UDK 636.082.2;636.4

Izlaganje sa znanstvenog skupa
Conference paper

STRATEGIES FOR MARKER-ASSISTED SELECTION IN PIG BREEDING PROGRAMMES

P. M. Visscher, C. S. Haley

Summary

We review marker-assisted selection strategies for genetic merit in pig breeding programmes. Markers are best used when normal selection is not very effective, e.g. for sex-limited or carcass traits. The appropriate strategy depends on the trait under consideration (e.g. age or sex-limited), the amount of linkage disequilibrium in the population and the marker technology. There is already much scope to utilise genetic markers in pig breeding programmes. Future selection schemes, using new reproductive techniques such as cloning and meiosis *in vitro*, may speed up genetic progress further.

Keywords: pigs, genetic markers, QTL, introgression, marker-assisted selection

Introduction

Traditionally, pig breeders have made genetic progress by using phenotypic information on animals available for selection and their relatives. The breeding goal may consist of a combination of traits, such as growth rate, backfat, and litter size and selection indices have been used to combine the information from several traits in order to maximize genetic progress. More recently, advanced statistical methodology such as BLUP has been adopted separate environmental from genetic effects. These methods all treat the true, unknown, genotype of an individual as a black box, and rely on average population correlation between phenotypes and genotypes to predict the genetic merit of selection candidates. However, these methods have worked well in sustaining a steady rate of improvement in pig breeding programmes. Can we do better still? With the advance of molecular techniques, we can now look at genetic variation directly. By identifying variation at the DNA level, we

Rad je priopćen na "6th World Congress on Genetics Applied to Livestock Production", Armidale, 1998.

P. M. Visscher, University of Edinburgh, IERM, West Mains Road, Edinburgh EH9 3JG, Scotland, C. S. Haley, Roslin Institute, Roslin, Midlothian EH25 9PS, Scotland

can, in principle, calculate the exact proportion of genetic material that two individuals share, rather than rely on their expected proportion of genes in common, as is used when we have only phenotypic information. Recently linkage maps of the porcine genome have been developed (e.g., Archibald et al., 1994). By putting the publicly available information together it is already possible to produce a map with over 1500 highly polymorphic (microsatellite) markers spaced throughout the genome (http://www.ri.bbsrc.ac.uk/genome mapping.html). Thus it is now possible to genotype pigs for evenly spaced polymorphic markers 10 centi Morgans (cM) or less apart.

To what uses could marker data be put? In this paper we will focus on the use of markers to enhance genetic progress in pig breeding programmes. Markers can also be used to control and confirm parentage and for individual, line, and product identification purposes, but these possibilities are outside the scope of this paper. In breeding programmes genetic markers allow the inheritance of whole segments of chromosomes to be traced from parents to offspring. If a particular segment contains a gene or genes that contribute significantly to variation between animals, then we may see an association between the particular segment an animal receives and its performance. In principal this allows progeny to be selected on the basis of which chromosome segments they have inherited from their parents, as well as on the basis of their performance and their relationships with other animals. Genetic markers usually have no effect on performance themselves, thus their value is solely that they mark chromosome segments containing genes affecting performance. Genes linked to markers might be major genes, causing qualitative differences between animals (e.g. coat colour, some disease genes) or they may be some of the many genes (quantitative trait loci or QTLs) which contribute to quantitative variation between animals for traits such as growth, fatness and litter size.

Genetic markers have a number of features which make their use in breeding programmes potentially attractive: (i) Codominant markers, i.e. markers for which heterozygotes can be distinguished from either homozygote, have a heritability of unity, because we have a direct measurement of the genotype. (ii) DNA-based genetic markers can be measured on animals of both sexes at any age, so that boars can be genotyped for litter size markers and markers for carcass quality can be measured on live animals. (iii) Genetic markers can explain some of the within family genetic variation. The latter is valuable because when using phenotypic information in a selection index or BLUP, the deviation of an animal's breeding value from its pedigree index (the average breeding value of the parents) can only be estimated using records on the animal itself or records on its progeny. Using genetic markers we may be

able to say which half or full sibs are likely to be superior, even if they do not have phenotypic observations themselves. Hence, markers that partly explain within family variance may be most beneficial for sex and age limited traits.

In this paper, we discuss strategies to utilize marker information in pig breeding programmes. Depending on the purpose of using the markers, we distinguish between (i) gene introgression, (ii) selection from synthetic populations, and (iii) within-line selection.

Methods

Marker-Assisted Introgression (MAI). The rationale for an introgression programme is that we want to move a gene or chromosome segment for a trait of interest from an otherwise inferior breed to a commercial breed. The aim of introgression is to fix the favourable allele in the commercial population with as little as is possible of the remainder of the genome from the inferior breed. Genetic markers could be used in two ways: (i) to help identify the gene which is to be introgressed (ii) to select for (or against) a particular background genotype. The route usually proposed (e.g. Smith et al., 1987) is a number of generations of backcrossing of a population which carries the allele to be introgressed (from the donor population, i.e. inferior breed) to a recipient population (i.e. the commercial breed). An *inter se* cross to make the desired allele homozygous follows this backcrossing.

The efficiency of the MAI programme depends on the frequency of the introgressed allele in the final population and genetic progress for traits of economic benefit. In a backcross programme with no selection practiced, the proportion of genes from the donor line would halve each generation. Most studies have assumed that the allele to be introgressed can be identified without error from the phenotype so the allele frequency during the backcross phase remains at 50%, i.e. half of the animals in the population carry one copy of the desired allele. In practice, a single marker or a pair of markers flanking the QTL or major gene may be used, so that the frequency of the allele can be substantially less than 50% after a few generations of backcrossing (Visscher et al., 1996). Van Heelsum et al. (1997) highlighted another potential inefficiency of a MAI scheme when markers are not fully informative.

Genetic lag. One criterion for the efficacy of an MAI programme is the performance of the animals carrying the introgressed allele compared with the mean of the recipient line at the start of the programme. However, during the backcrossing and intercrossing phase the recipient (commercial) line will undergo selection, so that a better comparison is to compare the population carrying the allele with the commercial population at the same point in time.

This is analogous to the problem studied by Gama et al., (1993), who calculated the genetic lag for economic performance for various backcross and intercross programmes when introgressing a transgene in a nucleus pig population. It was assumed that the transgene genotype was known and genetic markers were not used to distinguish between the background genome of the founder transgenic animal and the rest of the population. These authors concluded that a gene would need an economic effect equivalent to 1-2 generations of selection to make its introgression worthwhile. The overall genetic lag of the final new commercial product (in which animals carry two copies of the desired allele) is not greatly influenced by the initial breed difference or the selection lost during backcrossing. However, there is a major effect of the selection lost during last two generations of intercrossing to make the animals homozygous for the desired allele. Hence, to reduce that lag at this latter stage, increasing population size or selection intensity, for example through technologies such as embryo transfer, should be considered, even for a prolific species such as pigs.

During the backcrossing phase and intercrossing phases, candidates for selection i.e. animals with the allele to be introgressed, will vary in their proportion of the genome that is from the donor and recipient line. For example, the mean proportion of the genome which is from the recipient line is 87.5% in the second backcross generation, but the theoretical range between animals is 81 to 94% for the pig genome (from Hill, 1993). To speed up recovery of the recipient genome, selection can take place on this so-called background genotype. Using markers for this type of selection has the advantage that no phenotypic data needs to be collected, and that markers can be scored on individuals of both sexes, early in life. Hospital et al., (1992) and Visscher et al., (1996) showed that progress equivalent to 1-2 generations of backcrossing can be gained by using this approach.

Recently, we have attempted to model the efficiency of reducing genetic lag by marker-assisted selection on background genotype (Visscher and Haley, 1997), assuming a genetic model in which the total genetic variance at generation t is the sum of the within-line variance (var_w) and the between-line variance (var_b), $var_t = var_w + var_{b(t)}$ with the latter depending on the initial line difference (D) and the variation in genomic proportion at generation t, i.e. $var_{b(t)} = D^2\ var_{g(t)}$. Only three to four informative markers per chromosome are needed to explain most of variation in genomic proportion. It was concluded that phenotypic or BLUP selection is superior if the initial line difference is small (a few standard deviations), but selection solely on markers could be efficient for very large line differences (>10 standard deviations).

Selection from a synthetic population. The efficiency of MAS depends heavily on whether it can be assumed that the association between certain

markers and quantitative traits is the same in different families. If it can, we speak of linkage disequilibrium across the population, and animals can be selected on the basis of their marker genotypes across families. Smith and Smith (1993) argued that markers should be found which are so close to the QTL that recombination between them and the QTL can be ignored, so that selection can be across families. The obvious way to generate such disequilibrium is to cross two different lines, so we might be crossing two commercial lines and selecting a new commercial line from the intercross. Lande and Thompson (1990), using theoretical derivations, and Zhang and Smith (1992) and Gimelfarb and Lande (1994), using simulation results, all assumed linkage disequilibrium throughout the population and considered the simplest situation, a cross between inbred lines. MAS from the intercross (an F_2 in these studies) then proceeds in two phases, (i) markers are selected based on estimates from a marker-quantitative trait association experiment, and (ii) animals are selected based on their marker genotype and their phenotypes (and/or phenotypes of their relatives). Zhang and Smith (1992) compared the efficiency of MAS (using only marker information), with selection on BLUP breeding values and selection on an optimum combination of marker and phenotypic information, and showed an increased efficiency of 10 to 20% of combined selection over BLUP selection. The suggested selection scheme proposed by Lande and Thompson (1990) results in the markers receiving too much weight in an index combining phenotypic information with marker information, because markers are treated as fixed effects and the best markers are selected. Whittaker et al., (1997) recently tried to correct for the overestimation of combined marker effects, by a simple cross validation procedure. This procedure worked well, and the resulting genetic gain was essentially the same as if MAS was performed with unbiased marker effects.

Within-line selection. If the population under consideration has been selected for many generations without crossing to divergent lines or breeds, it is unreasonable to assume linkage disequilibrium between markers and QTL unless they are very close together. Hence unless we have a very high-density of markers, marker-QTL associations have to be estimated and utilized within families.

How should we approach within-line selection using markers? We distinguish between traditional breeding schemes, and novel breeding schemes. In the former, markers are used an additional source of information to select between and within families, and the basic breeding structure and selection strategy is unaltered. The key question for this approach is how to weight the marker information in an optimum way. Several methods have been suggested which incorporate marker information in a BLUP (e.g., Goddard, 1992).

However, the scope for using genetic markers is in our opinion much larger when novel breeding schemes are considered, because the advantages of genetic markers (easy to measure on animals of either sex at any time in their life) can be exploited to create faster genetic progress. For example, at present the selection of which litter mates go onto a performance test may be random, whereas with marker information which explains some of the within family genetic variation, the sibs with the highest EBV could be re-selected. This type of pre-selection is appealing because there is no risk attached to it, since selection would be at random otherwise. Meuwissen and Goddard (1996) suggested a selection scheme for carcass traits, in which parents are selected on their pedigree index and marker genotype, and phenotypic information, i.e. carcass information is recorded on the non-selected individuals. Using this approach, they found a sustainable increase in genetic gain of about 30%.

The benefit of MAS within outbred populations depends on the assumptions regarding the genetic model for quantitative traits, in particular in the number of QTL alleles and the distribution of effects across loci. At the extremes, models assume either bi-allelic QTLs in the population, or an infinite number of QTL alleles. With respect to the distributions of QTL effects, assumptions range from uniform distributions to geometric series. These assumptions will impact upon the extra progress to be made with MAS, in particular in the comparisons between short and long-term gains. For example, if we assume a few bi-allelic QTLs with large effects and many minor loci with small effects, MAS selection will fix the favorable QTL alleles relatively quickly, resulting in short-term gain and a large reduction in genetic variance. If we assume a continuous distribution of effects at a particular locus, then we would predict an advantage of MAS even in the long term, because the marked QTL will always explain some of the genetic variation. At present, it is not clear how much difference these assumptions make in terms of comparisons between BLUP and MAS, and between short and long term response to selection. However, it is probably fair to say that we tend to overestimate the extra progress from MAS in traditional breeding schemes.

Genomic selection. The ability to use markers to speed the recovery of the recipient genotype demonstrates that the standard relationship matrix describes an average appropriate for the infinitesimal model. Under the finite locus reality there is variation around this average and marker information can be used to estimate this variation. This is exactly what we do when implementing marker-assisted selection for individual loci - the marker information is used to predict how the inheritance of the locus deviates from the average expectation. At the level individual loci the deviation from the average expectation is more extreme (i.e. an allele is inherited or not) than across the genome as a whole.

However, using the actual relationship matrix might improve the accuracy of breeding value estimation and hence improve genetic progress even in models where large numbers of loci of small effect are acting.

Marker technology. The prospects of using markers depend to some extent on the state of marker technology (i.e. costs, number, and informativeness of markers). As stated earlier approaching 1500 microsatellite markers are available and these would be the current markers of choice and they do make the application of MAS feasible. However, typing such markers has a cost (around one or a few dollars per marker per individual typed) and hence typing a large number of individuals for 200-300 markers covering the entire genome would be expensive. Thus it may be feasible to use microsatellite markers to select on limited areas of the genome (e.g. to introgress a QTL from one breed to another) but using many such markers at the same time in a breeding programme may be impossible until costs are reduced.

Microsatellite markers are also not completely informative. Thus it is not possible to use them to follow the inheritance of the linked section of the chromosome from all individuals. However, it seems unlikely that types of markers will be found that are much more informative than microsatellite markers. Furthermore, the problem can be overcome to some extent by combining information from several microsatellites. Thus if one microsatellite flanking a QTL is not informative, it will often be possible to replace it with another linked microsatellite. However, this may not completely recover the lost information (as the replacement marker will be further away) and may also not always be possible. Many of the theoretical and simulation studies performed to date have assumed that markers were completely informative and thus are likely to be too optimistic in their predictions. Van Heelsum et al., (1997) recently showed that uninformative markers could cause a large reduction in the efficiency of marker assisted introgression programmes.

Future technological developments may create new opportunities by making very high densities of markers available cheaply. We know that at the DNA level there is no shortage of variation in outbred species, with polymorphisms every few hundred base pairs. Thus the potential exists to find a marker very close to any gene of interest, so close that population wide linkage disequilibrium may exist, allowing selection across families as envisaged by Smith and Smith (1993). Ways of tapping into this large amount of DNA level variation are already been devised. For example, a large number of RAPD (random amplified polymorphic DNA) markers can be produced very quickly. Combining such markers with bulked segregant analysis (where bulked DNA samples from individuals at the two extremes of the phenotypic distribution are compared) has been shown to be effective for

detecting marker-trait gene linkages in plants (Michelmore et al., 1991). RAPDs are less than ideal for animal studies because they are dominant markers (i.e. the heterozygote cannot normally be distinguished from one of homozygotes) and have a low reproducibility between populations and laboratories. Other types of markers and technologies for exploiting these are being developed and may be more suited to animal populations than are RAPDs. For example, a new class of markers, Amplified Fragment Length Polymorphism (AFLP), developed for use in plants has been successfully applied to chicken populations (Herbergers et al., 1997), and is being studied for its applicability in pig populations. If AFLPs were suitable for pig populations, it would be relatively easy to completely saturate a particular genomic region of interest with genetic markers, and hence move much closer to trait genes. Such technological developments could completely alter our perspective on the use of markers in pig breeding programmes.

New reproductive technologies. Georges and Massey (1991) suggested speeding genetic progress in cattle throught 'velogenetics' by harvesting oocytes from calves *in utero*, thereby reducing generation intervals substantially. Using markers in such a scheme, which could also be applied to pigs, would allow rapid selection based solely upon markers, for example in an introgression programme. The combination of MAS and such embryo technologies could be further enhanced by technologies currently under development, such as nuclear transfer. If we allow ourselves to imagine that the technology will develop to a stage where cell differentiation can be controlled *in vitro* we can imagine that *in vitro* meiosis followed by fertilization may become possible. Utilizing this would allow for very rapid introgression, or with high-density marker maps and knowledge of marker-QTL associations, more generalized selection objectives (e.g. Haley and Visscher, 1997).

Discussion and conclusions

We have discussed what is possible at present, with today's marker and trait gene information, and how this information might be incorporated into pig breeding programmes. Methods to utilize markers differ in the amount of linkage disequilibrium which they require, and the use of phenotypic-marker information. The list of potential useful markers in pig breeding programmes is impressive (Haley and Archibald, 1997), and growing continuously. Haley and Archibald (1997) listed 8 QTLs for growth, 3 QTLs for backfat, mayor genes for meat quality and diseases, and 3 QTLs for litter size. Furthermore,

data from experimental divergent crosses are being analysed world-wide at present, and many more significant chromosome regions for a plethora of traits are expected to be published soon. Hence, there already is plenty of scope of utilizing this information, and some pig breeding companies are marketing genetic tests for litter size genes (Rothschild et al., 1996), and coat color.

The main traits of economic importance in pigs may be categorised as (i) performance traits (ii) reproduction traits, and (iii) disease resistance traits. For the performance traits, traditional selection methods are likely to remain competitive and sustainable. However, in syntetic populations there is scope for using markers, because we still expect a large amount of linkage disequilibrium in such populations, and there is growing evidence that the 'inferior' breeds may harbour 'positive' alleles that increase performance. For example, in a synthetic Meishan population, markers can be used to select against Meishan genomic regions associate with increased fatness, and for regions that may increase growth. For the reproduction traits, markers could increase the rate of progress substantially, because the traits are sex-limited, and usually have low heritabilities. The estrogen receptor (ER) polymorphism has been shown to affect litter size both in synthetic Meishan populations and within Large White commercial lines (Rothschild et al., 1996). Hence, in this case markers can be used to introgress the favorable ER allele into commercial dam lines , and to select within existing population. The third category of traits is more difficult to deal with, because they are difficult to measure. It is the phenotypic measurement of disease resistance traits which causes a problem, and not the marker technology. However, even for these traits there is already scope for manipulation using markers. Porcine stress syndrome can be eliminated from the nucleus populations by identifying carriers of the ryanodine receptor mutation, and markers which identify resistance to two strains of E. Coli have been reported (e.g., Haley and Archibald, 1997).

For the future, the amount of cheap genomic information on pigs is likely to increase, and will be a challenge to utilize the abundance of information. Coupled with exciting new reproduction techniques such as nuclear transfer, the time may come when generation intervals will be so short that fast introgression programmes and genomic selection breeding scheme will be possible.

Acknowledgements

We are grateful to BBSRC, MAFF, EU and the UK pig breeding industry MAS Consortium. Thanks to Philippe Baret for helpful comments.

REFERENCES

1. Archibald, A. L., J. F. Brown, S. Couperwhite, et al. (1994): Mamm. Genome 6:157-175.
2. Gama, L. T., C. Smith, J. P. Gibson (1993): Anim. Prod. 54, 427-440.
3. Georges, M., J. M. Massey (1991): Theriogenology 35:151-159.
4. Gimelfarb, A., R. Lande (1994): Genet. Res. 63:39-47.
5. Goddard, M. E. (1992): Theor. Appl. Genet. 83:878-886.
6. Haley, C. S., A. Archibald (1997): Progress in Pig Science. Univ. of Nottingham.
7. Haley, C. S., P. M. Visscher (1997): Submitted.
8. Herbergers, J., Veenedaal, T, Crooijmans et al. (1997): Plant and Animal Genome V, San Diego.
9. Hill, W. G. (1993): J. Hered. 84:212-213.
10. Hospital, F., C. Chevalet, P. Mulsant (1992). Genetics 132:1199-1210.
11. Lande, R., R. Thompson (1990): Genetics 124:743-756.
12. Michelmore, R. W., I. Paran, R. V. Kesseli (1991): Proc. Nat. Acad. Sci. 88:9828-9832.
13. Meuwissen, T. H. E., M. E. Goddard (1996): Genet. Sel. Evol. 28:161-176.
14. Rothschild, M. F., C. Jacobson, D. A. Vaske et al. (1996): Proc. Nat. Acad. Sci. 93:201-205.
15. Smith, C., T. H. E. Meuwissen, J. P. Gibson (1987): Anim. Breed. Abstr. 55:1-10.
16. Smith, C., D. B. Smith (1993): Anim. Breed. Abstr. 61:197-204.
17. Van Heelsum, A. M., P. M. Visscher, C. S. Haley (1997): Anim. Genet. 28:181-187.
18. Visscher, P. M., C. S. Haley, R. Thompson (1996): Genetics 144:1923-1932.
19. Visscher, P. M., C. S. Haley (1997): Submitted.
20. Whittaker, J. C., C. S. Haley, R. Thompson (1997): Genet. Res. (in press)
21. Zhang, W., C. Smith (1992): Theor. Appl. Genet. 83:813-820.

STRATEGIJE ZA SELEKCIJU POMOĆU MARKERA U PROGRAMIMA UZGOJA SVINJA

Sažetak

Prikazujemo strategije selekcije pomoću markera za genetske odlike u programima uzgoja svinja. Markere je najbolje upotrijebiti kad normalna selekcija nije vrlo djelotvorna, npr. za spolom ograničene osobine i osobine polovica. Odgovarajuća strategija ovisi o razmatranoj osobini (npr. dob ili ograničenje spolom), veličini neravnoteže povezanosti u populaciji i tehnologiji markera. Postoji veliko područje za iskorištavanje genetskih markera u programima uzgoja svinja. Budući programi selekcije mogu primjenom novih reprodukcijskih postupaka kao što su kloniranje i meiosis in vitro ubrzati genetski napredak.

Ključne riječi: svinje, genetski markeri, QTL, introgresija, selekcija pomoću markera

Primljeno: 20. 11. 1998.