Quantitative Genetics

Bruce Walsh, *University of Arizona, Tucson, Arizona, USA*

Almost any trait that can be defined shows variation, both within and between populations. Quantitative genetics is concerned with the analysis of the genetic and environmental basis of this variation. Classical genetics typically deals with single genes of large effect, while quantitative genetics often assumes a large number of genes, each with small effects, influence trait variation.

Secondary article

Article Contents

Introduction

Essentially all characters show some phenotypic variation, either between individuals within a population and/or between populations. Such variation is probably due to both genetic and environmental differences. Sorting out the nature of these genetic differences, the relative importance of genetic versus environmental factors, and how this phenotypic variation translates into evolutionary change is the domain of quantitative genetics. As such, this branch of genetics provides the theoretical foundations for plant and animal breeding and for much of human and evolutionary genetics.

It is generally impossible to determine the genotypes at all relevant loci influencing a trait simply from an individual's phenotype. Fortunately, the relative importance of genetic and environmental factors for a particular trait can be estimated from the phenotypic resemblance between relatives. While quantitative genetics has historically relied entirely on phenotypic information, more recently the tools of molecular biology are being applied in attempts to locate quantitative trait loci (QTLs), loci at which segregation contributes to the observed character variance. Identifying QTLs will augment, rather than supplant, phenotypic measures, and it is straightforward to incorporate such genotypic information into a quantitative genetic analysis.

Quantitative traits

Most often, quantitative genetic analysis is performed on traits showing a continuous range of values, such as height and weight. However, traits displaying a discrete number of values (such as number of offspring) and even binary traits (such as disease presence or absence) are all amenable to quantitative genetic analysis.

Historical roots of quantitative genetics

The roots of quantitative genetics trace back to the work of Galton and Pearson in 1880–1900, who developed many of the basic statistical tools (such as regression and correlation) used in quantitative genetics. The formal beginning of the field starts with R. A. Fisher's 1918 treatment of the inheritance of quantitative characters, which showed how explicit mendelian genetic models of inheritance could account for the resemblance in continuous traits between relatives. Fisher's treatment introduced the powerful statistical method of analysis of variance, which is now applied widely outside the field of quantitative genetics. Indeed, many of the basic statistical tools now commonly in use were first introduced and developed in the context of quantitative genetics.

Genetic and environmental values

One of the fundamental ideas of quantitative genetics is that the phenotypic value P of an individual is the sum of that individual's genotypic value G plus its environmental value E:

$$P = G + E \qquad [1]$$

Given a large number of clones of a particular genotype, the value of G could be estimated from the average value of the clones, while the environmental value E is estimated by the difference between the observed phenotype and estimated value of G. Since clones are rare in most organisms, other types of relatives are much more commonly used to make inferences about G and E. Unless individuals are clones (such as monozygotic twins), the G values of even close relatives are different, and different types of relatives share different aspects of G. To account for this, the genotypic value is decomposed into additive (A), dominance (D) and epistatic (I) values:

$$G = A + D + I \qquad [2]$$

As detailed below, A accounts for the average effects of individual alleles, D for the interaction between alleles at each locus (dominance), and I for the interaction between genotypes at different loci (epistasis). The G values from different types of relatives share different amounts of these components, and it is these differences that allow for inferences about the amount of variation contributed by each of these components. Before proceeding, a few

remarks are in order on the basic statistical measures of variation and association which are widely used in quantitative genetics.

Variances: measures of variation

The standard measure of variation is the variance, Var, which quantifies how much the distribution of character values is spread around the mean μ. As such, the variance provides a measure of the amount of uncertainty in the data. A small variance implies that most values are close to the mean, and hence the mean provides a good predictor of the character value for a randomly chosen individual. Conversely, a large variance implies that there is considerable uncertainty in the value of a randomly chosen individual and the mean is a poor predictor.

If x represents the value of a character whose mean is μ_x, the variance of x is given by the average of the squared deviations about the mean:

$$\mathrm{Var}(x) = \mathrm{Ave}[(x - \mu_x)^2] \qquad [3]$$

Note that $\mathrm{Var} \geq 0$, with the variance being zero only if all individuals have exactly the same value. An important property of variances is that, if two variables are uncorrelated, the variance of their sum equals the sum of their variances:

$$\mathrm{Var}(x + y) = \mathrm{Var}(x) + \mathrm{Var}(y) \qquad [4]$$

In particular, since $P = G + E$, then (provided G and E are uncorrelated) the phenotypic variance is the sum of the genetic and environmental variances:

$$\mathrm{Var}(P) = \mathrm{Var}(G) + \mathrm{Var}(E) \qquad [5]$$

and the fraction of variation in trait values due to differences in the genetic values of individuals is $\mathrm{Var}(G)/\mathrm{Var}(P)$. Recalling eqn [2], the genetic variance can be further decomposed into the variances associated with the additive, dominance and epistatic components:

$$\mathrm{Var}(G) = \mathrm{Var}(A) + \mathrm{Var}(D) + \mathrm{Var}(I) \qquad [6]$$

As we will see shortly, it is these variance components, rather than $\mathrm{Var}(G)$ itself, that appear in the expressions for the resemblance between relatives.

Covariances: Measures of Association

While the variance provides a measure of variation, the covariance $\mathrm{Cov}(x,y)$ provides a measure of association (or covariation) between two traits (x and y) in the same individual, where:

$$\mathrm{Cov}(x,y) = \mathrm{Ave}[(x - \mu_x)(y - \mu_y)] \qquad [7]$$

Thus the covariance is the average value of the product of the deviations of x and y from their respective means. If $\mathrm{Cov}(x,y) > 0$, individuals with a large value for character x also tend to have a large value for y, and likewise individuals with small values of x tend to have small values of y. A negative covariance, $\mathrm{Cov}(x,y) < 0$, implies small values of x are associated with large values of y, and vice versa. A covariance of zero implies there is no (linear) association between the two variables, and the variables are said to be uncorrelated.

The covariance and variance are very closely related. Comparing eqns [3] and [7] shows that the covariance of a variable with itself is its variance:

$$\mathrm{Cov}(x,x) = \mathrm{Var}(x) \qquad [8]$$

Likewise, the covariance of a sum is the sum of the covariances:

$$\mathrm{Cov}(x + y,z) = \mathrm{Cov}(x,z) + \mathrm{Cov}(y,z) \qquad [9]$$

Resemblance Between Relatives

Consider two relatives, such as a parent and its offspring, two full sibs, or identical twins. The phenotypic values P_1 and P_2 of the two relatives are expected to be more similar to each other than either is to a random individual from the population, as relatives share genes and may also share similar environments. If trait variation has a large genetic component, the resemblance between relatives should increase as ever-closer pairs of relatives are considered, as these share more and more genes.

The resemblance between relatives is measured formally by the covariance between the phenotypic values of the two relatives, $\mathrm{Cov}(P_1,P_2)$. Recalling eqns [2] and [9], and assuming G and E are uncorrelated:

$$\begin{aligned}\mathrm{Cov}(P_1,P_2) &= \mathrm{Cov}(G_1 + E_1,G_2 + E_2) \\ &= \mathrm{Cov}(G_1,G_2) + \mathrm{Cov}(E_1,E_2)\end{aligned} \qquad [10]$$

showing that the phenotypic resemblance between relatives can be due to shared genes, $\mathrm{Cov}(G_1,G_2)$ and/or shared environmental values, $\mathrm{Cov}(E_1, E_2)$.

Similarity due to shared environmental factors

Resemblance between relatives due to an association between environmental values ($\mathrm{Cov}(E_1, E_2) \neq 0$) can arise in several ways. Shared maternal effects can generate a significant environmental covariance, as can shared family environments (beyond maternal effects). Full sibs are thus especially susceptible to having the phenotypic resemblance inflated by shared environmental factors. Similarly, an offspring and its mother may also have shared environmental values, owing to the environment influencing both the maternal contribution to the offspring and the mother's phenotype.

In controlled breeding programmes, individuals can be randomized across environments, reducing or eliminating shared environmental values. Human geneticists use a different (but related) approach, looking at twins separated at birth and adopted to different households. While both of these approaches can remove shared environmental effects following birth, there may still be a significant environmental covariance because of shared neonatal maternal effects.

Similarity due to shared genes

Since $G = A + D + I$, the covariance between genetic values can be further decomposed as:

$$\begin{aligned}\mathrm{Cov}(G_1, G_2) &= \mathrm{Cov}(A_1 + D_1 + I_1,\ A_2 + D_2 + I_2) \\ &= \mathrm{Cov}(A_1, A_2) + \mathrm{Cov}(D_1, D_2) + \mathrm{Cov}(I_1, I_2)\end{aligned} \quad [11]$$

The last step follows since A, D and I are (by construction) uncorrelated. Covariances between genotypic values are generated because relatives share genes (alleles shared by both relatives that descend from a single ancestral copy). At a given locus, two relatives can share zero, one or two alleles, with different sets of relatives having different probabilities of these events. For example, a parent and its offspring share exactly one allele (unless there is a history of inbreeding), as a parent contributes only a single allele to its offspring. Identical (monozygotic) twins share exactly two alleles, as they are genetically identical. Full sibs have probabilities of 1/4, 1/2 and 1/4 of sharing zero, one or two common alleles (as there is a 50% chance that two sibs share the same allele from their father, and likewise the same probabilities of sharing maternal alleles).

As we will see shortly, the additive (A) values are due to the effects of individual alleles, whereas the dominance (D) values are the residual values generated by interactions between the two particular alleles at a locus. Thus, if the relatives share only one allele at a locus, they share only half the additive variance (as the additive variance contributed by a locus is the sum of both allelic effects). If relatives share both alleles at this locus, they share $\mathrm{Var}(A) + \mathrm{Var}(D)$. Ignoring epistasis and letting π_i denote the probability that a relative pair shares i (= 0, 1 or 2) alleles at a random locus, the covariance due to shared genetic effects is given by:

$$\begin{aligned}\mathrm{Cov}(G_1, G_2) &= \pi_1 \mathrm{Var}(A)/2 + \pi_2[\mathrm{Var}(A) + \mathrm{Var}(D)] \\ &= \mathrm{Var}(A)[\pi_1/2 + \pi_2] + \mathrm{Var}(D)\pi_2\end{aligned} \quad [12]$$

The appropriate values for some common relative pairs, are given in **Table 1**.

With estimates of variance components in hand, prediction may be made of how similar relatives should be to one another. For example, what is the chance that a grandchild displays a disease that was seen in one of its grandparents? The predicted covariance in this case is $\mathrm{Var}(A)/4$, which translates into a probability of $K + \mathrm{Var}(A)/(4K)$, where K is the disease frequency in a random individual.

Since different pairs of relatives weight the genetic variance components differently, combinations of relatives can be used to estimate genetic variance components; for example:

$$4[\mathrm{Cov}(\text{parent, offspring}) - 2\,\mathrm{Cov}(\text{half-sibs})] = \mathrm{Var}(A) \quad [13]$$

Similarly, the dominance variance can be estimated by:

$$4[\mathrm{Cov}(\text{full sibs}) - 2\,\mathrm{Cov}(\text{parent, offspring})] = \mathrm{Var}(D) \quad [14]$$

As mentioned above, a complication with using full sibs is that shared environmental values can inflate the resemblance between relatives. With shared environmental values between full sibs, eqn [14] becomes $\mathrm{Var}(D) + 4\,\mathrm{Cov}(E_1, E_2)$. Likewise, fathers (rather than mothers) are preferred for single parent–offspring covariances, as this avoids potential inflation due to maternal environmental effects.

Epistasis

The epistatic component I can be decomposed into separate (and uncorrelated) factors, representing different levels of interactions between multiple loci. If only two-locus interactions are considered, then $I = AA + AD + DD$, and each of these epistatic variance terms itself has a different coefficient for the resemblance between different sets of relatives. For example, the shared covariance between parent–offspring due to shared

Table 1 Genetic variance components in relative pairs

Relative pair	π_0	π_1	π_2	$\mathrm{Var}(A)$	$\mathrm{Var}(D)$
Parent–offspring	0	1	0	1/2	0
Grandparent–grandchild	1/2	1/2	0	1/4	0
Full sibs	1/4	1/2	1/4	1/2	1/4
Half-sibs	1/2	1/2	0	1/4	0
Monozygotic twins	0	0	1	1	1

epistatic effects is Var(AA)/4, while it is Var(AA)/4 + Var(AD)/8 + Var(DD)/16 for full sibs.

Generally, epistatic variances are smaller than the additive and dominance variances, as even in the presence of very strong epistasis much of the genetic variation is still loaded into Var(A) and Var(D). Further, the coefficients associated with epistatic variances becomes smaller and smaller as higher-order epistatic interactions are considered.

Components of the Genetic Variance

As mentioned, A represents the contribution to the character from the effects of individual alleles, while D is the contribution from the interaction between these alleles. Likewise, I (which will not be considered further) represents the contribution from interactions between different loci.

To develop A and D formally, consider the genotypic value G_{kj} for an individual with alleles k and j at a particular locus. Assuming no epistasis, an individual's total genotypic value G is the sum of the genotypic values at all relevant loci. Fisher's insight was that the genotypic value for a given locus can be written as:

$$G_{kj} = \mu + \alpha_k + \alpha_j + \delta_{kj} \qquad [15]$$

Under random mating α_k (the average effect of allele k) is the average character value of an individual with a copy of allele k over and above the population mean μ. In the absence of dominance, the value of G_{kj} equals the sum of both allelic effects, $\alpha_k + \alpha_j$. The α values are determined by least-squares (best-fit) criteria and hence they can change as the frequencies of the genotypes change. If dominance is present, the genotypic value may differ from that predicted by the sum of the individual allelic effects. The amount of dominance for a particular genotype is measured by the dominance deviation, $\delta_{kj} = G_{kj} - (\alpha_k + \alpha_{j.})$. The least-squares fit ensures that α and δ are uncorrelated. As an example of α and δ, suppose the average character value (over and above the population mean μ) in an individual carrying allele Q_3 in the population of interest is 5, and the average value of an individual carrying allele Q_1 is 7, implying $\alpha_3 = 5$ and $\alpha_1 = 7$, and giving the predicted value of a Q_1Q_3 heterozygote as 12. However, if the actual character value of a Q_1Q_3 heterozygote is 10, then the dominance deviation for this genotype is $\delta_{13} = 10 - (5 + 7) = -2$.

The sum of the average effects over all loci contributing to the character gives the additive value A, while D is the sum of dominance deviations over all loci. Thus, the additive variance contributed by this locus is the variance of the allelic effects, $\mathrm{Var}(\alpha_k + \alpha_j) = 2\,\mathrm{Var}(\alpha_k)$, and the additive variance is the sum over all loci:

$$\mathrm{Var}(A) = 2\sum_{Loci} \mathrm{Var}(\alpha_k) \qquad [16]$$

Likewise, the contribution from this locus to the dominance variance is given by Var(δ_{kj}), with the dominance variance being the sum over all loci:

$$\mathrm{Var}(D) = \sum_{Loci} \mathrm{Var}(\delta_{kj}) \qquad [17]$$

As mentioned above, the α and δ values, and hence the additive and dominance genetic variances, depend on the frequencies of the various genotypes in the population. To see this, consider the contribution from a single QTL with two alleles (Q_1 and Q_2) whose frequencies are p and $1 - p$, respectively. Suppose the contributions to the character value from each the three genotypes at this locus are:

Q_1Q_1	Q_1Q_2	Q_2Q_2
m	$m + (1 + d)a$	$m + 2a$

Thus, the difference in character value between Q_1 Q_1 and Q_2 Q_2 individuals is $2a$, while the amount of dominance between these two alleles is measured by d. If $d = 0$, the mean phenotype of the heterozygote is exactly intermediate to the two homozygotes, while $d = -1$ implies Q_1 is completely dominant to Q_2 and $d = 1$ implies Q_2 is completely dominant over Q_1. The additive variance contributed by this locus becomes:

$$\mathrm{Var}(A) = 2p(1-p)a^2(1 + d[2p - 1])^2 \qquad [18]$$

showing that the additive variance is a function of not only a, but also the allele frequency (p) and the amount of dominance (d). The dominance variance for this locus is:

$$\mathrm{Var}(D) = [2p(1-p)ad]^2 \qquad [19]$$

If there is no dominance ($d = 0$), the dominance variance is zero. However, even when there is considerable dominance, Var(A) is usually larger than Var(D).

Heritability

The most common summary statistic in quantitative genetics is the heritability, h^2, of a character:

$$h^2 = \mathrm{Var}(A)/\mathrm{Var}(P) \qquad [20]$$

which is the fraction of the total variance due to additive genetic effects. A heritability close to 1 indicates that most of the observed variation is due to variation in the average effects of different alleles. Conversely, while a small heritability implies that Var(A) is small, it tells us little about Var(G), as genetic effects could be largely in nonadditive terms (D and I). Thus a character with $h^2 = 0$ can still have very considerable genetic variation at loci contributing to the observed character variation. A zero heritability simply means that all individual alleles at a

locus have the same average effect on a character. Interactions between alleles (either at the same locus or between different loci) can still generate considerable variance. Note that a character can be entirely genetically determined but show no genetic variation within a population. Hence, $\mathrm{Var}(G) = 0$ does not imply that the character lacks a genetic basis; it implies only that the observed trait variation within the population being considered is entirely environmental.

Heritabilities, being a function of the additive variance (and hence the particular allele frequencies), apply only to the population and environment in which they were measured. The same character measured in different populations thus can have rather different heritabilities. Typically, estimated h^2 values are much lower for characters thought to be closely associated with fitness (such as viability, fertility, fecundity) than for morphological characters thought to be less closely associated with fitness (such as bristle number or wing size). The likely explanation for this observation is that selection tends to reduce heritability, and selection is stronger on characters more closely associated with fitness.

While many sets of relatives can be used to estimate $\mathrm{Var}(A)$, and hence h^2, heritability is most commonly estimated from a parent–offspring regression. Here, the midparent value (the average value of both parents) is plotted against the mean value of their offspring. The resulting points scatter around a straight line whose expected slope is h^2. Mathematically, this implies that the expected offspring value O, given the average value of both its parents (MP), is given by the equation for a straight line:

$$O = \mu + h^2(MP - \mu) \qquad [21]$$

where μ is the population mean of the character. Thus, knowledge of h^2 is sufficient to predict the offspring mean, given the phenotypes of its parents. The actual value for any particular offspring is distributed around this predicted value, with variance:

$$\mathrm{Var}(O \mid MP) = (1 - h^4/2)\,\mathrm{Var}(P) \qquad [22]$$

For a character with $h^2 = 1$, the variance in offspring value about the expected value is $\mathrm{Var}(P)/2$. Thus, even if all the character variance is due to $\mathrm{Var}(A)$, the uncertainty in an offspring's value, given we know its parents, is reduced by only 50% relative to an individual whose parents are unknown.

An example of a midparent–offspring regression is Galton's original 1889 analysis of adult height in the English population. As shown in **Figure 1**, Galton plotted the average height of parents against the average height of their offspring. Rather than plotting each family separately, Galton lumped together all offspring whose parents had the same midparental height (using one-inch categories). The resulting data fell along a straight line, whose slope was 0.65. Thus, $h^2 = 0.65$, implying that 65% of the observed variation in height is due to variation in the

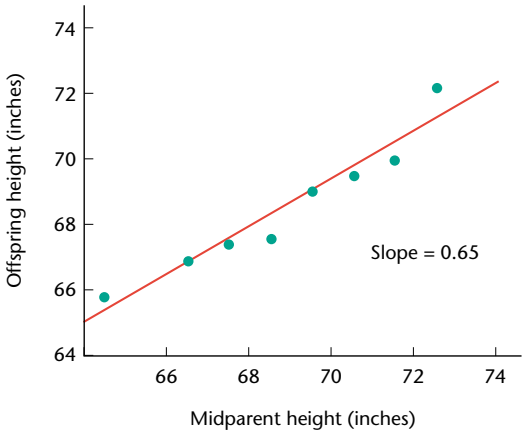

Figure 1 Galton's 1889 plot of average parental height versus average height of offspring.

average effects of individual alleles. Since $h^4/2 = 0.65^2/2 = 0.211$, eqn [22] implies that knowing the midparent value reduces the variance (and hence the uncertainty) in the offspring height by 21% relative to the variance associated with a random individual whose parental heights are unknown.

Selection of Quantitative Characters

Knowledge of the heritability is sufficient to predict the response to a single generation of selection. Defining the response R as the change in mean over one generation, and the selection differential S as the difference between the mean of selected parents and the population mean before selection, it follows from eqn [21] that:

$$R = h^2 S \qquad [23]$$

This is often referred to as the Breeders' equation. If heritability is close to zero, the population will show very little response to selection, no matter how strong the selection. For example, suppose the average value of a character in the population is 100, but only individuals with large values are allowed to reproduce, so that among the reproducing adults the average trait value is 120. This gives $S = 120 - 100 = 20$, and an expected mean in the next generation of $100 + 20h^2$. If $h^2 = 0.5$, the mean increases by 10, while if $h^2 = 0.05$ the mean increases by only 1.

Eqn [23] implies that the response to selection depends on only a part of the total genetic variation, namely $\mathrm{Var}(A)$. The reason for this is that parents pass on single alleles, rather than whole genotypes, to their offspring. Only the average effects of alleles influence the response: any dominance contributions due to interaction between alleles in a parent are not passed on to the next generation, as only a single parental allele is passed to its offspring.

The Breeders' equation (and its rather sophisticated modifications) form the basis for much of plant and animal breeding. One complication is that, since the heritability is a function of allele frequencies, h^2 changes as selection proceeds. While h^2 often declines throughout the course of selection, if rare alleles of large effect are favoured by selection h^2 may increase over several generations before ultimately declining. Eventually (in the absence of new variation), heritability approaches zero as alleles favoured by selection are fixed. Hence, while the short-term response to selection (that over five to ten generations) can be predicted fairly accurately by iterating the Breeders' equation using the original value of h^2, the long-term response is not predictable from the original variance components.

While directional selection (selection to increase the mean) is commonly the goal of plant and animal breeding, it is thought that stabilizing selection may be equally (or more) important than directional selection in natural populations. Under stabilizing selection, individuals with extreme (very large or very small) values are selected against, while individuals with intermediate values have the highest fitness. The net result is little change in the mean, but a decrease in the variance, especially Var(A). As with directional selection, stabilizing selection tends to drive the heritability to zero.

What Maintains Quantitative Variation?

Most characters examined in natural populations show nonzero heritabilities, with h^2 often around 0.2–0.5. Hence, most characters have the intrinsic variation to respond to directional selection. This is an extremely important observation, as in the absence of sufficient variation a population can go extinct if it is not able to respond quickly enough to an environmental change.

What maintains such high levels of additive variance? Mutation is one possibility. The input from new mutation can be substantial, increasing heritability by around 0.1% per generation (for a population with a heritability near zero). If the character is not under selection, the balance between mutation introducing variation and genetic drift removing it can maintain a substantial level of Var(A). If the character is under selection, reasonable levels of Var(A) can still be maintained by the balance between selection and mutation. As mentioned above, characters more closely related to fitness appear to show lower levels of additive variance. This is expected under the mutation–selection balance argument, as the stronger the amount of selection, the lower the equilibrium value of Var(A), and hence h^2.

A second hypothesis is that the nature of selection on the character is shifting over time and/or space. This could occur by frequency-dependent selection, spatial differences in fitness, or genotype–environment interactions. Under certain conditions, this can generate overdominance in fitness (heterozygotes having the highest fitness). Such loci have no additive variance in fitness at equilibrium, but still are segregating genetic variation and hence can have a significant effect on the Var(A) of a character.

The final hypothesis is that variation is maintained by pleiotropy (loci having effects of several characters at once). Under this hypothesis, while the character of interest may not itself be under direct selection, it is influenced by a number of pleiotropic loci under selection for other traits. For example, selection may be acting on some biochemical pathway in such a manner as to maintain variation. If loci in this pathway also influence the character of interest through pleiotropic interactions, variation in the character is maintained.

How Many Loci Underlie a Quantitative Character?

There are several related questions concerning the number of loci. How many loci are responsible for genetic differences between populations and for the variation in genotypes within a population? How many loci can potentially influence a character? The first question deals with the standing variation within (or between) populations, while the second concerns the appearance of new variation by mutation at QTLs. We have very little information on the latter.

Most studies examining the number of QTLs have involved crosses between distinct lines (usually inbred), and thus address only between-population differences. The logic behind such line-cross analysis using only phenotypic information is as follows: if the lines are fixed for alternate alleles at a given QTL (say QQ in one, qq in the other), then all F_1 individuals are Qq. In the F_2, however, 1/4 are QQ, 1/2 are Qq and 1/4 are qq. This F_2 segregation increases the phenotypic variance relative to the F_1, and this increase can be used to estimate a lower bound for the average number of loci contributing to the genetic differences between lines. Use of this approach generally yields lower-bound estimates of around 1–10 factors (single loci or groups of linked loci) accounting for the differences between inbred lines.

While phenotypic line-cross analysis can provide only crude estimates of the number of loci, if marker information is also available a greatly improved strategy is to use F_2 (or related) populations for QTL mapping by looking for associations between trait values and random marker loci (see below). This approach provides direct estimates of QTL positions and their effects. For results using line crosses (mostly involving crosses of different cultivars), a typical character shows about 5–20 detected QTLs, which together account for around 30–50% of the trait variance.

Often a few factors account for the majority of the detected variance.

Crosses between inbred lines do not directly address how many loci are responsible for the observed variation within a population. With an inbred line cross, all segregating loci in the F_1 and F_2 populations have two alleles, each with frequency 1/2. Loci in outbred populations have no such restrictions on either the number or frequency of alleles. Because of this, the number of segregating loci in an outbred population is much less important than the fraction of variation contributed by each. However, since the variance at each locus is a function of both the allelic frequencies and their effects, the contribution from any given locus can change substantially over time. For example, a locus may have a small effect on the character variance even if it is segregating an allele with a large effect, provided the allele is rare. As the rare allele increases, the variance contribution from this loci can increase dramatically. While there are very few studies of the QTLs underlying within-population variation, the general pattern is that many loci appear to have a small effect on the character, while just a few loci account for a significant fraction of the genetic variation.

Searching for Quantitative Trait Loci

An exciting development in quantitative genetics is the use of random deoxyribonucleic acid (DNA) polymorphisms to search for QTLs. It is a fairly routine matter to locate highly polymorphic marker loci, and, by choosing a set of these that span the genome, a search can be undertaken for marker–trait associations, indicating the presence of a QTL linked to the marker. The most straightforward approach is to cross two inbred lines. In the F_2 from such a cross, only those loci fixed for different alleles between the lines will be segregating. Let M/m and Q/q denote alternate marker and QTL alleles fixed in the two lines (say MQ/MQ in one line and mq/mq in the other). Provided that the marker and QTL are linked, M-bearing individuals are more likely to carry a Q allele at the QTL, and m-bearing individuals are more likely to contain a q allele. This results in a difference in the mean trait value between different marker genotypes. If the marker and QTL are separated by a recombination fraction of c, the strength of the marker–trait association scales as $(1 - 2c)$. Thus, a weak association can be generated by tight linkage to a QTL of small effect or loose linkage to a QTL of major effect. Using this basic idea, a variety of statistical approaches (such as analysis of variance and maximum likelihood) have been used not only to detect QTLs but also to estimate their effects and to map positions.

A similar approach applies to outbred populations, but with the restriction that marker–trait association analysis must be performed separately on each parent, as opposed to lumping all individuals together in a single analysis (as is done with line crosses). This restriction arises because the linkage phase can vary over parents. For example, one family might have a MQ/mq parent, another a Mq/mQ parent. If the offspring from both families are combined together in a single analysis, an M-bearing offspring has an equal chance of being associated with either Q or q, and no marker–trait association is observed. The analysis for each parent proceeds as follows. Suppose the father has genotype Mm at a particular marker locus. The presence of a linked QTL is indicated if offspring bearing the paternal M allele have a different mean from offspring bearing the paternal m allele. Since a parent must be a double (marker–QTL) heterozygote in order to observe a marker–trait association, only some parents show a marker–trait association. Because of the separate analyses required for each parent (only some of which will be informative), QTL mapping is far more powerful in inbred line crosses than in outbred populations. In spite of these limitations, a variety of tests for marker–trait association has proven quite successful in mapping QTLs for human diseases.

Further Reading

Barton NH and Turelli M (1989) Evolutionary quantitative genetics: how little do we know? *Annual Review of Genetics* **23**: 337–370.

Falconer DS and Mackay TFC (1996) *Introduction to Quantitative Genetics, 4th edn*. Essex, UK: Longman.

Lander ES and Schork NJ (1994) Genetic dissection of complex traits. *Science* **265**: 2037–2048.

Lynch M and Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer.

Mackay TFC (1996) The nature of quantitative genetic variation revisited: lessons from *Drosophila* bristles. *BioEssays* **18**: 113–121.

Mayo O (1987) *The Theory of Plant Breeding, 2nd edn*. Oxford: Clarendon Press.

Mitchell-Olds T and Rutledge JJ (1986) Quantitative genetics in natural plant populations: a review of the theory. *American Naturalist* **127**: 379–402.

Paterson AH, Tanksley SD and Sorrells ME (1991) DNA markers in plant improvement. *Advances in Agronomy* **46**: 39–90.

Turner HN and Young SSY (1969) *Quantitative Genetics in Sheep Breeding*. Ithaca, NY: Cornell University Press.

Weiss KM (1993) *Genetic Variation and Human Disease: Principles and Evolutionary Approaches*. Cambridge: Cambridge University Press.