

# Basics of Quantitative Genetics

Peter von Rohr

2022-09-30

# Background

- ▶ Central Dogma of Molecular Biology

→ Genotypes are the basis for phenotypic expression

- ▶ Start with simple model

→ one locus that affects quantitative trait

# Population

Population (N = 10)

$$\frac{G_1}{G_1}$$

$$\frac{G_1}{G_2}$$

$$\frac{G_2}{G_2}$$

$$\frac{G_2}{G_2}$$

$$\frac{G_1}{G_1}$$

$$\frac{G_2}{G_1}$$

$$\frac{G_1}{G_1}$$

$$\frac{G_1}{G_1}$$

$$\frac{G_2}{G_2}$$

$$\frac{G_2}{G_1}$$

# Terminology

- ▶ **alleles**: variants occurring at a given genetic Locus
- ▶ **bi-allelic**: only two alleles, e.g.,  $G_1$  and  $G_2$  at a given locus  $G$  in population
- ▶ **genotype**: combination of two alleles at locus  $G$  in an individual
- ▶ **homozygous**: genotypes  $G_1 G_1$  and  $G_2 G_2$  where both alleles identical
- ▶ **heterozygous**: genotype  $G_1 G_2$  different alleles

# Frequencies in Example Population

## ► genotype frequencies

$$f(G_1 G_1) = \frac{4}{10} = 0.4$$

$$f(G_1 G_2) = \frac{3}{10} = 0.3$$

$$f(G_2 G_2) = \frac{3}{10} = 0.3$$

## ► allele frequencies

$$f(G_1) = f(G_1 G_1) + \frac{1}{2} * f(G_1 G_2) = 0.55$$

$$f(G_2) = f(G_2 G_2) + \frac{1}{2} * f(G_1 G_2) = 0.45$$

# Hardy-Weinberg Equilibrium

► **allele frequencies**

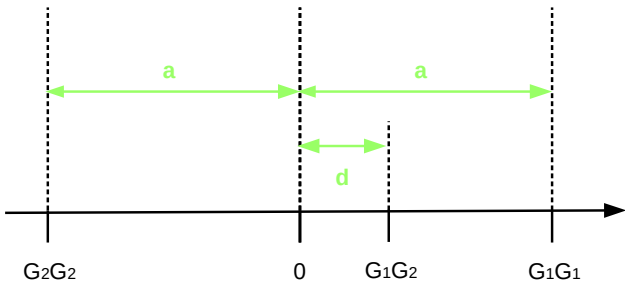
$$f(G_1) = p, f(G_2) = q = 1 - p$$

► **genotype frequencies**

Alleles	$G_1$	$G_2$
$G_1$	$f(G_1 G_1) = p^2$	$f(G_1 G_2) = p * q$
$G_2$	$f(G_1 G_2) = p * q$	$f(G_2 G_2) = q^2$

$$f(G_1 G_1) = p^2, f(G_1 G_2) = 2pq, f(G_2 G_2) = q^2$$

# Genotypic Values



# Population Mean

- ▶ Expected value of genotypic value  $V$  as discrete random variable

$$\begin{aligned}\mu &= V_{11} * f(G_1 G_1) + V_{12} * f(G_1 G_2) + V_{22} * f(G_2 G_2) \\ &= a * p^2 + d * 2pq + (-a) * q^2 \\ &= (p - q)a + 2pqd\end{aligned}$$



## Breeding Values Definition

The breeding value of an animal  $i$  is defined as two times the difference between the mean value of offsprings of animal  $i$  and the population mean.

## Derivation of Breeding value for $G_1 G_1$

	Mates of $S$	
	$f(G_1) = p$	$f(G_2) = q$
Parent $S$		
$f(G_1) = 1$	$f(G_1 G_1) = p$	$f(G_1 G_2) = q$

## Computation of Breeding value for $G_1 G_1$

$$\mu_{11} = p * a + q * d$$

The breeding value  $BV_{11}$  corresponds to

$$\begin{aligned} BV_{11} &= 2 * (\mu_{11} - \mu) \\ &= 2 (pa + qd - [(p - q)a + 2pqd]) \\ &= 2 (pa + qd - (p - q)a - 2pqd) \\ &= 2 (qd + qa - 2pqd) \\ &= 2 (qa + qd(1 - 2p)) \\ &= 2q (a + d(1 - 2p)) \\ &= 2q (a + (q - p)d) \end{aligned}$$

## Computation of Breeding value for $G_2G_2$

$$\mu_{22} = pd - qa$$

The breeding value  $BV_{22}$  corresponds to

$$\begin{aligned} BV_{22} &= 2 * (\mu_{22} - \mu) \\ &= 2 (pd - qa - [(p - q)a + 2pqd]) \\ &= 2 (pd - qa - (p - q)a - 2pqd) \\ &= 2 (pd - pa - 2pqd) \\ &= 2 (-pa + p(1 - 2q)d) \\ &= -2p(a + (q - p)d) \end{aligned}$$

## Computation of Breeding value for $G_1 G_2$

$$\mu_{12} = 0.5pa + 0.5d - 0.5qa = 0.5[(p - q)a + d]$$

The breeding value  $BV_{12}$  corresponds to

$$\begin{aligned} BV_{12} &= 2 * (\mu_{12} - \mu) \\ &= 2(0.5(p - q)a + 0.5d - [(p - q)a + 2pqd]) \\ &= 2(0.5pa - 0.5qa + 0.5d - pa + qa - 2pqd) \\ &= 2(0.5(q - p)a + (0.5 - 2pq)d) \\ &= (q - p)a + (1 - 4pq)d \\ &= (q - p)a + (p^2 + 2pq + q^2 - 4pq)d \\ &= (q - p)a + (p^2 - 2pq + q^2)d \\ &= (q - p)a + (q - p)^2d \\ &= (q - p)[a + (q - p)d] \end{aligned}$$

## Summary of Breeding Values

Genotype	Breeding Value
$G_1 G_1$	$2q\alpha$
$G_1 G_2$	$(q - p)\alpha$
$G_2 G_2$	$-2p\alpha$

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

## Allele Substitution

$$\begin{aligned}BV_{12} - BV_{22} &= (q - p)\alpha - (-2p\alpha) \\&= (q - p)\alpha + 2p\alpha \\&= (q - p + 2p)\alpha \\&= (q + p)\alpha \\&= \alpha\end{aligned}$$

$$\begin{aligned}BV_{11} - BV_{12} &= 2q\alpha - (q - p)\alpha \\&= (2q - (q - p))\alpha \\&= \alpha\end{aligned}$$

## Dominance Deviation I

$$\begin{aligned}V_{11} - BV_{11} &= a - 2q\alpha \\&= a - 2q[a + (q - p)d] \\&= a - 2qa - 2q(q - p)d \\&= a(1 - 2q) - 2q^2d + 2pqd \\&= [(p - q)a + 2pqd] - 2q^2d \\&= \mu + D_{11}\end{aligned}$$



## Dominance Deviation II

$$\begin{aligned}V_{12} - BV_{12} &= d - (q - p)\alpha \\&= d - (q - p)[a + (q - p)d] \\&= [(p - q)a + 2pqd] + 2pqd \\&= \mu + D_{12}\end{aligned}$$

$$\begin{aligned}V_{22} - BV_{22} &= -a - (-2p\alpha) \\&= -a + 2p[a + (q - p)d] \\&= [(p - q)a + 2pqd] - 2p^2d \\&= \mu + D_{22}\end{aligned}$$

## Summary of Values

Genotype $G_i G_j$	genotypic value $V_{ij}$	Breeding Value $BV_{ij}$	Dominance Deviation $D_{ij}$
$G_1 G_1$	$a$	$2q\alpha$	$-2q^2d$
$G_1 G_2$	$d$	$(q - p)\alpha$	$2pqd$
$G_2 G_2$	$-a$	$-2p\alpha$	$-2p^2d$

# Decomposition of Genotypic Values

$$V_{ij} = \mu + BV_{ij} + D_{ij}$$

## Variances

$$\text{Var}[X] = \sum_{x_i \in \mathcal{X}} (x_i - \mu_X)^2 * f(x_i)$$

where  $\mathcal{X}$ : set of all possible  $x$ -values  
 $f(x_i)$  probability that  $x$  assumes the value of  $x_i$   
 $\mu_X$  expected value  $E[X]$  of  $X$

## Variance Computation

$$\begin{aligned}\sigma_G^2 = Var[V] &= (V_{11} - \mu)^2 * f(G_1 G_1) \\ &\quad + (V_{12} - \mu)^2 * f(G_1 G_2) \\ &\quad + (V_{22} - \mu)^2 * f(G_2 G_2)\end{aligned}$$

where  $\mu = (p - q)a + 2pqd$  the population mean.

## Simplification

$$\begin{aligned}\sigma_G^2 = \text{Var}[V] &= (BV_{11} + D_{11})^2 * f(G_1 G_1) \\ &\quad + (BV_{12} + D_{12})^2 * f(G_1 G_2) \\ &\quad + (BV_{22} + D_{22})^2 * f(G_2 G_2)\end{aligned}$$

## Result

$$\begin{aligned}\sigma_G^2 &= 2pq\alpha^2 + (2pqd)^2 \\ &= \sigma_A^2 + \sigma_D^2\end{aligned}$$

## Two and more Loci

- ▶ Two loci  $G$  and  $H$  having an effect on the same quantitative trait.
- ▶ Effect of one locus can have an influence on the effect of the other locus

→ **Interaction** between loci.

- ▶ Interaction is quantified by

$$I_{GH} = V - V_G - V_H$$

where  $V$  is the total genotypic value,  $V_G$  and  $V_H$  correspond to the genotypic values due to loci  $G$  and  $H$ , respectively



# Decomposition and Collecting Terms

- ▶ Genotypic values can be decomposed as

$$V_G = \mu_G + BV_G + D_G$$

$$V_H = \mu_H + BV_H + D_H$$

- ▶ Collecting terms leads to

$$V = V_G + V_H + I_{GH} = \mu + U + D + I$$

with  $\mu = \mu_G + \mu_H$ ,  $U = BV_G + BV_H$ ,  $D = D_G + D_H$  and  $I = I_{GH}$

- ▶ Can be generalized to more than two loci.

## More Than Two Loci

- ▶ Genotypic value  $V$  influenced by an unknown number of loci:  
 $A, B, C, \dots$ ,
- ▶ Decomposition of  $V$

$$V = V_A + V_B + V_C + \dots + I_{ABC\dots}$$

where  $I_{ABC\dots}$  is a generic Interaction term which we do not specify further here.

# Decomposition

- ▶ Genotypic value of single loci

$$V_{A_i A_j} = \mu_A + BV_{A_i A_j} + D_{A_i A_j}$$

$$V_{B_i B_j} = \mu_B + BV_{B_i B_j} + D_{B_i B_j}$$

$$V_{C_i C_j} = \mu_C + BV_{C_i C_j} + D_{C_i C_j}$$

# Collecting Terms

- ▶ for a given animal  $k$ ,  $\mu$ ,  $BV$  and  $D$  terms are collected

$$\mu_k = \mu_A + \mu_B + \mu_C + \dots$$

$$U_k = BV_k = BV_{A_i A_j} + BV_{B_i B_j} + BV_{C_i C_j} + \dots$$

$$D_k = D_{A_i A_j} + D_{B_i B_j} + D_{C_i C_j} + \dots$$

# Phenotype

- ▶ Including phenotypic observations  $y$
- ▶ Central Dogma of Molecular Biology
- ▶ Decomposition

$$y = V + E$$

where  $V$  is the genotypic value and  $E$  is the non-genetic or environmental rest

- ▶ Insert decomposition of  $V$  as shown above

# Decomposition of Phenotype

- Insert  $V = V_A + V_B + V_C + \dots + I_{ABC\dots}$  into  $y = V + E$

$$y = V + E = V_A + V_B + V_C + \dots + I_{ABC\dots} + E$$

$$= (\mu_A + BV_{A_iA_j} + D_{A_iA_j}) + (\mu_B + BV_{B_iB_j} + D_{B_iB_j})$$

$$+(\mu_C + BV_{C_iC_j} + D_{C_iC_j}) + \dots I_{ABC\dots} + E$$

## Collecting terms

$$\begin{aligned} &= (\mu_A + \mu_B + \mu_C + \dots) + (BV_{A_i A_j} + BV_{B_i B_j} + BV_{C_i C_j} + \dots) \\ &\quad + (D_{A_i A_j} + D_{B_i B_j} + D_{C_i C_j} \dots) + I_{ABC\dots} + E \\ &= \mu + BV + D + I_{ABC\dots} + E = \mu + BV + E^* \end{aligned}$$