

# Lecture 4: Allelic Effects and Genetic Variances

Bruce Walsh lecture notes  
Tucson Winter Institute  
7 - 9 Jan 2013

# Basic model of Quantitative Genetics

Phenotypic value -- we will occasionally  
also use  $z$  for this value

Basic model:  $P = G + E$  ← Environmental value

↑  
Genotypic value

$G$  = average phenotypic value for that genotype  
if we are able to replicate it over the **universe**  
of environmental values,  $G = E[P]$

Hence, genotypic values are **functions of the  
environments experienced.**

# Basic model of Quantitative Genetics

$$\text{Basic model: } P = G + E$$

$G$  = average phenotypic value for that genotype if we are able to replicate it over the **universe** of environmental values,  $G = E[P]$

$G$  = average value of an inbred line over a series of environments

**$G \times E$  interaction** --- The performance of a particular genotype in a particular environment differs from the sum of the average performance of that genotype over all environments and the average performance of that environment over all genotypes.

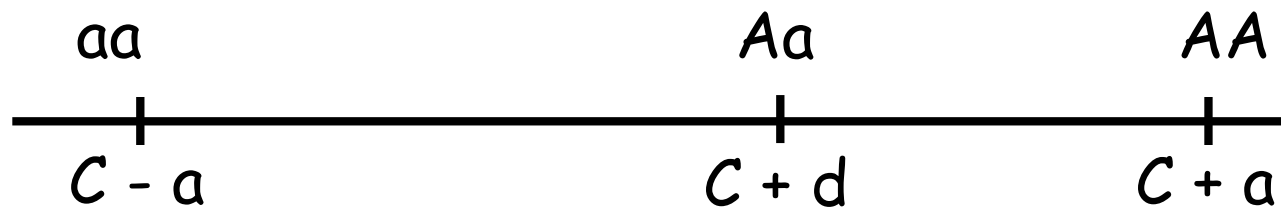
Basic model now becomes  **$P = G + E + GE$**

# The transmission of genotypes versus alleles

- With fully inbred lines, offspring have the same genotype as their parent, and hence the entire parental genotypic value  $G$  is passed along
  - Hence, favorable interactions between alleles (such as with dominance) are not lost by randomization under random mating but rather passed along.
- When offspring are generated by crossing (or random mating), each parent contributes a **single allele** at each locus to its offspring, and hence **only passes along a PART** of its genotypic value
- This part is determined by the **average effect of the allele**
  - Downside is that favorable interaction between alleles are NOT passed along to their offspring in a diploid (but, as we will see, are in an autoteraploid)

# Genotypic values

It will prove very useful to decompose the genotypic value into the difference between homozygotes ( $2a$ ) and a measure of dominance ( $d$  or  $k = d/a$ )



Note that the constant  $C$  is the average value of the two homozygotes.

If no dominance,  $d = 0$ , as heterozygote value equals the average of the two parents. Can also write  $d = ka$ , so that  $G(Aa) = C + a(1 + k)$

# Computing a and d

Suppose a major locus influences plant height, with the following values

Genotype	aa	Aa	AA
Trait value	10	15	16

$$C = [G(AA) + G(aa)]/2 = (16+10)/2 = 13$$

$$a = [G(AA) - G(aa)]/2 = (16-10)/2 = 3$$

$$d = G(Aa) - [G(AA) + G(aa)]/2 \\ = G(Aa) - C = 15 - 13 = 2$$

# Population means: Random mating

Let  $p = \text{freq}(A)$ ,  $q = 1-p = \text{freq}(a)$ . Assuming random-mating (Hardy-Weinberg frequencies),

Genotype	aa	Aa	AA
Value	$C - a$	$C + d$	$C + a$
Frequency	$q^2$	$2pq$	$p^2$

$$\text{Mean} = q^2(C - a) + 2pq(C + d) + p^2(C + a)$$

$$\mu_{\text{RM}} = C + a(q-p) + d(2pq)$$

Contribution from  
homozygotes

Contribution from  
heterozygotes

# Population means: Inbred cross $F_2$

Suppose two inbred lines are crossed. If  $A$  is fixed in one population and  $a$  in the other, then  $p = q = 1/2$

Genotype	$aa$	$Aa$	$AA$
Value	$C - a$	$C + d$	$C + a$
Frequency	$1/4$	$1/2$	$1/4$

$$\text{Mean} = (1/4)(C - a) + (1/2)(C + d) + (1/4)(C + a)$$

$$\mu_{RM} = C + d/2$$

Note that  $C$  is the average of the two parental lines, so when  $d > 0$ ,  $F_2$  exceeds this. Note also that the  $F_1$  exceeds this average by  $d$ , so only half of this passed onto  $F_2$ .



# Population means: RILs from an $F_2$

A large number of  $F_2$  individuals are fully inbred, either by selfing for many generations or by generating doubled haploids. If  $p$  and  $q$  denote the  $F_2$  frequencies of  $A$  and  $a$ , what is the expected mean over the set of resulting RILs?

Genotype	aa	Aa	AA
Value	$C - a$	$C + d$	$C + a$
Frequency	$q$	0	$p$

$$\mu_{RILs} = C + a(p-q)$$

Note this is independent of the amount of dominance ( $d$ )

# The average effect of an allele

- The average effect  $\alpha_A$  of an allele  $A$  is defined by the difference between offspring that get that allele and a random offspring.
  - $\alpha_A = \text{mean}(\text{offspring value given parent transmits } A) - \text{mean}(\text{all offspring})$
  - Similar definition for  $\alpha_a$ .
- Note that while  $C$ ,  $a$  and  $d$  (the genotypic parameters) do not change with allele frequency,  $\alpha_x$  is clearly a function of the frequencies of alleles with which allele  $x$  combines.

# Random mating

Consider the average effect of allele  $A$  when a parent is randomly-mated to another individual from its population

Suppose parent contributes  $A$

Allele from other parent	Probability	Genotype	Value
$A$	$p$	$AA$	$C + a$
$a$	$q$	$Aa$	$C + d$

$$\text{Mean}(A \text{ transmitted}) = p(C + a) + q(C + d) = C + pa + qd$$

$$\alpha_A = \text{Mean}(A \text{ transmitted}) - \mu = q[a + d(q-p)]$$

# Random mating

Now suppose parent contributes a

Allele from other parent	Probability	Genotype	Value
A	p	Aa	C + d
a	q	aa	C - a

$$\text{Mean}(a \text{ transmitted}) = p(C + d) + q(C - a) = C - qa + pd$$

$$\alpha_a = \text{Mean}(a \text{ transmitted}) - \mu = -p[a + d(q-p)]$$

# $\alpha$ , the average effect of an allelic substitution

- $\alpha = \alpha_A - \alpha_a$  is the average effect of an allelic substitution, the change in mean trait value when an  $a$  allele in a random individual is replaced by an  $A$  allele
  - $\alpha = a + d(q-p)$ . Note that
    - $\alpha_A = q\alpha$  and  $\alpha_a = -p\alpha$ .
    - $E(\alpha_X) = p\alpha_A + q\alpha_a = pq\alpha - qp\alpha = 0$ ,
    - The average effect of a random allele is zero, hence average effects are deviations from the mean

# Dominance deviations

- Fisher (1918) decomposed the contribution to the genotypic value from a single locus as  $G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij}$ 
  - Here,  $\mu$  is the mean (a function of  $p$ )
  - $\alpha_i$  are the average effects
  - Hence,  $\mu + \alpha_i + \alpha_j$  is the **predicted genotypic value** given the average effect (over all genotypes) of alleles  $i$  and  $j$ .
  - The **dominance deviation** associated with genotype  $G_{ij}$  is the difference between its true value and its value predicted from the sum of average effects (essentially a residual)

# Fisher's (1918) Decomposition of $G$

One of Fisher's key insights was that the genotypic value consists of a **fraction that can be passed from parent to offspring** and a **fraction that cannot**.

In particular, under sexual reproduction, parents only pass along **SINGLE ALLELES** to their offspring

Consider the genotypic value  $G_{ij}$  resulting from an  $A_i A_j$  individual

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Average contribution to genotypic value for allele  $i$

Mean value  $\mu_G = \sum G_{ij} \text{Freq}(A_i A_j)$

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Since parents pass along single alleles to their offspring, the  $\alpha_i$  (the **average effect** of allele  $i$ ) represent these contributions

The average effect for an allele is **POPULATION-SPECIFIC**, as it depends on the types and frequencies of alleles that it pairs with

The genotypic value predicted from the individual allelic effects is thus  $\hat{G}_{ij} = \mu_G + \alpha_i + \alpha_j$

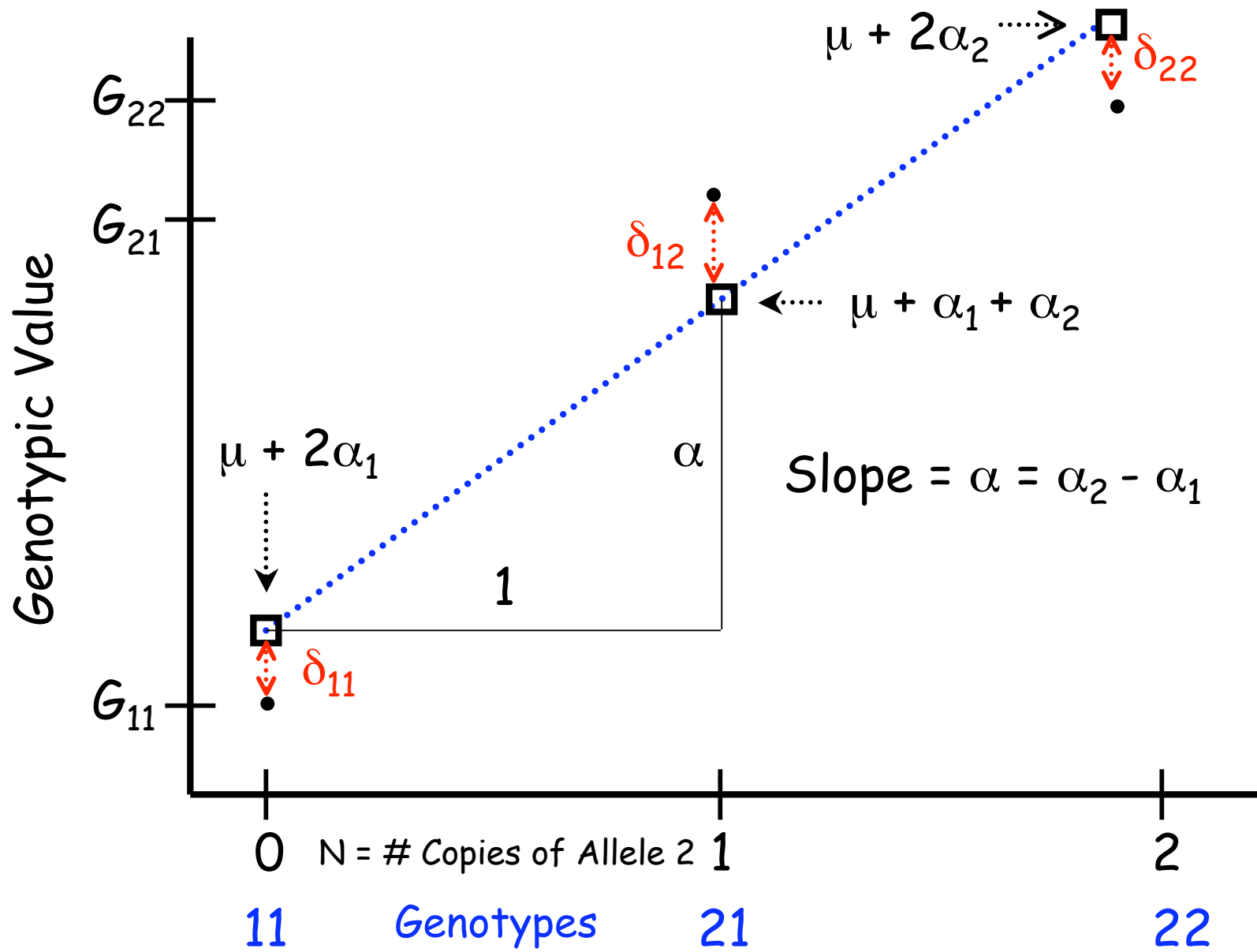


$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

The genotypic value predicted from the individual allelic effects is thus  $\hat{G}_{ij} = \mu_G + \alpha_i + \alpha_j$

**Dominance deviations** --- the difference (for genotype  $A_iA_j$ ) between the genotypic value predicted from the two single alleles and the actual genotypic value,

$$G_{ij} - \hat{G}_{ij} = \delta_{ij}$$



# Fisher's decomposition is a Regression

$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Predicted value

Residual error

A notational change clearly shows this is a regression,

$$G_{ij} = \mu_G + 2\alpha_1 + (\alpha_2 - \alpha_1)N + \delta_{ij}$$

Independent (predictor) variable  $N = \#$  of  $A_2$  alleles

Note that the slope  $\alpha_2 - \alpha_1 = \alpha$ , the average effect of an allelic substitution

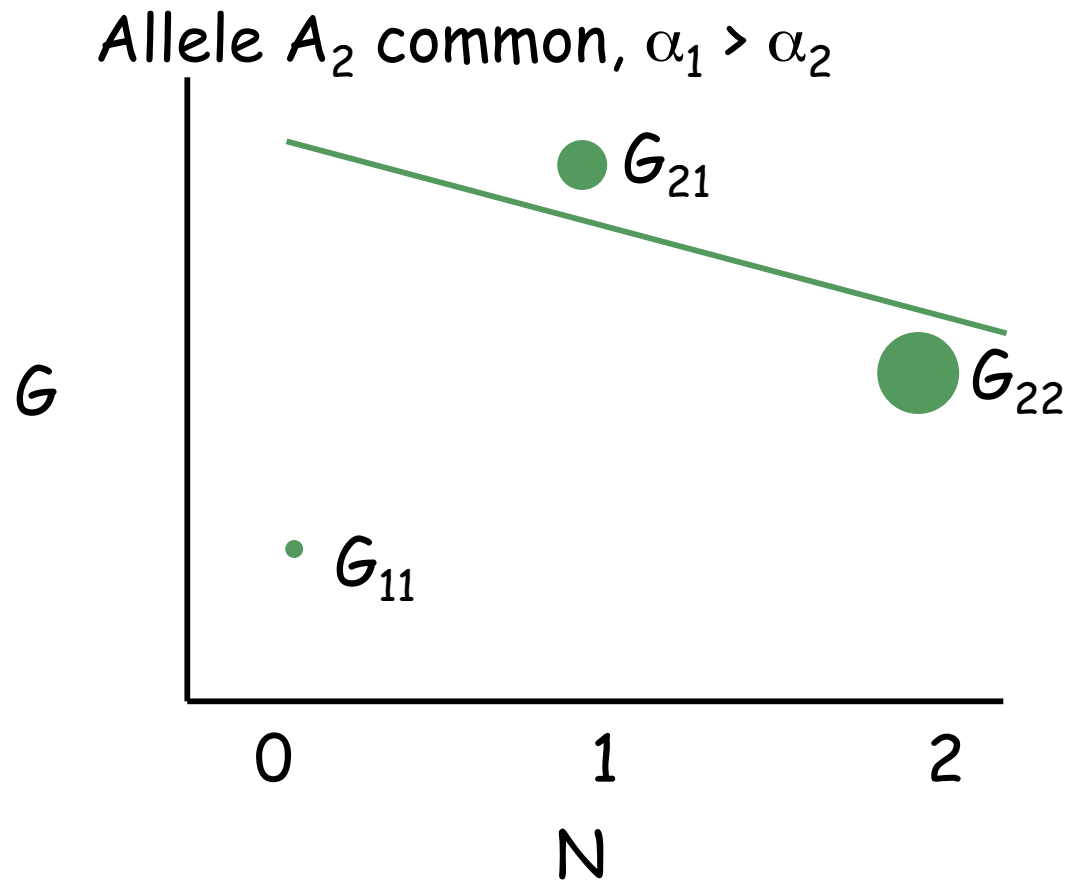
$$G_{ij} = \mu_G + 2\alpha_1 + (\alpha_2 - \alpha_1)N + \delta_{ij}$$

Intercept

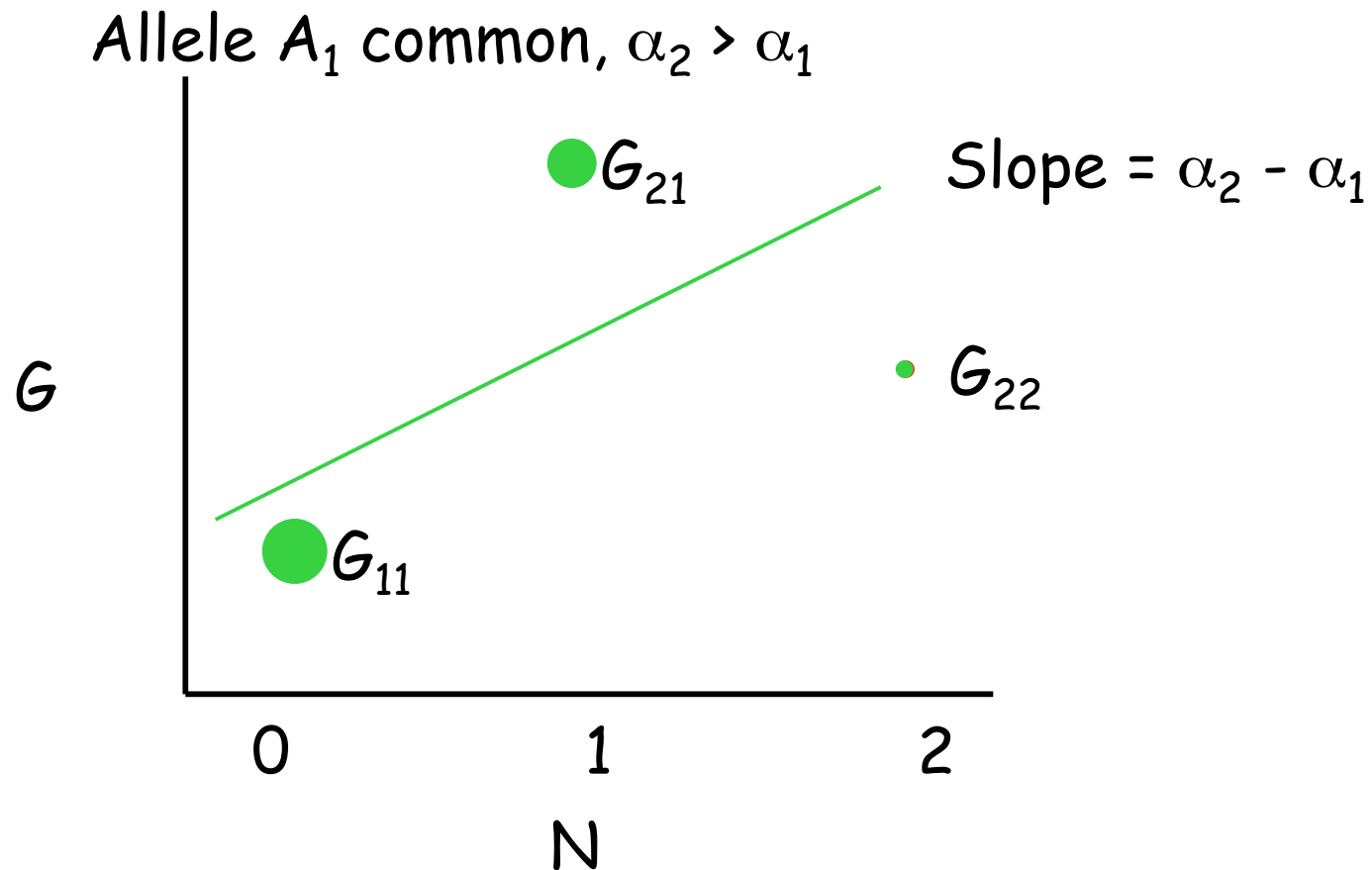
Regression slope

$$2\alpha_1 + (\alpha_2 - \alpha_1)N = \begin{cases} 2\alpha_1 & \text{for } N = 0, \text{ e.g., } A_1A_1 \\ \alpha_1 + \alpha_2 & \text{for } N = 1, \text{ e.g., } A_1A_2 \\ 2\alpha_2 & \text{for } N = 2, \text{ e.g., } A_2A_2 \end{cases}$$

A key point is that the average effects change with allele frequencies. Indeed, if overdominance is present they can change sign with allele frequencies.

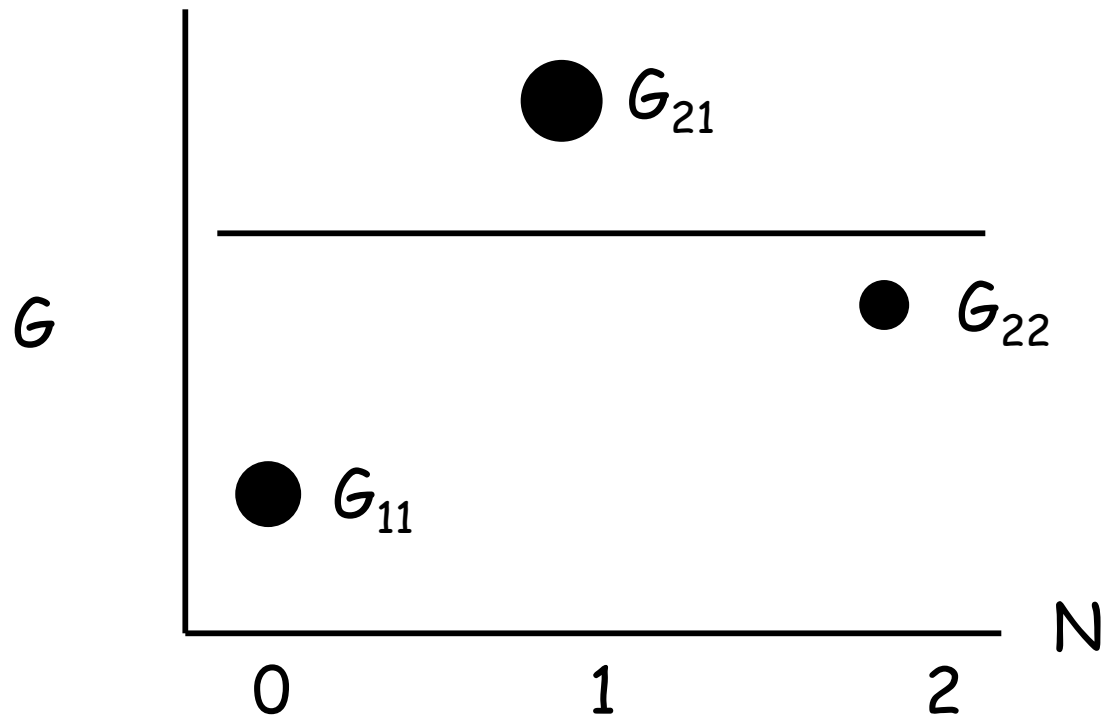


The size of the circle denotes the weight associated with that genotype. While the genotypic **values** do not change, their frequencies (and hence weights) do.



Again, same genotypic values as previous slide, but different weights, and hence a different slope (here a change in sign!)

Both  $A_1$  and  $A_2$  frequent,  $\alpha_1 = \alpha_2 = 0$



With these allele frequencies, both alleles have the same mean value when transmitted, so that all parents have the same average offspring value -- no response to selection

# Average Effects and Additive Genetic Values

The  $\alpha$  values are the **average effects** of an allele

A key concept is the **Additive Genetic Value (A)** of an individual

$$A(G_{ij}) = \alpha_i + \alpha_j$$

$$A = \sum_{k=1}^n \left( \alpha_i^{(k)} + \alpha_k^{(k)} \right)$$

$\alpha_i^{(k)}$  = effect of allele  $i$  at locus  $k$

$A$  is called the **Breeding value** or the **Additive genetic value**



$$A = \sum_{k=1}^n \left( \alpha_i^{(k)} + \alpha_k^{(k)} \right)$$

Why all the fuss over  $A$ ?

Suppose pollen parent has  $A = 10$  and seed parent has  $A = -2$  for plant height

Expected average offspring height is  $(10-2)/2 = 4$  units above the population mean. Offspring  $A =$  average of parental  $A$ 's

**KEY:** parents only pass single alleles to their offspring. Hence, they only pass along the  $A$  part of their genotypic value  $G$

# Genetic Variances

Writing the genotypic value as

$$G_{ij} = \mu_g + (\alpha_i + \alpha_j) + \delta_{ij}$$

The genetic variance can be written as

$$\sigma^2(G) = \sum_{k=1}^n \sigma^2(\alpha_i^{(k)} + \alpha_j^{(k)}) + \sum_{k=1}^n \sigma^2(\delta_{ij}^{(k)})$$

This follows since

$$\sigma^2(G) = \sigma^2(\mu_g + (\alpha_i + \alpha_j) + \delta_{ij}) = \sigma^2(\alpha_i + \alpha_j) + \sigma^2(\delta_{ij})$$

$$\text{As Cov}(\alpha, \delta) = 0$$

# Genetic Variances

$$\sigma^2(G) = \sum_{k=1}^n \sigma^2(\alpha_i^{(k)} + \alpha_j^{(k)}) + \sum_{k=1}^n \sigma^2(\delta_{ij}^{(k)})$$

Additive Genetic Variance  
(or simply Additive Variance)

Dominance Genetic Variance  
(or simply dominance variance)

Hence, total genetic variance = additive + dominance variances,

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2$$

# Key concepts (so far)

- $\alpha_i$  = average effect of allele  $i$ 
  - Property of a single allele in a particular population (depends on genetic background)
- $A$  = Additive Genetic Value ( $A$ )
  - $A$  = sum (over all loci) of average effects
  - Fraction of  $G$  that parents pass along to their offspring
  - Property of an Individual in a particular population
- $\text{Var}(A)$  = additive genetic variance
  - Variance in additive genetic values
  - Property of a population
- Can estimate  $A$  or  $\text{Var}(A)$  without knowing any of the underlying genetical detail (forthcoming)

$$\sigma_A^2 = 2E[\alpha^2] = 2 \sum_{i=1}^m \alpha_i^2 p_i$$

Q <sub>1</sub> Q <sub>1</sub>	Q <sub>1</sub> Q <sub>2</sub>	Q <sub>2</sub> Q <sub>2</sub>
0	a(1+k)	2a

Since  $E[\alpha] = 0$ ,

$$\text{Var}(\alpha) = E[(\alpha - \mu_a)^2] = E[\alpha^2]$$

One locus, 2 alleles:  $\sigma_A^2 = 2p_1 p_2 a^2 [1 + k (p_1 - p_2)]^2$

↑  
⋮  
Dominance alters  
additive variance

When dominance present, Additive variance is an asymmetric function of allele frequencies

Dominance variance

$Q_1Q_1$	$Q_1Q_2$	$Q_2Q_2$
0	$a(1+k)$	$2a$

$$\sigma_D^2 = E[\delta^2] = \sum_{i=1}^m \sum_{j=1}^m \delta_{ij}^2 p_i p_j$$

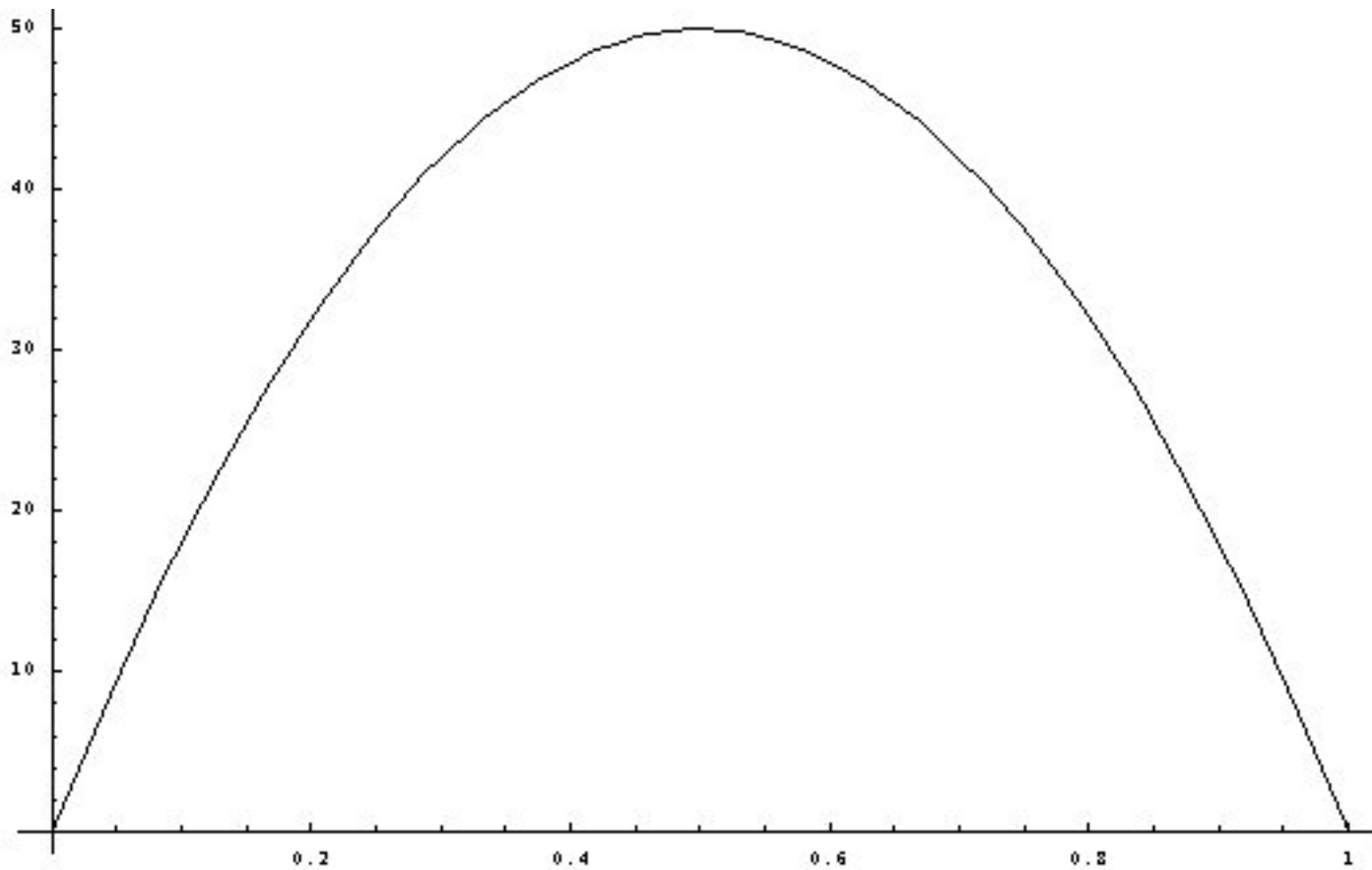
Equals zero if  $k = 0$

One locus, 2 alleles:  $\sigma_D^2 = (2p_1 p_2 a k)^2$

This is a symmetric function of allele frequencies

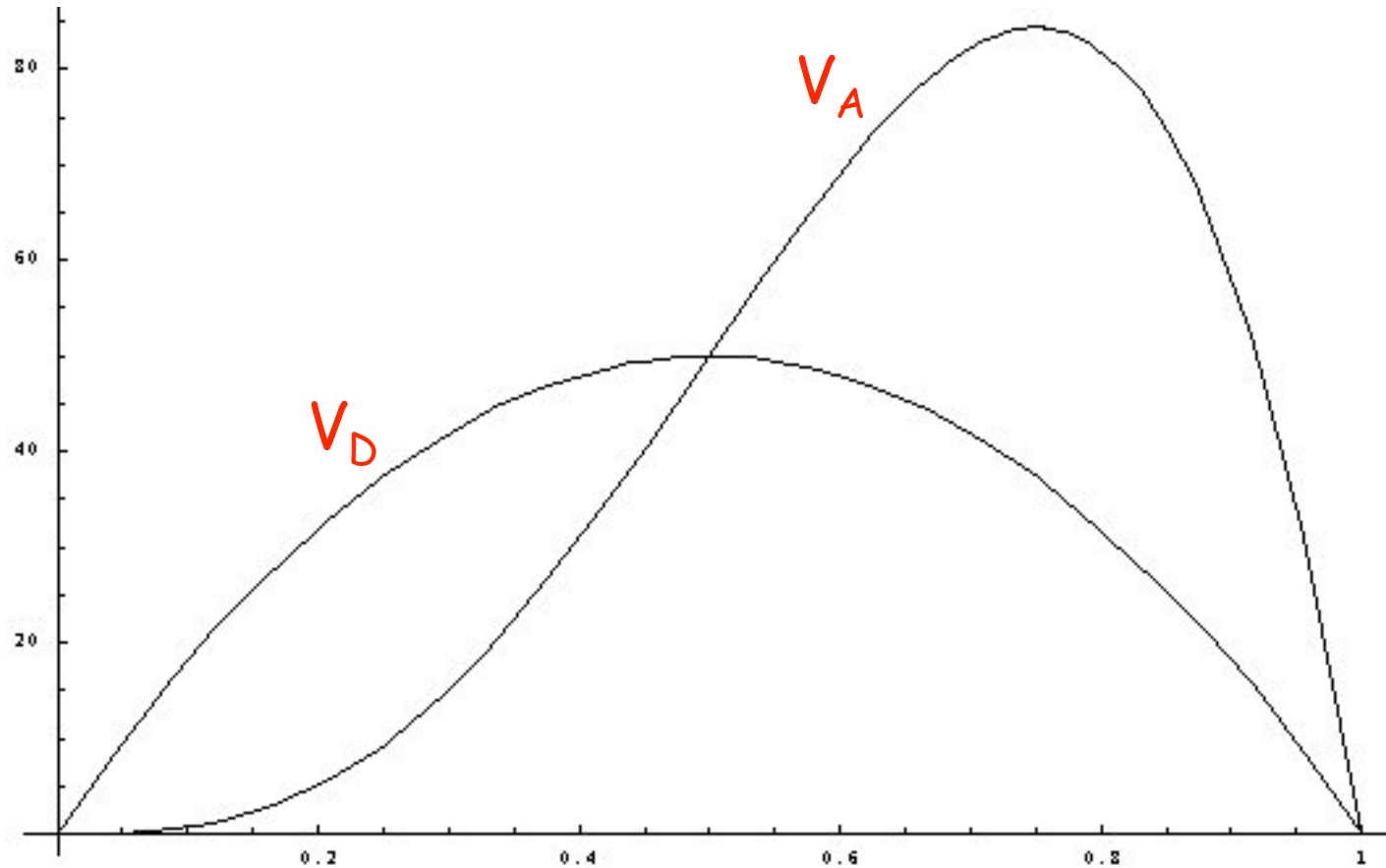
Can also be expressed in terms of  $d = ak$

Additive variance,  $V_A$ , with no dominance ( $k = 0$ )



Allele frequency,  $p$

## Complete dominance ( $k = 1$ )



Allele frequency,  $p$



# Epistasis

$$\begin{aligned}G_{ijkl} &= \mu_G + (\alpha_i + \alpha_j + \alpha_k + \alpha_l) + (\delta_{ij} + \delta_{kj}) \\ &\quad + (\alpha\alpha_{ik} + \alpha\alpha_{il} + \alpha\alpha_{jk} + \alpha\alpha_{jl}) \\ &\quad + (\alpha\delta_{ikl} + \alpha\delta_{jkl} + \alpha\delta_{kij} + \alpha\delta_{lij}) \\ &\quad + (\delta\delta_{ijkl}) \\ &= \mu_G + A + D + AA + AD + DD\end{aligned}$$

These components are defined to be uncorrelated, (or **orthogonal**), so that

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

$$\begin{aligned}
G_{ijkl} &= \mu_G + (\alpha_i + \alpha_j + \alpha_k + \alpha_l) + (\delta_{ij} + \delta_{kj}) \\
&\quad + (\alpha\alpha_{ik} + \alpha\alpha_{il} + \alpha\alpha_{jk} + \alpha\alpha_{jl}) \\
&\quad + (\alpha\delta_{ikl} + \alpha\delta_{jkl} + \alpha\delta_{kij} + \alpha\delta_{lij}) \\
&\quad + (\delta\delta_{ijkl}) \\
&= \mu_G + A + D + AA + AD + DD
\end{aligned}$$

**Additive x Additive** interactions --  $\alpha\alpha$ , AA  
interactions between a single allele  
at one locus with a single allele at another

**Additive x Dominance** interactions --  $\alpha\delta$ , AD  
interactions between an allele at one  
locus with the genotype at another, e.g.  
allele  $A_i$  and genotype  $B_{kj}$

**Dominance x dominance** interaction ---  $\delta\delta$ , DD  
the interaction between the dominance  
deviation at one locus with the dominance  
deviation at another.

# Effects and Variance when using a testor

- A common design in plant breeding is to cross members from a population to a **testor** to generate a **testcross**.
  - Testor can be either an inbred or an outcrossing population
  - Often from a different heteroic group from the population being tested
  - Often testor is an elite genotype
- The average effect of an allele in a testcross, its variance, and its additive (**General combining ability, GCA**) and interaction (**Specific combining ability, SCA**) effects all follow in analogous fashion to previous results for crosses within a population

## The average effect of an allele in a testcross

- The concept of the average effect of an allele when crossed within its population is easily extended to the average effect of an allele when crossed to a testor.
  - Called the **testcross average effect**.
- The average effect of allele X in this testcross,  $\alpha_x^T$ , is defined as difference between the mean value of offspring getting this allele from the population versus the mean value of a random offspring from this cross
  - Will turn out to be a function of the frequencies of alleles in both the tested and the testor population.

# Mean value for a testcross

Suppose the frequency of  $A$  is  $p$  in the population and  $p^T$  in the testor (with  $q$  and  $q^T$  similarly defined for  $a$ ).

		testor	
		$A (p^T)$	$a (q^T)$
Parental line	$A (p)$	$pp^T$ $C + a$	$pq^T$ $C + d$
	$a (q)$	$qp^T$ $C + d$	$qq^T$ $C - a$

$$\text{Mean of cross} = C + a(pp^T - qq^T) + d(pq^T + qp^T)$$

# Average testcross mean in a series of RILs

- Slide 9 gave an expression for the expected average performance from a series of RILs formed by crossing two populations.
- A similar expression exists for the average testcross performance for a series of RILs from a cross of A x B
  - Mean =  $(1/2) \mu_A^T + (1/2) \mu_B^T$ , namely the average of the testcross means for A and B
  - More generally (since lines can, by chance, given an equal contribution of alleles),
  - Mean =  $\pi_A \mu_A^T + \pi_B \mu_B^T$ , where  $\pi_A = (1 - \pi_B)$  is the fraction of alleles from A in the sample of RILs
  - Can use molecular markers to estimate the  $\pi_x$  directly.

# $\alpha_A^T$ , testcross effect of allele A

Suppose parent contributes A

Allele from other parent	Probability	Genotype	Value
A	$p^T$	AA	$C + a$
a	$q^T$	Aa	$C + d$

$$\text{Mean}(A \text{ transmitted}) = p^T(C + a) + q^T(C + d) = C + p^T a + q^T d$$

$$\alpha_A^T = \text{Mean}(A \text{ transmitted}) - \mu = q[a + d(q^T - p^T)]$$

Likewise,

$$\alpha_a^T = \text{Mean}(a \text{ transmitted}) - \mu = -p[a + d(q^T - p^T)]$$

# $\alpha^T$ , the average testcross effect of an allelic substitution

- $\alpha^T = \alpha_A^T - \alpha_a^T$  is the average testcross effect of an allelic substitution, the change in mean trait value when an  $a$  allele in a random testcrossed individual is replaced by an  $A$  allele
  - $\alpha^T = a + d(q^T - p^T)$ . Note that this is independent of the allele frequencies in the parental population, and depends ONLY on the testor allele frequencies.
    - $\alpha_A^T = q\alpha^T$ ,  $\alpha_a^T = -p\alpha^T$ , and  $E(\alpha_x^T) = 0$



# Testcross variance

- Just as the additive genetic variance was the population variance in the sum of the average effects of an allele, the testcross variance is variance in the average testcross effects of a random allele
  - $\text{Var}(A^T) = \text{Var}(\alpha_x^T) = \text{Var}(\alpha_x^T)$
  - $\text{Var}(\alpha_x^T) = p (\alpha_A^T)^2 + q (\alpha_a^T)^2 =$
  - $p(q[a + d(q^T-p^T)])^2 + q(-p[a + d(q^T-p^T)])^2$
  - $= pq[a + d(q^T-p^T)]^2$
  - Hence,  $\text{Var}(\alpha_x^T) = pq[a + d(q^T-p^T)]^2$

# GCS and SCA

- Consider a cross between individuals from population 1 and population 2
- Let  $\mu_{1 \times 2}$  denote the average value for all of these crosses, and let  $G_{ij}$  be the average genotypic value of an individual from a cross from individual  $i$  (or line) in population one and individual  $j$  (or line) from population two.
- Analogous to Fisher's decomposition, we can write this in terms of two additive effects and one interaction effect.

$$G_{ij} = \mu_{1 \times 2} + \alpha_i^2 + \alpha_j^1 + \delta_{ij}^{12}$$

$\alpha_i^2$  is the testcross average effect for allele  $i$  (more generally an allele from individual  $i$ ) when tested using population 2 as a testor, with  $\alpha_j^1$  similarly defined for allele  $j$  (from pop 2) using one as the testor

$\delta_{ij}^{12}$  is the interaction between allele  $i$  from and allele  $j$  in the testcross of 1 and 2

The sum over all loci of the  $\alpha_i^2$  values is the **general combining ability (GCA)** of line  $i$  when crossed to line 2 (note these are cross-specific)!

The sum of the  $\delta$  is the **specific combining ability (SCA)**

$$G_{ij} = \mu + GCA_i^2 + GCA_j^1 + SCA_{ij}^{12}$$

The superscripts denoting the population in which the allele is being tested is often suppressed

The *GCA* is akin to the breeding value from one parent, but now it is the testcross value of that parent

The predicted mean of a particular cross is the sum of the two *GCA*s for those individuals/lines

As with average effects and dominance deviations, these are only defined with respect to a particular reference set of crosses (I.e., lines from Pop 1 X lines from pop 2)

## Within-population crosses vs. testors

	Within-pop	testor
Allelic effects	$\alpha$	$\alpha^T$
Additive transmitting factor	Breeding value $A$	$GCA$
Predicting offspring mean	$A_1/2 + A_2/2$	$GCA_1 + GCA_2$
Nonadditive component	Dominance value	$SCA$
Genetic Variances	$Var(A),$ $Var(D)$	$Var(GCA),$ $Var(SCA)$