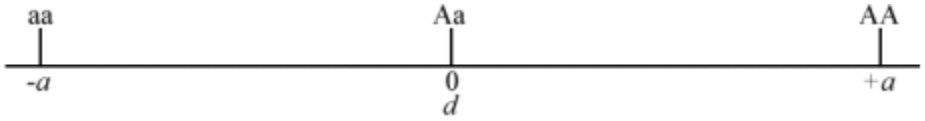
Coded genotypic value of one homozygote (AA) = +*a*

Coded genotypic value of the other homozygote (aa) = -*a*

Coded genotypic value of a heterozygote (Aa) = *d*

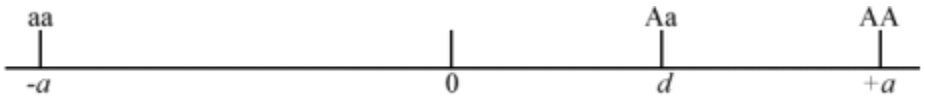
We can arbitrarily designate the A allele as the allele that increases the genotypic value. The genotypic value of the heterozygotes (*d*) depends on the level of dominance:

1. No dominance: *d* = 0

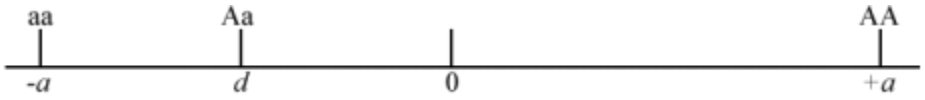


Degree of Dominance

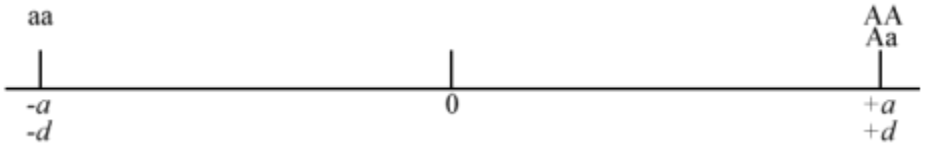
2. If A is dominant or partially dominant relative to the a allele, then d is positive



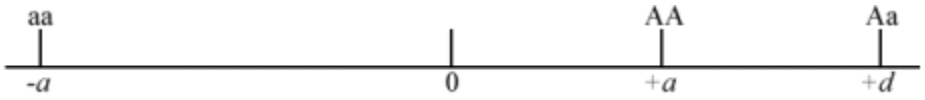
3. If a is dominant or partially dominant to the A allele, then d is negative



4. If dominance is complete: d = +a or a



5. If there is overdominance: d is greater than +a or less than -a



Allele Frequencies and Population Mean

Table 1 Influence of allele frequencies and dominance deviation on the average value of the trait in the population.

	Genotype			
	AA	Aa	aa	Total
Frequency	p^2	2pq	q^2	1
Genotypic Value	Y_{AA}	Y_{Aa}	Y_{aa}	
Coded GV	a	d	-a	
Freq. x Coded GV	p^2a	2pqd	$-q^2$	=a(p - q) + 2dpq

Table 2 An example of a genotype that controls the number of flowers and the expected population value for number of flowers.

	Genotype			
	AA	**Aa**	**aa**	**Total**
Frequency	0.64	0.32	0.04	1
Genotypic Value	16	12	0	
Coded GV	8	4	-8	
Freq. x Coded GV	5.8	1.28	-.32	14.08

Note. Coded Genotypic Values are obtained by subtracting the midparent value i.e., the midpoint between the genotypic values of the two homozygotes.

Population mean = $\bar{Y}$ = μ + Expected value of g_{ij}

mh = mid-homozygote value

$\bar{Y}$ = mh + a(p - q) + 2pqd

p = frequency of A

q = frequency of a

Additive Gene Action

Invalid Equation

$\bar{Y} = a(p - q) + 2pdq$: This is both the mean genotypic value and the mean phenotypic value of the population with respect to the trait.

Notice that if *d* = 0 the heterozygote genotype has no impact on the population mean and we say that completely additive gene action exists.

Two Loci

Next consider the contributions of alleles at more than one locus and find the joint effect on the mean.

Consider two single loci:

- Genotypic value of AA is a_A
- Genotypic value of BB is a_B

Consider multiple loci:

- Genotypic value of AABB is $a_A + a_B$
- $G_T = G_A + G_B$
- The mid-homozygote genotypic value is the average of double homozygotes (A1A1B1B1 and A2A2B2B2)

Table 3 Joint effects of coded genotypic values and frequencies of alleles at two loci.

Two-locus genotypic values and frequencies		A locus genotype		
		AA	Aa	aa
B locus genotype	Coded Genotypic Value/Freq.	a_A Invalid Equation	d_A $2p_Aq_A$	$-a_A$ ${q_A}^2$
BB	a_B ${p_B}^2$	Invalid Equation $p_A^2p_B^2$	Invalid Equation $2p_Aq_Ap_B^2$	$G_T = \mu + a_A + a_B$ $q_A^2p_B^2$
Bb	d_B $2p_Bq_B$	Invalid Equation $p_A^2 2p_Bq_B$	Invalid Equation $2p_Aq_A(2p_Bq_B)$	Invalid Equation $q_A^2 2p_Bq_B$
bb	$-a_B$ q_B^2	Invalid Equation $p_A^2q_B^2$	Invalid Equation $2p_Aq_Aq_B^2$	Invalid Equation $q_A^2q_B^2$

Population Mean

Population mean = $\bar{Y}$ = μ + Expected values of G_A and G_B

G_A and G_B are weighted averages based on allele frequencies and coded genotypic values.

$$\bar{Y} = \mu + E(G_A + G_B) = E(G_A + G_B)), (\bar{Y} = \mu + \{a_A(p_A - q_A) + 2p_A q_A d_A\} + \{a_B(p_B - q_B) + 2p_B q_B d_B\}$$

Table 4 A numeric example involving dominance and allele frequencies that are not equal at two loci.

Two-locus genotypic values and frequencies		A locus genotype		
		AA	Aa	aa
B locus genotype	Boded Genotypic Value/Freq.	8 0.64	4 0.32	-8 0.04
BB	4 0.04	12 0.0256	8 0.0128	-4 0.0016
Bb	2 0.32	10 0.2048	6 0.1024	-6 0.0128
bb	-4 0.64	4 0.4096	0 0.2048	-12 0.0256
Average at locus A		6.24	2.24	-9.76
Average at locus B		10.08	8.08	2.08

Extension to More Than 2 Loci

$$mh = \frac{12 + (-12) + -4}{4} = 0$$

$$\bar{Y} = \mu + E(G_A + G_B) = E(G_A + G_B))(\bar{Y} = \mu + \{a_A(p_A - q_A) + 2p_Aq_Ad_A\} + \{a_B(p_B - q_B) + 2p_Bq_Bd_B\})(\bar{Y} = 0 + \{8(0.8 - 0.2) + 2 * 0.8 * 0.2 * 4\} + \{4(0.2 - 0.8) + 2 * 0.8 * 0.2 * 2\})(\bar{Y} = 0 + 6.08 - 1.76 = 4.32$$

Extension To More Than 2 Loci

- $G_T = \sum G_i$
- midpoint is the average of the most extreme multi-locus-homozygotes
- Population mean = Invalid Equation

= Invalid Equation

= Invalid Equation

Average Genetic Effects

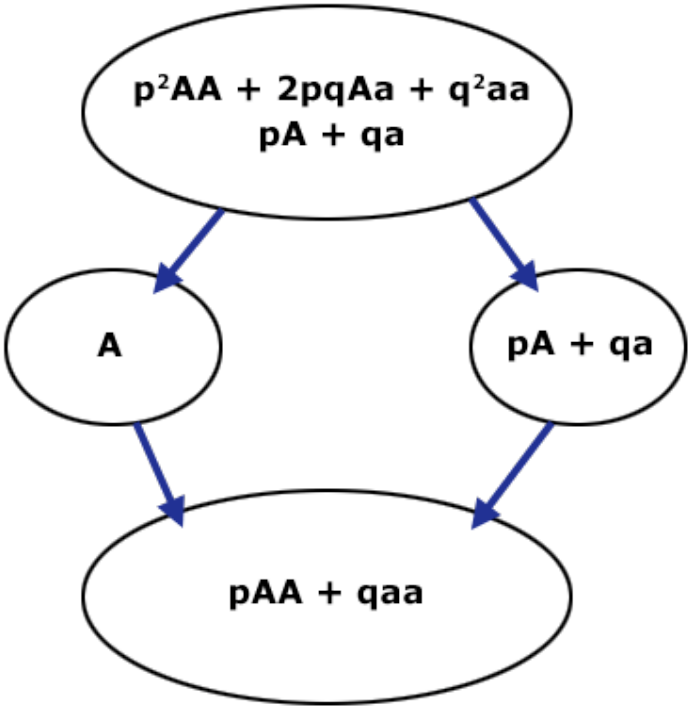
Average Genetic (Allelic) Effects

Individuals chosen as parents transmit only a sample consisting of ½ of its alleles. With selection we are concerned with the transmission of value from parent to offspring. This cannot be determined based on genotypic value alone. Parents pass on their genetic alleles, NOT their genotypes to the next generation. Genotypes are created anew in each generation. One result is that some aspects of the value of a particular genotype are unpredictable. Yet, selection theory can work only with the predictable aspects of the union of two gametes. Therefore, we introduce the *average effect of a gene (allele)* to represent this concept.

Key Concept: Although genotypes determine genotypic values, alleles and not genotypes are inherited by progeny.

Average effect of a gene (allele)

The mean deviation from the population mean of individuals that received the gene (allele) from one parent is the average effect of the gene (allele). The effect of the other gene (allele) received from the remaining parent is represented as a random allele from the population … for this concept.



Formula of Average Effect of an Allele

Conceptually, let a number of gametes carrying A allele unite at random with gametes from the population; then the mean of the genotypes deviates from the population mean by an amount that is the average effect of the A gene. This represents the average allele effect and is the average deviation from the population mean of individuals who received a specific allele from one parent and the other allele at random from the population.

Table 5 Average effect of an allele.

Alleles	**Genotypes, coded genotypic values and frequencies**			**Mean value of genotypes produced**	**Population mean to be deduced**	**Average effect of the allele**
	AA a	Aa d	aa -a	next gen	this gen	(next gen)−(this gen)
A	p	q		$pa + qd$	$[a(p-q) + 2dpq]$	$q[a + d(q-p)]$
a		p	q	$-qa + pd$	$[a(p-q) + 2dpq]$	$-p[a + d(q-p)]$

The average effect of the *A* allele (or the *a* allele) from a single locus is designated as α_A (or α_a) calculated for data presented in Table 5:

$$\alpha_A = q\,[a + d\,(q-p)] = 0.2\,[8 + (0.2-0.8)\,4] = 1.12)(\alpha_a = -p\,[a + (q-p)\,d] = -0.8\,[8 + (0.2-0.8)\,4] = -4.48$$

Allele Substitution Effect

The average effect of an allele substitution, often designated as α, is the difference between average effects of each allele.

$$\alpha = \alpha A - \alpha a = a + (p - q)\, d$$

For example, from table 5, α_1 and α_2 represent average allele effects of A and a respectively.

Average effects of each allele can be calculated as:

$$\alpha_A = q\alpha = 0.2\,(5.6) = 1.12)(\alpha_a = -p\alpha = (-0.8)\,(5.6) = -4.48$$

Thus, the average effect of allele substitution =

$$\alpha = \alpha_A - \alpha_a = 1.12 - (-4.48) = 5.6$$

Note the average genetic effect is:

- Dependent on genotypic value,
- Dependent on gene frequencies,
- A property of the population as well as the genes concerned.

Breeding Value

Breeding value is a concept that is based on the following:

- The average value of a parent is judged by its progeny.
- Alleles carried by an individual and transmitted to its offspring can be inferred from the progeny,
- Which represents the sum of the average effects of all alleles an individual carries.

Let's use the average effect of alleles to rewrite the equation:

$$P = G + E)(= \alpha_i + \alpha_j + \delta_{\ddot{y}} + E$$

where α_i is the average effect of allele *i* in a diploid individual and $\delta_{\ddot{y}}$ is the dominance deviation.

Breeding value is the value of an individual judged by the average value of its progeny. The breeding value of an individual is equal to the sum of the average effects of the alleles it carries. The summation is over pairs of alleles at a locus and over all loci (Table 5). It is defined as twice the expected deviation of the individual's progeny mean from the population mean when the individual is mated at random to other individuals from the same population (Tables 6 and 7).

Mean Breeding Value in Random Population

The mean breeding value in a random mating population is zero.

Table 6 Relationship of breeding values to genotypes.

Genotype	Breeding value
AA	$2a_A = 2qa$
Aa	$a_A + a_a = (q - p)a$
aa	$2a_a = -2pa$

Table 7 Theoretical example of calculations of breeding values.

Genotype	Breeding value
AA	$2a_A = 2(0.2)(5.6) = 2.24$
Aa	$a_A + a_a = (0.2 - 0.8)5.6 = -3.36$
aa	$2a_a = -2(0.8)(5.6) = -8.96$

Deviations for Average Genetic Effects

Dominance Deviation

For a single locus: the difference between the genotypic value and the breeding value of a particular genotype is known as the dominance deviation. It is associated with the Aa genotype. δ_{ij} represents the deviation of genotypic value (i.e., G_{ij}) from the regression-fitted genotypic value and is zero when dominance is absent ($d = 0$)

Consider Genotype AA. Recall that the Coded genotypic value of AA = +a, and the population mean

$$= a(p-q) + 2dpq$$

If a is expressed as deviation relative to the population mean, then

$$a - [a(p-q) + 2dpq] = a(1-p+q) - 2dpq = 2qa - 2dpq = 2q(a-dp)$$

Notice that if *d* is not 0, and *p* is not equal to *q*, then *a* is affected by d. Also, recall that *a* can be expressed as in terms of the average effect of an allele substitution,

$$a = \alpha - d(q-p)$$

thus

$$2q(a-dp) = 2q(\alpha - d(q-p) - dp) = 2q(\alpha - dq + dp - dp) = 2q(\alpha - dp)$$

Using similar algebra, the dominance deviation,

$$\delta_{ij} = 2q(\alpha - dp) - 2pq = -2q^2 d$$

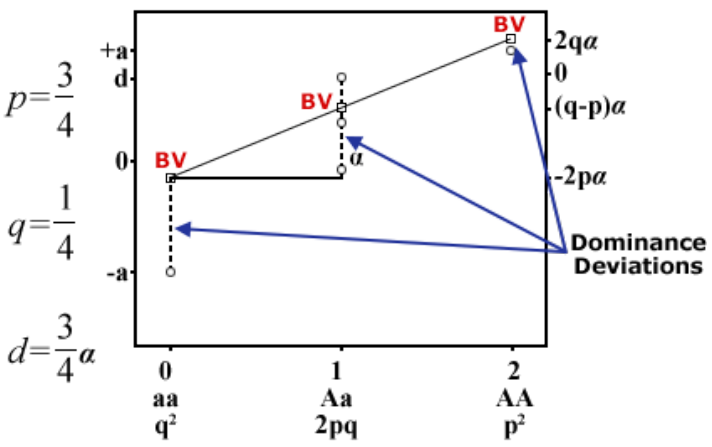
Observations About Dominance

Notice that

- If there is no dominance, d is zero and the dominance deviations are also zero.
- In the absence of dominance, breeding values and genotypic values are the same.
- Alleles involved with genotypes that show no dominance (d=0) are sometimes called ‘additive genes’, or are said to ‘act additively’.

Breeding Values and Dominance Deviations

Dominance Deviations



The algebra of breeding values and dominance deviations provide the theoretical basis for subdividing the G component of the Phenotypic model

$$P = G + E)(= \alpha_i + \alpha_j + \delta_{ij} + E$$

Thus, based on the algebra and substituting data from Table 7

$$\delta_{AA} = G_{AA} - \mu - \alpha_A - \alpha_A = a - \mu - 2(pa + qd - \mu) = -2q^2d)(= -2(0.2)2(4) = -0.32$$

Invalid Equation

Invalid Equation

Epistasis

Epistasis exists when genotypes at two or more loci result in a genotypic value that is greater or less than the sum of the average genotypic effects at each of the individual loci. For example,

Table 8 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**		
	BB	**Bb**	**bb**
AA	22	18	6
Aa	20	16	4
aa	14	10	-2

Table 9 Two-locus genotypic values that exhibit epistasis. The total genotypic value is NOT equal to the sum of the genotypic values at the loci ($G_A + G_B$).

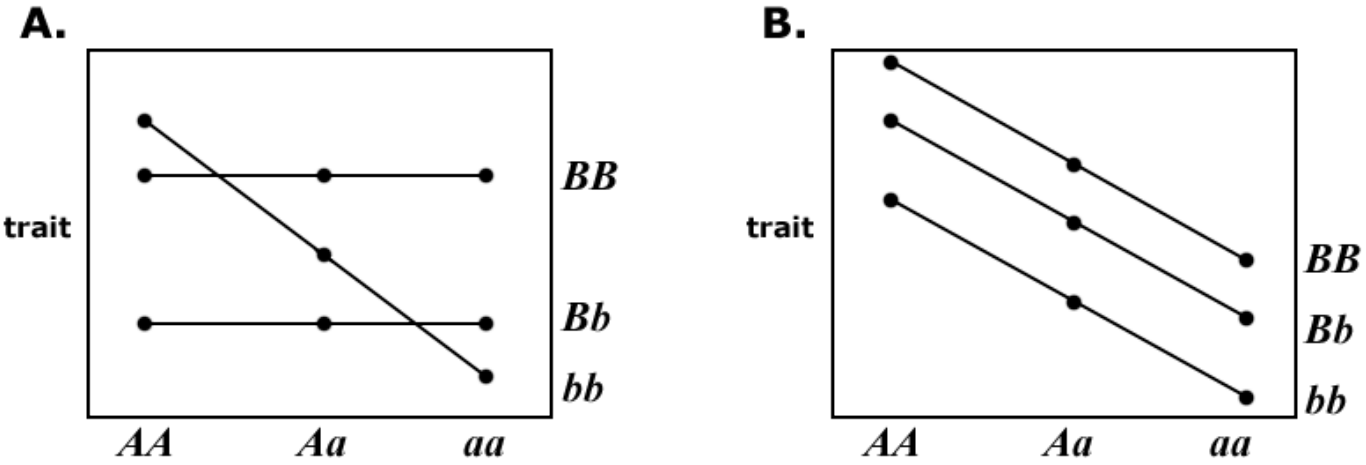
Genotype at Locus A	**Genotype at Locus B**		
	BB	**Bb**	**bb**
AA	24	18	6
Aa	20	16	4
aa	14	10	-2

Graphical View of Epistasis

Epistasis between loci within an individual can be represented as the reaction norm of different genotypes at one locus, plotted against the genotypes at a second locus.

Figure A: Epistasis between loci occurs because the reaction norms of the locus B genotypes differ in slope.

Figure B: the reaction norms are parallel, thus the effects of the two loci are independent and no epistasis is present.



Physiological and Statistical Dominance

Cheverud and Routman (1995) identified two concepts associated with the term epistasis: **physiological epistasis and statistical epistasis.** Genetics 139:1455.

The distinction between these two concepts is similar to that made between physiological and statistical dominance:

Physiological Dominance:

- Heterozygote is not midway between two homozygotes
- Values of a and d are not dependent on allele frequencies
- When $d \neq 0$ it reflects intralocus interaction is present
- Least squares solution of the unweighted regression of number of genotypic value on number of “a” alleles
- Physiological dominance contributes to both additive and dominance values and variances

Statistical Dominance Deviations:

- Deviations of single-locus from the additive combination of alleles contribute to the genotype
- Depend on allele frequencies and will change with changes in allele frequencies
- Least squares solution of a weighted (weighted by genotypic frequencies) regression of genotypic value on number of alleles

Physiological Epistasis

In physiological epistasis (or mechanistic epistasis):

- Interaction effects occur “within” genotypes, where genes expressed within a single genome interact.
- Simply recognizes that certain genotypes at two or more loci interact in the production of a phenotype.
- If all possible genotypic classes are equally frequent in a population the influence of genetic interactions on phenotypes will be directly observable.
- The contribution of physiological epistasis to populations is a function of the frequencies of interacting genotypes in a population.

Model for Physiological Epistasis

Let the phenotypic value of an individual be determined by the combination of the alleles present at two loci. This model is used to illustrate how physiologically based gene interactions map to components of genetic variation.

Consider the two loci each with two alleles per locus, the two locus (physiological) genotypic values, G_{ijkl}, are the average phenotype of individuals with the ij^{th} genotype at the first locus and the kl^{th} genotype at the second locus.

Table 10 Example of Epistasis using Genotypic Values.

	AA	**Aa**	**aa**	**Unweighted marginal mean**
BB	G_{AABB}=22	G_{AaBB}=18	G_{aaBB}=6	G_{BB}=15.33
Bb	G_{AABb}=20	G_{AaBb}=16	G_{aaBb}=4	G_{bb}=13.33
bb	G_{AAbb}=14	$G_{Aabb=10}$	G_{aabb}=-2	G_{bb}=7.33
Unweighted marginal means	G_{AA}=18.66	G_{Aa}=14.66	G_{aa}=2.66	G=12

Single Locus Genotype

The single locus genotype is defined as the unweighted average across the genotypes at the second locus:

Where at locus A,

$$G_{ij} = (G_{ij11} + G_{ij12} + G_{ij22})/3)(G_{AA..} = \frac{(22 + 20 + 14)}{3} = 18.66$$

and at locus B,

$$G_{..kt} = (G_{11kt} + G_{12kt} + G_{22kt})/3)(G_{..BB} = \frac{(22 + 18 + 6)}{3} = 15.33$$

Non-Epistatic Genotypic Value

Subscripts *A* and *a* or *B* and *b* refer to the two alleles at the interacting loci. The single locus values of *a* and *d* are:

Invalid Equation

The non-epistatic genotypic value is:

$$ne_{ijkl} = G_{ij..} + G_{..kl} - G_{....})(ne_{ijkl} = G_{AA..} + G_{..BB} - G....)(ne_{ijkl} = 18.66 + 15.33 - 12 = 21.99$$

Epistatic Genotypic Value

Table 11 Non-epistatic values.

	AA	**Aa**	**aa**
BB	21.99	17.99	5.99
Bb	19.99	15.99	3.99
bb	13.99	9.99	-2.01

Epistatic genotypic value:

$$e_{ijkl} = G_{ijkl} - ne_{ijkl})(e_{ijkl} = 22 - 21.99 = 0.01$$

e_{ijkl} : A value different from zero indicates that PHYSIOLOGICAL EPISTASIS is present. There is little evidence for epistasis for this cell. Is there evidence for epistasis in the other cells?

Statistical Epistasis

In statistical epistasis (or population epistasis):

- The term is used to refer to the amount of population variation in genotypic values associated with variation among loci.
- Notation that is often used includes V_I or $G_{A\times B}$ or I_{AB}
- The amount of statistical epistasis present in a population is a function of the frequencies of interacting multilocus genotypes and therefore is a function of population allele frequencies as it is for additive and dominance variance (V_A and V_D).

$$G_T = G_A + G_B + G_{A\times B}$$

Presence of epistasis between locus A and B changes the population mean (*M*), mid-homozygote value (μ), a (additive) and d (dominance) values.

Table 12 An example with Epistatic effects.

Two-locus genotypic values and frequencies		A locus genotype		
		AA	Aa	aa
B locus genotype	Coded Genotypic Value/Freq.	8 0.64	4 0.32	-8 0.04
BB	4 0.04	24 0.0256	18 0.0128	6 0.0016
Bb	2 0.32	20 0.2048	16 0.1024	4 0.0128
bb	-4 0.64	14 0.4096	10 0.2048	-2 0.0256
Average at locus A		16.32	12.24	0.24
Average at locus B		21.36	18.08	12.08

Epistasis Effects

Genotypic value of *AABB* has been increased from 22 to 24 due to epistatic effects.

There are changes in population mean, mid-homozygote values for *A* and *B* locus, and the average at *A* and *B* locus.

$$\text{Epistatic effects between } AABB = 2)(G_T = \mu + G_A + G_B = 10 + 8 + 4 = 22)(\text{With epistatic effects } G_T = \mu + G_A + G_B + G_{A\times B} = 10 + 8 + 4 + 2 = 24$$
$$\text{Changes in the mid-homozygote } (\mu) = 24 + (-2)/2 = 11$$
$$\text{Changes in the mean population } (Y) = 14.37)(\mu_A = 8.28)(\mu_B = 16.72)(G_A : +a = 8.04; d = 3.96; -a = -8.04)(G_B : +a = 4.64; d = 1.36; -a = -4.64)$$

Acknowledgements

This module was developed as part of the Bill & Melinda Gates Foundation Contract No. 24576 for Plant Breeding E-Learning in Africa.

Quantitative Genetics Gene Effects Author: William Beavis, Kendall Lamkey, and Katherine Espinosa (ISU)

Multimedia Developers: Gretchen Anderson, Todd Hartnell, and Andy Rohrback (ISU)

How to cite this module: Beavis, W., K. Lamkey, and K. Espinosa. 2016. Gene Effects. *In* Quantitative Genetics, interactive e-learning courseware. Plant Breeding E-Learning in Africa. Retrieved from https://pbea.agron.iastate.edu.