Chapter 2

Building Blocks of Quantitative Genetics

Brian Kinghorn and Julius van der Werf

GENETIC COMPONENTS OF MERIT

This lecture considers:

**** How genes are transmitted to the next generation.***

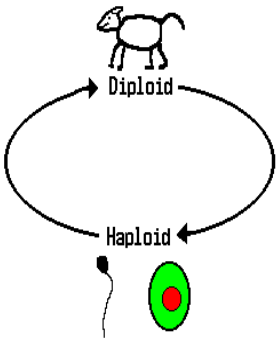
**** How useful an individual's genes are to its progeny.***

***** This lecture condenses a lot into a little.

***** It is needed to show that genes play a role in the selection and crossbreeding theory developed in this course.

Consider a single locus with two alleles segregating. A heterozygote is illustrated here, together with a reminder of the simple biology of transmission of genetic material.

A_1A_2 Heterozygote



Recall that under ***Hardy-Weinberg Equilibrium*** we can predict the frequencies of genotypes A_1A_1, A_1A_2 and A_2A_2 quite simply:

From genotype frequencies to allele frequencies (no HW required)

	Genotype	**A_1A_1**	**A_1A_2**	**A_2A_2**	
Diploid: to	Frequency	P	H	Q	$\Sigma = 1$
Haploid:	freq(A_1)	P	½ H	-	= p
	freq(A_2)	-	½ H	Q	= q
					$\Sigma = 1$

From allele frequencies to genotype frequencies (HW required)

				Eggs		
				A_1 p	A_2 q	
Haploid		A_1	p	p^2	pq	$P = p^2$
to Diploid	**Sperms**	A_2	q	qp	q^2	$H = 2pq$ $Q = q^2$

ASSUMPTIONS FOR HARDY-WEINBERG EQUILIBRIUM.

1. Equal survival of genotypes
2. Equal fertility of genotypes
3. Large sample of animals
4. Random mating of animals
5. Gene frequency same in each sex

1. and 2. together imply NO SELECTION.

Hardy Weinberg Equilibrium for allele frequency f(A1) = p and f(A2) = q.

Genotype:	**A_1A_1**	**A_1A_2**	**A_2A_2**
Frequency:	**p^2**	**2pq**	**q^2**

Whereas ***Population Genetics*** is concerned with the ***fitness*** of different genes (ie. their likelihood of surviving and increasing in frequency over generations), ***quantitative genetics*** is concerned with the ***merit*** of different genotypes (ie. their value to us in agricultural terms). The merit of different genotypes is addressed by considering a single locus (Falconer Ch. 7):

Single locus model of genotypic effects on merit.

The object of this section is to illustrate:

* The concept of ***Genetic value*** - *the value of an animal's genes to itself*. This will also help show the effects of gene frequency (p and q) on the population mean merit.

* The concept of ***Breeding value*** - *the value of an animal's genes to its progeny*. This is of greater interest to us, as it encompasses the basis of ongoing *genetic improvement*.

Consider a single locus with just two alleles segregating (A_1 and A_2). The merit with respect to a certain trait is only due to this particular locus. The values of the possible genotypes are in the Table. A_1A_2 has more merit than the average of the 2 homozygotes - i.e. showing some *dominance*.

Genotype:	A_1A_1	A_1A_2	A_2A_2
Genotype mean merit: [Example]	**$g_{1,1}$** 320	**$g_{1,2}$** 310	**$g_{2,2}$** 280
Frequency: [p= 0.8]	**p^2** 0.64	**2pq** 0.32	**q^2** 0.04
Genetic Value G: [= $g_{x,y}$ - 315.2*]	**$G_{1,1}$** +4.8	**$G_{1,2}$** -5.2	**$G_{2,2}$** -35.2

Population mean merit $= p^2.g_{1,1} + 2pq.g_{1,2} + q^2.g_{2,2} = 315.2$Kg

Population mean G $= p^2.G_{1,1} + 2pq.G_{1,2} + q^2.G_{2,2} = 0$Kg

We also introduce here the commonly used "Falconer notation" (Falconer and MacKay, 1996). The difference between the homozygous genotypes is symmetric around 0.

A2A2	0	A1A2	A1A1
-a		d	+a

Phenotypic mean of the homozygotes = 300

A_1A_1	=	+20	+a	half the difference between genotypic values of homozygote
A_1A_2	=	+10	d	dominance deviation of heterozygote from homozygote mean
A_2A_2	=	-20	-a	

If there is no dominance, d =0, and we have only additive genetic effects. If A1 is dominant over A2, then if 0<d<a, we have partial dominance, if d=1, we have complete dominance and if d>a we have overdominance. Value for d would be negative if A2 is dominant.

Notice that the population mean (M) and Genetic Values (G_{ij}) are population-dependent:

Population mean $M = p^2.a + 2pq.d + q^2.(-a) = a(p-q) +2pqd.$

Genetic value	$G_{11} = a-M =$	$2q(a -pd)$
	$G_{21} = d – M =$	$a(q-p)+d(1-2pq)$
	$G_{22} = -a-M =$	$-2p(a+qd)$

BREEDING VALUE - the sum of average effects of genes.

§ **GENETIC VALUE and BREEDING VALUE - the difference.**

Consider genotype A_1A_2:

Its heterozygosity means its carrier enjoys the effect of dominance in its ***GENETIC VALUE*** - *the value of its genes to itself.*

Its heterozygosity cannot be transmitted to its progeny - because it cannot give both alleles to any one progeny. Thus *the value of its genes to its progeny* is different from *the value of its genes to itself.*

Its ***BREEDING VALUE*** - *the value of its genes to its progeny,* depends on the single genes it can transmit, A_1 and A_2. Each of these has an average effect on progeny. Its BREEDING VALUE is thus the sum of average effects of the genes it carries.

Consider the average effect of gene A_1:

Sperm	Egg	Freq.	Diploid	GeneticValue	Mean
A_1	A_1	p	A_1A_1	$G_{1,1}$	$pG_{1,1}+qG_{1,2}$
	A_2	q	A_1A_2	$G_{1,2}$	

Thus the average effect of A_1 is $\alpha_1 = pG_{1,1} + qG_{1,2} = (.8 * 4.8) + (.2 * -5.2) = 2.8$

And, the average effect of A_2 is $\alpha_2 = pG_{1,2} + qG_{2,2} = (.8 \times -5.2) + (.2 \times -35.2) = -11.2$

Breeding Value (BV) = Sum of Average effects

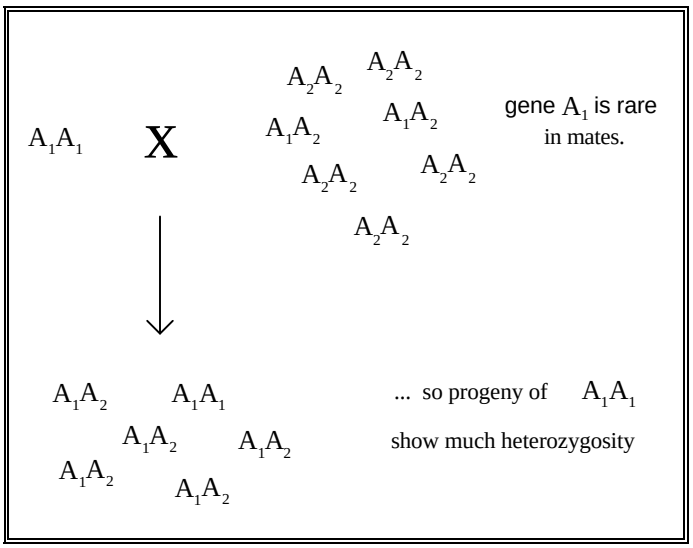
$BV(A_2A_2)$	$= a_2 + a_2$	$= -11.2 + -11.2$	$= -22.4$
$BV(A_1A_2)$	$= a_1 + a_2$	$= 2.8 + -11.2$	$= -8.4$
$BV(A_1A_1)$	$= a_1 + a_1$	$= 2.8 + 2.8$	$= +5.6$

Breeding Values are additive effects. The breeding value of the heterozygote is always halfway the two homozygotes, irrespective of dominance or not.

The average effect of a gene is larger (either positive or negative) when the gene is more rare!

Note that the average effect of a gene involves more heterozygous progeny when it is rarer - as you would expect.

An animal's breeding value depends on population gene frequencies.



<u>In Falconer notation</u>:

The average effect of A_1 is

$$\alpha_1 = pG_{1,1} + qG_{1,2} = p[2q(a - pd)] + q[a(q-p)+d(1-2pq)] = q[a+d(q-p)]$$

The average effect of A_2 is

$$\alpha_2 = pG_{1,2} + qG_{2,2} = p[a(q-p)+d(1-2pq)] + q[-2p(a+qd)] = -p[a+d(q-p)]$$

The difference between the average effects (for a model with only two alleles) is indicated a *average effect of the gene substitution*

$\alpha_1 - \alpha_2 = \alpha = a+d(q-p)]$.

The breeding value is the sum of the average effects

Genotype	Breeding Value
A1A1	$\alpha_1 + \alpha_1 = 2q\alpha$
A1A2	$\alpha_1 + \alpha_2 = (q-p)\alpha$
A2A2	$\alpha_2 + \alpha_2 = -2p\alpha$

Summarising our example:

Genotype:	A_1A_1	A_1A_2	A_2A_2	Mean	$A_1A_2 - \frac{A_1A_1 + A_2A_2}{2}$
Freq.	0.64	0.32	0.04	-	-
Effects:					
Genetic	+4.8	-5.2	-35.2	0	+10
Addit.Genetic	+5.6	-8.4	-22.4	0	0
Dominance	-0.8	+3.2	-12.8	0	+10
Falconer not^{n}.					
G	$2q(a-pd)$	$a(q-p)+d(1-2pq)$	$-2p(a+qd)$	0	d
A	$2q\alpha$	$(q-p)\alpha$	$-2q\alpha$	0	0
D	$-2q^2d$	$2pqd$	$-2p^2d$	0	d

Note that in the above example the mean Genetic value and the mean Breeding value both equal zero. [Remember to use genotype frequencies to give a properly weighted average]. Thus all individuals' values reflect their ***superiority or inferiority*** compared to their contemporaries. ***This makes the subject much easier to handle***. So, from now $\overline{G} = \overline{A} = \mathbf{0}$

Note that ***Dominance deviation*** (D) is simply the difference between G and A. You can check that D values all equal zero (i.e. A = G) whenever there is no heterozygote advantage. Note also that breeding value is additive: $A_{1,2}$ is the average of $A_{1,1}$ and $A_{2,2}$

As A is the sum of the effects of 2 genes, and as only 1 gene can be passed on to each progeny, breeding values must be halved when used to predict progeny performance. For example, if a ram with a high breeding value is used over randomly selected ewes, his progeny show only half of his breeding value superiority in their genetic values ...

$$\hat{G}_o = \frac{\hat{A} + 0}{2} \qquad \text{eg.} \quad \text{Progeny of } A_1A_1: \quad \frac{+5.6 \ + \ 0}{2} \ = \ 2.8$$

- where $\hat{G}_o$ is the predicted (hat, ^) genetic value of offspring (*o*). The 0 reflects the 'averageness' of the randomly selected ewes.

In our example, the predicted value of progeny of A_1A_1 is ½.5.6 above the population mean (2.8 + 315.2 = 318 Kg), if s/he had been allocated mates of average breeding value. To check this is easy, by looking at the frequencies and values of the progeny of an A_1A_1 individual.

For the offspring of an A_1A_1 parent:

MATE GENOTYPE	MATE FREQUENCY	PROGENY GENOTYPE	PROGENY FREQUENCY	PROGENY VALUE	FREQ. x VALUE
A_1A_1	p^2	A_1A_1	p^2=.64	320	204.8
A_1A_2	2pq	A_1A_1 A_1A_2	pq=.16 pq=.16	320 310	51.2 49.6
A_2A_2	q^2	A_1A_2	q^2=.04	310	12.4
Sum the products of progeny frequencies and values to give the predicted					318

Variances

Additive genetic variance: sum of frequency * value^2

$$V_A = p^2(2q\alpha)^2 + 2pq(q-p)^2\alpha^2 + q^2(-2p\alpha)^2 = 2pq\alpha^2$$

Dominance variance: $V_D = d^2(4q^4p^2 + 8p^3q^3 + 4p^4q^2) = (2pqd)^2$

Total genetic variance $V_G = V_A + V_D$

Two-locus models:

The genetic basis of heterosis can be divided into two components - Dominance and Epistasis.

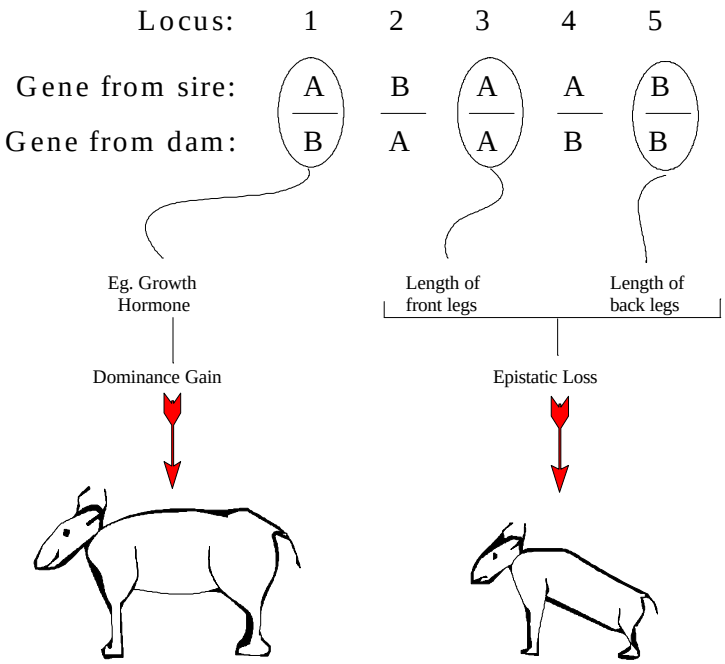


Figure 1: Mixing genes from different breeds leads to dominance gain and epistatic loss.

DOMINANCE An individual carries two copies of each gene, one from each of its parents. They are both designed to do the same job, but they may be slightly different and do the job in slightly different ways or with different effectiveness. Where the individual's parents come from two different breeds the individual will carry a wider range of genes, sampled from two breeds rather than just one. It is thought that this better equips the individual to perform well, especially under a varying or stressful environment. The classical meaning of dominance is that the better gene of each pair dominates in its effect on performance, and this may also be involved. We would thus expect dominance to be a positive effect, and there is much evidence to support this.

EPISTASIS Epistasis is the interaction between genes which are not partners, and which do different jobs. Generations of selection in pure breeds have ensured that these genes cooperate well in carrying out their tasks. It is difficult to give an example here, as we know relatively little about genes of importance in domestic animals - however it seems quite evident that life processes are complex, and there must be cooperation and coordination in the way genes act. When we cross breeds, genes find themselves having to cooperate with other genes that they are not used to. The crossbred animal may thus be out of harmony with itself, and we expect that epistasis, if important, is a negative effect. This has been found most notably in egg production, and milk production in the tropics.

Models of epistasis

When considering degree of expression of dominance, heterozygosity is taken into account – either known heterozygosity, or probable level of heterozygosity from incomplete information, such as genetic marker information in pedigreed data sets. However, epistasis can be classified in two general categories:

- Interaction between single genes and the total genotype at all other loci. This is seen as a scale effect. Here is an example from crossbreeding: If milk yield per day and lactation length showed zero heterosis, then total lactation volume would show heterosis because of the multiplicative nature of these component traits. With this type of interaction, multi-locus QTL detection methods will not give benefit over single locus QTL methods – but the effect of QTL will differ between genetic backgrounds (typically different breeds). Single-QTL effects will tend to be higher in the populations in which they are detected – this is actually an effect of selection.
- Interaction within small groups of loci whose products are interdependent in function (eg. Kinghorn, 1987). Such interactions the subject of a number of models. These fall into two categories: general and specific.

A General models of epistasis

Here is a simple one-locus model of genetic effects, similar to that found in all texts in this area. *II*, *Ii* and *ii* are the genotype values for combinations of the two alleles *I* and *i*, μ is a general mean, A_i is the additive affect and D_i the dominance effect at locus *i*. We can now expand this to cater for effects at two loci. The classical statistical approach (eg. Jana 1971) is typified as follows:

$$\begin{pmatrix} II \\ Ii \\ ii \end{pmatrix} = \begin{pmatrix} \mu + A_i \\ \mu + D_i \\ \mu - A_i \end{pmatrix}$$

$$\begin{pmatrix} II\,JJ & II\,Jj & II\,jj \\ Ii\,JJ & Ii\,Jj & Ii\,jj \\ ii\,JJ & ii\,Jj & ii\,jj \end{pmatrix} = \begin{pmatrix} \mu + A_i + A_j + AA_{ij} & \mu + A_i + D_j + AD_{ij} & \mu + A_i - A_j - AA_{ij} \\ \mu + D_i + A_j + AD_{ji} & \mu + D_i + D_j + DD_{ij} & \mu + D_i - A_j - AD_{ji} \\ \mu - A_i + A_j - AA_{ij} & \mu - A_i + D_j - AD_{ij} & \mu - A_i - A_j + AA_{ij} \end{pmatrix}$$

The number of parameters to handle has increased from three (μ, A_i and D_i) to nine (μ, A_i, A_j, D_i, D_j, plus interaction terms AA_{ij}, AD_{ij}, AD_{ji}, and DD_{ij}).

Specific models of epistasis

Many specific epistatic interactions can be described in the classical patterns: complementary, dominant, duplicate, recessive and inhibitory epistasis (Jana 1971). Carlborg et al. (2000) describe these: “Complementary epistasis is observed when a defect in either of two genes gives the same mutant phenotype, giving an expected Mendelian segregation ratio of 9:7 (Table 1). In this case functional copies of both genes must be present to produce the dominant phenotype. Duplicate epistasis is observed when a defect in two genes gives a mutant phenotype and the expected segregation ratio will be 15:1. In this case a functional copy of only one of the two genes must be present to produce the dominant phenotype. Dominant, recessive and inhibitory epistasis occurs when one gene blocks the phenotypic expression of a second gene. For dominant epistasis, the dominant allele at the first locus is also dominant over the alleles at the second locus. The phenotypic effects of the second locus are therefore only expressed when the individual is recessive homozygote at the first locus. This gives an expected segregation ratio of 12:3:1. Recessive epistasis occurs when the recessive

homozygote at one locus is dominant over the alleles at the other locus. The phenotypic effects for the second locus is therefore only expressed when the individual is dominant homozygous or heterozygous at the first locus. The expected segregation ratio will in this case be 9:3:4. Inhibitory epistasis works in the same way as dominant epistasis and is the special case when the two genes have equal sized effects with opposite signs. The expected segregation ratio is here 13:3. The relationships among the genetic parameters for these five genetic models are given in Table 1. The translation of the genetic parameters to the genotypic effects of the two interacting QTL are given <above>."

Table 1. The relationships among the eight genetic parameters producing digenic segregation ratios in the F2 generation characteristic of classical epistasis (from Jana 1971).

Nature of epistasis	Relationship among parameters*	F2 Ratio
Complementary	$A_1 = A_2 = D_1 = D_2 = AA_{12} = AD_{12} = AD_{21} = DD_{12}$	9:7
Duplicate	$A_1 = A_2 = D_1 = D_2 = -AA_{12} = -AD_{12} = -AD_{21} = -DD_{12}$	15:1
Dominant	$A_1 = D_1 \neq A_2$, $A_2 = D_2 = -AA_{12} = -AD_{12} = -AD_{21} = -DD_{12}$	12:3:1
Recessive	$A_1 = D_1 \neq A_2$, $A_2 = D_2 = AA_{12} = AD_{12} = AD_{21} = DD_{12}$	9:3:4
Inhibitory	$A_1 = -A_2 = D_1 = -D_2 = -AA_{12} = -AD_{12} = -AD_{21} = -DD_{12}$	13:3

*A_1 is the additive effect at locus 1, A_2 is the additive effect at locus 2, D_1 is the dominance effect at locus 1, D_2 is the dominance effect at locus 2, AA_{12} is the interaction between A_1 and A_2 , AD_{12} is the interaction between A_1 and D_2 , AD_{21} is the interaction between A_2 and D_1 and DD_{12} is the interaction between D_1 and D_2

We suspect that many loci affect most traits. The rest of this lecture illustrates the build-up of normal distribution of genetic merit, assuming that we are dealing with many unknown genes, each of small effect. Later on we will look at breeding strategies for when we have at least some knowledge (genotype probabilities) about known Quantitative Trait Loci.

From genes to distributions

Most traits do not show such distinct classes of expression as in the one-locus model. For most quantitative traits, we usually observe a continuous variation and the observed values follow a *normal distribution*. There are two explanations for this:

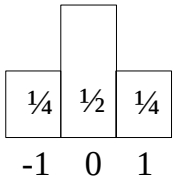
1. Many loci affect the trait. The distribution of genetic effects becomes normal if traits are influenced by genes at many loci, possibly with more than two alleles at each locus.

2. The phenotypic expression of traits is not only due to genotype, but also due to environment (generally a larger part of the differences in observed phenotypes can be attributed to variation in environmental effects).

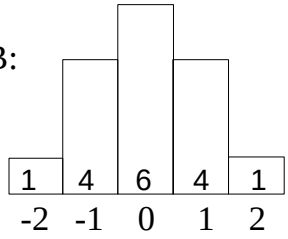
Assume gene frequencies p = q = ½ at all loci, and contributions to genetic value as shown. As more loci are added, the distribution of genetic values becomes more normal:

		A_1A_1 -1	A_1A_2 0	A_2A_2 1
B_1B_1	-1	-2	-1	0
B_1B_2	0	-1	0	1
B_2B_2	1	0	1	2

Locus A:



Loci A & B:



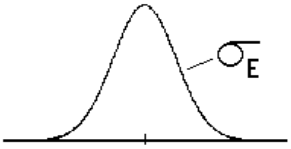
This example, with intermediate allele frequencies and no dominance, may seem like an ideal situation, giving a symmetric distribution even for a single locus. However, as more loci are added, the distribution of genetic values becomes more normal. Even with more extreme frequencies, and with large dominance effects, the distribution of the action of many genes working together will follow a normal distribution (can be illustrated with GENUP-module LOCI).

A genetic model that assumes the action of very many genes, each with a small effect, can therefore explain traits for which we observe a normal distribution of genotypic values. This genetic model is indicated as *polygenic model*. One version of this model postulates that effects at individual loci are so small that allele frequencies do not significantly change with selection. This is indicated as *infinitesimal model*.

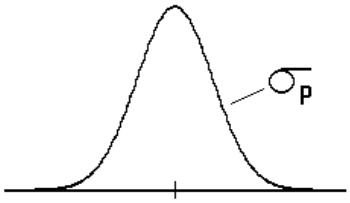
Polygenic effects result from the action and interaction of genes at a large number of loci, each with a small effect. The resulting effects are predicted to follow a normal distribution.

Many loci, Genetic value G:

Environmental deviation E:



Phenotype, P = G + E :



$V_P = V_G + V_E$

P = G + E and G = A + D

gives:

P = A + D + E

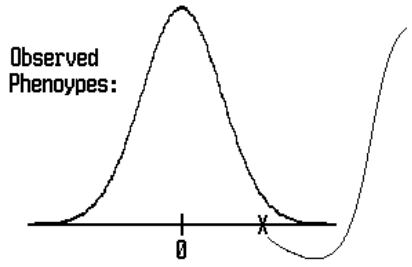
Example

(+100) = (+90) + (+40) + (-30) - note all elements are deviations.

Estimation of Breeding Value

IN REAL LIFE

We can only see P **We want to estimate A.**



Individual X is superior in Phenotype (P).

It is also ***likely*** to be superior in A, D and E: Since we don't know to what extent each of those effects have contributed, we give them the *most likely* values: the contribution of each effect is proportional to the variance explained by that effect.
For example, for a superior phenotype of + 100, we expect the additive genetic value to be + 25 if 25% of the total variance is due to additive genetic effects.

Hence, the *estimated* breeding value (EBV, or A-hat) is $\hat{A} = \frac{V_A}{V_P} P$:

similarly we can also estimate the effect of dominance: $\hat{D} = \frac{V_D}{V_P} P$

and environment $\hat{E} = \frac{V_E}{V_P} P$

and all of these estimated effects should add up to P (as the proportions of each of the variance components add up to 1)

Breeding values are estimated from regression of breeding value on phenotype. Of course, in practical animal breeding we extend this to use information from relatives, and cater for fixed effects, using BLUP.

Should we estimate Breeding Values or Genetic Values?

Genetic value is the value of an animal's genes to itself. Breeding value is the value of an animal's genes to its progeny. In general, breeding value has been of much more importance to animal breeders - it reflects the merit that can be transmitted to the next generation. It is the sum of the average effects of alleles carried by the animal, and because of the large number of loci classically assumed, there is no power to capitalize on anything but the average effects of these alleles, as dominance deviations in progeny cannot be predicted under normal circumstances.

However, when dealing with individual QTL we have the power to set up matings designed to exploit favourable non-additive interaction in the progeny. This means that prediction of

breeding value at individual QTL (***average*** effects of QTL alleles) will only be of partial value in many circumstances. Therefore, later in this course we will consider both prediction of breeding values and prediction of QTL genotypes, and therefore genetic values, at individual QTL.

Of course prediction of QTL genotype of candidates is only of real value in helping to predict genetic values of their progeny - because the object is to improve performance of descendants. This in turn means that the evaluation system should be intimately associated with the mate allocation process, wherever non-additive effects (dominance and/or epistasis) are to be exploited. The combination of animal selection and mate allocation can be termed *mate selection*. Application of evaluation systems to exploit individual QTL will thus frequently involve mate selection strategies in addition to the simpler ranking processes we are used to with selection.

One extreme example of this is where we manage to use genetic markers to identify QTL and chromosomal regions which can contribute strongly to increased expression of heterosis in crossbred progeny. Recurrent selection of purebreds on the performance of their crossbred progeny has not been of great practical value - however now with extra information from genetic markers and known QTL we have some power to breed for increased heterosis in a systematic manner.

References

Carlborg, O., Andersson, L. and Kinghorn, B.P. 2000. The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics.* In Press

Falconer, D.S. and MacKay, T.F.C. 1996. Introduction to quantitative genetics. (4th Ed.) Longham.

Jana, S. (1971). Simulation of quantitative characters from qualitatively acting genes. I. Nonallelic gene interactions involving two or three loci. *Theoretical and Applied Genetics*. **41**, 216-226.

Kinghorn, B.P. (1987). The nature of 2-locus epistatic interactions in animals: evidence from Sewall Wright's guinea pig data. *Theoretical and Applied Genetics*. **73**, 595-604.