

INTRODUCTION TO QUANTITATIVE GENETICS

Introduction

So far in this course we have dealt almost entirely with the evolution of characters that are controlled by simple Mendelian inheritance at a single locus. We talked some about gametic disequilibrium and how allele frequencies change at two loci simultaneously, but in every other example we've considered we've imagined that we could understand something about evolution by examining the evolution of a single gene. That's the domain of classical population genetics.

For the next few weeks we're going to be exploring a field that's actually older than classical population genetics, although the approach we'll be taking to it involves the use of population genetic machinery. If you know a little about the history of evolutionary biology, you may know that after the rediscovery of Mendel's work in 1900 there was a heated debate between the "biometricians" (e.g., Galton and Pearson) and the "Mendelians" (e.g., de Vries, Correns, Bateson, and Morgan).

Biometricians asserted that the really important variation in evolution didn't follow Mendelian rules. Height, weight, skin color, and similar traits seemed to

- vary continuously,
- show blending inheritance, and
- show variable responses to the environment.

Since variation in such *quantitative traits* seemed to be more obviously related to organismal adaptation than the "trivial" traits that Mendelians studied, it seemed obvious to the biometricians that Mendelian geneticists were studying a phenomenon that wasn't particularly interesting.

Mendelians dismissed the biometricians, at least in part, because they seemed not to recognize the distinction between genotype and phenotype. It seemed to at least some of them that traits whose expression was influenced by the environment were, by definition, not inherited. Moreover, the evidence that Mendelian principles accounted for the inheritance of many discrete traits was incontrovertible.

Woltereck's experiments on *Daphnia* helped to show that traits whose expression is environmentally influenced may also be inherited. He introduced the idea of a *norm of reaction* to describe the observation that the same genotype may produce different phenotypes in different environments. When you fertilize a plant, for example, it will grow larger and more robust than when you don't. The phenotype an organism expresses is, therefore, a product of *both* its genotype and its environment.

Nilsson-Ehle's experiments on inheritance of kernel color in wheat showed how continuous variation and Mendelian inheritance could be reconciled. He demonstrated that what appeared to be continuous variation in color from red to white with blending inheritance could be understood as the result of three separate genes influencing kernel color that were inherited separately from one another. Fisher, in a paper that grew out of his undergraduate Honors thesis at Cambridge University, set forth the mathematical theory that describes how it all works. That's the theory of *quantitative genetics*, and it's what we're going to spend the next three weeks discussing.

An overview of where we're headed

Woltereck's ideas force us to realize that when we see a phenotypic difference between two individuals in a population there are three possible explanations for that difference:

1. The individuals have different genotypes.
2. The individuals developed in different environments.
3. The individuals have different genotypes *and* they developed in different environments.

This leads us naturally to think that phenotypic variation consists of two separable components, namely genotypic and environmental components.¹ Putting that into an equation

$$\text{Var}(P) = \text{Var}(G) + \text{Var}(E) \quad ,$$

where $\text{Var}(P)$ is the *phenotypic variance*, $\text{Var}(G)$ is the *genetic variance*, and $\text{Var}(E)$ is the *environmental variance*.² As we'll see in just a moment, we can also partition the genetic variance into components, the *additive genetic variance*, $\text{Var}(A)$, and the *dominance variance*, $\text{Var}(D)$.

¹We'll soon see that separating genotypic and environmental components is far from trivial.

²Strictly speaking we should also include a term for the interaction between genotype and environment, but we'll ignore that for the time being.

There's a surprisingly subtle and important insight buried in that very simple equation: Because the expression of a quantitative trait is a result both of genes involved in that trait's expression and the environment in which it is expressed, it doesn't make sense to say of a particular individual's phenotype that genes are more important than environment in determining it. You wouldn't have a phenotype without both. What we might be able to say is that when we look at some population of organisms some fraction of the phenotypic differences among them is due to differences in the genes they carry and that some fraction is due to differences in the environment they have experienced.³

It's often useful to talk about how much of the phenotypic variance is a result of additive genetic variance or of genetic variance.

$$h_n^2 = \frac{\text{Var}(A)}{\text{Var}(P)}$$

is what's known as the *narrow-sense heritability*. It's the proportion of phenotypic variance that's attributable to differences among individuals in their additive genotype,⁴ much as F_{st} can be thought of as the proportion of genotypic diversity that attributable to differences among populations. Similarly,

$$h_b^2 = \frac{\text{Var}(G)}{\text{Var}(P)}$$

is the *broad-sense heritability*. It's the proportion of phenotypic variance that's attributable to differences among individuals in their genotype. It is *not*, repeat *NOT*, a measure of how important genes are in determining phenotype. Every individual's phenotype is determined both by its genes and by its environment. It measures how much of the difference among individuals is attributable to differences in their genes.⁵ Why bother to make the distinction? Because, as we'll see, it's only the additive genetic variance that responds to natural selection. In fact,

$$R = h_n^2 S \quad ,$$

where R is the *response to selection* and S is the *selective differential*.

As you'll see in the coming weeks, there's a lot of stuff hidden behind these simple equations, including a lot of assumptions. But quantitative genetics is very useful. Its principles have been widely applied in plant and animal breeding for almost a century, and they have been increasingly applied in evolutionary investigations in the last thirty years. Nonetheless, it's useful to remember that quantitative genetics is a lot like a bikini. What it reveals is interesting, but what it conceals is crucial.

³When I put it this way, I hope it's obvious that I'm neglecting genotype-environment interactions, and that I'm oversimplifying quite a bit.

⁴Don't worry about what I mean by *additive genotype*—yet. We'll get to it soon enough.

⁵As we'll see later it can do this only for the range of environments in which it was measured.

Genotype	A_1A_1	A_1A_2	A_2A_2
Frequency	p^2	$2pq$	q^2
Genotypic value	x_{11}	x_{12}	x_{22}
Additive genotypic value	$2\alpha_1$	$\alpha_1 + \alpha_2$	$2\alpha_2$

Table 1: Fundamental parameter definitions for quantitative genetics with one locus and two alleles.

Partitioning the phenotypic variance

Before we worry about how to estimate any of those variance components I just mentioned, we first have to understand what they are. So let's start with some definitions (Table 1).⁶

You should notice something rather strange about Table 1 when you look at it. I motivated the entire discussion of quantitative genetics by talking about the need to deal with variation at many loci, and what I've presented involves only two alleles at a single locus. I do this for two reasons:

1. It's not too difficult to do the algebra with multiple alleles at one locus instead of only two, but it gets a little messy, and I'd rather avoid the mess.
2. Doing the algebra with multiple loci involves a *lot* of assumptions, which I'll mention when we get to applications, and the algebra is even worse than with multiple alleles at a single locus.

Fortunately, the basic principles extend with little modification to multiple loci, so we can see all of the underlying logic by focusing on one locus with two alleles where we have a chance of understanding what the different variance components mean.

Two terms in Table 1 will almost certainly be unfamiliar to you: *genotypic value* and *additive genotypic value*. Of the two, *genotypic value* is the easiest to understand (Figure 1). It simply refers to the average phenotype associated with a given genotype.⁷ The *additive genotypic value* refers to the average phenotype associated with a given genotype, as would be inferred from the *additive effect* of the alleles of which it is composed. That didn't help

⁶Warning! There's a lot of algebra between here and the end. It's unavoidable. You can't possibly understand what additive genetic variance is without it. I'll try to focus on principles, but a lot of the algebra that follows *is* necessary. Sorry about that.

⁷Remember. We're now considering traits in which the environment influences the phenotypic expression, so the same genotype can produce different phenotypes, depending on the environment in which it develops.

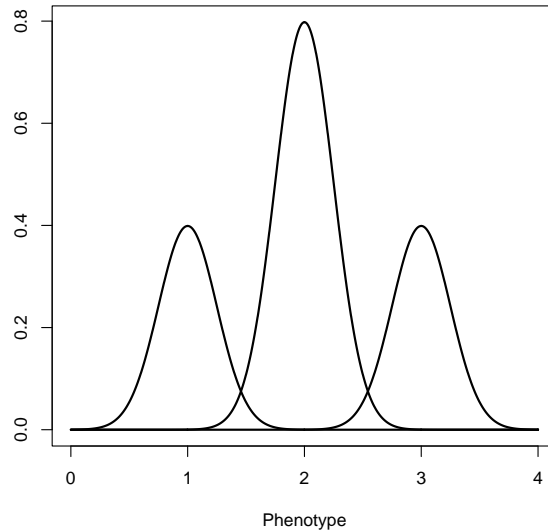


Figure 1: The phenotype distribution in a population in which the three genotypes at a single locus with two alleles occur in equal frequency. The A_1A_1 genotype has a mean trait value of 1, the A_1A_2 genotype has a mean trait value of 2, and the A_2A_2 genotype has a mean trait value of 3, but each genotype can produce a range of phenotypes with the standard deviation of the distribution being 0.25 in each case.

much, did it? That’s because I now need to tell you what we mean by the *additive effect* of an allele.⁸

The additive effect of an allele

In constructing Table 1 I used the quantities α_1 and α_2 , but I didn’t tell you where they came from. Obviously, the idea should be to pick values of α_1 and α_2 that give additive genotypic values that are reasonably close to the genotypic values. A good way to do that is to minimize the squared deviation between the two, weighted by the frequency of the genotypes. So our first big assumption is that genotypes are in Hardy-Weinberg proportions.⁹

The objective is to find values for α_1 and α_2 that minimize:

$$a = p^2[x_{11} - 2\alpha_1]^2 + 2pq[x_{12} - (\alpha_1 + \alpha_2)]^2 + q^2[x_{22} - 2\alpha_2]^2 \quad .$$

⁸Hold on. Things get even more interesting from here.

⁹As you should have noticed in Table 1.

To do this we take the partial derivative of a with respect to both α_1 and α_2 , set the resulting pair of equations equal to zero, and solve for α_1 and α_2 .¹⁰

$$\begin{aligned}\frac{\partial a}{\partial \alpha_1} &= p^2\{2[x_{11} - 2\alpha_1][-2]\} + 2pq\{2[x_{12} - (\alpha_1 + \alpha_2)][-1]\} \\ &= -4p^2[x_{11} - 2\alpha_1] - 4pq[x_{12} - (\alpha_1 + \alpha_2)] \\ \frac{\partial a}{\partial \alpha_2} &= q^2\{2[x_{22} - 2\alpha_2][-2]\} + 2pq\{2[x_{12} - (\alpha_1 + \alpha_2)][-1]\} \\ &= -4q^2[x_{22} - 2\alpha_2] - 4pq[x_{12} - (\alpha_1 + \alpha_2)]\end{aligned}$$

Thus, $\frac{\partial a}{\partial \alpha_1} = \frac{\partial a}{\partial \alpha_2} = 0$ if and only if

$$\begin{aligned}p^2(x_{11} - 2\alpha_1) + pq(x_{12} - \alpha_1 - \alpha_2) &= 0 \\ q^2(x_{22} - 2\alpha_2) + pq(x_{12} - \alpha_1 - \alpha_2) &= 0\end{aligned}\tag{1}$$

Adding the equations in (1) we obtain (after a little bit of rearrangement)

$$[p^2x_{11} + 2pqx_{12} + q^2x_{22}] - [p^2(2\alpha_1) + 2pq(\alpha_1 + \alpha_2) + q^2(2\alpha_2)] = 0 \quad .\tag{2}$$

Now the first term in square brackets is just the mean phenotype in the population, \bar{x} . Thus, we can rewrite equation (2) as:

$$\begin{aligned}\bar{x} &= 2p^2\alpha_1 + 2pq(\alpha_1 + \alpha_2) + 2q^2\alpha_2 \\ &= 2p\alpha_1(p + q) + 2q\alpha_2(p + q) \\ &= 2(p\alpha_1 + q\alpha_2) \quad .\end{aligned}\tag{3}$$

Now divide the first equation in (1) by p and the second by q .

$$p(x_{11} - 2\alpha_1) + q(x_{12} - \alpha_1 - \alpha_2) = 0\tag{4}$$

$$q(x_{22} - 2\alpha_2) + p(x_{12} - \alpha_1 - \alpha_2) = 0 \quad .\tag{5}$$

Thus,

$$\begin{aligned}px_{11} + qx_{12} &= 2p\alpha_1 + q\alpha_1 + q\alpha_2 \\ &= \alpha_1(p + q) + p\alpha_1 + q\alpha_2 \\ &= \alpha_1 + p\alpha_1 + q\alpha_2 \\ &= \alpha_1 + \bar{x}/2 \\ \alpha_1 &= px_{11} + qx_{12} - \bar{x}/2 \quad .\end{aligned}$$

¹⁰We won't bother with proving that the resulting estimates produce the minimum possible value of a . Just take my word for it.

Similarly,

$$\begin{aligned}
px_{12} + qx_{22} &= 2q\alpha_2 + p\alpha_1 + p\alpha_2 \\
&= \alpha_2(p + q) + p\alpha_1 + q\alpha_2 \\
&= \alpha_2 + p\alpha_1 + q\alpha_2 \\
&= \alpha_2 + \bar{x}/2 \\
\alpha_2 &= px_{12} + qx_{22} - \bar{x}/2 \quad .
\end{aligned}$$

α_1 is the additive effect of allele A_1 , and α_2 is the additive effect of allele A_2 . If we use these expressions, the additive genotypic values are as close to the genotypic values as possible, given the particular allele frequencies in the population.

Components of the genetic variance

Let's assume for the moment that we can actually measure the genotypic values. Later, we'll relax that assumption and see how to use the resemblance among relatives to estimate the genetic components of variance. But it's easiest to see where they come from if we assume that the genotypic value of each genotype is known. If it is then, writing V_g for $\text{Var}(G)$

$$\begin{aligned}
V_g &= p^2[x_{11} - \bar{x}]^2 + 2pq[x_{12} - \bar{x}]^2 + q^2[x_{22} - \bar{x}]^2 & (6) \\
&= p^2[x_{11} - 2\alpha_1 + 2\alpha_1 - \bar{x}]^2 + 2pq[x_{12} - (\alpha_1 + \alpha_2) + (\alpha_1 + \alpha_2) - \bar{x}]^2 \\
&\quad + q^2[x_{22} - 2\alpha_2 + 2\alpha_2 - \bar{x}]^2 \\
&= p^2[x_{11} - 2\alpha_1]^2 + 2pq[x_{12} - (\alpha_1 + \alpha_2)]^2 + q^2[x_{22} - 2\alpha_2]^2 \\
&\quad + p^2[2\alpha_1 - \bar{x}]^2 + 2pq[(\alpha_1 + \alpha_2) - \bar{x}]^2 + q^2[2\alpha_2 - \bar{x}]^2 \\
&\quad + p^2[2(x_{11} - 2\alpha_1)(2\alpha_1 - \bar{x})] + 2pq[2(x_{12} - \{\alpha_1 + \alpha_2\})(\{\alpha_1 + \alpha_2\} - \bar{x})] \\
&\quad + q^2[2(x_{22} - 2\alpha_2)(2\alpha_2 - \bar{x})] \quad . & (7)
\end{aligned}$$

There are two terms in (7) that have a biological (or at least a quantitative genetic) interpretation. The term on the first line is the average squared deviation between the genotypic value and the additive genotypic value. It will be zero only if the effects of the alleles can be decomposed into strictly additive components, i.e., only if the phenotype of the heterozygote is exactly intermediate between the phenotype of the two homozygotes. Thus, it is a measure of how much variation is due to non-additivity (dominance) of allelic effects. In short, the *dominance genetic variance*, V_d , is

$$V_d = p^2[x_{11} - 2\alpha_1]^2 + 2pq[x_{12} - (\alpha_1 + \alpha_2)]^2 + q^2[x_{22} - 2\alpha_2]^2 \quad . \quad (8)$$

Similarly, the term on the second line of (7) is the average squared deviation between the additive genotypic value and the mean genotypic value in the population. Thus, it is a measure of how much variation is due to differences between genotypes in their additive genotype. In short, the additive genetic variance, V_a , is

$$V_a = p^2[2\alpha_1 - \bar{x}]^2 + 2pq[(\alpha_1 + \alpha_2) - \bar{x}]^2 + q^2[2\alpha_2 - \bar{x}]^2 \quad . \quad (9)$$

What about the terms on the third and fourth lines of the last equation in 7? Well, they can be rearranged as follows:

$$\begin{aligned} & p^2[2(x_{11} - 2\alpha_1)(2\alpha_1 - \bar{x})] + 2pq[2(x_{12} - \{\alpha_1 + \alpha_2\})(\{\alpha_1 + \alpha_2\} - \bar{x})] \\ & \quad + q^2[2(x_{22} - 2\alpha_2)(2\alpha_2 - \bar{x})] \\ & = 2p^2(x_{11} - 2\alpha_1)(2\alpha_1 - \bar{x}) + 4pq[x_{12} - (\alpha_1 + \alpha_2)][(\alpha_1 + \alpha_2) - \bar{x}] \\ & \quad + 2q^2(x_{22} - 2\alpha_2)(2\alpha_2 - \bar{x}) \\ & = 4p^2(x_{11} - 2\alpha_1)[\alpha_1 - (p\alpha_1 + q\alpha_2)] \\ & \quad + 4pq[x_{12} - (\alpha_1 + \alpha_2)][(\alpha_1 + \alpha_2) - 2(p\alpha_1 + q\alpha_2)] \\ & \quad + 4q^2(x_{22} - 2\alpha_2)[\alpha_2 - (p\alpha_1 + q\alpha_2)] \\ & = 4p[\alpha_1 - (p\alpha_1 + q\alpha_2)][p(x_{11} - 2\alpha_1) + q(x_{12} - \{\alpha_1 + \alpha_2\})] \\ & \quad + 4q[\alpha_2 - (p\alpha_1 + q\alpha_2)][p(x_{11} - 2\alpha_1)p + q(x_{12} - \{\alpha_1 + \alpha_2\})] \\ & = 0 \end{aligned}$$

Where we have used the identities $\bar{x} = 2(p\alpha_1 + q\alpha_2)$ [see equation (3)] and

$$\begin{aligned} p(x_{11} - 2\alpha_1) + q(x_{12} - \alpha_1 - \alpha_2) & = 0 \\ q(x_{22} - 2\alpha_2) + p(x_{12} - \alpha_1 - \alpha_2) & = 0 \end{aligned}$$

[see equations (4) and (5)]. In short, we have now shown that the total genotypic variance in the population, V_g , can be subdivided into two components—the additive genetic variance, V_a , and the dominance genetic variance, V_d . Specifically,

$$V_g = V_a + V_d \quad ,$$

where V_g is given by the first line of (6), V_a by (9), and V_d by (8).

An alternative expression for V_a

There's another way to write the expression for V_a when there are only two alleles at a locus. I show it here because it comes in handy some times.

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

Genotype	A_1A_1	A_1A_2	A_2A_2
Genotypic value	0	1	2

Table 2: A set of perfectly additive genotypic values. Note that the genotypic value of the heterozygote is exactly halfway between the genotypic values of the two homozygotes.

$$\begin{aligned}
&= 4p^2\alpha_1^2 + 2pq(\alpha_1 + \alpha_2)^2 + 4q^2\alpha_2^2 - 4(p^2\alpha_1^2 + 2pq\alpha_1\alpha_2 + q^2\alpha_2^2) \\
&= 2pq[(\alpha_1 + \alpha_2)^2 - 4\alpha_1\alpha_2] \\
&= 2pq[(\alpha_1^2 + 2\alpha_1\alpha_2 + \alpha_2^2) - 4\alpha_1\alpha_2] \\
&= 2pq[\alpha_1^2 - 2\alpha_1\alpha_2 + \alpha_2^2] \\
&= 2pq[\alpha_1 - \alpha_2]^2 \\
&= 2pq\alpha^2
\end{aligned}$$

An example: the genetic variance with known genotypes

We've been through a lot of algebra by now. Let's run through a couple of numerical examples to see how it all works. For the first one, we'll use the set of genotypic values in Table 2

For $p = 0.4$

$$\begin{aligned}
\bar{x} &= (0.4)^2(0) + 2(0.4)(0.6)(1) + (0.6)^2(2) \\
&= 1.20
\end{aligned}$$

$$\begin{aligned}
\alpha_1 &= (0.4)(0) + (0.6)(1) - (1.20)/2 \\
&= 0.0
\end{aligned}$$

$$\begin{aligned}
\alpha_2 &= (0.4)(1) + (0.6)(2) - (1.20)/2 \\
&= 1.0
\end{aligned}$$

$$\begin{aligned}
V_g &= (0.4)^2(0 - 1.20)^2 + 2(0.4)(0.6)(1 - 1.20)^2 + (0.6)^2(2 - 1.20)^2 \\
&= 0.48
\end{aligned}$$

$$\begin{aligned}
V_a &= (0.4)^2[2(0.0) - 1.20]^2 + 2(0.4)(0.6)[(0.0 + 1.0) - 1.20]^2 + (0.6)^2[2(1.0) - 1.20]^2 \\
&= 0.48
\end{aligned}$$

$$V_d = (0.4)^2[0 - 2(0.0)]^2 + 2(0.4)(0.6)[1 - (0.0 + 1.0)]^2 + (0.6)^2[2 - 2(1.0)]^2$$

Genotype	A_1A_1	A_1A_2	A_2A_2
Genotypic value	0	0.8	2

Table 3: A set of non-additive genotypic values. Note that the genotypic value of the heterozygote is closer to the genotypic value of A_1A_1 than it is to the genotypic value of A_2A_2 .

$$= 0.00 \quad .$$

For $p = 0.2$, $\bar{x} = 1.60$, $V_g = V_a = 0.32$, $V_d = 0.00$. You should verify for yourself that $\alpha_1 = 0$ and $\alpha_2 = 1$ for $p = 0.2$. If you are ambitious, you could try to prove that $\alpha_1 = 0$ and $\alpha_2 = 1$ for *any* allele frequency.

For the second example we'll use the set of genotypic values in Table 3.

For $p = 0.4$

$$\begin{aligned} \bar{x} &= (0.4)^2(0) + 2(0.4)(0.6)(0.8) + (0.6)^2(2) \\ &= 1.104 \end{aligned}$$

$$\begin{aligned} \alpha_1 &= (0.4)(0) + (0.6)(0.8) - (1.104)/2 \\ &= -0.072 \end{aligned}$$

$$\begin{aligned} \alpha_2 &= (0.4)(0.8) + (0.6)(2) - (1.104)/2 \\ &= 0.968 \end{aligned}$$

$$\begin{aligned} V_g &= (0.4)^2(0 - 1.104)^2 + 2(0.4)(0.6)(0.8 - 1.104)^2 + (0.6)^2(2 - 1.104)^2 \\ &= 0.5284 \end{aligned}$$

$$\begin{aligned} V_a &= (0.4)^2[2(-0.072) - 1.104]^2 + 2(0.4)(0.6)[(-0.072 + 0.968) - 1.104]^2 \\ &\quad + (0.6)^2[2(0.968) - 1.104]^2 \\ &= 0.5192 \end{aligned}$$

$$\begin{aligned} V_d &= (0.4)^2[0 - 2(-0.072)]^2 + 2(0.4)(0.6)[0.8 - (-0.072 + 0.968)]^2 \\ &\quad + (0.6)^2[2 - 2(0.968)]^2 \\ &= 0.0092 \quad . \end{aligned}$$

To test your understanding, it would probably be useful to calculate \bar{x} , α_1 , α_2 , V_g , V_a , and V_d for one or two other allele frequencies, say $p = 0.2$ and $p = 0.8$. Is it still true that α_1 and α_2 are independent of allele frequencies? If you are *really* ambitious you could try to

prove that α_1 and α_2 are independent of allele frequencies if and only if $x_{12} = (x_{11} + x_{12})/2$, i.e., when heterozygotes are exactly intermediate.

Creative Commons License

These notes are licensed under the Creative Commons Attribution-NonCommercial-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.