Chapter 16

Implementing direct and indirect markers.

Brian Kinghorn

University of New England

Some Definitions

Allele	An allele is a variant of a gene. Genes might have favourable and unfavourable alleles. An individual carries two alleles for each gene (except for alleles on the X chromosome in males), one inherited from each of its parents. In this review, the term 'variant' will be used for allele in most cases, for ease of reading.
Candidate gene	A gene whose effect is known to be related to the biological systems which might affect the trait(s) of interest. This information usually comes from work in other species such as humans and mice.
Comparative mapping	The study of the pattern of genomic locations of known genes in different species. Usually it is practiced by mapping genes in one species and then comparing the locations of the same genes in other species and drawing conclusions on genome structure and evolution.
Direct marker	A genetic marker within a (major) gene, but generally a marker whose DNA variations are not a cause of functional differences in the host gene.
Estimated Breeding Value (EBV)	An estimate of the value of an animals alleles to its progeny. This is usually taken as EBV across all genes involved (as in Breedplan EBV's), but can also refer to the breeding value at an individual QTL or major gene.
Functional marker	A direct marker whose DNA variations are a cause of functional differences in the host gene. This means that the marker is a causal mutation. Of course it is possible that other regions in the gene also contain functional DNA variations. The only 'foolproof' marker is a complete sequence of the full gene.
Genetic marker	A section of DNA which differs between animals, and can be tested for easily in the laboratory. Genetic markers are almost invariably not genes, but they can be contained in genes (whence they are direct markers).
Genome	The full set of chromosomes. Strictly speaking, DNA in the mitochondria is included.
Genome scan	A test of association between target trait(s) and a typically large number of genetic markers, covering at least 90-95% of the genome.
Linkage disequilibrium	At equilibrium, different variants at linked marker will have no systematic association with variants at a linked QTL. However, linkage disequilibrium can occur, for example when there has been recent crossing of *Bos indicus* and *Bos taurus*. The *taurus* versions of the marker and the QTL will tend to be associated, and transmitted together in sperm or in eggs. This association will break down over time with recombination (mixing-up of chromosome segments).
Linked marker	A genetic marker which is linked to a QTL (i.e. on the same chromosome) but not part of that QTL.
Major gene	A gene with a major impact on an observable trait(s). Examples are dwarfism genes, the double muscling gene and the Booroola gene.
Marker Assisted selection (MAS)	Use of information from genetic markers to help make selection decisions. This will usually be done in a manner that exploits both known major genes and all unknown genes. The latter are accommodated now by Breedplan EBV's.
Positional candidate gene	A candidate gene which is in the region of the genome identified from a genome scan as likely to host a QTL.
QTL	Quantitative Trait Locus. A locus is a place in the genome. So a *trait locus* is a place (region) which affects that trait. This could span several genes, one or more of which affects the trait. Each gene is located at a *gene locus*. These definitions are adopted here following confusion at the review itself. It has become common to take 'QTL' as equal to 'major gene', and these two terms will be used with equivalence in this review.

Directly and indirectly marked genes

Directly marked genes. If we have a genetic marker directly inside a gene, then we have great power to exploit this gene in breeding programs. This is because in most cases we can know with 100% confidence which animals have the good variants of this gene. Before proceeding to application, we must first discover the effect of each variant of the gene on all traits of commercial importance, in each production environment, and possible in different genetic backgrounds (eg. different breeds). Then we can proceed to use this marker information to help identify which animals to select for breeding.

Of course there is the potential problem that the desired variant of a gene is already widespread in the breed, with little to be gained from a genetic marker program. Alternatively, there may be no good variants within the breed, such that some form of crossbreeding or gene transfer is required to import them.

However, a big advantage with directly marked genes is that we do not need trait or pedigree recording for application in the field. Unfortunately, this is not the case where the gene is marked indirectly – by one or more genetic markers that are close, but not at the same location on the chromosome as the gene itself.

Indirectly marked genes.   The process of sexual reproduction serves to increase genetic variation between individuals, and this makes both evolution and our animal breeding programs work more effectively. However, this has also given problems when it comes to using genetic markers. The sexual process involves a mixing of DNA and the genes it forms. Thus if good gene variant **G** in a heterozygous sire is associated with marker **M**, this association can be broken down such that in some progeny, marker **M** marks the bad gene variant **g.**

This causes a big problem. Recall that for indirectly marked genes we only have knowledge of each animal's markers, not their genes. So we know the markers, but which animals have the good gene variants? To answer this, we need trait measurements and pedigree information in each herd, and even then we will not have full confidence about where the good variants are. However, our confidence can be high (well into the 95% to 99% range) under good conditions – deep pedigree, large gene effects and large sire families.

To initiate a marker assisted selection (MAS) breeding program for a known but indirectly marked gene, the breeder needs to acquire appropriate genetic marker information and trait recordings on a pedigreed cohort of animals. Within each half-sib group, the marker variants will be associated with trait merit in a particular direction (eg. depending on whether **M** is linked to **G** or **g** in the sire). Thus within that family, animals inheriting the appropriate variant of the marker have a high probability of carrying the favourable variant of the gene. Half sib families need to be large for this to work with reasonable confidence. This means that the best candidates for a MAS program are the progeny of widely used AI sires, which should ideally be of high genetic merit to capitalise on the costs involved.

However, the need for large half-sib families is reduced over time, as marker and trait information is gathered on a deeper pedigree. This is because we now have methods (outlined in Chapters 17 and 18) to use information from all relatives to make inference about which marker variant is linked to the superior gene variants in each animal.

Despite this, it is quite clear that directly marked genes are much easier and cheaper to exploit. They are also much better suited to seedstock sales, due to the possibility of giving guarantee about gene status to customers, rather than just high probability of gene status.

The potential commercial value of detected QTL

It should be recognised that there is potential for profit in the seedstock marketplace through the progressive image of MAS, in addition to the value of its true benefit at the commercial level. There will also be power to label seedstock in a much more tangible manner (genetic status guaranteed) than through use of performance figures and EBVs.

Will the observed QTL effects translate to expression in commercial animals?

QTL segregation in industry animals

QTL detected in experimental herds will be of lower immediate value if no favourable variants of these QTL exist in industry animals. In such cases favourable variants need to be introgressed from other populations (i.e. several cycles of backcrossing to the home population, with selection).

On the other hand, QTL detected in experimental herds will be of no immediate value if (sufficiently) favourable variants of these QTL are already at a high frequency in industry animals. In such cases there is still potential for gains through finding even more favourable variants to be introgressed from other populations.

QTL effects in industry animals

In this type of work there can be considerable difficulties in making correct inference about the significance of results. This works in both directions - there is danger of declaring presence of a QTL when in fact it does not exist, and there is danger of missing valuable QTL because of conservative inference.

There is also a potential problem in this area for candidate genes that look promising, if the number of candidate genes is not low, but this appears to have been handled properly. However, it is more difficult to correct the estimated effects of QTL for the number of the effective tests made, especially in genome scanning work where this number is large.

There is also the potential danger that QTL directly detected in experimental data sets will have a smaller true effect in other herds when it comes to be exploited commercially. The extent to which this is really a problem depends on the strength of interactions between different QTL, and is essentially not predictable from a narrow information base. There is a statistical tendency for QTL to have stronger effects in the genetic backgrounds (and production/management environments) in which they are detected. Re-evaluation of putative QTL in unrelated herds is an important step here.

The same comments also apply to inference about the effect of candidate genes, especially where these are evaluated in the herds in which positional inference has been made.

An important point is that for both directly and indirectly marked QTL, there should be evaluation of effects on all traits of commercial importance, not just the target trait(s). It could be argued that QTL variants that are favourable for the measured target trait(s) would have been brought to a high level in the population by artificial selection if it were not for unfavourable effects on other objective traits, or on fitness traits affecting reproduction and/or survival. Natural selection may play a similar role for target traits that have not been measured and selected on artificially.

Can the benefits be reaped in a commercially profitable manner?

In general, work needs to be done to quantify the likely costs and benefits of use of marker technology. This section makes some observations but supplies no real answers.

At the breeding level

Rates of genetic improvement over a good normal (BLUP) genetic improvement program have been estimated at different levels, but typically about 15% for a simple trait such as body weight.

However, the real benefit of genetic markers in breeding is for making change in traits that are difficult to shift by normal breeding methods. No attempt is made here to put a value on this.

At the commercial level

Genetic markers have an additional potential role to play in making drafting decisions at the production level. This will rarely involve the luxury of pedigree information, and will thus be largely restricted to use of direct genetic markers.

*Costs of using genetic markers.*

DNA tests for known genes.

The red/black coat color marker test in cattle is currently priced at $US160 (with costs probably a good bit lower than this). Other specific marker tests range about $25 - $100. The test for the 'halothane' locus in pigs has been as low as $25 per animal for privately negotiated large runs. Sequence specific DNA tests in humans currently run as high as $3000. However these prices reflect what the market will bear rather than the cost of the test.

Marker assisted selection (MAS).

Let us assume that evaluation of a sire is a separate cost to ongoing testing of his progeny. Evaluation basically means genotyping and phenotyping enough progeny to establish QTL status and QTL-marker linkage pattern in the sire. Assuming price of about $15,000 for genotyping, phenotyping and analysis for sire evaluation - with say 10 chromosomal regions being targeted, with five markers per region (phase 1 in section 'Implementation of linked markers', below). Assume $120-$150 to test any of the progeny following evaluation of the sire (phase 2 in section 'Implementation of linked markers', below). If there were significant volume this is likely to be substantially cheaper, and with technical development of current systems it should be possible to make further reductions. This all assumes current technology and throughput. Improvements are on the horizon and may well bring prices down, indeed this may already be the case. Tests based on fewer markers should also cost less. It should be noted that these costs may also contribute to a larger program, eg. testing key sires by a breed society, thus spreading costs.

Response to MAS over time

Response due to genetic change in individual QTL is limited – stopping when ideal variants are widespread throughout the population. This happens more quickly where just one or two closely marked QTL are involved for the trait of interest, and frequencies of favourable variants are not low to start with. Of course the actual benefits of favourable variants will continue. It is the level of impact in the population that is limited - in contrast to 'normal' selection where genetic gains usually appear to be unlimited.
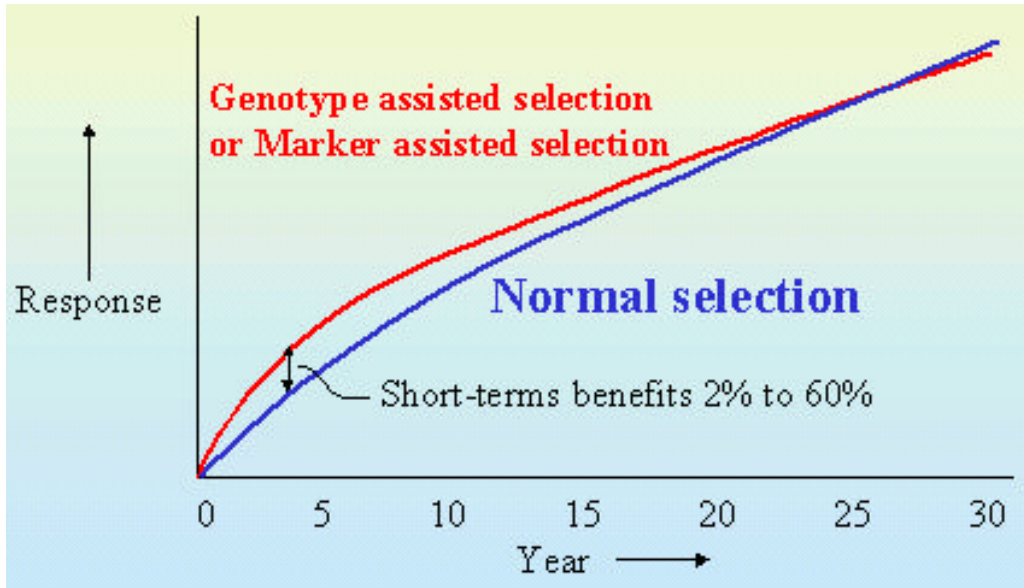
There is an argument that normal selection would increase the frequency of favourable QTL variants anyway. This can be true - with studies showing little long-term benefit of MAS over normal selection *under certain simple conditions.* The real conclusion from this work is that MAS is most useful for traits that are difficult or costly to measure - typically carcass and disease traits. Moreover, QTL detection work is likely to be ongoing, such that new QTL take over the focus of MAS programs.

Over time, with many QTL being tracked, there will be power to model the biology of what is happening to the traits of interest, and therefore power to predict the best combination of variants across QTL. This may not occur within the next ten years, but when it does, we will have considerable flexibility to specify beef genotypes that will be highly focussed on production in defined environments and defined markets.

Modes of implementation of genetic markers

We can exploit genetic markers to make inference about QTL genotypes, and thus undertake marker assisted selection (MAS). This is most useful in cases where the trait(s) of interest have low heritability, are difficult to measure, are expressed in one sex only, and/or are expressed late in life.

The best mode of implementation depends on the way in which the genetic markers give us information about the QTL that we want to exploit. Five approaches to marker-based QTL evaluation of industry animals can be identified:

1. Simple marker association with merit across herds and families. This relies on population-level linkage disequilibrium, and can be carried out without any inference to specific QTL. This approach is seen as unfocussed and costly, and is not under consideration in Australia, except possibly in the dairy industry.

2. Within-family analysis, making inference about sires' QTL heterozygosity and marker-QTL linkage phases (as in Chapter 6). This leads to information for selection between progeny of each sire. This is seen as the basis of an approach to be taken for use of linked markers from the CSIRO group, extending to use of the next two categories are technically involved but useful refinements of this approach.

3. Use of markers to infer probability of identity by descent of contributing QTL alleles, with QTL effects treated as random and no assumption about number of alleles at each QTL. This effectively extends 2. above to use all pedigree information and give QTL EBV's. (See Chapter 17).

4. Use of markers to modify transmission probabilities in segregation analysis to calculate QTL genotype probabilities. Typically two QTL alleles are involved and QTL genotype effects are treated as fixed. This is probably preferable where few effectively distinct alleles are known to be segregating, and where the value of a QTL variant depends on what other variants are inherited - at the same QTL or at different QTL. (See Chapter 18).

5. Use of genetic markers located within target QTL. This removes the need for trait measurements and pedigree information to evaluate animals at QTL of known effect. However, multiple alleleism means that only complete sequence markers are fully reliable, as otherwise QTL alleles of identical marker type can have different effects.

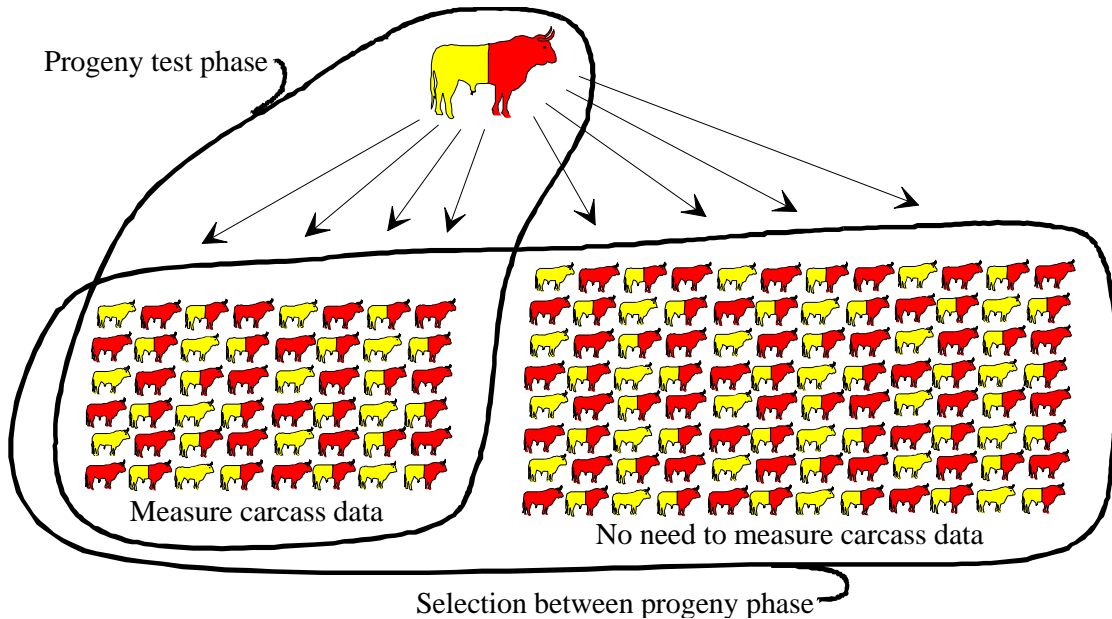## *Implementation of linked markers*

Progeny testing phase

The simple approach to exploiting an indirectly marked gene operates first at the level of sire families (approach 2. above). The sire and sufficient progeny (say, 50 progeny) are genotyped for markers flanking the known QTL - this probably means an A.I. program and/or repeated use of a sire over years. The progeny are measured for the target traits, most probably carcass traits.

Sacrifice of these progeny is not critical as long as there are or will be a good number of progeny left to select amongst. Sacrificed progeny do not have to be produced out of top quality cows, as long as this does not compromise the normal progeny testing which parallels this work.

There are two key objectives for the 'genetic marker' aspect of this progeny test. These are to infer whether the sire is heterozygous at the QTL (two separate variants of the QTL), and to infer the linkage relationship between genetic markers and QTL in the sire. In addition to this is the desirability of estimating the QTL effect as expressed in the offspring. This is not essential if the effects of the QTL are well established as being robust across different breeding lines and environments.

The sires for this phase should ideally be of the highest possible genetic merit, and (likely to be) widely used, to capitalise on the costs involved.

Progeny test phase

Measure carcass data

No need to measure carcass data

Selection between progeny phase

Selection between progeny phase

Once this has been done, it is a relatively simple task to predict which variant of the QTL each progeny carries, and this is where the value of the program is reaped. This includes future progeny as well as progeny in the progeny test, such that overall value is increased it the original sires are widely used. Accuracy is improved by gathering genetic marker information on dams as well, because this helps to resolve which marker variants have been inherited from the sire.

Improvements in power

One major problem with this 'approach 2.', as listed above, is that the sires used are heterozygous - carrying only one copy of the favourable variant of the major gene. This is because we need both *haves* and *have nots* among the progeny to be able to pick out the ones which *have* inherited the good variant from the sire. If the sire had two copies of the favourable variants, all progeny would inherit the good variant from the sire.

So we need power to detect when the sire has two copies of the favourable variant. Approaches 3 and 4 listed above give some power to do this. They both use any marker and trait information that is available, on all relatives. They also manage to make value out of breeding designs without these large half-sib families - because of use of information from all relatives, not just half-sibs. In practice, these approaches should come into play over time, as this information accumulates.

Extensions to these approaches will help us to infer major gene variants inherited from both sire and dam, across a number of different major genes. This could become of extra benefit, as it will allow us to plan matings to give progeny predicted to have the best genetic constitution across all major genes involved. This is important wherever there are interactions between genes - eg the best 'marbling gene' variant might depend on what 'total fat gene' variants are carried by the animal.

Implementation of direct markers

Direct markers are potentially much more simple to implement. At the most simple level, a DNA test can be used to determine the full QTL status of each animal, whether it is to be used for breeding or commercial production.

There is the potential for direct markers to be unreliable, especially if these are not functional markers (see 'Definitions' if needed). This was made evident in work on the myostatin gene for double muscling - if certain single direct markers had been adopted, there could have been false negatives in diagnostic tests.

The contrast between linked and direct markers

As implied above, direct markers are generally much preferred, if they are truly markers for major gene effects. Their biggest benefit is that they can 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.

However, there is some potential to incorrectly identify a candidate gene as a major gene directly affecting the trait of interest, because of linkage disequilibrium with the true causative gene (a reasonably consistent linkage on the chromosome with that gene) in both the original experimental population and in re-evaluation populations. This highlights the value of re-evaluation in distinctly different stock. There is a tendency for linkage of genes that are related in function. [In many cases this is derived from local DNA duplication along the chromosome, followed by some divergence in function between the resulting genes]. This means that there is more that a random chance of close linkage and thus linkage disequilibrium. There is also the potential, albeit small, for true direct markers to be unreliable, as noted in the last section. These may turn out to be small problems in practice, and they should be identifiable during application if a reasonable amount of trait recording is maintained for monitoing purposes.

In contrast, there is considerable need to gather trait and pedigree information for use of linked genetic markers. However, trait measurement is not required for selection between progeny of sires that have already been tested (phase 2 in 'Implementation of linked markers'). It may also be more difficult to market the concept that bull 'X' has a 95% chance of carrying this major gene 'Y', as opposed to a virtual guarantee from a direct marker test. Moreover, the direct marker test will usually tell the variant of major gene inherited from each parent, which is even less reliable with use of linked markers.

However, the fact that linked markers cover a region of chromosome means that they could be more robust in some ways. They will be more likely to properly track a major gene than a direct marker that turns out to be only closely linked to the causative gene. They may also give information about a number of QTL, whether or not this is known to be the case. Moreover, the information gathered in linked marker programs can be of direct benefit in verifying parentage, finding direct markers, and detecting other QTL affecting the measured traits.

In conclusion, we do not need to set up a contest between these two approaches. They should go hand-in-hand in application, driven by commercial demands, with a natural progression from linked markers to direct markers as more information becomes available for each case.

A guide to the usefulness of different types of marker

This guide is very approximate, and is included to help reader orientation alone. The percentage figures quoted are not reliable in themselves, they depend on a range of factors such as closeness of linkage, allele frequencies and amount of pedigree information, and they are useful for approximate ranking purposes alone.

Marker type	Mode of use	Value as a 'handle' on the QTL	Value as a 'handle' on the target trait	Collateral Value**
Linked marker†	Single marker close to QTL	90% for detecting QTL variant inherited from a known heterozygous sire. 50%-80% for an average animal's full QTL status.	40%-70%	Moderate to high for many markers and QTL
Linked Marker†	Close markers flanking QTL, sufficient in number to give good information.	98% for detecting QTL variant inherited from a known heterozygous sire. 70%-90% for an average animal's full QTL status.	60%-75%	Moderate to high for many markers
Direct marker	Direct DNA test	97 - 99%*	80%*	Low
Functional marker	Direct DNA test	99 - 100%	82%	Low
Full suite of functional markers for the QTL	Direct DNA test	100%	83%	Low
Full suite of functional markers for the QTL	Direct DNA test plus trait monitoring in target breeds and environments	100%	100%	Low

*   Considerably less if the direct marker has been mis-diagnosed as the causative gene.
**   Collateral value relates to the potential value of genotype information gathered during industry implementation for verification of parentage and detecting other QTL. It does not include the value of trait and pedigree recordings for classical aspects of genetic improvement. Direct markers (including functional markers) have collateral value at time of detection, through the pointers they might give to further candidate genes that are related in biological function.
†   The higher values in the percentage ranges for linked markers relate to populations with deep pedigrees, and having both marker and trait information recorded over generations.