Conditions of efficiency of marker-assisted selection - benefits and limits in sheep and goat.

Didier Boichard

Proceedings of the International Workshop on Major Genes and QTL in Sheep and Goat
Toulouse (France) 8-11 dec. 2003
CD-ROM communication n°3-01

CONDITIONS OF EFFICIENCY OF MARKER - ASSISTED SELECTION – BENEFITS AND LIMITS IN SHEEP AND GOAT

D. Boichard

INRA, Animal Genetics Department, 78352 Jouy en Josas, France

INTRODUCTION

In the last decade, advances in molecular genetics have made it possible to dissect the genetic variability of complex traits into quantitative trait loci (QTL) and many large QTL detection experiments have been carried out all over the world in farm animals, particularly in cattle, pig and chicken, but also in sheep. Different strategies were used according to the species, the traits, and the available funding : within *vs* between populations experiments ; general purpose experiments aiming to analyse many traits *vs* experiments primarily targeted to one trait of major interest (for instance disease resistance) or the identification of one major gene (for instance Booroola) ; using experimental facilities allowing specific traits to be recorded *vs* exploiting the population structure and the recording systems existing in large commercial populations, as with the granddaughter design. As a result, many studies provided strong evidence for many QTL affecting important traits but in most situations genes involved are still unknown and usually are very inaccurately mapped. An increasing number of genes of interest, however, are identified, as illustrated in this workshop.

The final goal of these studies in farm animals is usually to account for these results in marker-assisted selection (MAS). In this paper, we call MAS any selection procedure incorporating molecular information about QTL or known genes. Of course, MAS is easier to implement when the genes involved and the causal mutations are known, but one may need to apply MAS before they are discovered. Indeed, whereas a QTL detection experiment has a well defined framework in term of required time and budget, the identification of a gene is a much less predictable task and one cannot always afford to wait. This general synthesis will show the possible uses of QTL for genetic improvement of animal populations and the interest and specific limits in small ruminants.

CONDITIONS OF APPLICATION OF MAS

The basic principles of MAS were proposed many years ago (Smith, 1967). Whatever the knowledge about the genes involved, QTL information could theoretically enhance selection efficiency by decreasing generation interval and/or increasing selection pressure and index accuracy. In most situations, however, identified or marked genes explain only a fraction of the total genetic variability of all selected traits and, consequently, MAS cannot simply replace classical selection. The best results are usually obtained by adding new early selection steps in the breeding scheme and using all available information by combining molecular and phenotypic informations.

MAS is known to be particularly beneficial when the traits of interest are difficult or expensive to measure (trait not expressed, sex-limited or expressed late in life, invasive measure such as disease challenge or recording after slaughtering), when each individual performance brings little information in breeding value prediction (trait with low heritability, recessive or low penetrance genetic determinism), or, more generally, when the polygenic approach has limited efficiency or a high cost. In contrast, when the classical selection is very efficient and cost effective with a cheap and easy phenotype recording system, MAS may be less competitive, at least because of the additional genotyping cost.

Therefore, we do expect a selective use of MAS in the conditions where it is the most favourable, and probably not a generalisation of MAS in replacement of classical selection. In practice, however, MAS development still remains very limited, due to its cost, to a lack of confidence of some users, and to the limited number of genes of importance fully characterised.

THE THREE DIFFERENT MAS APPROACHES

Three steps could be distinguished in the characterisation of a QTL.

1) A primo-localisation of the QTL is first obtained, with a quite large confidence interval and no information on the gene(s) involved. This large region is characterised by a set of markers. Because the true location of the QTL in this region is unknown, the distance between these markers and the QTL could be large (>10 cM) and linkage disequilibrium present in a given generation (within family, or resulting from crossbreeding…) is not expected to be maintained over many generations.

2) The fine-mapping step of the QTL narrows the confidence interval and provides new markers in the vicinity of the QTL, which is still unknown. Recombination events between these markers and the QTL are rare and any linkage disequilibrium could be maintained over many generations.

3) The gene involved is identified, as well as the DNA polymorphism responsible for the genetic variability of the phenotype.

These three situations provide opportunities for MAS, but with very different cost and efficiency. Situation 3 is the most favourable, the simplest to implement but also the least frequent, at least today. This situation is called MAS3. The genetic merit of the candidates is simply predicted from their genotyping results, with a straightforward interpretation. In the simple case of one gene and two alleles, a single test is required and this considerably limits the lab cost. The total cost, however, should include possible intellectual property fees, which are a major challenge in the future.

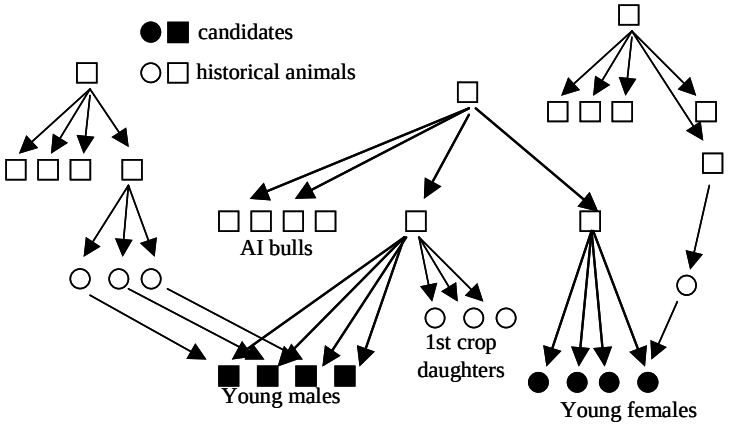
For a trait with a mixed inheritance (identified gene(s) + polygene), a total breeding value could be predicted by combining molecular and own performance into a molecular score, as proposed by Lande and Thompson (1990). More generally, all the information can be combined in a BLUP statistical model where the phenotype is described by the fixed effect of the genotype at the major locus, the random polygenic fraction and a residual. Such a model, however, requires to infer the genotype, at least in probability, of all individuals with record in the evaluation. This limitation probably explains why the molecular information is frequently used in practice in a very crude way in selection, as an additional and independent information.

In contrast, the first generation MAS, called MAS1, takes advantage from the QTL primo-detection results. A known linkage disequilibrium, for instance after a crossbreeding, could be used to select candidates during few generations but the efficiency is rapidly decreased by the accumulation of recombination events. This is the most common way to apply MAS in plants resulting from advanced crossbreeding from two inbred parent lines. In an outbred population in (assumed) linkage equilibrium, there is no preferential association between marker alleles and QTL alleles and it is not possible to select for given marker alleles. Linkage disequilibrium, however, is still present within-family and could be used in selection. The principle is as follows. The marker information is used to estimate the relationship matrix between relatives at the QTL locus or, equivalently, to estimate the probability of identity-by-descent (PID) of chromosome segments of two related individuals (Wang *et al*, 1995 ; Pong Wong *et al*, 2002),. In the case of a major gene, these PID could be used to infer the unknown genotype of the candidates from the known genotype of their relatives. This approach was used in France to eradicate the Achondroplasia (Bulldog) defect in Holstein cattle breed and to introgress the Polled gene in a Charolais elite population. In the case of a QTL, it is highly preferable to combine the QTL information with the phenotypic and pedigree information into a genome merit index estimated by marker-assisted BLUP (Fernando and Grossman, 1989 ; Goddard, 1992). The total breeding value is decomposed into a polygenic value and the value of the paternal and maternal chromosome segments. Each genetic component is assumed to be a random effect with an additive effect. Covariances between polygenic values are assumed to be proportional to the conventional relationship coefficients, whereas the covariances between QTL values are assumed to be proportional to PID.

In contrast to MAS3, MAS1 presents severe limitations and a high cost. The information is not included in the nature of the marker alleles but in the QTL transmission from parent to progeny they reveal. In other words, each QTL effect is estimated from the phenotype of the individual and of all relatives carrying this QTL in probability, with appropriate weights proportional to PID. Therefore, for each QTL, several markers need to be genotyped in order to accurately trace chromosomal segments and balance the incomplete marker informativity. Moreover, flanking markers are useful to assess

recombination events in the traced chromosomal segment. These markers should be genotyped not only for the candidates but also for a number of relatives, including relatives with phenotypic information. Figure 1 shows the strategy used by Boichard *et al* (2002) in the MAS1 design implemented in French dairy cattle. This strategy is a mixture of the top-down and bottom-up approaches of Mankinnon and Georges (1996). Not only young male and female candidates were genotyped but also many non-candidates relatives with phenotypic information ("historicals"). Last but not least, only a part of the QTL genetic variability could be used by MAS1. It is often argued that only Mendelian sampling within heterozygous parents is available for selection. It is not entirely true because the genotype of homozygous parents could be inferred if the design is strongly connected over a number of genotyped generations, allowing for across families selection. When the latter condition is not fulfilled, however, a large amount of information is lost.

Figure 1. Representation of genotyped animals in a given family



As a summary, because several markers are required to trace each QTL and because many non-candidates should be genotyped, MAS1 is more expensive than MAS3. Because of the multiple sources of potential information loss (recombinations, non-informativity of markers, poor inference of the genotype of homozygous parents…), its technical efficiency is much lower than MAS3 and it appears to be superior to conventional BLUP-based selection only in some favourable conditions (Ruane and Colleau, 1995 ; Spelman and Bovenhuis, 1998). Consequently, its economic efficiency is rather poor and its profitability is positive only when the cost of the classical breeding scheme is really very high.

When the QTL are accurately mapped with very close markers, recombination events are rare and these markers are maintained in linkage disequilibrium with the QTL at the population level. This information used in MAS2 can enhance MAS efficiency. In a crossbred population, this information can be used over more generations than in MAS1. In an outbred populations, the interpretation is more complex and, in practice, MAS2 could be implemented as follows. MAS1 already accounts for within-family linkage disequilibrium in a similar way as in MAS2, except that MAS2 PIDs are likely to be higher. But in MAS1, founders QTL are assumed to be unrelated and with the same zero expected effect. MAS2 differs from MAS1 by relaxing these assumptions. A first approach is to assume founders QTL are related. Several methods have been proposed to estimate the PID of founders QTL from marker haplotype information. These methods were developed for QTL fine-mapping but they can also be used for genetic evaluation. Farnir *et al* (2002) assumed a biallelic QTL with one allele appeared after a mutation of the other allele. Meuwissen and Goddard's approach (2000, 2001), based on genetic drift or coalescence, does not make any assumption on the number of QTL alleles number. Computational difficulties may arise because of the large number of non-zero terms generated by this approach in the mixed model equations system. These methods are well suited to closed populations. In the case of recent crossbreeding between populations with very different QTL alleles, an interesting

alternative is to assume groups of founders QTL with different expected effects, and to define these groups according to the known or inferred origin of the marker alleles.
The efficiency of MAS2 is much higher than that of MAS1, for three different reasons: it uses a much better basic information (QTL are accurately located), recombination events are rare (PIDs are close to one or the zero), and across families information is used. As a result, MAS2 could be nearly as efficient as MAS3, without intellectual property limitation. It is, however, much more complex to use than MAS3.

LONG TERM VS SHORT TERM GENETIC TREND

Gibson (1994) first presented that increased gain in the first generation due to MAS is at the expense of the gain in the long-term. This result has been confirmed by other authors and is due to the polygenic value of the animals selected by MAS, lower than after a selection by conventional BLUP. This is true if selection intensity remains the same. In practice, however, markers provide new and early information and the breeding scheme remains rarely the same. When new selection steps are added, particularly on young animals without phenotype, MAS leads to an overall increase in selection intensity (Manfredi *et al*, 1998).

USE OF MAS IN THE DIFFERENT SPECIES

In plants, MAS is applied in a crossbred population originating from a usually small number of inbred parents. The strong linkage disequilibrium generated by such a procedure makes it possible to select directly on the markers linked to the best QTL alleles, over several generations. The selection criterion is the molecular score obtained by the sum over the selected QTL of allele values (estimated by regression of the phenotype on the marker information) times the number of carried copies. More complex approaches have been proposed to avoid any loss of favourable alleles by genetic drift. This plant situation is not directly transposable to animal populations with many parents and no inbred lines.
In animal populations, MAS is particularly studied in dairy cattle and seems to be particularly profitable in this species. Although very efficient, dairy cattle classical selection is very costly. Indeed, this species concentrates many conditions unfavourable to classical selection and, therefore, favourable to MAS. First of all, most traits of interest are sex-limited and generation interval is long. Artificial insemination bulls should be progeny tested, which is a long and costly step. MAS could be a profitable alternative if it can avoid this progeny test step, particularly for the poorest candidate bulls. Furthermore, in some countries as in France, most bull dams are selected before their first lactation on pedigree information only, in order to reduce generation interval. Consequently, not only male but also female candidates are lacking accurate phenotypic information. Finally, functional traits, such as disease resistance or fertility are more and more important in the breeding objective but are difficult to select with classical tools due to their low heritability. All these reasons explain why several attempts have been made to implement MAS in practice, sometimes at a large scale.
In French dairy cattle, large scale MAS1 was implemented since 2000 (Boichard *et al*, 2002). Twelve QTL are traced simultaneously by 43 microsatellite markers (*i.e.* 3-4 markers per QTL) and 8000 animals are genotyped each year. As 800 bulls are progeny tested each year in the French population and 100 are selected as elite sire, the cost of each elite sire should include the genotyping of 80 animals, *i.e.* around 5,000€. According to Colleau (1999ab), the optimum design would involve 16,000 genotyped animals and the corresponding additional cost for each elite sire would reach 10,000€. This apparently very large extra-cost, in fact, is rather small for an elite bull and is easily balanced by a small reduction (<5%) of the number of bulls entering progeny-test. As a reduction of 10-20% in progeny-test is allowed without any loss in genetic trend, MAS1 appears to be highly profitable in dairy cattle. Moreover, it provides a high-quality resource population for QTL fine-mapping, and this synergy prepares the future implementation of MAS2 and MAS3. Clearly, this favourable situation is rather specific to dairy cattle, due to the high cost of its breeding scheme and the extremely high value of the elite sires. I believe that MAS1 cannot be applied to any other species.

In the other species, MAS, if any, is limited to few usually fully characterised genes with a major effect. Selection pressure is variable according to the practical importance and the allele frequency of these genes. In pig, Halothane and RN genes, affecting meat quality, have been eradicated in most populations where it was feasible. In French goat, α_{s1}-casein alleles increasing protein content in milk have been selected (Piacere *et al*, 1996). In sheep, a large-scale selection programme has been implemented to eradicate alleles of scrapie susceptibility in all European populations (Palhière *et al*, this workshop)

WHICH MAS FOR SMALL RUMINANTS ?

In contrast to MAS1, MAS3 does not require any complex family structure, it is simple to implement and lab costs remains relatively inexpensive. Including information of major genes of economic interest in selection procedures is feasible, at least for breeding males. In AI dairy bucks, such a procedure was implemented in 1995 for the α_{s1}-casein locus, which affects protein content but also protein yield and fat content in milk. It should be kept in mind, however, that the total cost also includes an increase in the number of candidates (in order to allow for some additional selection) and, therefore, in the number of dams. Some studies are needed to find the optimal structure of the breeding scheme, which is different from the structure without major gene information (Manfredi *et al*, 1998 ; Villanueva *et al*, 1999). In most situations, MAS is more efficient when the parents of the candidates are also genotyped and selected because it increases the frequency of the favourable alleles in the candidate population. Such a strategy has been applied for α_{s1}-casein in the French buck dams population.

Although rather low, the cost of MAS3 can be still too high for natural mating rams or bucks. In that case, it is necessary to organise a pyramidal breeding scheme with a top-down gene flow and to concentrate MAS3 in the breeding population.

The selection intensity (and therefore the cost) and the expected gain of MAS are obviously highly dependent on the initial allele frequencies in the populations. In the case of a rare genetic defect, few candidates are found to carry the defect and this additional selection step has a cost nearly limited to the genotyping work. In contrast, when the unfavourable alleles are frequent, eradicating these alleles may use all available selection intensity. Some contrasted examples can be observed in sheep breeds regarding the Prp allele frequencies.

As a conclusion, MAS1 appears to be feasible only in dairy cattle because of the very high cost of each AI elite bull. In contrast, use of MAS in small ruminants is limited to MAS3 when the genes and the causal polymorphism are fully identified. Even in these conditions, genotyping cost under current economic conditions is still limiting and restrict use of MAS to males and possibly parents of males, in the selected nucleus of the population.

REFERENCES

Boichard D., Fritz S., Rossignol M.N., et al (2002). *7th World Congr. Genet. Appl. Livest. Prod.*, 22.03.

Colleau J.J. (1999a). *6th World Congr Genet Appl Livest Prod*, **25**:419-422

Colleau J.J. (1999b). *6èmes Renc Rech Rum*, 231-233

Farnir, F., Grisart B., Coppieters W., et al. (2002). *Genetics*, **161:** 275–287

Fernando R., Grossman M. (1989) *Genet. Sel. Evol*, **21**:467-477

Gibson J.P. (1994). *5th World Congr Genet Appl Livest Prod*, **21**:201-204

Goddard M.E. (1992) *Theor. Appl. Genet*., **83**: 878-886

Lande R., Thompson R. (1990). *Genetics*, 124: 743-756

Mackinnon M.J., Georges M. (1998). *Livest. Prod. Sci.*, **54**:229-250

Manfredi E, Barbieri ME, Fournet F, Elsen JM (1998). *Genet Sel Evol*, **30**, 127-148.

Meuwissen T.H.E., Goddard M.E. (2000) *Genetics*, **155**: 421-430

Meuwissen T.H.E., Goddard M.E. (2001) *Genet Sel Evol*, **33**: 605-634

Piacère A, Bouloc-Duval N, Sigwald JP, Larzul C, Manfredi E (1997). *Renc. Rech. Ruminants*, 4, 187-190.

Pong-Wong R., George A.W., Woolliams J.A., et al(2001) *Genet Sel Evol*, **33**:453-472
Smith C. (1967) *Anim. Prod*., **9**:349-358
Spelman R.J., Bovenhuis H. (1998). *Genetics*, 148:1389-1396
Ruane J., Colleau J.J. (1995*). Genet. Res.*, **66**:71-83
Villanueva B, Png Wong R., Grundy B., et al (1999). *Genet. Sel. Evol*., **31**:115-133
Wang,T., Fernando R.L., van der Beek S.,Grossman M., van Arendonk J.A.M. (1995). *Genet. Sel. Evol*. **27**:251–274.