

Chapter 15

Basics of Marker Assisted Selection

*Julius van der Werf, Department of Animal Science
 Brian Kinghorn, Twynam Chair of Animal Breeding Technologies
 University of New England*

Basics of Marker Assisted Selection	119
Introduction	119
Exploiting major genes with Genetic Markers	121
How can a marker genotype be helpful in selection?.....	123
How important is the marker information?.....	124
Direct markers	124
Linked Markers.....	125
Selecting for QTL genotypes	126
Non-additive genetic effects and mate selection	126

Introduction

The idea behind marker assisted selection is that there may be genes with significant effects that may be targeted specifically in selection. Some traits are controlled by single genes (e.g. hair colour) but most traits of economic importance are quantitative traits that most likely are controlled by a fairly large number of genes. However, some of these genes might have a larger effect. Such genes can be called major genes located at QTL. Although the term QTL strictly applies to genes of any effect, in practice it refers only to major genes, as only these will be large enough to be detected and mapped on the genome. Following the pattern of inheritance at such QTL might assist in selection.

Figure 1 illustrates that QTL constitute only some of the many genes that affect phenotype. The other relevant genes are termed *polygenes*. Variation at the polygenes jointly with polymorphism at the QTL determines total genetic variation. Although QTL effects explain only a part of genetic differences between animals, knowledge of the genes located at QTL could greatly assist in estimating an animal's true genotype. Information available at QTL therefore adds to accuracy of estimation of breeding value. If genetic effects at QTL are really large, such genes could be more specifically exploited in breeding programs, e.g. to target specific production circumstances or

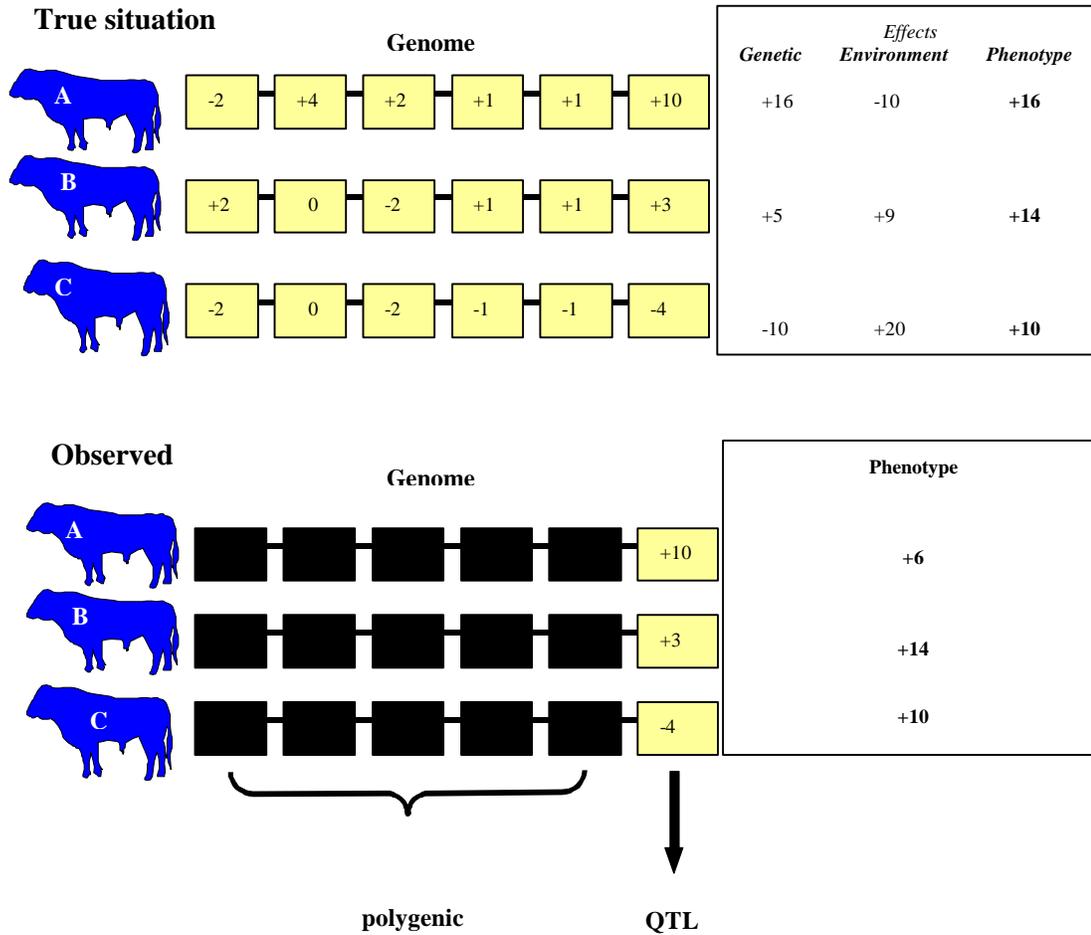


Figure 1. Illustration of three bulls with different phenotypes. The top drawing gives the true allelic values at the different genes affecting body weight, the bottom situation illustrates what would be observed if QTL could be identified in addition to phenotype, adding significant information about true genotype

specific markets.

Figure 1 suggests that the value or the allelic forms at individual QTL could be known. In practice, this is rarely the case. That is, currently there are few examples where QTL effects can be directly determined, but knowledge in this area is rapidly developing. Most QTL known today can only be targeted by *genetic markers*.

Genetic markers are “landmarks” at the genome that can be chosen for their proximity to QTL. We cannot actually observe inheritance at the QTL itself, but we observe inheritance at the marker, which is close to the QTL.

Exploiting major genes with Genetic Markers

Until recently, we have had to resort to information about how animals and their relatives perform to make inference about how good their genes are. However, a direct handle on the genes conferring merit could lead to much faster genetic progress. Such handles are becoming available as ‘genetic markers’. These are a bit like genes with no function and no impact on animal performance. However, they can be easily identified in the laboratory – we can tell which variant of each genetic marker an animal carries.

As with genes, genetic markers are located on chromosomes, like the proverbial ‘beads on a string’. Well-designed experiments can be used to find genetic markers that are located close on their chromosome to major genes. Figure 2 shows a genetic marker that is ‘linked’ to a gene.

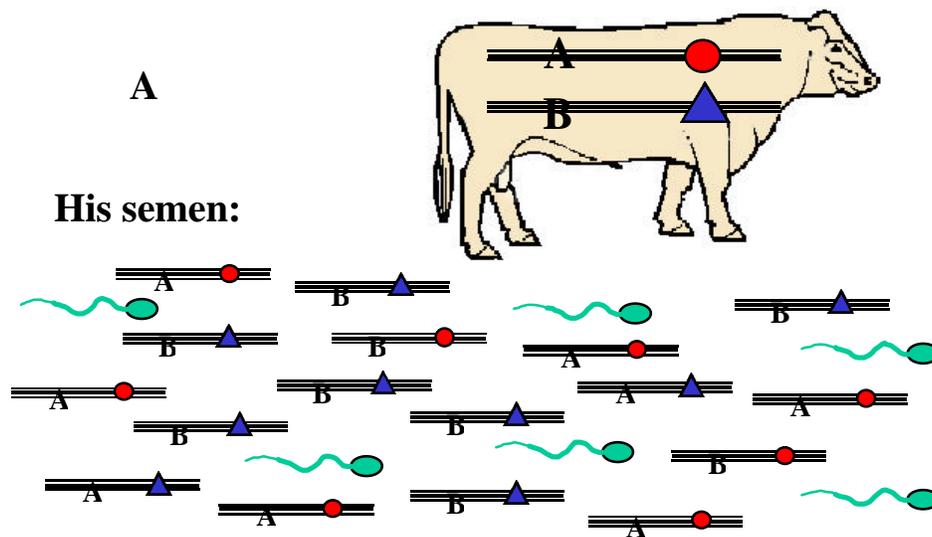


Figure 2 An indirect genetic marker (variants A and B) linked to a major gene (variants circle and triangle).

If we 'organise' this bull's semen we get Figure 3.

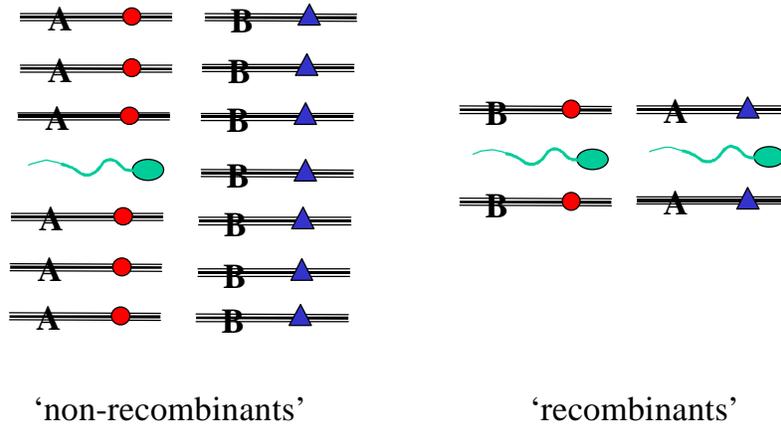
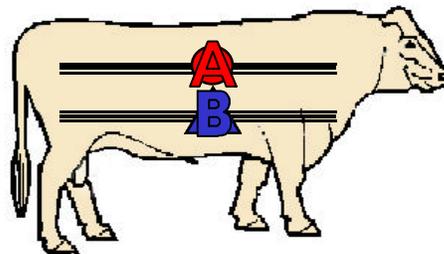


Figure 3.

Notice that the bull can make 4 ‘types’ of sperm. For example, marker A can end up together with the triangle variant of the major gene. This is because of ‘recombination’, or ‘crossing-over’. The machinery that makes the sperm can ‘accidentally’ jump tracks and start copying DNA from the partner chromosome instead. The recombination fraction above is 4/17 or about 24%. The more loosely linked the marker and gene are - the more far apart they are - the higher the recombination fraction. The maximum recombination fraction is 50%.

Recombination is a real problem. Because of recombination we cannot be sure which marker variant is associated with each gene variant in an animal - hence the term ‘indirect’ genetic marker. We have to record pedigree and make trait measurements in order to work with indirect markers.

If a marker is located within a major gene, then recombination is no longer the problem it was. In this case, knowledge of marker variants tells us directly about the variants of the gene carried by the animal, as in figure 11. We only need to measure the trait(s) involved for monitoring purposes.



- A** - always circle, always good
- B** - always triangle, always bad

Figure 11 A direct genetic marker.

How can a marker genotype be helpful in selection?

When making selection decisions based on marker genotypes, it is important to know what information can be inferred from the marker genotypes. Use of direct markers is straightforward, as the marker genotype give a direct indication of the QTL genotype. The problem is how to base decisions based on indirect markers.

Figure 2 shows the principle of inheritance of a marker and a linked QTL. We can identify the marker genotype (Mm) but not the QTL genotype (Qq). The last is really what we want to know because of its effect on economically important traits. Let the Q allele have a positive effect, therefore being the preferred allele. In the example, the M marker allele is linked to the Q in the sire. Progeny that receive the M allele from the sire, have a high chance of having also received the Q allele, and are therefore the preferred candidates in selection.

As shown in Figure 2 there are 4 types of progeny. All progeny will inherit m-alleles and q-alleles from the mother. The sire will provide them with either an M allele or an m allele and either Q or q. In the figure, 90% of the progeny that receive an M allele has also received a Q allele, because M and Q alleles are linked on the same chromosome in the sire. However, in 10% of the cases after the sire reproduced, there has been a *recombination* between the two loci, and animals that inherited an M allele from their father have received a q allele rather than a Q allele. Therefore, marker alleles do not always provide certainty about the genotype at the linked QTL.

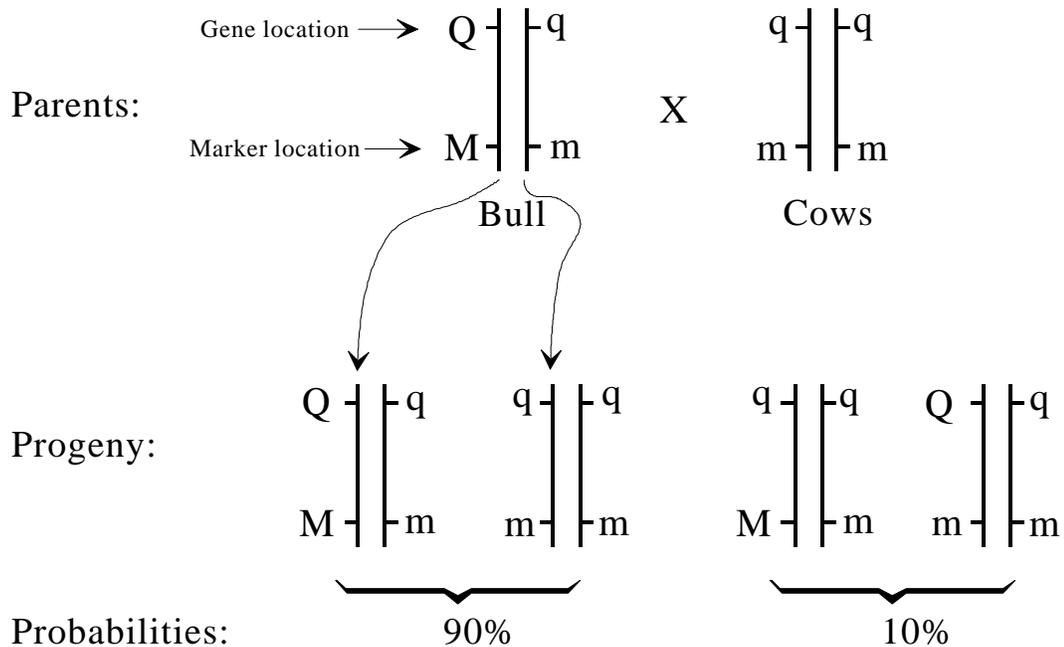


Figure 2: Following the inheritance of a major genotype affecting a quantitative trait (Q locus) with a genetic marker (M locus) closely linked to the Q locus. The sire is heterozygous for either locus and the dam is homozygous. For this example, we can determine for each progeny whether they received M or m allele from their sire. The recombination rate (10%) determines how often Q alleles join M alleles.

Selection of animals could be based on genetic marker information only. However, in that case the effect of other genes not covered by the marker would be ignored. Optimal selection should aim for QTL as well as for polygenes. It should be based on information from markers genotypes combined with information on animals' phenotype. The first aims to get the good QTL, the second aims at getting also good 'other genes'. Selection with the aid of information at genetic markers is termed *marker assisted selection (MAS)*.

How important is the marker information?

The question now arises: what value to give to a particular marker genotype? Should we prefer an animal having a desired marker-genotype over one that that doesn't, but that performs better phenotypically? The answer is that the information from these different sources should be weighted optimally, as in BLUP. The value of having the right marker genotype (Mm in our example) depends on three things:

- 1) The size of the QTL effect
- 2) The frequency of the Q-allele. If it is nearly fixed, MAS will not add a lot to genetic improvement, whereas the opposite will be true if it is found in low frequency.
- 3) The probability that an Mm animal indeed has inherited the Q allele.

With regard to the probability of how sure we can be that an *M* animal indeed has a Q-allele, there is a distinction between *direct markers* and *linked markers*. If there is no recombination between marker and QTL, i.e. the marker exactly identifies the gene, then finding an *M* implies finding a *Q*. However, if *M* is only near *Q* on the genome, *M* and *Q* have a possibility to break up at each meiosis. Meiosis is the process that produces sperm or eggs and determines the genetic make up of offspring. This has the important consequence that finding an *M* is not a guarantee of finding a *Q*. Not only is the chance of having a *Q* allele lower than 1, but for a randomly chosen sire, we have no idea whether *M* or *m* is linked to *Q*. The implication is that for each family this *linkage phase* needs to be determined based on data. For example, in the family in Figure 2 we would see that *M*-progeny perform better than *m*-progeny, hence *M* would most probably be linked to *Q* (the positive allele). But some progeny will have *m* linked to *Q*, so in their offspring, *m* will be the preferred marker genotype.

Direct markers

The easy scenario is when the marker allele *M* and the QTL allele *Q* are always together. This is only the case if the marker is actually measuring the relevant polymorphism within the gene that causes the effect. Such a *direct marker* is very convenient, because the marker genotype will directly inform us about the QTL genotype. However, there are currently only a few direct genetic markers for economically important traits. Examples are:

- The Halothane gene in pigs (giving increased lean percent but also stress susceptibility). This gene has been found - it is the ryanodine receptor gene.
- The double muscling gene in cattle (giving increased muscle mass). This gene has been found - it is the myostatin gene.

These are typically markers for genes that were known to exist before they were mapped, and had a large effect.

Direct markers are generally much preferred to linked markers, if they are truly markers for major gene effects. Their biggest benefit is that they can even be used without trait measurement or pedigree recording. Despite this, there is value in having such information, to monitor the effect of the major gene in different breeds/lines and production systems, and exploit it accordingly.

Possible risks with using direct markers are:

- § There can be more than one mutation causing the desired genetic effect. A DNA test for only one of those mutations would not pick up all the animals with the desired effect. An example of this case was the myostatin gene for double muscling, where several mutations within the gene caused the same desired effect. If only some of the single direct markers had been adopted, there could have been false negatives in diagnostic tests.
- § There is also some potential to incorrectly identify a candidate gene as a major gene directly affecting the trait of interest, because it is only near the true causative gene. In that case there is a risk of false positives: we pick the ‘positive gene’ but it turns out to be an indirect marker and recombination might have made it linked to the ‘negative allele’.

This highlights the value of re-evaluation of markers in distinctly different stock and a continuous need for trait recording for monitoring purposes.

Linked Markers

Linked markers are only near QTL on the genome and not the causative mutation in the gene concerned. For a randomly chosen animal in the population, we have no clue whether one or another marker allele is associated with a preferable QTL allele. If we observe within the progeny of one sire a difference in performance between different marker alleles (as M and m in Figure 2) we can determine which of the marker alleles is associated with the preferred QTL-allele. But this information is only useful for this particular sire, and its family! The information on *linkage phase* is also useful for its progeny, since for them we can determine the probability of whether or not they inherited the preferred QTL allele (depending on the recombination rate). Within the family, the marker and the QTL are said to be in *linkage disequilibrium*. Notice that we can only detect linkage in heterozygous sires. When sires don’t show a difference in progeny for the different marker genotypes, they may be homozygous for the QTL. Sires should also be heterozygous for the markers, and in addition, it will be useful to genotype dams, since otherwise it can be unclear which marker allele an animal received from its sire. However, with markers abundantly available, animals could be genotyped for a panel of markers, therefore reducing the need for genotype information on dams.

With linked markers, the information on which marker genotype is linked to the positive QTL allele is family specific. This linkage phase has to be determined by genotyping at least 2 generations (a sire and its progeny), and using phenotypic information on the progeny. In a number of families we cannot detect linkage, because the sires are homozygous, either for the marker or the QTL, or both.

It may be obvious that there is a considerable need to gather trait and pedigree information for use of linked genetic markers because for each family the linkage phase between marker and QTL needs to be established. However, many breeding populations already have a performance and pedigree recording system in place. Furthermore, the need for large half-sib families is also reduced over time, as marker and trait information is gathered on a deeper pedigree. This is because we now have methods to use information from all relatives to make inference about which marker variant is linked to the superior gene variants in each animal.

Selecting for QTL genotypes

Where a direct marker (DNA-test) exists for a QTL, we can use direct indirect Marker Assisted Selection, sometimes known as Genotype Assisted Selection (GAS). Where only linked markers exist for a QTL, we must use indirect Marker Assisted Selection (MAS). In either case, the aim is to determine QTL genotypes to assist selection decisions, either to increasing the frequency of favourable QTL alleles, or targeting their introgression into other lines. The value of this depends on a number of factors:

- a. Where heritability is low, the value of information on individual QTL tends to be higher because accuracy of breeding values is increased by a relatively larger amount.
- b. Where the trait(s) of interest cannot be measured on one sex, marker information gives a basis to rank animals of that sex. This is particularly useful when to determine which males should be progeny tested.
- c. If the trait is not measurable before sexual maturity, marker information can be used to select at a juvenile stage.
- d. If a trait is difficult to measure or requires sacrifice (as with many carcass traits) marker information can be used instead.

Non-additive genetic effects and mate selection

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. These two are not the same when there is dominance. The effect of dominance is illustrated in Table 1. When the Q allele is dominant over the q allele, the genetic value of the heterozygote Qq is close to the effect of the homozygous QQ. The breeding value of QQ is twice the value of giving a Q allele to offspring. The breeding value of Qq is intermediate the breeding values of QQ and qq, as a heterozygous animal will give 50% of the times a Q and it will pass on a q otherwise. Therefore, the breeding value of QQ and Qq animals is quite different. Such differences depend on the degree of dominance (the difference between genetic value of Qq and the average of the homozygotes QQ and qq) and on the frequency of Q-alleles in the population. Dominance is quite common. It is indicated as a *non-additive genetic effect* as it causes genetic values of Q- and q alleles to be non-additive. Dominance forms the genetic basis of the existence of heterosis.

Table 1: Hypothetical example of a dominance effect on eye muscle area. Suppose the Q allele has a major positive effect and the difference between QQ and qq genotypes is 20 cm². The heterozygote's mean is close to the QQ genotype, as the Q allele is dominant over the q allele. The genetic values of QQ and Qq are similar, but breeding values (the value of the alleles that are passed on) are quite different. Breeding values are always additive. The example is for a population frequency of the Q allele of 0.325.

Genotype	Average eye muscle area (cm ²)	Genetic Value	Breeding Value
QQ	80	+10	+17.3
Qq	78	+8	+4.5
qq	60	-10	-8.3

In general, breeding value has been of much more importance to animal breeders - it reflects the merit which 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 will only be of partial value in many circumstances. Therefore we should consider both prediction of breeding values and prediction of QTL genotypes, and therefore genetic values, at individual QTL.

Another consequence of dominance is that the difference between marker genotypes will be very dependent on the dam population. If a marker was tested using a sire on a dam population with only q alleles, a large QTL (Q-allele – q-allele) effect would be found (basically the difference between Qq and qq genotypes). However, if the same sire was tested on a population with predominantly Q alleles, the difference between marker genotypes would be the difference between QQ and Qq genotypes, which could be quite small indeed.