Chapter 2

Basics in Quantitative Genetics

As already mentioned in section 1.3.1, the central dogma of molecular biology tells us that the genotype is the basics of any phenotypic expression. The genotype of an individual is composed of a number of genes which are also called **loci**. In this section, we start with the simplest possible genetic architecture where the genotype is composed by just one locus. The connection between the genotype and the phenotype is modeled according to equation (1.1). The phenotype is assumed to be a quantitative trait. That means we are not looking at binary or categorical traits. Categorical traits can just take a limited number of different levels. Examples of categorical traits are the horn status in cattle or certain color characteristics. Quantitative traits do not take discrete levels but they show specific distributions.

2.1 Single Locus - Quantitative Trait

In Livestock there are not many examples where a quantitative trait is influenced by just one locus. But this case helps in understanding the foundation of more complex genetic architectures. We start by looking at the following idealized population (Figure 2.1).

2.1.1 Terminology

The different genetic variants that are present at our Locus G are called **alleles**. When looking at all individuals in the population for our locus, we have two different alleles G_1 and G_2. Hence, we call the locus G to be a **bi-allelic** locus. In any given individual of the population, the two alleles of the locus G together are called the individuals **genotype**. All possible combinations of the two alleles at the locus G leads to a total number of three genotypes. It is important to mention that the order of the alleles in a given genotype is not important. Hence, G_1G_2 and G_2G_1 are the same genotype. The two genotypes G_1G_1 and G_2G_2 are called **homozygous** and the genotype G_1G_2 is called **heterozygous**.

2.2 Frequencies

To be able to characterize our population with respect to the locus of interest, we are first looking at some frequencies. These are measures of how often a certain allele or genotype does occur in our population. For our example population shown in Figure 2.1, the **genotype frequencies** are

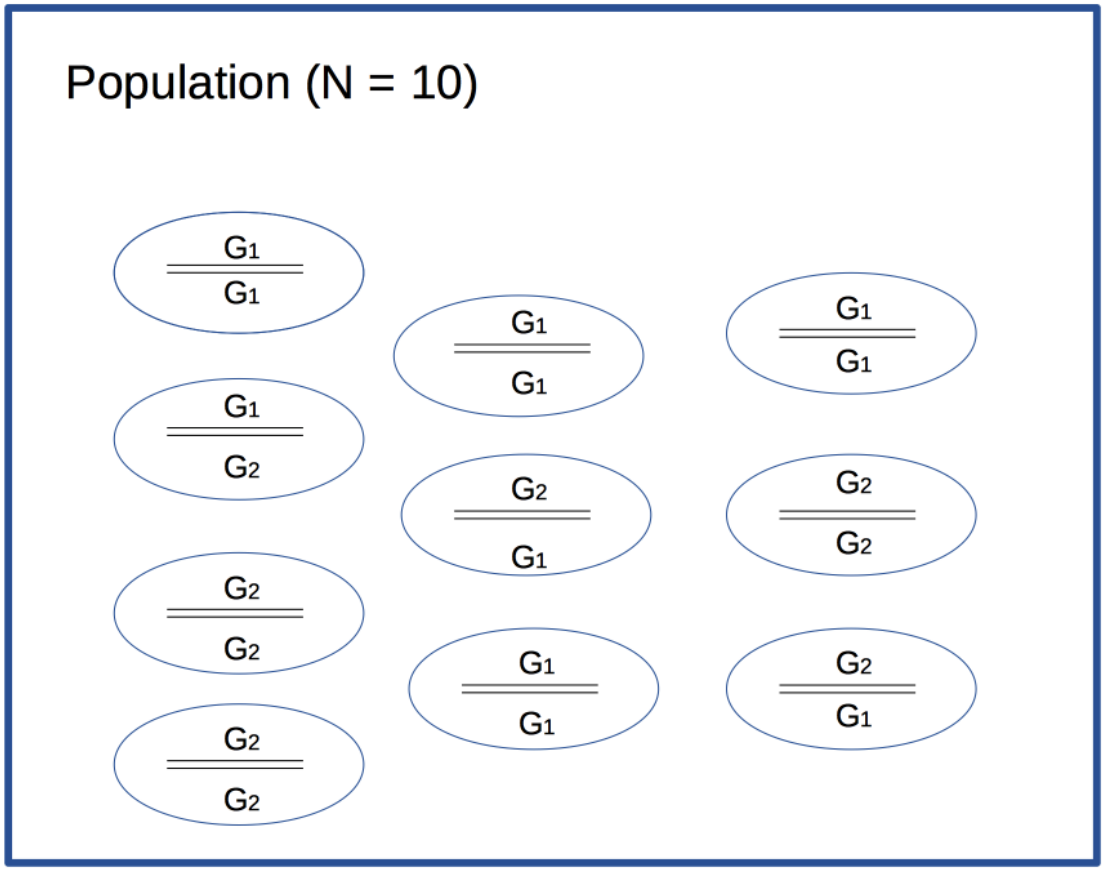

Figure 2.1: Idealized Population With A Single Locus

Table 2.1: Genotype Frequencies under Hardy-Weinberg equilibrium

Alleles	G_1	G_2
G_1	$f(G_1G_1) = p^2$	$f(G_1G_2) = p * q$
G_2	$f(G_1G_2) = p * q$	$f(G_2G_2) = q^2$

$$\begin{aligned} f(G_1G_1) &= \frac{4}{10} = 0.4 \\ f(G_1G_2) &= \frac{3}{10} = 0.3 \\ f(G_2G_2) &= \frac{3}{10} = 0.3 \end{aligned} \tag{2.1}$$

The **allele frequencies** can be determined either by counting or they can be computed from the genotype frequencies.

$$\begin{aligned} f(G_1) &= f(G_1G_1) + \frac{1}{2} * f(G_1G_2) = 0.55 \\ f(G_2) &= f(G_2G_2) + \frac{1}{2} * f(G_1G_2) = 0.45 \end{aligned} \tag{2.2}$$

2.3 Hardy-Weinberg Equilibrium

The Hardy-Weinberg equilibrium is the central law of how allele frequencies and genotype frequencies are related in an idealized population. Given the allele frequencies

$$\begin{aligned} f(G_1) &= p \\ f(G_2) &= q = 1 - p \end{aligned} \tag{2.3}$$

During mating, we assume that in an idealized population alleles are combined independently. This leads to the genotype frequencies shown in Table 2.1.

Summing up the heterozygous frequencies leads to

$$\begin{aligned} f(G_1G_1) &= p^2 \\ f(G_1G_2) &= 2pq \\ f(G_2G_2) &= q^2 \end{aligned} \tag{2.4}$$

Comparing these expected genotype frequencies in a idealized population under the Hardy-Weinberg equilibrium to what we found for the small example population in Figure 2.1, we can clearly say that the small example population is not in Hardy-Weinberg equilibrium.

2.4 Value and Mean

Our goal is still to improve our population at the genetic level. The term improvement implies the need for a quantitative assessment of our trait of interest. Furthermore, we have to be able to associate the genotypes in the population to the quantitative values of our trait.

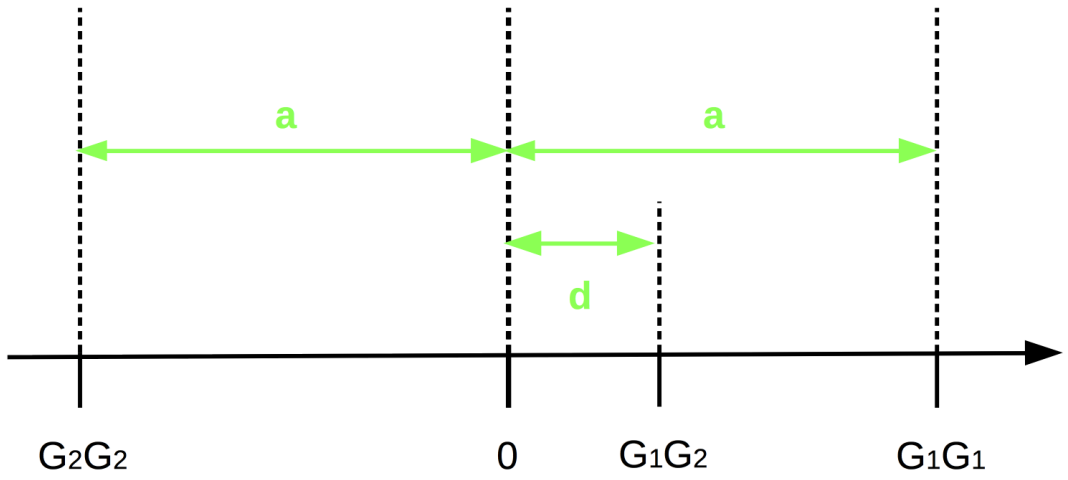

Figure 2.2: Genotypic Values

2.4.1 Genotypic Values

The values V_{ij} to each genotype G_iG_j are assigned as shown in Figure 2.2.

The origin of the genotypic values is placed in the middle between the two homozygous genotypes G_2G_2 and G_1G_1. Here we are assuming that G_1 is the favorable allele. This leads to values of $+a$ for genotype G_1G_1 and of $-a$ for genotype G_2G_2. The value of genotype G_1G_2 is set to d and is called dominance deviation. Table 2.2 summarizes the values for all genotypes.

2.4.2 Population Mean

For the complete population, we can compute the **population mean** (μ) of all values at the locus G. This mean corresponds to the expected value and is computed as

Table 2.2: Values for all Genotypes

Variable	Genotype	Values
V_{11}	G_1G_1	a
V_{12}	G_1G_2	d
V_{22}	G_2G_2	-a

$$\begin{aligned}\mu &= V_{11} * f(G_1G_1) + V_{12} * f(G_1G_2) + V_{22} * f(G_2G_2) \\ &= a * p^2 + d * 2pq + (-a) * q^2 \\ &= (p-q)a + 2pqd \end{aligned} \tag{2.5}$$

The population mean depends on the values a and d and on the allele frequencies p and q. The larger the difference between p and q the more influence the value a has in μ, because for very different p and q the product $2pq$ is very small. On the other hand, if $p = q = 0.5$, then $\mu = 0.5d$. For loci with $d = 0$, the population mean $\mu = (p - q)a$ and hence, if in addition we have $p = q$, then $\mu = 0$.

2.4.3 Breeding Values

The term **breeding value** is defined as shown in Definition 2.1.

Definition 2.1 (Breeding Value)**.** 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.

Applying this definition and using the parameters that we have computed so far leads to the following formulas for the breeding value of an animal with a certain genotype.

2.4.3.1 Breeding value for G_1G_1

Assume that we have a given parent S with a genotype G_1G_1 and we want to compute its breeding value. Let us further suppose that our single parent S is mated to a potentially infinite number of animals from the idealized population, then we can deduce the following mean genotypic value for the offspring of parent S.

	Mates of S	
	$f(G_1) = p$	$f(G_2) = q$
Parent S		
$f(G_1) = 1$	$f(G_1G_1) = p$	$f(G_1G_2) = q$

Because parent S has genotype G_1G_1, the frequency $f(G_1)$ of a G_1 allele coming from S is 1 and the frequency $f(G_2)$ of a G_2 allele is 0. The expected genetic value (μ_{11}) of the offspring of animal S can be computed as

$$\mu_{11} = p * a + q * d \tag{2.6}$$

Applying definition 2.1, we can compute the breeding value (BV_{11}) for animal S as shown in equation (2.7) while using the results given by equations (2.6) and (2.5).

$$\begin{aligned} BV_{11} &= 2 * (\mu_{11} - \mu) \\ &= 2\,(pa + qd - [(p-q)a + 2pqd]) \\ &= 2\,(pa + qd - (p-q)a - 2pqd) \\ &= 2\,(qd + qa - 2pqd) \\ &= 2\,(qa + qd(1-2p)) \\ &= 2q\,(a + d(1-2p)) \\ &= 2q\,(a + (q-p)d) \end{aligned} \tag{2.7}$$

Breeding values for parents with genotypes G_2G_2 and G_1G_2 are derived analogously.

2.4.3.2 Breeding value for G_2G_2

First, we determine the expected genotypic value for offsprings of a parent S with genotype G_2G_2

	Mates of parent S	
	$f(G_1) = p$	$f(G_2) = q$
Parent S		
$f(G_2) = 1$	$f(G_1G_2) = p$	$f(G_2G_2) = q$

The expected genetic value (μ_{22}) of the offspring of animal S can be computed as

$$\mu_{22} = pd - qa \tag{2.8}$$

The breeding value BV_{22} corresponds to

$$\begin{aligned} BV_{22} &= 2 * (\mu_{22} - \mu) \\ &= 2\,(pd - qa - [(p-q)a + 2pqd]) \\ &= 2\,(pd - qa - (p-q)a - 2pqd) \\ &= 2\,(pd - pa - 2pqd) \\ &= 2\,(-pa + p(1-2q)d) \\ &= -2p\,(a + (q-p)d) \end{aligned} \tag{2.9}$$

2.4.3.3 Breeding value for G_1G_2

The genotype frequencies of the offsprings of a parent S with a genotype G_1G_2 is determined in the following table.

	Mates of parent S	
	$f(G_1) = p$	$f(G_2) = q$
Parent S		
$f(G_1) = 0.5$	$f(G_1G_1) = 0.5p$	$f(G_1G_2) = 0.5q$
$f(G_2) = 0.5$	$f(G_1G_2) = 0.5p$	$f(G_2G_2) = 0.5q$

The expected mean genotypic value of the offsprings of parent S with genotype G_1G_2 is computed as

$$\mu_{12} = 0.5pa + 0.5d - 0.5qa = 0.5\left[(p-q)a + d\right] \tag{2.10}$$

The breeding value BV_{12} corresponds to

$$\begin{aligned} ZW_{12} &= 2 * (\mu_{12} - \mu) \\ &= 2\left(0.5(p-q)a + 0.5d - [(p-q)a + 2pqd]\right) \\ &= 2\left(0.5pa - 0.5qa + 0.5d - pa + qa - 2pqd\right) \\ &= 2\left(0.5(q-p)a + (0.5 - 2pq)d\right) \\ &= (q-p)a + (1 - 4pq)d \\ &= (q-p)a + (p^2 + 2pq + q^2 - 4pq)d \\ &= (q-p)a + (p^2 - 2pq + q^2)d \\ &= (q-p)a + (q-p)^2 d \\ &= (q-p)\left[a + (q-p)d\right] \end{aligned} \tag{2.11}$$

2.4.3.4 Summary of Breeding Values

The term $a + (q-p)d$ appears in all three breeding values. We replace this term by α and summarize the results in the following table.

Genotyp	Zuchtwert
G_1G_1	$2q\alpha$
G_1G_2	$(q-p)\alpha$
G_2G_2	$-2p\alpha$

2.4.4 Allele Substitution

Comparing the genotype G_2G_2 with the genotype G_1G2, one of the differences is in the number of G_1-alleles. G_2G_2 has zero G_1-alleles and G_1G_2 has one G_1-allele. They also have different breeding values. Because the breeding values are to be used to assess the value of a given genotype, any difference between the breeding values BV_{12} and B_{22} can be associated to that additional G_1 allele in G_1G_2 compared to G_2G_2.

The computation of the difference between the breeding value BV_{12} and B_{22} results in

$$\begin{aligned} BV12 - BV_{22} &= (q-p)\alpha - (-2p\alpha) \\ &= (q-p)\alpha + 2p\alpha \\ &= (q-p+2p)\alpha \\ &= (q+p)\alpha \\ &= \alpha \end{aligned} \tag{2.12}$$

The analogous computation can be done by comparing the breeding values BV_{11} and BV_{12}.

$$\begin{aligned} BV_{11} - BV_{12} = & \quad 2q\alpha - (q-p)\alpha \\ = & \quad (2q - (q-p))\,\alpha \\ = & \quad \alpha \end{aligned} \tag{2.13}$$

The breeding values themselves depend on the allele frequencies. But the differences between breeding values are linear while replacing G_2 alleles by G_1 alleles. This replacement is also called allele-substitution and the term *alpha* is called **allele-substitution effect**.

2.4.5 Dominance Deviation

When looking at the difference between the genotypic value V_{ij} and the breeding value BV_{ij} for each of the three genotypes, we get the following results.

$$\begin{aligned} V_{11} - BV_{11} &= a - 2q\alpha \\ &= a - 2q\,[a + (q-p)d] \\ &= a - 2qa - 2q(q-p)d \\ &= a(1-2q) - 2q^2 d + 2pqd \\ &= [(p-q)a + 2pqd] - 2q^2 d \\ &= \mu + D_{11} \end{aligned} \tag{2.14}$$

$$\begin{aligned} V_{12} - ZW_{12} &= d - (q-p)\alpha \\ &= d - (q-p)\,[a + (q-p)d] \\ &= [(p-q)a + 2pqd] + 2pqd \\ &= \mu + D_{12} \end{aligned} \tag{2.15}$$

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

The difference all contain the population mean μ plus a certain deviation. This deviation term is called **dominance deviation**.

2.4.6 Summary of Values

The following table summarizes all genotypic values all breeding values and the dominance deviations.

Genotyp	genotypic value	Breeding Value	Dominance Deviation
G_iG_j	V_{ij}	BV_{ij}	D_{ij}
G_1G_1	a	$2q\alpha$	$-2q^2d$
G_1G_2	d	$(q-p)\alpha$	$2pqd$
G_2G_2	$-a$	$-2p\alpha$	$-2p^2d$

The formulas in the above shown table assume that G_1 is the favorable allele with frequency $f(G_1) = p$. The allele frequency of G_2 is $f(G_2) = q$. Since we have a bi-allelic locus $p + q = 1$.

Based on the definition of dominance deviation, the genotypic values V_{ij} can be separated according to equation (2.16).

$$V_{ij} = \mu + BV_{ij} + D_{ij} \tag{2.16}$$