

Lecture 2: Introduction to Quantitative Genetics

Bruce Walsh lecture notes
Introduction to Quantitative Genetics
SISG, Seattle
16 – 18 July 2018

1

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.**

2

Basic model of Quantitative Genetics

Basic model: $P = G + E$

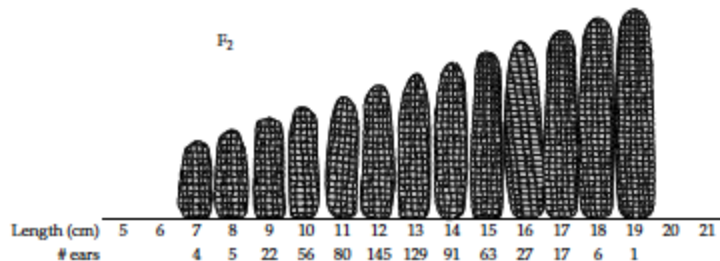
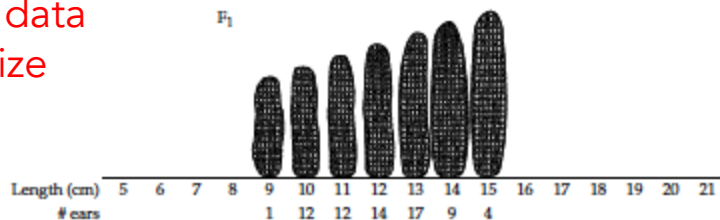
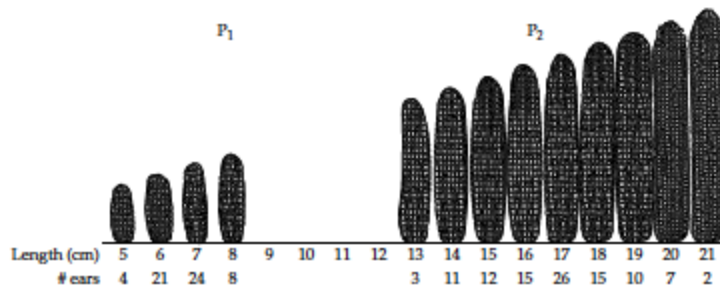
G = average phenotypic value for that genotype if we are able to replicate it over the **universe** of environmental values, $G = E[P]$

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

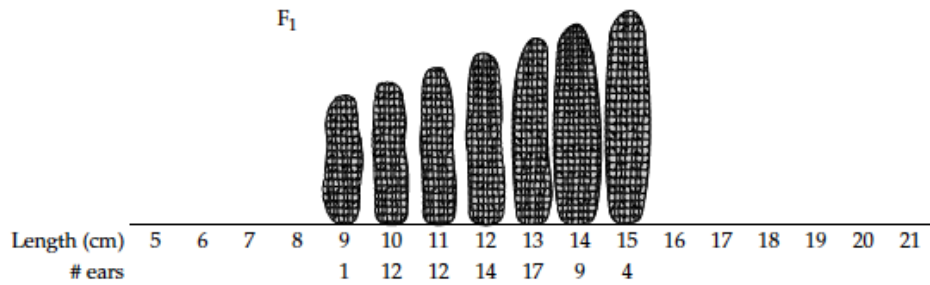
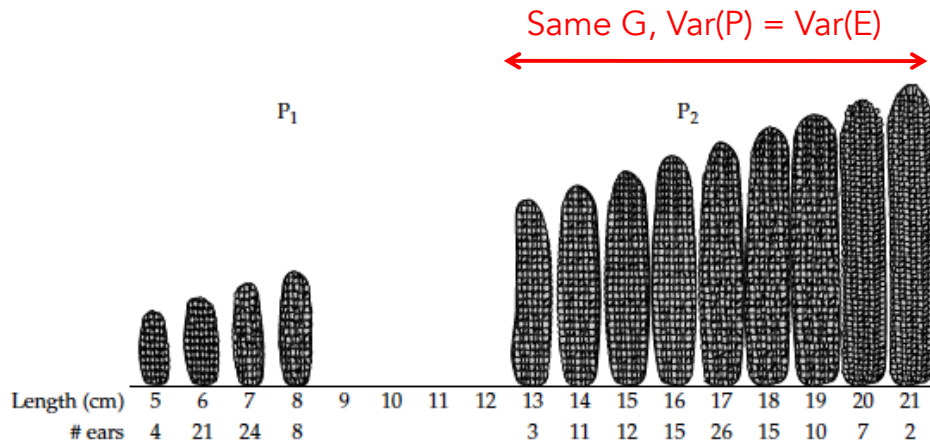
G x E interaction --- The performance of a particular genotype in a particular environment differs from the sum of the average performance of that genotype over all environments and the average performance of that environment over all genotypes. Basic model now becomes $P = G + E + GE$

3

East (1911) data
on US maize
crosses

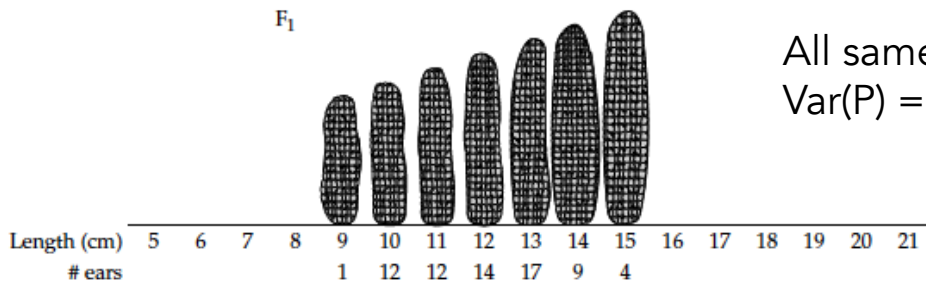


4

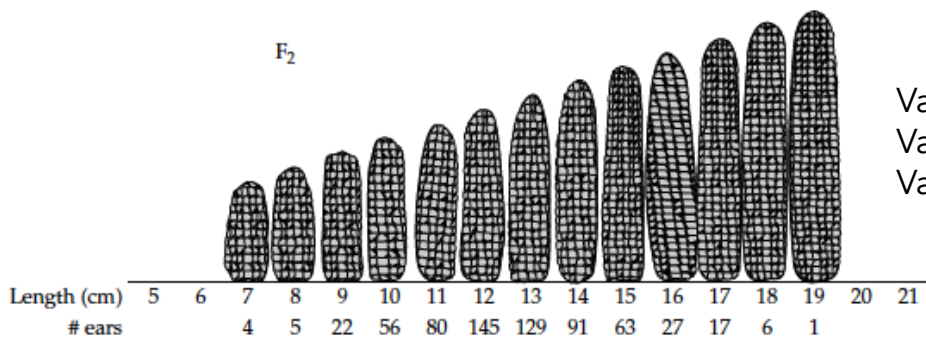


Each sample (P₁, P₂, F₁) has same G, all variation in P is due to variation in E

5



All same G, hence $\text{Var}(P) = \text{Var}(E)$



Variation in G
 $\text{Var}(P) = \text{Var}(G) + \text{Var}(E)$

$\text{Var}(F_2) > \text{Var}(F_1)$ due to Variation in G

6

Johannsen (1903) bean data

- Johannsen had a series of fully inbred (= pure) lines.
- There was a consistent between-line difference in the mean bean size
 - Differences in G across lines
- However, within a given line, size of parental seed independent of size of offspring seed
 - No variation in G within a line

7

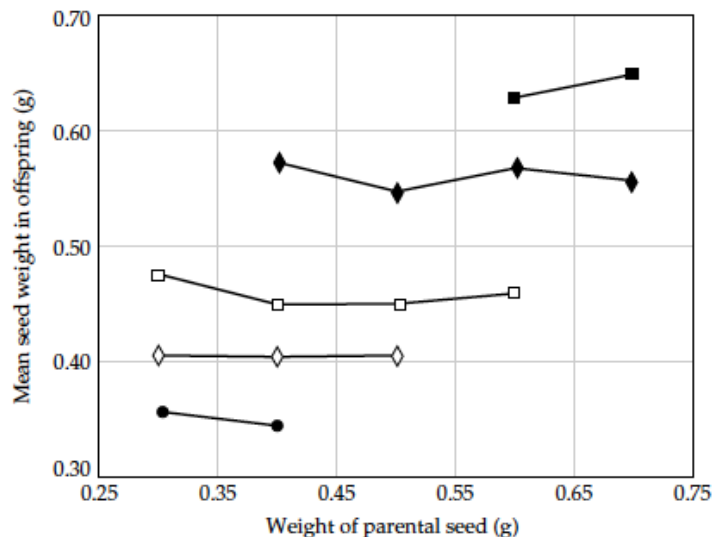


Figure 1.4 Mean offspring seed size as a function of parental seed size for some of Johannsen's pure lines. The data for the different lines are denoted by different symbols. If there is a heritable component to seed weight within a pure line, a line with positive slope is expected — larger parents should yield larger offspring. However, within each line, mean offspring size is essentially independent of the parental phenotype. (Data from Johannsen 1903.)

8

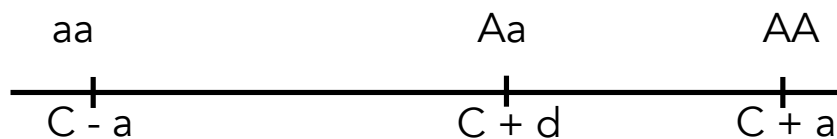
The transmission of genotypes versus alleles

- 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)

9

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 $G(Aa) = C + ak$

10

Computing a and d

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

| | | | |
|-------------|----|----|----|
| Genotype | aa | Aa | AA |
| Trait value | 10 | 15 | 16 |

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

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

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

11

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 |

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

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

Contribution from
homozygotes

Contribution from
heterozygotes

12

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 .

13

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)

14

The average effect of an allele

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

15

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 | $C + d$ |

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

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

16

Random mating

Now suppose parent contributes a

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

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

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

17

α , 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

18

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, μ is the mean (a function of p)
 - α_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)

19

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

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

20

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

Since parents pass along single alleles to their offspring, the α_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 $\hat{G}_{ij} = \mu_G + \alpha_i + \alpha_j$

21

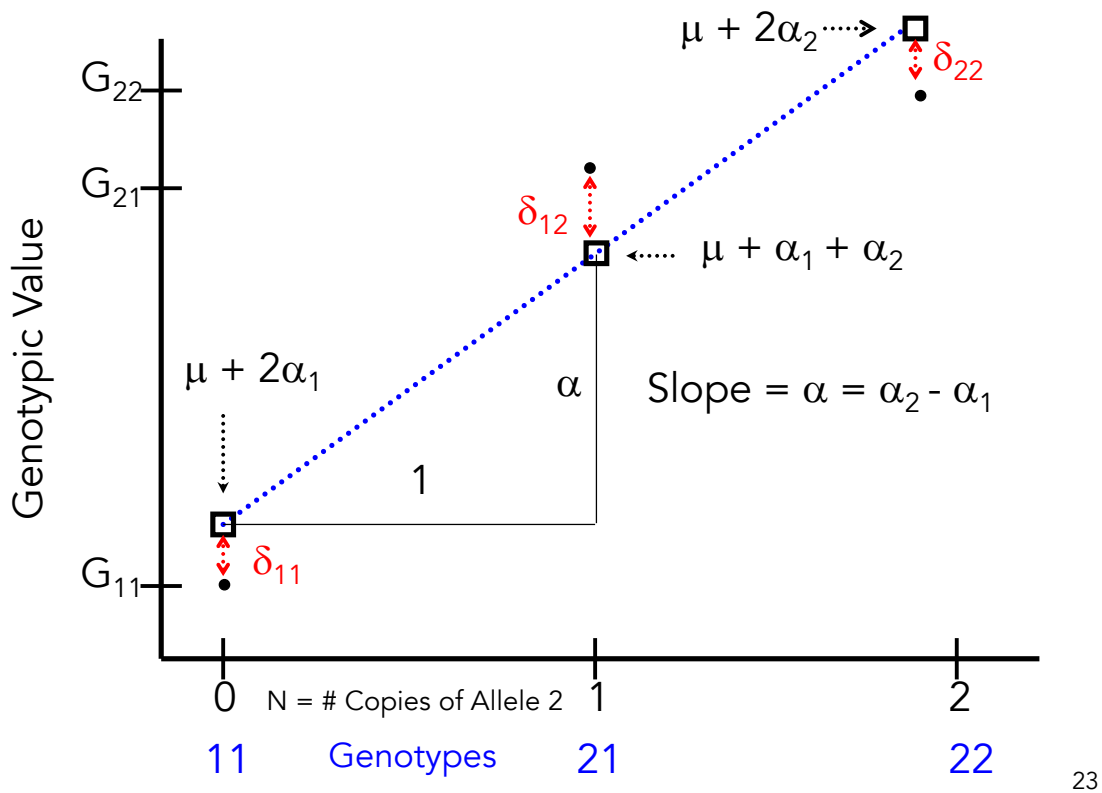
$$G_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

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,

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

22



Fisher's decomposition is a Regression

$$G_{ij} = \underbrace{\mu_G + \alpha_i + \alpha_j}_{\text{Predicted value}} + \underbrace{\delta_{ij}}_{\text{Residual error}}$$

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

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

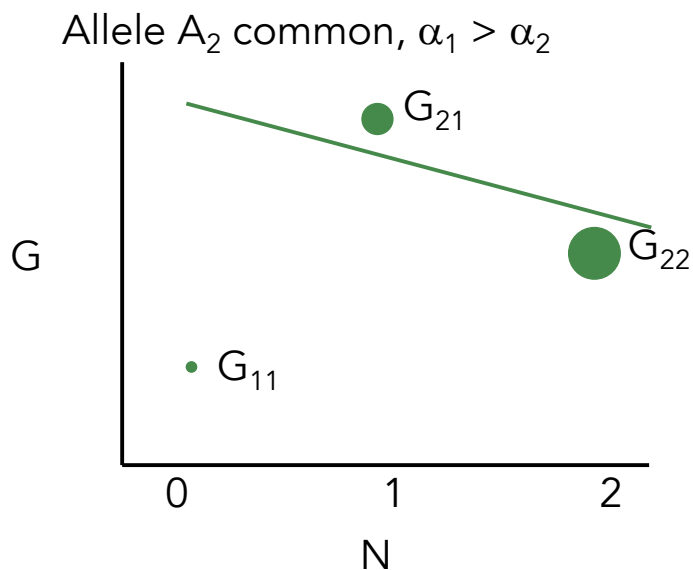
Intercept

Regression slope

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

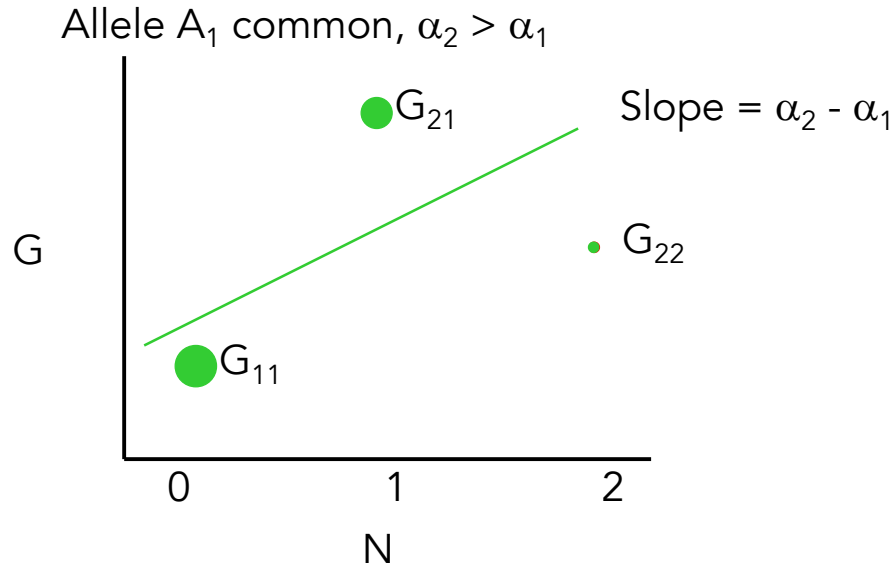
A key point is that the average effects change with allele frequencies. Indeed, if overdominance is present they can change sign with allele frequencies.

25



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.

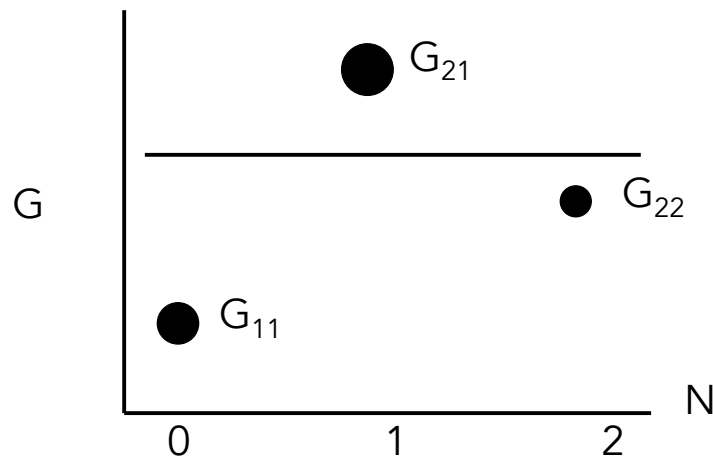
26



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

27

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

28

Average Effects and Additive Genetic Values

The α values are the **average effects** of an allele

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

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

$$A = \sum_{k=1}^n (\alpha_i^{(k)} + \alpha_j^{(k)})$$

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

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

29

$$A = \sum_{k=1}^n (\alpha_i^{(k)} + \alpha_j^{(k)})$$

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

30

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

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

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

31

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,

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

32

Key concepts (so far)

- α_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)

33

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

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

Since $E[\alpha] = 0$,
 $\text{Var}(\alpha) = E[(\alpha - \mu_a)^2] = E[\alpha^2]$

One locus, 2 alleles:

$$\sigma_A^2 = 2p_1 p_2 a^2 [1 + k(p_1 - p_2)]^2$$

\uparrow
 Dominance alters additive variance

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

34

| | | | |
|--------------------|----------|----------|----------|
| Dominance variance | Q_1Q_1 | Q_1Q_2 | Q_2Q_2 |
| | 0 | $a(1+k)$ | 2a |

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

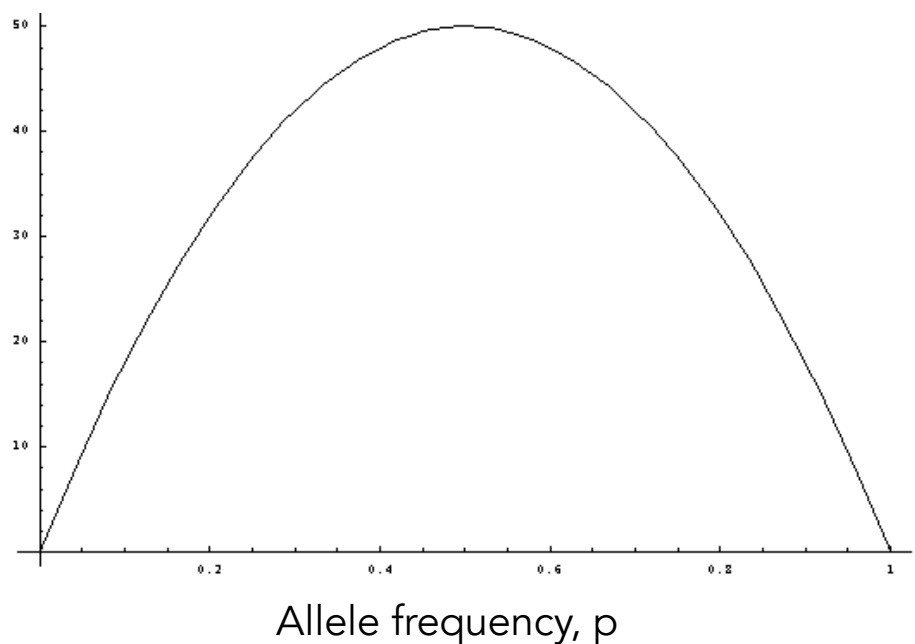
Equals zero if $k = 0$

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

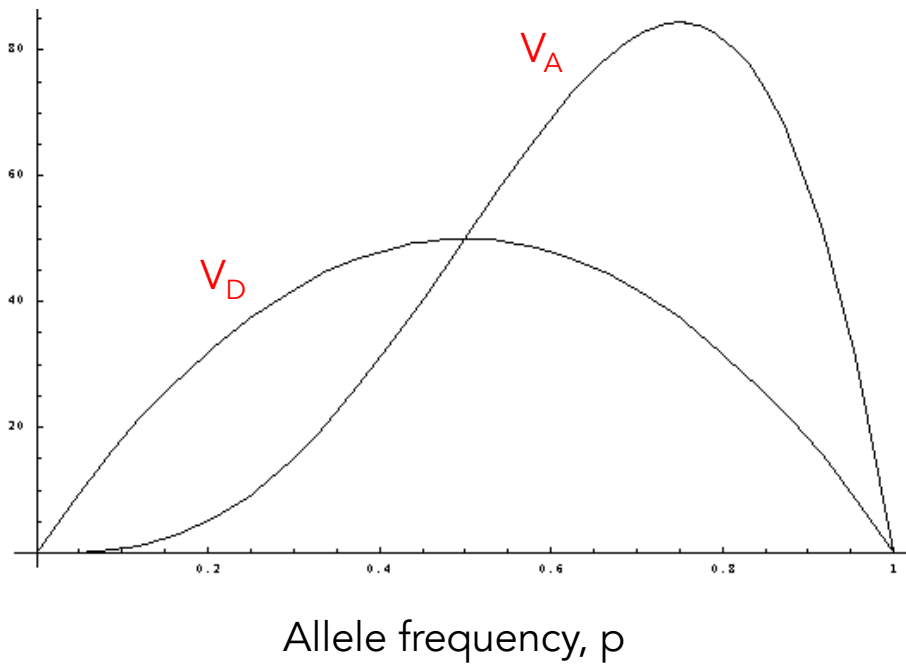
This is a symmetric function of allele frequencies

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

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



Complete dominance (k = 1)



37

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

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

38

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

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.