

Lecture 18. Genetics of complex traits (quantitative genetics)

PHENOTYPES ARE NOT ALWAYS A DIRECT REFLECTION OF GENOTYPES

Some alleles are only expressed in some environments, or have variable expression for other reasons..

Mendelian Traits that are also Affected by the Environment

The effects of these mutations are usually only apparent at high temperatures. Siamese cats have a mutation in the *C* gene controlling dark pigment formation. The c^h allele of this gene is heat sensitive. The c^h allele can make dark pigment at low but not high temperatures. The **permissive temperature** for dark-color occurs at the extremities, and a **restrictive temperature** occurs in the body core.



Nutritional effects

Phenylketonuria is a human nutritional defect that can lead to severe physical and mental disorders in children, but only if they consume phenylalanine. The mutation prevents individuals from metabolizing this amino acid. The disease phenotype can be avoided by eliminating phenylalanine from the diet.

18.1. QUANTITATIVE TRAITS

Most phenotypic traits in plants and animals are affected by many genes (size, weight, shape, lifespan, physiological traits, fecundity). Often, it is not feasible to determine the number of genes affecting a particular trait, and the individual effects of genes on the phenotype. Many of these traits can be measured on a quantitative, rather than a qualitative, scale. This is where the terms quantitative trait and quantitative genetics come from.

Quantitative traits:

1. Have continuous distributions, not discrete classes
2. Are usually affected by many genes (**polygenic**)
3. Are also affected by environmental factors

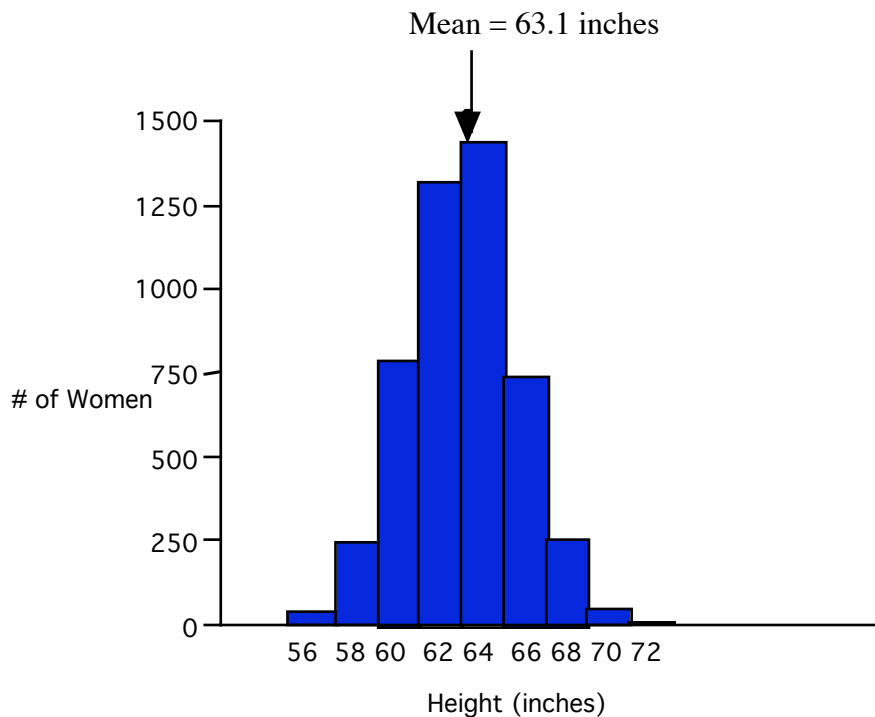
EXAMPLE: COLOR OF WHEAT KERNELS

This trait is determined by **two genes** that contribute “doses” of red pigment, and display partial dominance (heterozygotes intermediate). Each allele with the subscript “1” contributes 1 dose of red pigment. This trait demonstrates **additive effects** among different alleles at a single locus, and among alleles at different genetic loci. Genes with subscript “2” don't contribute any red pigment.

With **two additive genes** you have **five phenotypic classes** in the F2 offspring of true-breeding strains, instead of the three classes you would see if you had only one additive gene. Some classes are composed of several genotypes that are indistinguishable in phenotype. The number of phenotypic classes expected when there are n diallelic additive loci is $(2n+1)$. So, if you have four additive genes, you should expect nine phenotypic classes in the F2 offspring of pure-breeding strains.

As the number of additive genes increases, the **distribution of phenotypes** becomes more **continuous**. In addition, as stated above, most quantitative traits are also affected by the environment. Environmental effects may obscure genetically-caused differences between phenotypic classes. For example, nutrition affects the adult size in many organisms. The distribution of phenotypes then becomes even more continuous. The distribution of quantitative traits often approximates a **bell-shaped curve** when you plot phenotypic value (height, for example) against the frequency of individuals in particular phenotypic classes. Such a plot is called a **frequency histogram**.

EXAMPLE: DISTRIBUTION OF HEIGHT IN 5000 BRITISH WOMEN:



In this graph, the column designated "62" includes all individuals with heights between 61 and 63 inches, "64" includes all individuals with heights between 63 and 65 inches, and so on.

This curve has two easily measured properties--the **MEAN** (average), and the **VARIANCE** (variation about the mean). Curves with the same mean may have very different variances.

If you were to measure the heights of a sample of about 5000 different British women, you would get a curve similar to the one shown above. To calculate the mean of a distribution, you need to sum up all the different heights, then divide by the number of individuals:

mean = average = sum of heights / number of women: $\bar{X} = \sum (x_i) / N = 63.1$ inches. Another way of calculating the mean directly from a histogram (if you don't have available a list of the heights of all 5000 women, for example) is

$$\bar{x} = \frac{\sum_{i=1}^{n=\text{\#classes}} f_i x_i}{\sum_{i=1}^n f_i}$$

The variance is calculated from the sum of **squared deviations of individuals from the mean**:

$$\text{Var} = \frac{\sum (x_i - \bar{x})^2}{(N-1)} = 7.24 \text{ in}^2 \quad \text{or}$$

$$\text{Var} = \frac{\sum_{i=1}^n f_i (x_i - \bar{x})^2}{(N-1)}$$

where N = total number of individuals (5000). We will use this concept of variance extensively in our discussion of quantitative genetics.

Quantitative traits are influenced by genetics and by the environment. Under some circumstances, we can partition the phenotypic variance in quantitative traits into variance that is associated with genetic effects, and variance that is associated with environmental effects:

$$V_P = V_G + V_E$$

Where V_P is the total **phenotypic variance**, V_G is the **genetic variance**, and V_E is the **environmental variance**.

18.2. MODELS OF QUANTITATIVE INHERITANCE

HERITABILITY, GENETIC AND PHENOTYPIC VARIANCE

Some genetic variation is **heritable** because it can be **passed from parent to offspring**. Some genetic variation is not strictly heritable, because it is due to dominance or epistatic interactions that are not directly passed from parent to offspring. For example, if one allele is dominant to another, the phenotype of a heterozygous parent is determined in part by the dominance interaction between the two alleles. Since a sexually-reproducing parent only passes on a single allele to its offspring, the offspring does not inherit its whole genotype from a single parent. So it does not inherit the dominance interaction, just the effect of a single allele.

For example, assume that the size of a bird (measured as wing span) is determined by two genes, **one with complete dominance of one allele over the other, and one with additive effects**. Interactions *between* the two different genes are additive. Birds of genotype *aabb* have 16 cm long wing spans, and birds with other genotypes are somewhat larger, depending upon which alleles they have at each locus. The following table shows the increase in wing span conferred by different genotypes at each locus:

	Dominance effects			Additive effects					
	AA	Aa	aa		BB	Bb	bb		
	+2	+2	+0		+2	+1	+0		
F1:				Aa Bb	x	Aa Bb			
				19 cm		19 cm			
F2 Genotypes:	AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb
Genotypic Effects	+4	+3	+2	+4	+3	+2	+2	+1	+0
Phenotype (cm)	20	19	18	20	19	18	18	17	16
F2 proportions:	1/16	2/16	1/16	2/16	4/16	2/16	1/16	2/16	1/16

What is the phenotypic mean and variance, assuming we have exactly 16 F2 offspring in the proportions given above?

$$\bar{x} = \frac{\sum_{i=1}^{n=\text{#classes}} f_i x_i}{\sum_{i=1}^n f_i} \quad \text{Var} = \frac{\sum_{i=1}^n f_i (x_i - \bar{x})^2}{(N-1)}$$

Mean = [20*(3) + 19*(6) + 18*(4) + 17*(2) + 16*(1)]/16 = 296/16 = 18.5 inches

Variance in phenotypes = [(20-18.5)²*(3) + (19-18.5)²*(6) + (18-18.5)²*(4) + (17-18.5)²*(2) + (16-18.5)²*(1)]/(16-1) = 20/15 = 1.333 in²

SOURCES OF VARIANCE

In this case, all the **phenotypic variance** is due to differences in genotypes, so it is all **genetic variance**. However, some of the difference between the genotypes is due to additive effects of alleles, and some are due to dominance effects of alleles. For example, the difference in phenotype between *aabb*, *aaBb*, and *aaBB* (16, 17, and 18 respectively) are only due to the additive interactions between different alleles at the B locus. However, the difference in phenotype between *AABB*, *AaBB*, and *aaBB* (20, 20, and 18, respectively) is due to both **additive effects** (difference between aa and AA is 2 “units” of the A allele and the difference in height is 2 inches--so the average effect of a “unit” of A is one inch) and **dominance effects at the A locus**. Therefore, some of the genetic variation is due to additive effects and some is due to dominance effects. **Therefore, some of the genetic variance in the F2 is additive variance, and some is dominance variance.** In fact, although the derivation of the equation is beyond the scope of this course, the amount of additive (V_A) vs. dominance variance (V_D) is easily calculated in this example: $V_A = 1.0$ and $V_D = 0.333$; ($V_A = 2\sum p_i q_i a_i^2$; $V_D = \sum (2pqd^2)$, where the sum is over each locus contributing to the trait). In this simplified example, there are no environmental effects, so the **environmental variance (V_E) is zero**. If there were environmental effects, the phenotypic variance would be the sum of the genetic and environmental variance.

The additive and non-additive genetic variance (V_A and V_D) together are known as the genotypic or genetic variance, which is abbreviated V_G . In the bird-wing example, $V_G = 1.333$, which is also the total phenotypic variance is abbreviated $V_P = V_G + V_E$.

These considerations (and the mathematical fact that variances due to independent sources of variation can be summed) lead to the following equations:

$$V_P = V_G + V_E$$

$$V_P = V_A + V_D + V_E$$

In the bird-wing example, all the phenotypic variance was due to genetic variance (there were no environmental effects on the trait). We can also imagine a trait that has no genetic variance, so that the differences between individuals are due only to environmental effects. For example, assume that you grow a clonal strain of corn in a field, so that every individual has exactly the same genotype. At the end of the growing season, there will be differences in height (V_P) among different plants, but all the differences will be due to local soil, moisture, temperature, and light conditions. Thus, all the height differences are due to environmental variance (V_E).

A useful thing for plant and animal breeders to know is, for any trait of interest, how much of the phenotypic variability of that trait is due to genetic variance, and how much is due to non-genetic environmental factors.

This is the **broad-sense heritability**: $H^2 = V_G/V_P$

It is even more useful to know what proportion of the phenotypic variation is due to additive genetic effects.

The **heritability (narrow-sense)** of a trait is defined as the proportion of the total phenotypic variation that is due to heritable (additive genetic) effects:

$$V_A/V_P = h^2$$

h^2 is the proportion of variability that can be passed on from parent to offspring, and this is why this quantity is of interest to animal and plant breeders. They want to know whether selection on the parents will produce inherited changes in the offspring. **The degree to which selection on parents will produce inherited changes in the offspring is determined by the narrow sense heritability of the trait.**

When $h^2 = 0$, none of the phenotypic variance among individuals is due to additive genetic differences ($V_A=0$) and offspring will not closely resemble their parents for the trait of interest for genetic reasons. Can you think of any traits in humans, other animals, or plants that probably have zero heritability?

Note: Offspring can resemble their parents for non-genetic reasons. E.g., consider the fact that human offspring tend to have the same religious affiliation as their parents. What do you think is the source of this kind of resemblance between parents and offspring?

When $h^2 = 1$, all the variation among individuals is due to heritable genetic differences ($V_P=V_A$) and offspring will resemble their parents very closely. Can you think of any traits that probably have high heritability?

18.3 HERITABILITY AND THE RESPONSE TO SELECTION

h^2 is a very useful number, because it allows us to predict how a population will respond to artificial or natural selection. Animal and plant breeders rely heavily on the methods of quantitative genetics to improve economically-important traits such as milk yield in dairy cows, speed and endurance in race horses, protein and oil content in corn, or size, color, and ripening time in tomatoes.

The relationship between h^2 and the response to selection, is given by:

$$R = h^2S$$

R is the **response to selection**, given by the difference between **the population mean before selection** and the **mean of the offspring of selected parents** after one generation of selection.

S is the **selection coefficient**, given by the difference between **the unselected population mean**, and the **mean of the selected parents**.

If we know the heritability of a trait, and the strength of artificial selection applied to it, we can predict the response to selection. For example, in a population of the tobacco plant, *Nicotiana longiflora*, the mean flower length of the unselected population is 83 cm, and the mean of selected parents is 90 cm. If the heritability of the trait is known from other studies to be 0.64, then the predicted response to selection is given by $R = h^2S$, where $S = 90-83=7$ cm.

$$R = 0.64 * 7 \text{ cm} = 4.48 \text{ cm.}$$

The mean of the population after one generation of selection is therefore predicted to be:

$$\text{New mean} = \text{Unselected mean} + \text{predicted response} = 83 \text{ cm} + 4.48 \text{ cm} = 87.48 \text{ cm}$$

The actual result of this selection experiment was that the mean corolla length after one generation of selection was **87.9 cm**.

Because of this relationship between R, S, and h^2 , one way to estimate an unknown heritability is to measure the response to a known amount of artificial selection.

$$R = h^2 S \qquad h^2 = R/S$$

In fact, the value of $h^2 = 0.64$ used above was calculated from the results of a previous episode of artificial selection in tobacco. The unselected population mean was 70 cm, and the mean of the selected parents was 81 cm. Thus, the difference between the mean of selected parents and the unselected mean (S) was $81-70=11$ cm. The mean of the offspring of the selected parents was 77 cm. So the response (R) to a single generation of selection was $77-70=7$ cm. The estimate of h^2 is therefore

$$h^2 = R/S = 7/11 = 0.636.$$

So about **64 percent** of the total phenotypic variation in corolla length is due to heritable genetic differences. The remainder of the variation (**36 percent**) is due to environmental (or possibly to non-additive genetic) sources of variation.

18.4 INTERACTIONS BETWEEN GENOTYPE AND THE ENVIRONMENT (GXE INTERACTION)

So far, we've only considered cases in which the phenotype is affected only by the genotype. Sometimes the phenotype of a given genotype may also be affected by the environment in which an individual develops. **Most quantitative traits are influenced by both genes and the environment.** Quantitative traits are influenced by **genetic factors** in the form of alternate genotypes for one or more genes. They are also influenced by **environmental factors** such as nutrition, climate, density, or social interactions.

Some organismal traits are mainly determined by the genotype and are very little affected by the environment, and some are mainly determined by the environment, and little affected by genes. Quantitative geneticists can often determine the relative contribution of genetic and environmental factors by raising individuals with the same genotypes in different environments. For example, identical twins separated at birth represent individuals with the same genotype that experience different environments. In humans, however, adopted children are often placed with adoptive parents of same social, economic, racial, and religious backgrounds as their biological parents, and they are sometimes even placed with relatives. Also, identical twins look alike, so if a human's social environment depends at all on appearance, then identical twins may experience more similar social environments than unrelated individuals. So the environments of separated twins are not identical, but they may be similar in many ways.

Less problematic results can be obtained by using organisms that can be raised in large numbers, for which the genotype and environment of the organism can be controlled.

In the 1940's, researchers collected cuttings from *Potentilla glandulosa* (cinquefoil) at different altitudes. Some were collected in the Pacific coast range (**elevation 100 m**), some in mid-altitude in the Sierra Nevada mountains (**elevation 1400 m**), and some in the high Sierra Nevada (**elevation 3000 m**). They then grew **clones** from each elevation in gardens at the same sites. So genotypes representative of each elevation were grown at all three different elevations. This figure suggests that plant height is affected by both genotype and environment, since each genotype has a different height in each environment, and the effects of the environment are different in the three genotypes.

Another example of a trait that is affected by both genotype and environment is maze-brightness in rats. See class overhead for the results of maze-running tests in two different strains of lab rat. One strain had been selected for many generations to be very good at running mazes, and the other had been selected for poor performance in maze-running. Rats from both strains were then raised in three different environments. One was the standard lab environment that their ancestors had been raised in, one was an environment in which visual stimuli were reduced compared with a normal lab, and the third was an environment which contained greater than normal amounts of visual stimulation. The figure shows the number of errors made by rats of the two strains when they were run on the same maze. Obviously, both the strain of the rat, and the environment that it was raised in had very strong effects on the ability of the rats to run the maze.

Question: If two populations have different means (phenotypic), is the difference due to different environment or to different genetic factors? If individuals from the 2 populations can be raised in the same environment, we can answer this question, like we did in the cinquefoil example. If phenotypic differences remain when individuals from different populations are raised in the same environment, then differences between the populations are due, in part, to genetic differences. If

phenotypic differences vanish when individuals are raised in the same environment, then the differences are due entirely to environmental differences.

Question: Could you perform such an experiment on humans? Does this give you some insight into the “nature vs. nurture” debate about human behavior, IQ, etc.

18.5. QUANTITATIVE TRAITS IN HUMANS

TWIN STUDIES

Obviously, selection experiments cannot be performed in humans, but in some cases we would like to know if diseases or other traits are affected by genetic factors or by environmental factors, or both. One method that has frequently been used to try to address these questions is the **twin-study** approach.

Identical twins arise from the splitting of a single fertilized egg, and are genetically identical. They have the same alleles at all genetic loci. Theoretically, any phenotypic differences between identical twins are environmental.

Fraternal twins arise from two fertilized eggs, and have the same genetic relatedness as ordinary siblings (on average they share 50% of all alleles at all loci because of independent assortment of chromosome pairs at meiosis--review meiosis if you don't understand this). Therefore, phenotypic differences between fraternal twins can be due to both environmental and genetic differences.

One way to estimate heritability from twin data is to measure a trait in a large number of identical and fraternal twin pairs. One can then use the **correlation** between twin pairs for the trait concerned. In this case the correlation between co-twins gives an estimate of the heritability of a trait. If the heritability is high, identical twins will normally be very similar for a trait, and fraternal twins will be less similar. If the heritability is low, identical twins may not be much more similar than fraternal twins. If variation for a trait is completely heritable, then identical twins should have a correlation near 1, and fraternal twins should have a correlation near 0.5, since they will be similar for about 50% of the genetically variable loci.

If we can measure a trait in many sets of identical and fraternal twins, we can calculate the correlation for both kinds of twins

A measure of the broad-sense heritability, H^2 , is given by the difference between the two correlations: $H^2 = 2 * (r_i - r_f)$. Actually, this calculation **overestimates** H^2 , by an amount equal to $1/2 (V_D/V_P)$, so the heritabilities calculated by this formula should be interpreted as maximal possible values. The table below gives correlations between identical and fraternal human twins for several different traits.

Human Traits	Correlations		Heritability = H^2
	<i>Identical Twins</i>	<i>Fraternal Twins</i>	
Fingerprint ridge count	0.96	0.47	0.98
Height	0.90	0.57	0.66
IQ score	0.83	0.66	0.34
Social maturity score	0.97	0.89	0.16

PROBLEMS WITH USING TWIN STUDIES TO ESTIMATE HERITABILITY IN HUMANS

In addition to the fact that $2*(r_i - r_f)$ overestimates H^2 by a factor that is usually unknown, there are several additional sources of error in such studies.

Genotype-environment interaction (some genotypes are affected differently by the environment than are other genotypes--remember the rat maze-brightness example). Since different genotypes are lumped together in the category 'fraternal twins', but not in 'identical twins', GxE interaction lowers the correlation between fraternal twins, but not identical twins (**r_f is underestimated**).

An important assumption of these calculations is that identical twins and fraternal twins experience the same degree of environmental similarity. This assumption is likely to be violated because of:

Greater similarity in the treatment of identical twins by parents, teachers, and peers, resulting in inflated correlations between identical twins (**r_i is overestimated**).

Frequent sharing of embryonic membranes between identical twins, resulting in a more similar intrauterine **environment** (**r_i is overestimated**).

Different sexes in half of the pairs of fraternal twins, in contrast with the same sex of identical twins (r_f is underestimated).

Finally,

Many of these studies are based on small samples, and so the estimates are subject to large deviations from the 'true' value due to chance.

Therefore, estimates of heritability derived from human twin studies should be considered very approximate, and probably too high.

HERITABILITY CANNOT BE USED TO MAKE INFERENCES ABOUT DIFFERENCES BETWEEN POPULATIONS.

Another important fact to keep in mind is that measurements of heritability within populations do not give us any information about the causes of differences between populations.

Example: There are two farms where dairy cattle are raised, and h^2 has been estimated **within** each of these populations. h^2 tells us about variation among individuals **within** each farm, because only individuals within a farm were used in each estimate.

	Jones's Farm	Smith's Farm
h^2	1.0	1.0
milk yield	3 qts per day	2 qts per day

What does this tell you about the difference in milk yield between Smith's and Jones's cows?

18.5 MOLECULAR BIOLOGY AND QUANTITATIVE GENETICS

Many (probably most) organismal traits of medical, economic, and evolutionary interest are quantitative traits: susceptibility to heart disease, stroke, cancer, and diabetes in humans, yield, drought tolerance, and pest and disease resistance in crops, productivity, performance and disease resistance in farm animals.). Because quantitative traits are influenced by many genes, geneticists who study these traits would like to know the where these genes are located and what the genes do. Despite the fact that

the human genome has been sequenced, we do not know the complete function of most of the genes in the genome, and we do not yet know the complete genetic basis for any quantitative trait.

However, molecular biology has provided the tools necessary to locate at least some of the genes affecting quantitative traits. The process of finding these genes is called **quantitative trait locus mapping (QTL mapping)**. The basic idea is similar to the kinds of genetic mapping that we have already discussed in this class. In two and three-locus mapping problems, we were dealing with **discrete** traits, while in QTL mapping, we deal with **quantitative** traits.

Example: QTL mapping of fruit weight in tomato.

P: Large fruit strain X Small fruit strain the strains differ at **DNA-based genetic markers** such as RFLPs or microsatellites which can be scored on a gel

F1: F1 hybrid (has intermediate resistance) F1 is heterozygous for the DNA markers

F2: Some individuals have the genetic markers from small strain, while other individuals have the markers from the large strain. If we measure fruit weight and score the molecular markers in enough individuals, most of the possible combinations of small vs. large strain alleles will be found in different individuals. Then we can use a statistical procedure called **multiple regression** to look for associations between the fruit size of an individual and the markers that individual has. Markers that are statistically associated with high (or low) fruit weight are markers that are close to genes that affect fruit weight. If we have a large number of markers, this kind of mapping can be very precise, and we can eventually identify specific genes that affect fruit weight.

A series of such experiments in tomato has revealed at least 28 genes affecting the fruit weight. Of these, one has been cloned and transferred into different strains. When the allele from the “large” variety is transferred to the “small” variety, fruit weight increases by 30%.