Reading: Chapter 5 and 6.

Extensions to Multilocus traits

Recall that quantitative genetics is based on the extension of Mendelian principles to polygenic traits.

- Until now we modeled our expected phenotype assuming a single loci influences the trait. We had two types of effect: additive and dominance.

- To extend our model to include multilocus traits, we need to consider not only the effects corresponding to each locus but also the effects corresponding to interactions between loci.

Epistasis: an example

There are two loci in mice that correspond to two color genes that affect the pigment granules in mouse coat hair (example from Falconer and Mackay 1996).

		Locus 1	
		B-	bb
Locus 2	C-	95	90
	cc	38	34

Table 2.1: Mean number of melanin granules per unit volume of hair. B and C are dominant alleles for the two loci.

		Locus 1	
		B-	bb
Locus 2	C-	1.44	0.77
	cc	0.94	0.77

Table 2.2: Mean size of melanin granules. B and C are dominant alleles for the two loci.

Epistasis: the model for two loci

Consider two loci such that an individual has alleles A_i and A_j at the first locus and alleles B_k and B_l at the second locus. Let G_{ijkl} represent the genotypic value for this individual. More rigorously,

$$G_{ijkl} =$$

The basic model decomposes G_{ijkl} into a component due to additivity between loci and a component due to deviation from this additivity (sound familiar?)

$$G_{ijkl} = \mu_G + (\alpha_i + \alpha_j + \delta_{ij}) + (\alpha_k + \alpha_l + \delta_{kl}) + \varepsilon_{ijkl}$$

Here, ε_{ijkl} is the deviation from additivity of the effects for the two loci. This could correspond to any of the possible types of interaction: additive×additive, additive×dominance or dominance×dominance.

It is straightforward to extend the model to three loci:

$$G_{ijklmn} =$$

There are then more types of possible interaction: additive×additive, additive×dominance, dominance×dominance, additive×additive×additive, additive×additive×dominance, additive×dominance×dominance, dominance×dominance×dominance

A General Least-Squares Model for Genetic Effects

We have the following model:

$$G_{ijkl} = \mu_G + (\alpha_i + \alpha_j + \delta_{ij}) + (\alpha_k + \alpha_l + \delta_{kl}) + \varepsilon_{ijkl}$$

Here, ε_{ijkl} represents all interactions between loci. If we expand ε_{ijkl}, then our model becomes

$$G_{ijkl} = \mu_G + (\alpha_i + \alpha_j + \delta_{ij}) + (\alpha_k + \alpha_l + \delta_{kl}) + (\alpha\alpha)_{ik} + (\alpha\alpha)_{il} + (\alpha\alpha)_{jk} + (\alpha\alpha)_{jl}$$
$$+ (\alpha\delta)_{ikl} + (\alpha\delta)_{jkl} + (\alpha\delta)_{ijk} + (\alpha\delta)_{ijl} + (\delta\delta)_{ijkl}$$

We can then proceed as before using least-squares to derive formulae for each component. Denote the conditional mean genotypic value for individuals with allele i at the first locus by

$G_{i...}$=

and similarly let

$G_{i.k.}$ =

denote the conditional mean genotypic value for individuals with allele i at the first locus and allele k at the second locus and so forth for other conditional mean genotypic values.

The additive effect for allele i is defined as

α_i=

The dominance effect for genotype G_{ij} of locus one is

δ_{ij}=

Higher interactions are defined as follows:

$(\alpha\alpha)_{ik}$=

$(\alpha\delta)_{ikl}$=

$(\delta\delta)_{ijkl}$=

Thus we have partitioned the genotypic value beginning with the lowest order terms, accounting for as much variation as possible. This continues dealing with progressively more complex interactions.

Given that there is random mating and that the loci are independent (unlinked), there is no statistical relationship between the genes found within or among loci.

Why? Which assumption gives us the within independence? the among independence?

Using this fact, the total genetic variance is the sum of the variance of the individual effects:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2$$

Example 5.1 of the Text contains measurements of the average length of vegetative internodes in the lateral branch for a corn variety. Measurements are given by genotype class for two diallelic loci.

BV302	U_MU_M	U_MU_T	U_TU_T	**a**	**k**
B_MB_M	18.0	40.9	61.1	21.6	0.06
B_MB_T	54.6	47.6	66.5	6.0	-2.17
B_TB_T	47.8	83.6	101.7	27.0	0.33
a	14.9	21.4	20.3		
k	1.46	-0.69	-0.73		

The different variance components are calculated to be:

$$\sigma_A^2 = 303.6428$$
$$\sigma_D^2 = 11.7615$$
$$\sigma_{AA}^2 = 1.8225$$
$$\sigma_{AD}^2 = 43.6191$$
$$\sigma_{DD}^2 = 20.3627$$
$$\sigma_G^2 = 381.2086$$

Note that 303.6428/381.2086=0.7965266 of the variation is accounted for by the additive effects of alleles.

Things to keep in mind:

Linkage Disequilibrium

We have discussed the possibility of dependence between alleles within a locus (Hardy-Weinberg Disequilibrium). A dependence can exist for alleles between loci as well. We will begin by considering gamete frequencies for two diallelic loci:

If the allele at one locus occurs independently of the allele at the other locus, then we would expect:

However, dependence can be caused by many factors such as natural selection, founder effects, migration, mutation, random drift etc. A measure of this dependence is the linkage disequilibrium coefficient or the coefficient of gametic phase disequilibrium:

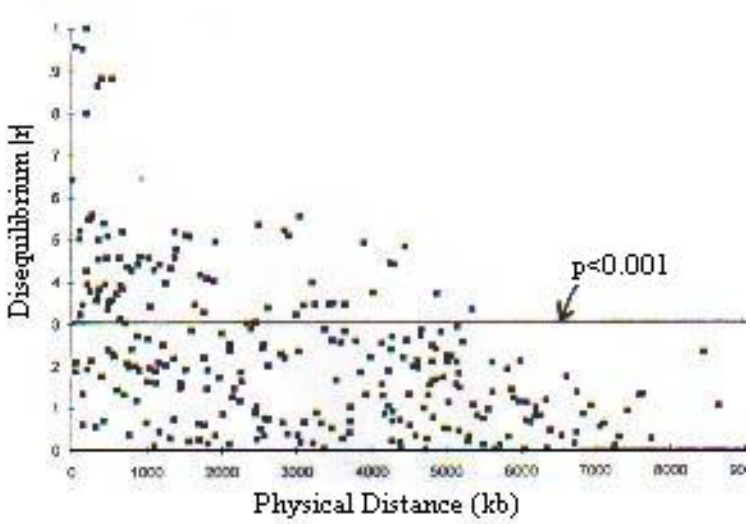
The relationship between D_{AB} and c_{AB}

Example of linkage disequilibrium: (Ajioka et al. 1997)

Hemochromatosis is an autosomal recessive disorder that results in a build-up of iron. The responsible locus has been mapped to 6p.

> From OMIM: "The features of hemochromatosis include cirrhosis of the liver, diabetes, hypermelanotic pigmentation of the skin, and heart failure. Primary hepatocellular carcinoma (HCC; 114550), complicating cirrhosis, is responsible for about one-third of deaths in affected homozygotes. Since hemochromatosis is a relatively easily treated disorder if diagnosed, this is a form of preventable cancer."

Ajioka and collegues studied disequilibrium for 24 polymorphisms in an 8 Mb region of 6p. The following plot shows the absolute value of the linkage disequilibrium coefficient for all pairs of these loci:



For additional information on measures of disequilibrium see Weir BS (1996) Genetic Data Analysis II. There are tons of tests of disequilibrium. Try a Medline search for "linkage disequilibrium" or "genetic association" or "gametic phase disequilibrium" or "transmission disequilibrium test" or any similar combination.

We will come back to such tests but for now we need to note that the presence of disequilibrium will affect our additive and dominance variances. Consider *n* diallelic loci:

$$\sigma_A^2 = 2\sum_{i=1}^{n} \alpha(i)^2 p_i q_i + 2\sum_{i=1}^{n}\sum_{j\neq i}^{n} \alpha(i)\alpha(j) D_{ij}$$

$\alpha(i)$ represents the average effect of allelic substitution at the i^{th} locus.

$$\sigma_D^2 = 4\sum_{i=1}^{n} \left(a_i k_i p_i q_i\right)^2 + 4\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_i a_j k_i k_j D_{ij}^2$$

Estimation of D_{AB}

Assuming random mating, an unbiased estimator of D_{AB} can be found by

$$\hat{D}_{AB} = \frac{N}{N-1}\left[\frac{4N_{AABB} + 2(N_{AABb} + N_{AaBB}) + N_{AaBb}}{2N} - 2\hat{p}_A\hat{p}_B\right]$$

with an estimate of variance

$$\hat{Var}(\hat{D}) = \frac{\hat{p}_A\hat{p}_a\hat{p}_B\hat{p}_b}{N-1} + \frac{(2\hat{p}_A - 1)(2\hat{p}_B - 1)\hat{D}}{2N} + \frac{\hat{D}^2}{N(N-1)}$$

Where are we in the grand scheme of things? We can describe variation in a trait as a function of a multilocus genotype. This includes being able to

- detect non-random assortment of alleles at a single locus
- detect non-random assortment of alleles between two loci
- describe the genotypic effects for a single locus
- describe the interaction between genetic information at multiple loci
- derive the variance corresponding to the additive effects of alleles, dominance effects, and interaction terms

Sources of Environmental Variation

- General environmental effects - influential factors that are shared by groups of individuals

- Special environmental effects - residual deviations from the phenotype expected on the basis of genotype and general environmental effects

- Genotype×environment interaction – the process by which genotypes respond to environmental change in different ways

Extension of linear model:

$$z_{ijk} = G_i + I_{ij} + E_j + e_{ijk}$$

for individual k with genotype i in environment j.

$G_i =$

$\mu_G =$

$\mu_G + E_j =$

$G_i + I_{ij} + E_j =$

$e_{ijk} =$

We now can decompose the phenotypic variation into components for genotype, environment (both general and specific), gene×environment interaction and genotype-environment covariance:

$$\sigma_P^2 = \sigma_G^2 + \sigma_I^2 + 2\sigma_{G,E} + \sigma_E^2 + \sigma_e^2$$

Example: Find G_i, E_j, I_{ij} and μ_G for the following data. Assume that there are an equal number of individuals in each genotype-environment class.

	Environment		
Genotype class	**1**	**2**	G_i
1	110	30	
2	60	30	
3	10	0	
E_j			

μ_G =

I_{11}=

I_{12}=

I_{21}=

I_{22}=

I_{31}=

I_{32}=

e_{ijk}=

Interpretation

Suppose that the three genotypic classes are defined by a diallelic locus with classes 1, 2 and 3 corresponding to genotypes AA, Aa and aa, respectively. Further, suppose that this locus has a major gene affecting our trait of interest. In other words, all variation due to other genetic causes is negligible. Answer the following questions.

Suppose you are unaware of the environmental effect.

- What would be our three mean genotypic values?

- How would you describe the mode of inheritance?

- If high trait values (>35) are detrimental and demand an individual undergo invasive treatment, what would you recommend to a heterozygous individual?

- How would knowledge of the environmental factor influence this advice?

Now it has come to your attention that there is indeed an environmental effect. Further, the exposure to environment one and the presence of an A allele increase an individual's level of the trait.

- Is this relationship purely additive? How do you support your answer?

Suppose I tell you that the locus doesn't correspond to a functional gene but a marker in a non-coding region of the chromosome.

- Give an explanation for why we could still see this relationship.

- What step might you take next?

Suppose you were only given information on the environmental variable and the trait.

- If the only other contribution to variation in our trait was random (normally distributed) noise, how would the overall trait distribution look?

- How would the trait distribution look with-in each environmental class?