

Industry Use of Marker-Assisted Selection

Michael M. Lohuis

Monsanto Company, St. Louis, MO 63198

Introduction

Several publications from simulation experiments predict that marker-assisted selection (MAS) can increase the rate of genetic response in the range of 5% to 64% in animal breeding populations, depending on the trait being selected and several assumptions about marker and quantitative trait loci (QTL) information (Ruane and Colleau, 1995; Meuwissen and Goddard, 1996; Hayes and Goddard, 2003). Additional publications have also described the application of industry operated MAS programs (Dekkers and Hospital, 2002; Boichard et al., 2002; Spelman, 2002; Dekkers, 2003). These papers identify several factors that influence how well animal breeders have been able to deliver on the theoretical potential of MAS. Several of these same factors have been identified in the application of MAS in Monsanto Choice Genetics swine breeding populations, which have undergone MAS for up to 26 QTL. This paper will attempt to describe some of those factors and how they can impact the degree of success of MAS applications.

Mastering the basics

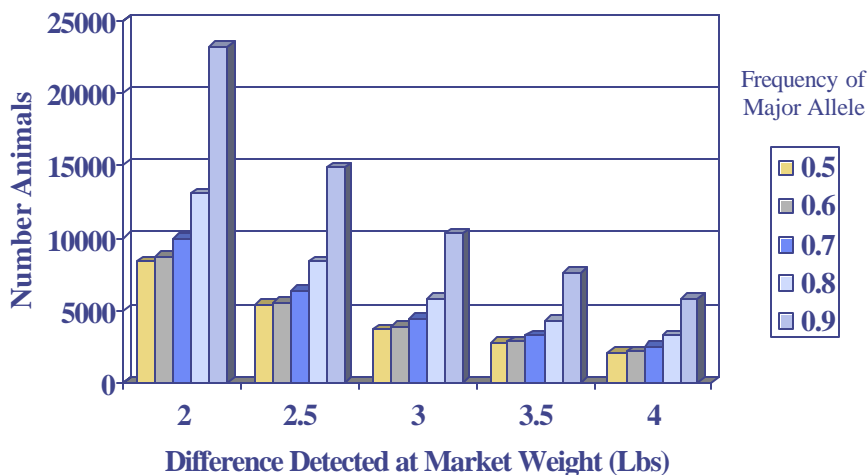
Accurate animal identification and complete phenotypic data collection have always been important in conventional swine breeding programs. However, best linear unbiased prediction (BLUP) tools make allowance for some of these errors in the error variance component and through reduced heritability. However, the forgiving nature of BLUP may also allow some data recording problems to continue relatively undetected. In contrast, MAS programs quickly point out problems in parentage misidentification that can be detected in marker data that do not agree with Mendelian inheritance rules. Although collecting genotypic information will not necessarily reduce the frequency of errors in phenotypic data recording, the addition of marker data can help offset these errors. As breeding programs begin to incorporate genetic marker information, the collection and tracking of DNA samples plus associated genotypes require increasingly sophisticated operating procedures. This naturally leads to fortification or development of quality control (QC) protocols and standard operating procedures (SOP) for parentage identification as well as sample handling. Due to the intensive nature of pig production, identification, DNA collection, and sample tracking need to take advantage of bar coding and scanning technology to minimize opportunities for transcription errors inherent in hand recorded data. Laboratory information management (LIMS) has become the standard in modern biotechnology companies and would also be a wise investment in swine MAS programs for tracking DNA from farm to laboratory. Similarly, well-organized relational databases for phenotypic and genotypic data have become absolutely essential.

The need for QTL validation

QTL discovery is expensive in terms of animal handling, phenotyping and genotyping costs. Consequently, several research groups have used carefully designed experimental populations (e.g. F2 designs) that maximize differences between animals and enable QTL detection with fewer animals. For example, calculations show that a typical outbred population would need to be 8-9 times larger than a F2 population to achieve equivalent statistical power. The drawback, however, is that the size of the QTL effect in experimental populations can differ significantly from the effect in targeted commercial breeding populations. This could occur for several reasons: 1) the breeding population may be fixed at the QTL location or not segregating for valuable alleles, 2) the favorable allele may be at extremely high (or low) frequencies in the breeding population making further selection difficult, and 3) the QTL alleles could have different effects in the genetic background of the breeding population.

The solution to this challenge is to validate QTL in outbred breeding populations prior to beginning MAS on them. This process is straightforward, but requires additional time and expense. Evans et al. (2003) described a broad industry validation of QTL, but the number of animals used in that study provided only minimal statistical power in each population. Calculations have shown that in order to validate a 3 lb. difference due to a QTL at market weight (assuming 90% probability of detection and 5% error rate), 3,700 to 10,000 observations are required depending on frequencies of QTL alleles (Figure 1). Approaches exist that can help optimize the validation process to reduce costs, nevertheless, this is a costly but fundamental step to ensure that value is actually being produced in the breeding program and transferred to swine producers.

Figure 1. Number of animals required to detect differences due to a QTL effect.



Assumptions: false positive rate = 5%, prob. of detecting difference = 90%

A shortcut approach involves the identification of several known genes, based on biological function, which could reasonably be expected to affect an economic trait. However, most known gene information tends to result from research in human and mouse studies and unique polymorphisms found in one species (and the related trait differences, if any) will almost certainly not be represented in other species. Also, one of the drawbacks of large scale testing of candidate genes is the high probability of finding false positives, i.e. when several candidate genes are tested, it becomes difficult to predict which candidate genes are truly associated with a QTL and which show association merely by chance. Due to the high degree of linkage disequilibrium in swine breeding populations and the large number of QTL that likely exist for economic traits, it should not be surprising that candidate genes can be associated with trait effects in one population, only to see those effects disappear (or perhaps even affecting different traits) in other populations. Another factor that reduces the effectiveness of relying solely on a candidate gene strategy is that although a gene (product) is essential to a given biological activity, there is no guarantee that in any given population there are genetic variations that cause observable phenotypic variation. For example, no one can argue the importance of the growth hormone gene for growth in animals, but most investigators find little or no association between polymorphisms in the GH gene in commercial cattle and swine populations and growth traits. This does not mean that the candidate gene approach will not yield associations that are useful in MAS, but great care must be used in interpreting the results. To ensure that candidate associations are real, validation in multiple populations is warranted. In addition, statistical thresholds must be more stringent to account for the increased probability of detecting an association when multiple candidate genes are tested. We expect that the likelihood of improving product performance with this approach alone is not as great as a more comprehensive approach that begins with the knowledge of the existence of QTL from genome scans, followed by linkage disequilibrium analysis using genotypes from markers densely blanketing the QTL region. The testing of candidate genes positioned within a known QTL region could be a useful addition to the strategy (a “positional candidate” strategy).

Optimizing linkage between markers and QTL

Common measures of the value of a genetic marker are the size of the linked QTL effect and proximity to the actual QTL location. However, in regard to proximity, how close is close enough? Several genomics labs have attempted, with varying degrees of success, to find the actual genetic mutations responsible for a specific trait or defect. Finding causative mutations is exciting and valuable research, but it can also be expensive and time-consuming and may not always be justifiable for purposes of swine breeding alone. Hayes and Goddard (2001) predicted that the majority of QTL that control traits of economic importance likely have small to moderate effects. It is expected that these types of mutations will often be traced to locations that enhance or modulate the expression of a gene or a biochemical pathway rather than change the structure and function of the gene product. QTL with small effects can be detected but are difficult to verify conclusively and, by definition have the least to contribute to genetic improvement. On the other hand, Hayes and Goddard (2001) also predicted that there are few QTL with very large effect. However, it is expected that favorable alleles of those genes are often already at high

frequencies in breeding populations due to the intense selection already occurring. Some exceptions may exist if QTL result from relatively recent mutations, migration between populations, or if the QTL is counterbalanced by unfavorable correlations with other economic traits in the selection index. Therefore, it is logical that QTL causing effects of moderate size may yield the most additional benefit to a swine breeding program. These QTL explain a relatively large portion of the genetic variation for economic traits while, with appropriate genomics tools, detection and incorporation into a MAS breeding program is possible.

Due to the intensive selection and migration of pig breeding populations, linkage disequilibrium between neighboring loci is more extensive than in human populations. This results in coinheritance of alleles from different loci spanning millions of base pairs (perhaps up to 5 to 10% of a typical chromosome) for several generations. As a result, finding the exact location of the causative mutation is not necessary for MAS in swine. Instead, combinations of relatively anonymous markers (marker haplotypes) can be used to identify and select animals with chromosomal regions that confer greatest benefit to the next generation. Despite coinheritance of sizable sections of the genome, relatively dense marker coverage of the QTL region is still required to maximize the impact of MAS in that region. This is difficult to achieve with publicly available markers, but with a platform dedicated to developing dense coverage of QTL regions with single nucleotide polymorphism (SNP) markers, we have been able to achieve exceptional marker information content throughout QTL regions. This allows the identification of the very best marker haplotypes, i.e. those with the strongest link to the economic trait and the fewest deleterious effects on related traits. The most valuable marker haplotypes are determined not only by proximity to the QTL, but also by the relative information content and degree of linkage disequilibrium between the marker and QTL alleles.

Ease of use

When attempting to perform MAS in a commercial breeding population, a factor that becomes important is ease of use. Several factors must be considered simultaneously, such as variable marker information content from animal to animal, and the different effects on multiple traits. In Monsanto's experience, the tool that has made this practical on a day-to-day basis is proprietary marker-assisted BLUP (MA-BLUP) software. Similar to conventional BLUP, this software package corrects the data for systematic environmental effects and the relationships between animals, but, in addition, also tracks the inheritance and effects of QTL. The result is a single animal ranking that accounts for both phenotypic information and molecular marker data, thus allowing on-farm breeding personnel to transition to MAS in a seamless fashion requiring no more effort than required for conventional selection.

The ultimate arbiter of success

At the end of the day, what's most important is the impact of MAS in customer herds. The degree of impact will be a function of quality of data used, validity of QTL associations in the breeding population, strength of marker-QTL linkages and quality of

MAS implementation tools. What matters most is that markers are found that allow reliable and cost-effective population-wide MAS to take place such that the product is improved as predicted. Several other questions must be addressed at the outset and tracked throughout the implementation of MAS protocols. Are the traits affected contributing to a relevant economic breeding goal? Is the breeding population monitored for unintended side effects in other traits? How well is the MAS program aligned with business strategy? Do customers of the improved line capture value from this improvement?

Acknowledgements

Helpful comments on this paper were provided by John Byatt, Archie Clutter, Michael Grosz and Sam Buttram, Monsanto Company, MO.

References

Boichard, D. et al. 2002. Implementation of marker-assisted selection in French Dairy Cattle. Proc. of 7th World Congress on Genetics Applied to Livestock Production. August 19-23, 2002. Montpellier, France.

Dekkers, J.C.M. and F. Hospital, 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews: Genetics* 3:22-32.

Dekkers, J.C.M, 2003. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. Proceedings of EAAP, Rome, Italy.

Evans, G.J., et al., 2003. Identification of quantitative trait loci for production traits in commercial pig populations. *Genetics*. 164:621-627.

Hayes, B. and M.E. Goddard. 2001. The distribution of effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.* 33:209-229.

Hayes, B. and M.E. Goddard. 2003. Evaluation of MAS in Pig Enterprises. *Livestock Production Science* 81:197-211.

Meuwissen, T.H.E. and M.E. Goddard. 1996. The Use of Marker Haplotypes in Animal Breeding Schemes *Genet. Sel. Evol.* 28:161-176.

Ruane, J. and J.J. Colleau. 1995. Marker assisted selection for genetic improvement of animal populations when a single QTL is marked.

Spelman, R.J. Utilization of molecular information in dairy cattle breeding. Proc. of 7th World Congress on Genetics Applied to Livestock Production. August 19-23, 2002. Montpellier, France.