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

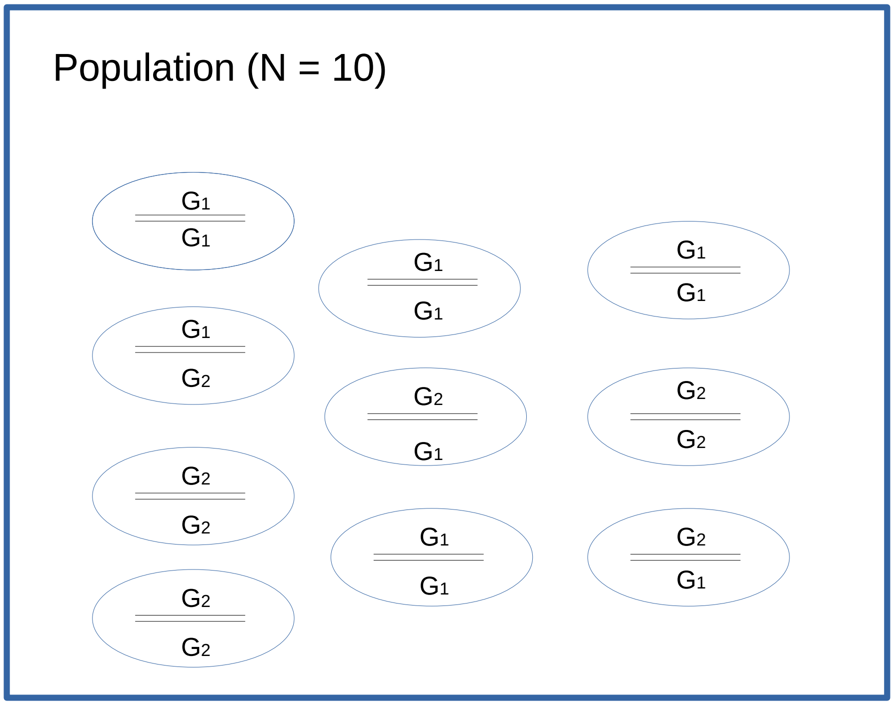

Figure 2.1: Idealized Population With A Single Locus

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

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.

2.4.1 Genotypic Values

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

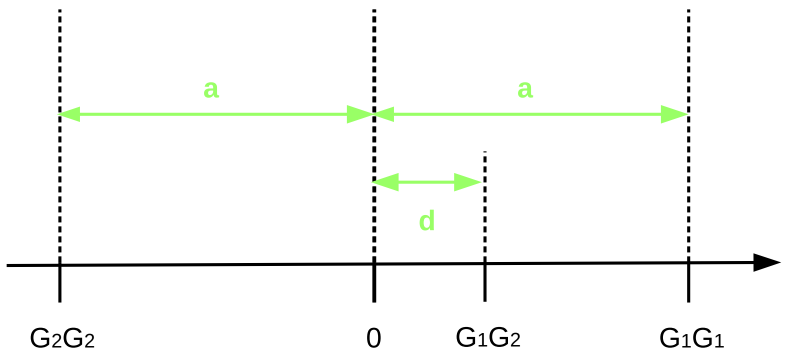

Figure 2.2: Genotypic Values

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.

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

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

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

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 **??**.

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 **??**, 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).

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

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

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

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

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

The breeding value BV_{12} corresponds to

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

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.

Genotype	Breeding Value
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.

Let us imagine that we take animal i with a G_2G_2 genotype and use the CRISPR-CAS genome editing technology to replace one of the G_2 alleles in animal i by a G_1 allele (see Figure 2.3). After applying the gene editing procedure to animal i at locus G, animal i would have genotype G_1G_2.

Due to the application of genome editing at locus G of animal i the breeding value changed. Before the genome editing procedure it was BV_{22} and after genome editing the breeding value of animal i corresponds to BV_{12}. So the effect of replacing a G_2 allele by a G_1 allele on the breeding value corresponds to the difference $BV_{12} - BV_{22}$. The computation of this difference between the breeding value BV_{12} and B_{22} results in

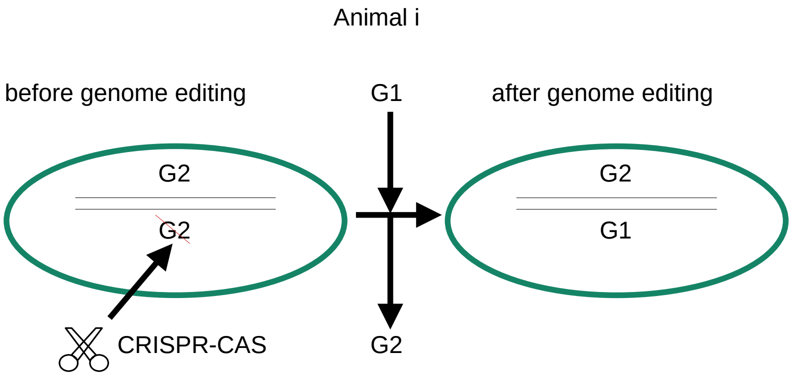

Figure 2.3: Schematic Depiction of Genome Editing on Animal i

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

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

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

Because the differences between breeding values computed in (2.12) and (2.13) are equal, we can conclude that the breeding values show a linear dependence on the number of G_1 alleles. This is the reason why the breeding values are also called additive effects, because adding a further G_1 allele instead of a G_2 allel has always the same effect on the breeding values, namely just adding the constant allele substitution effect α.

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.

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

$$\begin{aligned}
V_{12} - BV12 &= d - (q-p)\alpha \\
&= d - (q-p)\left[a + (q-p)d\right] \\
&= \left[(p-q)a + 2pqd\right] + 2pqd \\
&= \mu + D_{12} \qquad (2.15)
\end{aligned}$$

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

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 G_iG_j	genotypic value V_{ij}	Breeding Value BV_{ij}	Dominance Deviation 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 decomposed into the components population mean (μ), breeding value (BV_{ij}) and dominance deviation (D_{ij}) according to equation (2.16).

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

Taking expected values on both sides of equation (2.16) and knowing that the population mean μ was defined as the expected value of the genotypic values in the population, i.e. $E\left[V\right] = \mu$, it follows that the expected values of both the breeding values and the dominance deviations must be 0. More formally, we have

$$\begin{aligned}
E\left[V\right] &= E\left[\mu + BV + D\right] \\
&= E\left[\mu\right] + E\left[BV\right] + E\left[D\right] \\
&= \mu
\end{aligned} \tag{2.17}$$

From the last line in equation (2.17), it follows that $E\left[BV\right] = E\left[D\right] = 0$. This also shows that both breeding values and dominance deviations are defined as deviation from a given mean.

2.5 Variances

The population mean μ and derived from that the breeding values were defined as expected values. Their main purpose is to assess the state of a given population with respect to a certain genetic locus and its effect on a phenotypic trait of interest. One of our primary goals in livestock breeding is to improve the populations at the genetic level through the means of selection and mating. Selection of potential parents that produce offspring that are closer to our breeding goals is only possible, if the selection candidates show a certain level of variation in the traits that we are interested in. In populations where there is no variation which means that all individuals are exactly at the same level, it is not possible to select potential parents for the next generation.

In statistics the measure that is most often used to assess variation in a certain population is called **variance**. For any given discrete random variable X the variance is defined as the second central moment of X which is computed as shown in equation (2.18).

$$Var\left[X\right] = \sum_{x_i \in \mathcal{X}} (x_i - \mu_X)^2 * f(x_i) \tag{2.18}$$

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

In this section we will be focusing on separating the obtained variances into different components according to their causative sources. Applying the definition of variance given in equation (2.18) to the genotypic values V_{ij}, we obtain the following expression.

$$\begin{aligned}\sigma_G^2 = Var\left[V\right] = &(V_{11} - \mu)^2 * f(G_1G_1) \\ + &\ (V_{12} - \mu)^2 * f(G_1G_2) \\ + &\ (V_{22} - \mu)^2 * f(G_2G_2)\end{aligned} \tag{2.19}$$

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

Based on the decomposition of the genotypic value V_{ij} given in (2.16), the difference between V_{ij} and μ can be written as the sum of the breeding value and the dominance deviation. Then σ_G^2 can be written as

$$\begin{aligned}\sigma_G^2 = Var\left[V\right] = &(BV_{11} + D_{11})^2 * f(G_1G_1) \\ + &\ (BV_{12} + D_{12})^2 * f(G_1G_2) \\ + &\ (BV_{22} + D_{22})^2 * f(G_2G_2)\end{aligned} \tag{2.20}$$

Inserting the expressions for the breeding values BV_{ij} and for the dominance deviation D_{ij} found earlier and simplifying the equation leads to the result in (2.21). A more detailed derivation of σ_G^2 is given in the appendix (2.6) of this chapter.

$$\begin{aligned}\sigma_G^2 &= 2pq\alpha^2 + \left(2pqd\right)^2 \\ &= \sigma_A^2 + \sigma_D^2\end{aligned} \tag{2.21}$$

The formula in equation (2.21) shows that σ_G^2 consists of two components. The first component σ_A^2 is called the **genetic additive variance** and the second component σ_D^2 is termed **dominance variance**. As shown in equation (2.23) σ_A^2 corresponds to the variance of the breeding values. Because we have already seen that the breeding values are additive in the number of favorable alleles, σ_A^2 is called genetic additive variance. Because σ_D^2 corresponds to the variance of the dominance deviation effects (see equation (2.25)) it is called dominance variance.

2.6 Appendix: Derivations

This section shows how the genetic variance in equation (2.21) is computed.

$$
\begin{aligned}
\sigma_G^2 &= (BV_{11} + D_{11})^2 * p^2 \\
&+ \ (BV_{12} + D_{12})^2 * 2pq \\
&+ \ (BV_{22} + D_{22})^2 * q^2 \\
&= \left(2q\alpha - 2q^2 d\right)^2 * p^2 \\
&+ \ \left((q-p)\alpha + 2pqd\right)^2 * 2pq \\
&+ \ \left(-2p\alpha - 2p^2 d\right)^2 * q^2 \\
&= \left(4q^2\alpha^2 - 8q^3 d\alpha + 4q^4 d^2\right) * p^2 \\
&+ \ \left(q^2\alpha^2 - 2pq\alpha^2 + p^2\alpha^2 - 4(q-p)pqd\alpha + 4p^2q^2d^2\right) * 2pq \\
&+ \ \left(4p^2\alpha^2 + 8p^3 d\alpha + 4p^4\alpha^2\right) * q^2 \\
&= 4p^2q^2\alpha^2 - 8p^2q^3 d\alpha + 4p^2q^4d^2 \\
&+ \ 2pq^3\alpha^2 - 4p^2q^2\alpha^2 + 2p^3q\alpha^2 \\
&- \ 8p^3q^2d\alpha + 8p^2q^3d\alpha + 8p^3q^3d^2 \\
&+ \ 4p^2q^2\alpha^2 + 8p^3q^2d\alpha + 4p^4q^2d^2 \\
&= 4p^2q^2\alpha^2 + 4p^2q^4d^2 \\
&+ \ 2pq^3\alpha^2 + 2p^3q\alpha^2 \\
&+ \ 8p^3q^3d^2 \\
&+ \ 4p^4q^2d^2 \\
&= 2pq\alpha^2\left(p^2 + 2pq + q^2\right) \\
&+ \ \left(2pqd\right)^2\left(p^2 + 2pq + q^2\right) \\
&= 2pq\alpha^2 + \left(2pqd\right)^2 \\
&= \sigma_A^2 + \sigma_D^2 \qquad (2.22)
\end{aligned}
$$

From the last two lines of (2.22) it follows that $\sigma_A^2 = 2pq\alpha^2$ and $\sigma_D^2 = \left(2pqd\right)^2$. It can be shown that σ_A^2 corresponds to the squared breeding values times the associated genotype frequencies. Because the expected values of the breeding values is zero, σ_A^2 is equivalent to the variance of the breeding values.

$$
\begin{aligned}
\sigma_A^2 = Var\left[BV\right] &= (BV_{11} - E\left[BV\right])^2 * f(G_1G_1) \\
&+ (BV_{12} - E\left[BV\right])^2 * f(G_1G_2) \\
&+ (BV_{22} - E\left[BV\right])^2 * f(G_2G_2) \\
&= BV_{11}^2 * f(G_1G_1) + BV_{12}^2 * f(G_1G_2) + BV_{22}^2 * f(G_2G_2) \\
&= \left(2q\alpha\right)^2 * p^2 + \left((q-p)\alpha\right)^2 * 2pq + \left(-2p\alpha\right)^2 * q^2 \\
&= 4p^2q^2\alpha^2 + \left(q^2\alpha^2 - 2pq\alpha^2 + p^2\alpha^2\right) * 2pq + 4p^2q^2\alpha^2 \\
&= 8p^2q^2\alpha^2 + 2pq^3\alpha^2 - 4p^2q^2\alpha^2 + 2p^3q\alpha^2 \\
&= 4p^2q^2\alpha^2 + 2pq^3\alpha^2 + 2p^3q\alpha^2 \\
&= 2pq\alpha^2\left(2pq + q^2 + p^2\right) \\
&= 2pq\alpha^2 \qquad (2.23)
\end{aligned}
$$

In the above derivation in (2.23) of the variance of the breeding values, we were using the fact that the expected value $E\left[BV\right] = 0$. This can be shown more formally as follows

$$
\begin{aligned}
E\left[BV\right] &= BV_{11} * f(G_1G_1) + BV_{12} * f(G_1G_2) + BV_{22} * f(G_2G_2) \\
&= 2q\alpha * p^2 + (q-p)\alpha * 2pq + (-2p\alpha) * q^2 \\
&= 2p^2q\alpha + 2pq^2\alpha - 2p^2q\alpha - 2pq^2\alpha \\
&= 0 \qquad (2.24)
\end{aligned}
$$

Similarly to (2.23) we can show that σ_D^2 corresponds to the squared dominance deviations times the frequencies of the corresponding genotypes. That is the reason why σ_D^2 is called dominance variance.

$$
\begin{aligned}
\sigma_D^2 &= D_{11}^2 * f(G_1G_1) + D_{12}^2 * f(G_1G_2) + D_{22}^2 * f(G_2G_2) \\
&= (-2q^2d)^2 * p^2 + (2pqd)^2 * 2pq + (-2p^2d)^2 * q^2 \\
&= 4p^2q^4d^2 + 8p^3q^3d^2 + 4p^4q^2d^2 \\
&= 4p^2q^2d^2\left(q^2 + 2pq + p^2\right) \\
&= 4p^2q^2d^2 \qquad (2.25)
\end{aligned}
$$