Partitioning the Genetic Variance

Partitioning the Genetic Variance

- In lecture 2, we showed how to partition genotypic values G into their expected values based on additivity (G^A) and deviations from the additivity as a result of dominance (δ), where

$$G = G^A + \delta$$

- We will now focus on partitioning the genetic variance into additive and dominance variance components.

Partitioning the Genetic Variance

- In lecture 2, we showed how to partition genotypic values G into their expected values based on additivity (G^A) and deviations from the additivity as a result of dominance (δ)

- For this decomposition, individuals with alleles A_i and A_j at a locus have a mean genotypic value of

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

- We will now focus on partitioning the genetic variance into the additive and dominance components of variance: $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$

- σ_A^2 is the **additive genetic variance**. It is the genetic variance associated with the average additive effects of alleles

- σ_D^2 is the **dominance genetic variance**. It is the genetic variance associated with the dominance effects.

Heritability defined in terms of the variance components

- Note that for a quantitative phenotype Y, the quantity $h^2 = \frac{\sigma_A^2}{\sigma_Y^2}$ is referred to as the **narrow sense heritability** (or simply heritability)
  - proportion of the total phenotypic variance that is due to additive effects. Can also be viewed as the extent to which phenotypes are determined by the alleles transmitted from the parents.

- In lecture 1, we defined that the quantity $H^2 = \frac{\sigma_G^2}{\sigma_Y^2} = \frac{\sigma_A^2 + \sigma_D^2}{\sigma_Y^2}$ to be the **broad sense heritability**
  - proportion of the total phenotypic variance that is due to all genetic effects.

Decomposition of Genotypic Values

- We previously decomposed the genotypic value at a locus into additive effects and dominance deviations:

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

- For a locus with two allelic types, A_1 and A_2, we showed that the model can be given in terms of a linear regression of genotypic values on the number of copies of the A_1 allele such that:

$$G_{ij} = \beta_0 + \beta_1 X_1^{ij} + \delta_{ij}$$

where X_1^{ij} is the number of copies of the type A_1 allele in genotype G_{ij}, and with $\beta_0 = \mu_G + 2\alpha_2$ and $\beta_1 = \alpha_1 - \alpha_2 = \alpha$, the average effect of allele substitution.

Components of Genetic Variance

- The original definition of the average effect of allele substitution, α, was actually given by Ronald Fisher (1918) in terms of linear regression!
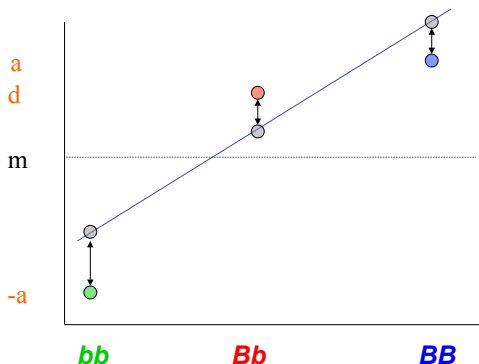
$$G_{ij} = \beta_0 + \beta_1 X_1^{ij} + \delta_{ij}$$

- Note that $G_{ij}^A = \beta_0 + \beta_1 X_1^{ij}$, the genotypic values predicted by the additive linear regression model for genotype $A_i A_j$, is the fitted value. We also have that δ_{ij}, the deviation from the additive model due to dominance, is the residual.

Linear Regression Figure for Predicting Additive Genetic Values

Falconer model for single biallelic QTL

Var (X) = Regression Variance + Residual Variance
= Additive Variance + Dominance Variance

15

Components of Genetic Variance

- Now $Var(G) = Var(G^A) + Var(\delta) + 2Cov(G^A, \delta)$
- What is $Cov(G^A, \delta)$?

Components of Genetic Variance

- From the properties of least squares, the residuals are orthogonal to the fitted values, and thus $Cov(G^A, \delta) = 0$. So we have that

$$Var(G) = Var(G^A) + Var(\delta)$$

or

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

- Let's first calculate $\sigma_A^2 = Var(G^A)$

Representation of Genotypic Values

- We previously represented genotypic values for the three genotypes as follows

$$\text{Genotype Value} = \begin{cases} -a & \text{if genotype is } A_2 A_2 \\ d & \text{if genotype is } A_1 A_2 \\ a & \text{if genotype is } A_1 A_1 \end{cases}$$

- If p and q are the allele frequencies of the A_1 and A_2 alleles, respectively in the population, we previously showed that

$$\mu_G = a(p-q) + d(2pq)$$

Additive Variance Component

- For the additive genetic component, we have that

$$G_{ij}^A = \begin{cases} \mu_G + 2\alpha_2 & \text{if genotype is } A_2 A_2 \\ \mu_G + \alpha_1 + \alpha_2 & \text{if genotype is } A_1 A_2 \\ \mu_G + 2\alpha_1 & \text{if genotype is } A_1 A_1 \end{cases}$$

where

$$\alpha_1 = pa + qd - \mu_G = q[a + d(q - p)]$$

and

$$\alpha_2 = -qa + pd - \mu_G = -p[a + d(q - p)]$$

- $\sigma_A^2 = E[(G^A - E[G^A])^2] = E[(G^A)^2] - (E[G^A])^2$, and since μ_G is a constant value that will be canceled in the variance calculation, we can essentially remove it by subtracting it out and calculating $\sigma_A^2 = E[(G^A - \mu_G)^2] - (E[G^A - \mu_G])^2$.

Additive Variance Component

- Now, $E[G^A - \mu_G]$ is

$$2\alpha_1 p^2 + (\alpha_1 + \alpha_2)2pq + 2\alpha_2 q^2 = 2\alpha_1 p(p+q) + 2\alpha_2 q(q+p)$$
$$= 2p\alpha_1 + 2q\alpha_2 = 2(p\alpha_1 + q\alpha_2) = 2(0) = 0$$

Additive Variance Component

- For $E[(G^A - \mu_G)^2]$, we have

$$
\begin{aligned}
E[(G^A - \mu_G)^2] &= p^2(2\alpha_1)^2 + 2pq(\alpha_1 + \alpha_2)^2 + q^2(2\alpha_2)^2 \\
&= p^2(2\alpha_1)^2 + 2pq(\alpha_1^2 + 2\alpha_1\alpha_2 + \alpha_2^2) + q^2(2\alpha_2)^2 \\
&= \alpha_1^2(4p^2 + 2pq) + \alpha_2^2(4q^2 + 2pq) + 4\alpha_1\alpha_2 pq \\
&= \alpha_1^2(2p(1+p)) + \alpha_2^2(2q(1+q)) + 4\alpha_1\alpha_2 pq
\end{aligned}
$$

[since
$4p^2 + 2pq = 4p^2 + 2p(1-p) = 4p^2 + 2p - 2p^2 = 2p^2 + 2p = 2p(1+p)$
]

Additive Variance Component

$$= 2\alpha_1^2 p(1+p) + 2\alpha_2^2 q(1+q) + 4\alpha_1\alpha_2 pq$$

$$= 2\alpha_1^2 p(q+2p) + 2\alpha_2^2 q(p+2q) + 4\alpha_1\alpha_2 pq$$

$$= 2\alpha_1^2 pq + 4\alpha_1^2 p^2 + 2\alpha_2^2 pq + 4\alpha_2^2 q^2 + 4\alpha_1\alpha_2 pq$$

$$= 2\alpha_1^2 pq + 2\alpha_2^2 pq + 4(\alpha_1^2 p^2 + \alpha_2^2 q^2 + \alpha_1\alpha_2 pq)$$

$$= 2\alpha_1^2 pq + 2\alpha_2^2 pq + 4(\alpha_1^2 p^2 + \alpha_2^2 q^2 + 2\alpha_1\alpha_2 pq - \alpha_1\alpha_2 pq)$$

$$= 2\alpha_1^2 pq + 2\alpha_2^2 pq - 4\alpha_1\alpha_2 pq + 4(\alpha_1^2 p^2 + \alpha_2^2 q^2 + 2\alpha_1\alpha_2 pq)$$

$$= 2\alpha_1^2 pq + 2\alpha_2^2 pq - 4\alpha_1\alpha_2 pq + 4(\alpha_1 p + \alpha_2 q)^2$$

$$= 2\alpha_1^2 pq + 2\alpha_2^2 pq - 4\alpha_1\alpha_2 pq + 0$$

$$= 2pq(\alpha_1 - \alpha_2)^2 = 2pq\alpha^2$$

Additive Variance Component

- So we have that

$$\sigma_A^2 = E[(G^A - \mu_G)^2] - (E[G^A - \mu_G])^2 = 2pq\alpha^2$$

Dominance Variance Component

- We will now obtain σ_D^2
- We have that $\delta_{ij} = G_{ij} - G_{ij}^A$ where

$$G_{ij} = \begin{cases} -a & \text{if genotype is } A_2 A_2 \\ d & \text{if genotype is } A_1 A_2 \\ a & \text{if genotype is } A_1 A_1 \end{cases}$$

and

$$G_{ij}^A = \begin{cases} \mu_G + 2\alpha_2 & \text{if genotype is } A_2 A_2 \\ \mu_G + \alpha_1 + \alpha_2 & \text{if genotype is } A_1 A_2 \\ \mu_G + 2\alpha_1 & \text{if genotype is } A_1 A_1 \end{cases}$$

- Now calculate $\sigma_D^2 = E(\delta^2) - (E(\delta))^2$

Dominance Variance Component

- We have that $\delta_{11} = a - \mu_G - 2\alpha_1$. Now

$$a - \mu_G = a - a(p - q) - d(2pq)$$

$$= a(1 - p + q) + 2dpq = 2aq - 2pdq = 2q(a - dp)$$

- So,

$$\delta_{11} = (a - \mu_G) - 2\alpha_1$$

$$= 2q(a - dp) - 2q[a + d(q - p)] = 2q(-dq) = -2q^2 d$$

- Now calculate δ_{12} and δ_{22}.
- What is $E[\delta]$?
- What is $E[\delta^2]$?
- What is $\sigma_D^2 = E(\delta^2) - (E(\delta))^2$?

Dominance Variance Component

- $\delta_{12} = 2pqd$
- $\delta_{22} = -2p^2d$
- $E[\delta] = 0$
- $E[\delta^2] = (2q^2d)^2p^2 + (2pqd)^2 2pq + (2p^2d)^2q^2 = (2pqd)^2q^2 + (2pqd)^2 2pq + (2pqd)^2p^2 = (2pqd)^2$
- So $\sigma_D^2 = (2pqd)^2$