Chapter 5
Basic Concepts of Quantitative Genetics

Quantitative genetics is a special branch of genetics, which is concerned with the inheritance of the differences between individuals that are measured in degree rather than in kind. These individual differences are referred to as quantitative differences or quantitative traits. Formally, a quantitative trait is defined as a trait whose value varies continuously across individuals (Falconer and Mackay 1996; Lynch and Walsh 1998). The phenotype of a quantitative trait measured from an individual is not determined by genes alone; it is also determined by environmental variants. The proportion of the phenotypic variance explained by the segregation of a single gene is usually small. However, the contribution of all these small-effect genes collectively is significant to the variation of the phenotype. Genes controlling the variation of a quantitative trait are called quantitative trait loci (QTL). Note that the term QTL defined in this book is used for both the singular and plural forms, e.g., one QTL for weight and two QTL for height. In the quantitative genetics literature, QTL represents the singular form and QTLs is used as the plural form. No matter how small a QTL is, it segregates just like a regular Mendelian locus. For small-effect QTL, we simply cannot observe the segregation and must resort to statistical methods to infer the segregation. Most statistical methods applied in quantitative genetics require specific genetic models, which will be the focus of this chapter.

5.1 Gene Frequency and Genotype Frequency

Throughout the entire book, we consider only diploid organisms. A diploid organism carries two copies of the homologous genome, one from the paternal parent and the other from the maternal parent. Each copy of the genome is called a haploid. Each locus of the genome, therefore, contains two alleles, one from each parent. Although each individual carries at most two different alleles, the entire population may have many different alleles, called multiallelic population. For simplicity, we only consider a population that contains two different alleles, called a biallelic

S. Xu, *Principles of Statistical Genomics*, DOI 10.1007/978-0-387-70807-2_5,

population. Let A_1 and A_2 be the two alleles of locus A in the population of interest. In the biallelic population, there are only three possible genotypes, denoted by A_1A_1, A_1A_2 and A_2A_2, respectively. Depending on the structure and the mating system of the population, the population may have different proportions of the three genotypes. Let $P_{11} = \Pr(A_1A_1)$, $P_{12} = \Pr(A_1A_2)$, and $P_{22} = \Pr(A_2A_2)$ be the frequencies of the three genotypes. The genotypes A_1A_2 and A_1A_1 contain one and two copies of allele A_1, respectively. Therefore, the frequency of allele A_1 in the population is $p_1 = \Pr(A_1) = P_{11} + \frac{1}{2}P_{12}$. The allelic frequency for A_2 is then $p_2 = \Pr(A_2) = P_{22} + \frac{1}{2}P_{12}$. These relationships hold regardless the population history and structure. However, to express genotypic frequencies as functions of the allele frequencies, some assumptions are required. In a large random-mating population, there is a unique relationship between genotype frequencies and gene frequencies, which is represented by $P_{11} = p_1^2$, $P_{12} = 2p_1p_2$, and $P_{22} = p_2^2$. This can be interpreted as independence of the two alleles joining together to form the genotype. The frequency of the heterozygote is $2p_1p_2$ because it contains two configurations of the same genotype, that is, A_1A_2 and A_2A_1, representing two different origins of the gametes. This particular relationship is represented by the binomial expansion,

$$(p_1 + p_2)^2 = p_1^2 + 2p_1p_2 + p_2^2 \tag{5.1}$$

corresponding to the event of

$$(A_1 + A_2)^2 = A_1A_1 + 2(A_1A_2) + A_2A_2 \tag{5.2}$$

If such a population undergoes no selection, no mutation, and no migration, the gene frequencies and genotypic frequencies will remain constant from generation to generation. Such a population is said to be in Hardy–Weinberg equilibrium (Hardy 1908; Weinberg 1908; Li 1955). If a large population is not in Hardy–Weinberg equilibrium, one generation of random mating will suffice to lead the population to Hardy–Weinberg equilibrium.

The Hardy–Weinberg equilibrium for a population with $k(k > 2)$ alleles is represented by $P_{ij} = 2p_ip_j$ for $i \neq j$ and $P_{ii} = p_i^2$ for $i = j$, where $P_{ij} = \Pr(A_iA_j)$ and $p_i = \Pr(A_i)$ for $i, j = 1, \ldots, k$.

Gene frequencies and genotypic frequencies are properties of a population. The genes studied are usually related to fitness and thus determine the adaption of the population to environmental changes and the evolution of the population. These are contents of population genetics. In quantitative genetics, we are interested in genes that determine the expression of quantitative traits. Therefore, we must first assign some value to a genotype and a value to an allele. These values are called genetic effects.

5.2 Genetic Effects and Genetic Variance

Each individual of a population has a phenotypic value for a particular quantitative trait. Assume that we can observe the genotypes of all individuals in the population. The genetic effect for genotype A_1A_1 is defined as the average phenotypic value of all individuals bearing genotype A_1A_1. This genotypic value is denoted by G_{11}. Similar notation applies to genotypes A_1A_2 and A_2A_2. The reason that G_{11} takes the average phenotypic value is explained as follows. Let Y_{11} be the phenotypic value for an individual with genotype A_1A_1, which can be expressed as

$$Y_{11} = G_{11} + E_{11}. \tag{5.3}$$

where E_{11} is a random environmental deviation. The environmental deviation varies from one individual to another, even though all the individuals have the same genotypic value. When we take the average value across all individuals of type A_1A_1, the equation becomes

$$\bar{Y}_{11} = G_{11} + \bar{E}_{11}. \tag{5.4}$$

For a sufficient number of individuals collected from this genotype, we have $\bar{E}_{11} \approx 0$ because positive and negative deviations tend to cancel out each other. This leads to $\bar{Y}_{11} = G_{11}$.

We now define three parameters as functions of the three genotypic values,

$$\begin{aligned} \mu &= \frac{1}{2}(G_{11} + G_{22}) \\ a &= G_{11} - \frac{1}{2}(G_{11} + G_{22}) \\ d &= G_{12} - \frac{1}{2}(G_{11} + G_{22}), \end{aligned} \tag{5.5}$$

where μ is called the midpoint value, a the additive effect, and d the dominance effect. The three genotypic values are then expressed as

$$\begin{aligned} G_{11} &= \mu + a \\ G_{12} &= \mu + d \\ G_{22} &= \mu - a \end{aligned} \tag{5.6}$$

We then express each genotypic value as a deviation from the midpoint value

$$\begin{aligned} \Phi_{11} &= G_{11} - \mu = a \\ \Phi_{12} &= G_{12} - \mu = d \\ \Phi_{22} &= G_{22} - \mu = -a \end{aligned} \tag{5.7}$$

Under Hardy–Weinberg equilibrium, the population mean of the genotypic values (expressed as deviations from the midpoint value) is

$$\begin{aligned}\mu_G = \mathrm{E}(\Phi) &= P_{11}\Phi_{11} + P_{12}\Phi_{12} + P_{22}\Phi_{22} \\ &= p_1^2 a + 2p_1p_2d + p_2^2(-a) \\ &= (p_1 - p_2)a + 2p_1p_2d \end{aligned} \tag{5.8}$$

and the variance of the genotypic values is

$$\sigma_G^2 = \mathrm{var}(\Phi) = \mathrm{E}(\Phi^2) - \mathrm{E}^2(\Phi), \tag{5.9}$$

where

$$\begin{aligned}\mathrm{E}(\Phi^2) &= P_{11}{\Phi_{11}}^2 + P_{12}{\Phi_{12}}^2 + P_{22}{\Phi_{22}}^2 \\ &= p_1^2 a^2 + 2p_1p_2d^2 + p_2^2(-a)^2 \\ &= (p_1^2 + p_2^2)a^2 + 2p_1p_2d^2 \end{aligned} \tag{5.10}$$

After some algebraic manipulations, we have

$$\sigma_G^2 = 2p_1p_2[a + (p_2 - p_1)d]^2 + (2p_1p_2d)^2. \tag{5.11}$$

5.3 Average Effect of Allelic Substitution

A single locus of an individual consists of two alleles, one from each of the two parents. When the individual reproduces, the two alleles will go to different gametes. The gametes will reunite in the next generation to form the genotypes of the next generation. Therefore, a genotype cannot be inherited from generation to generation. It is the allele (haplotype) that is passed from one generation to another. Therefore, we need to define the effect of an allele. Let us look at the following 2×2 table for the definition of the allelic effect (Table 5.1).

Table 5.1 Definitions for allelic effects and dominance deviations

	A_1 (p_1)	A_2 (p_2)	
A_1 (p_1)	A_1A_1 (p_1^2)	A_1A_2 (p_1p_2)	α_1
	$\Phi_{11} - \mu_G = \alpha_1 + \alpha_1 + \delta_{11}$	$\Phi_{12} - \mu_G = \alpha_1 + \alpha_2 + \delta_{12}$	
A_2 (p_2)	A_2A_1 (p_2p_1)	A_2A_2 (p_2^2)	
	$\Phi_{12} - \mu_G = \alpha_2 + \alpha_1 + \delta_{12}$	$\Phi_{22} - \mu_G = \alpha_2 + \alpha_2 + \delta_{22}$	α_2
	α_1	α_2	

Table 5.2 Breeding values and dominance deviations of the three genotypes in a Hardy–Weinberg population

Genotype	Frequency	Genotypic value	Breeding value	Dominance deviation
A_1A_1	p_1^2	a	$2p_2\alpha$	$-2p_2^2d$
A_1A_2	$2p_1p_2$	d	$(p_2-p_1)\alpha$	$2p_1p_2d$
A_2A_2	p_2^2	$-a$	$-2p_1\alpha$	$-2p_1^2d$

The effect of allele A_1 is defined as

$$\alpha_1 = \frac{(\Phi_{11}-\mu_G)\,p_1^2 + (\Phi_{12}-\mu_G)\,p_2p_1}{p_1} = p_2[a + d(p_2-p_1)] \tag{5.12}$$

and the effect of allele A_2 is defined as

$$\alpha_2 = \frac{(\Phi_{12}-\mu_G)\,p_1p_2 + (\Phi_{22}-\mu_G)\,p_2^2}{p_2} = -p_1[a + d(p_2-p_1)] \tag{5.13}$$

The difference between the two allelic effects is called the average effect of allelic substitution, denoted by α,

$$\alpha = \alpha_1 - \alpha_2 = a + (p_2-p_1)d \tag{5.14}$$

The sum of the two allelic effects included in a genotype is called the "breeding value," which is the expected genotypic value of the progeny of an individual bearing this genotype. Therefore, the breeding value for genotype A_1A_1 is $A_{11} = 2\alpha_1 = 2p_2\alpha$. The breeding values for the other two genotypes are $A_{12} = \alpha_1 + \alpha_2 = (p_2-p_1)\alpha$ and $A_{22} = 2\alpha_2 = -p_1\alpha$, respectively. The deviations of the actual genotypic values from the breeding values are called dominance deviations. The three dominance deviations are

$$\begin{aligned}
\delta_{11} &= \Phi_{11} - \mu_G - A_{11} = -2p_2^2d \\
\delta_{12} &= \Phi_{12} - \mu_G - A_{12} = 2p_1p_2d \\
\delta_{22} &= \Phi_{22} - \mu_G - A_{22} = -2p_1^2d
\end{aligned} \tag{5.15}$$

The genotypic values (G), the breeding values (A), and the dominance deviations (D) for the three genotypes are listed in Table 5.2.

5.4 Genetic Variance Components

One can verify that the expectations of both the breeding values and the dominance deviations are zero, i.e.,

$$\begin{aligned}
\mathrm{E}(A) &= P_{11}A_{11} + P_{12}A_{12} + P_{22}A_{22} \\
&= p_1^2(2p_2\alpha) + 2p_1p_2[(p_2-p_1)\alpha] + p_2^2(-2p_1\alpha) = 0
\end{aligned} \tag{5.16}$$

and

$$\begin{aligned} \mathrm{E}(D) &= P_{11}\delta_{11} + P_{12}\delta_{12} + P_{22}\delta_{22} \\ &= p_1^2(-2p_2^2 d) + 2p_1p_2(2p_1p_2 d) + p_2^2(-2p_1^2 d) = 0. \end{aligned} \tag{5.17}$$

This leads to

$$\sigma_A^2 = \mathrm{E}(A^2) = p_1^2(2p_2\alpha)^2 + 2p_1p_2[(p_2 - p_1)\alpha]^2 + p_2^2(-2p_1\alpha)^2 = 2p_1p_2\alpha^2 \tag{5.18}$$

and

$$\sigma_D^2 = \mathrm{E}(D^2) = p_1^2(-2p_2^2 d)^2 + 2p_1p_2(2p_1p_2 d)^2 + p_2^2(-2p_1\alpha)^2 = (2p_1p_2 d)^2 \tag{5.19}$$

Looking at the genetic variance given in (5.11), we found that the first part is σ_A^2 and the second part is σ_D^2. Therefore,

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2, \tag{5.20}$$

i.e., the total genetic variance has been partitioned into an additive variance component and a dominance variance component.

5.5 Heritability

The phenotypic value Y can be expressed by the following linear model:

$$Y = G + E = A + D + E, \tag{5.21}$$

where E is an environmental error with mean zero and variance σ_E^2. The phenotypic variance $\sigma_P^2 = \mathrm{var}(Y)$ is

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2 = \sigma_A^2 + \sigma_D^2 + \sigma_E^2. \tag{5.22}$$

The phenotypic variance contributed by the total genetic variance is called broad-sense heritability, denoted by

$$H^2 = \frac{\sigma_A^2 + \sigma_D^2}{\sigma_A^2 + \sigma_D^2 + \sigma_E^2}, \tag{5.23}$$

while the proportion contributed by the additive variance is called narrow-sense heritability, denoted by

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2 + \sigma_E^2}. \tag{5.24}$$

The narrow-sense heritability reflects the proportion of variance that is heritable. Therefore, it is a very important parameter to consider in developing a selection program for genetic improvement.

5.6 An F_2 Family Is in Hardy–Weinberg Equilibrium

An F_2 family initiated from the cross of lines A_1A_1 and A_2A_2 is in Hardy–Weinberg equilibrium, and thus, the theory developed in this chapter applies to an F_2 family. The allele frequencies are $p_1 = p_2 = \frac{1}{2}$, and the genotypic frequencies are $P_{11} = p_1^2 = \frac{1}{4}$, $P_{12} = 2p_1p_2 = \frac{1}{2}$, and $P_{22} = p_2^2 = \frac{1}{4}$. The average effect of allelic substitution is $\alpha = a + (p_2 - p_1)d = a$. Therefore,

$$\sigma_G^2 = 2p_1p_2\alpha^2 + (2p_1p_2d)^2 = \frac{1}{2}a^2 + \frac{1}{4}d^2. \tag{5.25}$$

The same result can be obtained from a different perspective. The genotypic value of an F_2 individual can be expressed as

$$G = \mu + Za + Wd, \tag{5.26}$$

where

$$Z = \begin{cases} +1 & \text{for } A_1A_1 \text{ with probability } \frac{1}{4} \\ 0 & \text{for } A_1A_2 \text{ with probability } \frac{1}{2} \\ -1 & \text{for } A_2A_2 \text{ with probability } \frac{1}{4} \end{cases} \tag{5.27}$$

and

$$W = \begin{cases} 0 & \text{for } A_1A_1 \text{ with probability } \frac{1}{4} \\ 1 & \text{for } A_1A_2 \text{ with probability } \frac{1}{2} \\ 0 & \text{for } A_2A_2 \text{ with probability } \frac{1}{4} \end{cases} \tag{5.28}$$

The genotypic variance is partitioned into

$$\sigma_G^2 = \sigma_Z^2 a^2 + \sigma_W^2 d^2 = \frac{1}{2}a^2 + \frac{1}{4}d^2. \tag{5.29}$$

where

$$\begin{aligned} \sigma_Z^2 &= \mathrm{E}(Z^2) - \mathrm{E}^2(Z) \\ &= \left[\frac{1}{4}(+1)^2 + \frac{1}{2}(0)^2 + \frac{1}{4}(-1)^2\right] - \left[\frac{1}{4}(+1) + \frac{1}{2}(0) + \frac{1}{4}(-1)\right]^2 \\ &= \frac{1}{2} - 0 = \frac{1}{2} \end{aligned} \tag{5.30}$$

and

$$\begin{aligned}\sigma_W^2 &= \mathtt{E}(W^2) - \mathtt{E}^2(W) \\ &= \left[\frac{1}{4}(0)^2 + \frac{1}{2}(1)^2 + \frac{1}{4}(0)^2\right] - \left[\frac{1}{4}(0) + \frac{1}{2}(1) + \frac{1}{4}(0)\right]^2 \\ &= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}\end{aligned} \tag{5.31}$$

Note that the covariance between Z and W is zero. Otherwise, a term $2\mathtt{cov}(Z, W)ad$ should be added to the genotype variance.