

POPULATION GENETICS
Winter 2005
Lecture 18
Quantitative genetics and QTL mapping

- from Darwin's time onward, it has been widely recognized that natural populations harbor a considerable degree of genetic variation.
- Darwin came to this conclusion from the experience of animal and plant breeders of his day and he relied on it heavily when developing his theory of evolution by natural selection.
- the form of variation envisaged by Darwin to be of fundamental importance for evolutionary change was “continuous” or what we now call “**polygenic**” or “**quantitative**”.
- the statistical approach to studying such traits is referred to as **quantitative genetics**.
- “polygenic” simply means that there are a number of genes affecting the trait.
- the vast majority of morphological, physiological, and behavioral characters are polygenic and thus understanding this class of variation is fundamental for evolutionary biologists.
- polygenic traits commonly exhibit a normal distribution that are influenced by both genes and the environment.
- what is the nature of the genes affecting quantitative characters?
- quantitative geneticists believe that they are no different from any other genes – they may possess multiple alleles, exhibit varying degrees of dominance, mutate, and undergo changes in allele frequencies as we’ve discussed so far in class.
- the numbers of genes affecting quantitative traits is largely unknown.
- what we have learned is that individual genes affect quantitative traits to a variable extent.
- some exert much stronger effects than others – these are called **major effect** and **minor effect genes**, respectively.
- understanding the total number of genes controlling a quantitative character and determining the relative roles of major and minor effect genes is referred to as understanding the **genetic architecture** of the trait.

A quantitative genetics model

- in order to construct this model we need to “decompose” phenotypic values into genetic and environmental components.
- let the phenotypic value of individual i in environment j be represented as:

$$P_{ij} = G_i + E_j$$

- here G_i is the genetic contribution of the i th genotype and E_j is the environmental deviation resulting from the j th environment.
- if we were capable of cloning the i th individual and measuring the phenotypic value of the phenotype in m environments, we could obtain a mean phenotype:

$$P_i = G_i + 1/m \sum_{j=1} E_j$$

$$= G_i$$

- in other words, the average phenotypic value becomes the genotypic value because the environmental deviations over all environments cancel out.

- the genetic contribution can be partitioned into different components due to additive, dominance, and epistatic effects.

$$G_i = A_i + D_i + I_i$$

- here, $A_i + D_i + I_i$ refer to the additive, dominance, and interactive (or epistatic) components, respectively.

- thus

$$P_{ij} = A_i + D_i + I_i + E_j$$

- a more useful way to examine the evolution of phenotypic traits is not to deal with individual phenotypic values (as done above) but variances.

- thus, we can use the following

$$V_P = V_A + V_D + V_I + V_E + 2COV_{GE}$$

- where the V terms refer to variances as described above, and COV refers to the genotype by environment covariance.

- in most plant and animal breeding studies it is usually assumed that the covariance is near zero and is thus ignored.

- from this it is possible to describe two forms of **heritability** called **broad-sense** and **narrow-sense**.

$$\text{broad sense heritability} = h^2_B = V_G/V_P$$

$$\text{narrow sense heritability} = h^2_N = V_A/V_P$$

What is additive gene action?

- alleles that act in an additive fashion exhibit no dominance.

- heterozygotes thus fall exactly intermediate between the two alternate homozygotes.

- genes that act in an additive fashion allow populations to evolve far beyond the current limit of phenotypic variation.

- an example of additive gene action:

$$B_1B_1 \quad B_1B_2 \quad B_2B_2$$

A ₁ A ₁	2	3	4
A ₁ A ₂	4	5	6
A ₂ A ₂	6	7	8

- it is the narrow sense heritability that is important in determining both the rate and the response to selection.
- the expression of heritability is complicated – it depends on the number of genes influencing the trait, the frequencies of alleles at these loci, and the extent of dominance and epistasis.
- there is another fundamental problem with the concept of heritability – **it is specific to the population and environment in which it was measured.**
- if populations differ in the types and frequencies of alleles controlling quantitative traits then h^2 may differ.
- if exactly the same alleles are present at the same frequencies in two populations then h^2 can differ if the environmental conditions differ.
- these problems severely limit the generality of estimates of heritability.

Estimating heritability

- how do we typically measure heritability?
- Hedrick reviews three ways in which heritability can be estimated:

1. Reduction/elimination of variance components
2. Resemblance among relatives
3. Realized heritability

Resemblance among relatives

- suppose we are interested in obtaining an estimate of heritability for growth rate for a salamander.
- one way to do this would be to perform pair matings between different individuals and then rear the progeny of these crosses under similar conditions (to hold the environmental effect constant among families).
- we allow the progeny of these matings to reach the same age as the parents.
- we then take the mean of each pair of parents - this is called the **midparent value**.
- we then take the mean of each set of progeny - we can call this the **offspring value**.

Cross	Midparent value (cm/day)	Offspring value (cm/day)
F1 x M2	2.34	3.65
F3 x M4	3.56	2.89
F2 x M1	2.03	2.45

- we can then estimate heritability by regressing the offspring values against the midparent values.

- the slope of this regression line is an estimate of heritability.
- other comparisons are possible that give rise to modified estimates of h^2 .

Comparison	Correlation
Midparent-offspring	h^2
Parent-offspring	$1/2h^2$
Half-sibs	$1/4h^2$
First cousins	$1/8h^2$

- when comparisons are undertaken among individuals that are more distantly related, the precision of the heritability estimate is reduced.

Realized heritability

- we can obtain an estimate of what is called realized heritability by estimating the response to selection.
- consider the following example.
- suppose we wanted to estimate the heritability of body size in mice.
- our starting population had a mean size of 124.5 g and exhibited a normal distribution.
- suppose we selected the top 15% of the population as parents for the next generation and the mean of this group was 168.4 g.
- we breed the individuals in this selected group and rear the mice in the same environment as the parents until they reach the same age.
- suppose the offspring had a mean size of 154.2 g.
- the realized heritability can be estimated from the relative response of the selected group

$$\text{Realized heritability} = h^2 = R/S$$

- where R is the **response to selection** and S is called the **selection differential**.
- in this example

$$S = 168.4 - 124.5 = 43.9$$

$$R = 154.2 - 125.4 = 28.8$$

$$\text{thus } h^2 = 28.8/43.9 = 0.656$$

- this tells us that 65.6% of the variance in the starting population was attributable to additive genetic variation.
- this realized heritability allows us to predict how the population will respond to further selection (at least in the short term).

- what about the long-term response?
- will the population continue to evolve under the imposed selection regime or will the genetic variation become depleted?

What maintains genetic variation for quantitative characters?

- accounting for the persistence of genetic variation underlying polygenic traits remains controversial.
- many models have been proposed.
- one of the most influential is Russ Lande's which postulated the existence of a mutation-selection balance at these genes.
- according to this model, the amount of variation affecting the trait reflects an equilibrium balance between the erosion of variation by stabilizing selection and the input of variation by mutation.
- the overall balance depends on four parameters:

- | | |
|--|-------------------------|
| 1. the strength of selection on the trait | s |
| 2. the number of loci involved | n |
| 3. the mutation rate per locus | μ |
| 4. the average effect of an allele on the phenotype. | a |

- the adequacy of this model has been the topic of many studies and is widely recognized as being inaccurate in of its most important postulates, namely, the ability of mutation to replenish variation lost by either drift or stabilizing selection.

- another model first developed by Michael Turelli postulates an equilibrium between the effects of mutation introducing variation and its loss by drift – this is mutation-drift equilibrium.
- this idea may seem rather silly – it is unlikely that most polygenic traits are neutral.
- however, the model has been useful in showing that over macroevolutionary time periods, the rate of evolution of polygenic characters is generally slow.
- so slow, in fact, that divergence of the trait by random drift cannot be excluded.

- a number of factors have been proposed to account for the maintenance of genetic variation in quantitative traits.
- unfortunately, there is little empirical evidence to evaluate their relative importance.
- the list includes:

1. Overdominance

- there is little support that overdominance is so widespread to maintain variation of this magnitude.

2. Pleiotropy

- the idea here is that a genotype at a certain locus may be beneficial for quantitative trait A and bad for trait B.
- another genotype would have impart a high fitness for trait B and low fitness for trait A.

- although there is some evidence suggesting life-history characters may trade-off in an antagonistic fashion, there is no evidence that genotypes at any locus do the same.
- furthermore, the circumstances allowing antagonistic pleiotropy to operate are rather restrictive.

3. Epistasis

- in theory epistatic interactions among genotypes at different loci can maintain variation.
- although historically there has been little support for epistasis, there is growing evidence that it could play some role in maintaining variation for quantitative traits.

4. Variable selection

- fluctuations in the optimal phenotype from one generation to the next can certainly delay the loss of variation.
- however, random fluctuations in selection will eventually result in the loss of the variation and thus this explanation cannot account for the long-term persistence of polygenic variation.

5. Gene flow

- gene flow between populations that differ in optimum phenotypes can maintain variation in each population.
- even if the same optimum phenotype is favored in the two populations, the genes contributing the trait may differ and the frequencies of alleles at the same loci may differ.
- thus, gene flow can maintain allelic variation underlying polygenic traits.
- as for epistasis, there is little empirical support for this possibility.

QTL mapping

- traditionally, there have been two ways to estimate the numbers of genes underlying quantitative traits.
- one has been to examine the response to selection experiment.
- making a few (questionable) assumptions that the effects of all loci are equal and that there is no linkage or epistasis, one can estimate the number of loci underlying a selection response.
- for example, Dudley and Lambert (2004) estimated that 56 genes were contributing to the response to selection on oil content in corn.
- alternatively, one can examine the phenotypic variation in the F₂ generation after crossing two parental lines (P₁ and P₂).
- a statistical estimate of the effective number of loci underlying the trait is:

$$n_e = (\bar{P}_1 - \bar{P}_2)^2 / 8V$$

- here \bar{P}_1 and \bar{P}_2 are the phenotypic means of the two parents and V is the difference in the genetic variance between the F₁ and F₂ populations ($V = V_{F1} - V_{F2}$).

- new insights are being provided by the ability to map **QTL's**, or "**quantitative trait loci**".
- this is done by melding traditional breeding experiments of quantitative genetics with fine-resolution linkage maps.

- a linkage map is a detailed "map" of the location of genetic markers dispersed throughout the genome of the species.
- identifying and characterizing QTL loci is going to dominate the field of quantitative genetics for many years to come.
- mapping QTL's in "non-model organisms" is still a long way off.
- to perform QTL mapping one needs a high resolution linkage map of the target species.
- these markers can be either microsats, SNPs or even AFLPs.
- developing a high resolution map for a species is a considerable undertaking – it requires the developing and mapping of hundreds to thousands of polymorphisms.
- to map QTLs one also needs to generate linkage disequilibrium between the markers and the quantitative traits of interest.
- why is this?
- well, if the QTLs are in linkage equilibrium with the markers our ability to identify their numbers and locations is impossible!
- to generate linkage disequilibrium, it is common to cross lines that have been made homozygous by inbreeding.
- these are crossed to form F1 and F2 populations that have highly significant levels of linkage disequilibrium.
- the premise of identifying QTLs is based on what is called a **LOD score**.
- a LOD score is the likelihood ratio of the probability of having an association between a marker and a QTL assuming genetic linkage (with a recombination rate between the genes of \hat{r}), divided by the probability of having an association assuming no linkage:

$$L = \frac{\text{Pr (association | } r = \hat{r}\text{)}}{\text{Pr (association | } r = 0.5\text{)}}$$

- this ratio is also called the **odds**.
- the LOD score is equal to $\log(L)$ is the **logarithm of the odds**.
- a LOD score of 3 or greater is considered as statistically significant evidence for linkage between a marker and a QTL.
- a LOD score of -2 or less is significant evidence of no linkage.