Vol. 63 January, 1995 No. 1

Animal Breeding Abstracts

Utilizing genetic markers in pig breeding programmes

P.M. Visscher and C.S. Haley
Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, UK

ABSTRACT

Highly polymorphic genetic markers covering most of the porcine genome are now available for use by the pig breeding industry. Breeders could potentially use these markers (1) to combine genes from different breeds, (2) to select animals within commercial lines, and (3) in novel applications such as controlling inbreeding or heterosis. There are at present relatively few candidate genes available for introgression, although major genes for coat colour, litter size, meat quality and intramuscular fat have been reported. The efficiency of marker-assisted introgression and marker-assisted selection depends on the nature of genetic variability within and between breeds and commercial lines. Experimental crosses between breeds and between commercial lines should provide parameters, such as the distribution and effects of genes affecting economically important traits, needed to assess the efficiency of using genetic markers.

INTRODUCTION

Recently linkage maps of the porcine genome have been developed (Archibald *et al.*, 1994; Ellegren *et al.*, 1994; Rohrer *et al.*, 1994). By putting the publicly available information together it is already possible to produce a map with approximately 1000 highly polymorphic (microsatellite) markers spaced throughout the genome. Hence, genotyping pigs for evenly spaced polymorphic markers 20 centimorgans (cM) or less apart is now possible. (The centimorgan is one hundredth of a morgan, the unit of distance on a linkage map. The porcine genome is around 26 morgans in length, so around 130 evenly spaced markers are needed to produce a map with 20 cM between markers. Twenty centimorgans corresponds approximately to a recombination fraction of 0.17. The recombination fraction measures how closely linked genes or markers are, and varies between 0 for two genes at the same position on a chromosome and 0.5 for two unlinked genes).

To what uses could marker data be put? In this paper we will focus on the use of markers to enhance genetic progress in breeding programmes. Markers can also be used to control and confirm parentage and for individual and line identification purposes, but we consider these possibilities no further. The essence of using genetic markers in breeding programmes is that they allow the inheritance of whole segments of chromosomes to be traced from parents to offspring. If a particular segment contains a gene or genes which contribute(s) 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. It is likely that genetic markers will have no effect on performance themselves, thus their value is solely that they mark chromosome segments containing genes affecting performance. The genes that are followed using markers may be so-called major genes, causing qualitative differences between animals (e.g. genes for coat colour and some disease genes), or they may be some of the many genes (quantitative trait loci, QTL) which contribute to quantitative variation between animals for performance traits such as growth, fatness and litter size.

Genetic markers have a number of features which make their use in breeding programmes potentially attractive: (1) 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; (2) DNA-based genetic markers can be measured on animals of both sexes at any age, so that, for example, boars can be genotyped for litter size markers, and markers for carcass quality can be measured on live animals; and (3) genetic markers can explain some of the within-family genetic variation. This last is of value because when using phenotypic information (or Best Linear Unbiased Prediction, 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 there are no phenotypic observations on them. Hence, markers that partly explain within-family variance may be most beneficial for sex- and age-limited traits.

In this article the use of markers in practical pig breeding programmes is addressed by reviewing the literature on this subject and adding some points for discussion. We do not intend to discuss the use of markers in a more theoretical and general framework. See, for example, the review by Soller (1990) for a more general discussion.

Broadly speaking, three types of applications for genetic markers in pig breeding programmes can be identified:

(1) Using markers to combine alleles from different breeds or lines.

(2) Marker-assisted selection (MAS) within a breed or commercial line.

(3) Novel applications which do not fit into either of the above categories.

COMBINING ALLELES FROM DIFFERENT BREEDS

No single commercial line or breed is likely to contain all of the best alleles for all traits of economic importance. If alleles of superior value for one or a few traits can be identified in breeds which are inferior in overall economic performance, an efficient crossing programme using genetic markers might increase overall economic performance substantially. Although the starting point of any such programme may be the same, i.e. a cross between two breeds or lines, the end-point (a commercial product) can be reached through many different routes. For example, if one line is inferior for all but one or a few alleles, we might use marker-assisted introgression (MAI). This is the process of introgressing particular alleles for major genes or QTL from one breed (or line) into another with the aid of genetic markers, by repeated backcrosses to the superior line. On the other hand, if the two lines are of similar genetic merit we might consider selecting directly from the F_2 intercross with the aid of genetic markers. Between these two extremes there is a continuum of possibilities with varying numbers of rounds of backcrossing prior to intercrossing the animals and selecting within the intercross. The two extreme options have received some attention, but little attention has been paid to which point in the continuum is optimal for particular circumstances.

Marker-assisted introgression

Using introgression, genetic markers could be used in two ways: (1) using markers for the gene which is to be introgressed, and (2) using markers to select for (or against) a particular background genotype. In most recent studies (Hillel *et al.*, 1990, 1993; Hospital *et al.*, 1992; Groen and Timmermans, 1992; Groen and Smith, 1995) it has been assumed that the genotype at the gene to be introgressed was known exactly, hence there was no need to use markers to aid its introgression, and attention was focused on the background genotype. In fact the genotype of genes to be introgressed will not usually be known as these will often be QTL, which can only be genotyped imprecisely, if at all, using phenotypic information, or they may be major loci (e.g. coat colour), which more often than not have dominance/recessive relationships and hence heterozygotes can not be distinguished from one of the homozygotes.

Markers can be used to select for a certain genetic background because whole segments of chromosomes are inherited. Hence, in the progeny of a particular breed cross the part of the genome around a marker which originates from one breed is also more likely to come from that breed. However, the association between a marker and genes linked to it will be eroded over time due to recombination.

Introgression of a desired gene

Having located a desired gene and genetic markers associated with it in an inferior breed, the aim of the introgression phase is to fix the favourable alleles in the commercial population with as little as possible of the remainder of the genome from the inferior breed. Fixation of favourable alleles refers to the QTL or major gene and not to the markers, because fixing marker alleles is not an aim in itself. The route usually proposed (e.g. Smith *et al.*, 1987; Hillel *et al.*, 1990; Hospital *et al.*, 1992; Groen and Timmermans, 1992) is to use a number of generations of backcrossing of a population which carries the allele to be introgressed (from the donor population, i.e. the inferior breed) to a recipient population (i.e. the commercial breed), followed by an *inter se* cross to make the desired allele homozygous.

The efficiency of the MAI programme depends on (1) the frequency of the introgressed allele in the final population, and (2) the genetic progress for traits of economic benefit. As most studies have assumed that the allele to be introgressed can be identified without error by a single marker, the frequency of the allele during the backcross phase remains at 50%. In practice, a single marker or a marker bracket (a pair of markers flanking the region of the gene of interest) associated with the QTL or major gene is likely to be used, so that the frequency of the allele may be substantially less than 50% after a few generations of backcrossing (Visscher *et al.*, 1995). The donor line (or breed) may be inferior for other traits of economic importance. For example, although the Meishan is superior to commercial European breeds for litter size and related traits, it is inferior with respect to lean growth and fatness traits. One criterion for comparison is the performance of the animals carrying the introgressed allele and 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 of the population carrying the allele with the commercial population at the same point in time. This is analogous to the problem studied by Haley (1991) and extensively by Gama *et al.* (1992), who calculated the genetic lag for economic performance for various backcross and intercross programmes when introgressing a transgene in a nucleus pig population. In that study 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 one or two generations of selection to make its introgression worthwhile.

Selection for the background genotype

Hillel *et al.* (1990, 1993) proposed that DNA fingerprints could be used for introgression of alleles in backcross populations by selecting for or against a certain genomic background. Their theory is based on a number of 'chromosome segments', which effectively are unlinked loci. However, there are some problems with their theory and results. In general, the class of minisatellite markers (fingerprints) is not very suitable for use in introgression programmes (the markers are dominant and the fingerprint loci will often not be mapped, but are known to be non-randomly distributed across the genome). Hillel *et al.* (1990, 1993) implicitly assume that the proportion of the genome from one line (or breed) in a (back)cross is the same as the proportion of unlinked markers from the same line. This is

not correct, because it ignores recombination around the marked loci, as was also noted by Hospital *et al.* (1992) and Groen and Smith (1995), hence the results from Hillel *et al.* (1990, 1993) are unrealistic.

Groen and Timmermans (1992) presented a simulation study on the use of genetic markers to increase the efficiency of introgression through selection for the background genotype. Unfortunately, their results are difficult to read from the graphs presented. When comparing phenotypic selection and selection using markers in a backcross programme, they conclude that not much benefit is to be expected from using markers. However, as pointed out by the authors, this conclusion depends on the parameters used (in particular on how effective phenotypic selection is). Using the parameters of Groen and Timmermans (1992), selection using markers has a small advantage over phenotypic selection for at least the first three backcross generations. Groen and Timmermans (1992) assume that the allele to be introgressed is assumed to be known, i.e. is identifiable in progeny. In practice this is unlikely to happen, and some recombination will occur between the QTL to be introgressed and a marker for it. The authors simulated a 'linkage block' with alternating QTL (for some quantitative trait we want to select for) and marker loci (i.e. QTL1 – marker 1 – QTL2 – marker 2, etc.). However, the recombination fraction between adjacent loci was 0.2. Assuming Haldane's mapping function without interference (Haldane, 1919), the recombination rate between adjacent marker loci was approximately 0.32, corresponding to a marker distance of 51 cM. This distance is rather large for the linkage maps currently available, and makes selection on markers seem less efficient.

The excellent study of Hospital *et al.* (1992) deals in detail with introgression in backcross breeding programmes. As in the other studies, genotypes for the introgressed allele are assumed to be known. The authors do not consider selection on a quantitative trait. For the chromosome with the allele to be introgressed, the authors only consider two markers. One of the main conclusions of Hospital *et al.* (1992) was that retrieving the recipient's genome was approximately two generations faster if markers were used.

Groen and Smith (1995) followed on from the study of Groen and Timmermans (1992), and studied the efficiency of phenotypic selection and selection on markers in backcross and intercross programmes. Selection was in two stages: firstly, animals were selected carrying the allele to be introgressed, and among these animals, those with the best phenotype or those with the largest number of markers from the recipient line were selected. Groen and Smith (1995) concluded that selection using markers is always inferior to selection for phenotypes. However, the assumption with regard to the distribution of gene effects for the quantitative trait under consideration in the donor and recipient line is important. Groen and Smith (1995) assumed QTL of equal size, with a frequency of 0.7 in the recipient line, and 0.6 in the donor line. In this case most of the genetic variation is *within* lines, so that selection on markers is not expected to contribute much.

Selection from the intercross

The efficiency of MAS in commercial populations depends heavily on whether it can be assumed that the association between certain markers and quantitative traits is the same in different families (Beckmann and Soller, 1983; Smith and Simpson, 1986). If this is the case we speak of linkage disequilibrium throughout 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, 1993) 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: (1) markers are selected based on estimates from a marker-quantitative trait association experiment, and (2) animals are selected based on their marker genotype and their phenotypes (and/or the phenotypes of their relatives).

For the first phase, an important question is how to select the markers which explain some of the genetic variance in the population. A simple multiple regression approach will result in an overestimate of the total variance explained by the markers (Wishart, 1931). Therefore, Lande and Thompson (1990) proposed a two-stage procedure. In the first stage, a set of promising markers is selected from all available markers. In the second stage, estimates of the regression coefficients for the selected markers are obtained from a new (independent) sample of animals. In this case, the marker effects are unbiased (Lande and Thompson, 1990). However, as pointed out by Visscher (1994), the total amount of variance explained by the selected group of markers may still be biased upwards.

Lande and Thompson (1990) give examples of the theoretical relative efficiency of MAS (including both marker information and phenotypic observations) compared to traditional index selection, and show that MAS performs much better for traits with low heritability (h^2) and large common family effects (c^2 = proportion of total phenotypic variance due to common family effects). For example, for a trait with $h^2 = 0.10$ and $c^2 = 0.25$, MAS is approximately 50% more efficient than a selection index based on individual and full-sib performance if the markers explain 20% of the genetic variance. If the markers explain 40% of the genetic variance, MAS is about 3.5 times more efficient than index selection for $h^2 = 0.05$ and $c^2 = 0.25$.

A number of simulation studies suggest that the theoretical predictions of Lande and Thompson (1990) are too optimistic, especially if the time horizon is more than one generation. Gimelfarb and Lande (1994) contrasted MAS with just phenotypic selection by simulation and found that MAS was more efficient for at least five generations of selection for a single trait. If the regression coefficients for selected markers were re-estimated each generation, MAS was more efficient for the first ten generations of selection. Surprisingly, practically all the weight in the index combining marker information and phenotypes was put on the marker information (Gimelfarb and Lande, 1994). Even with re-estimation of the marker effects, this is unexpected, since after a few generations of selection most genetic variance will be within rather than between marker genotypes.

Zhang and Smith (1992, 1993) 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. Although the authors clearly showed increased efficiency of 10–20% for combined selection compared with BLUP selection, they concluded that MAS using linkage disequilibrium has limitations until close linkages of markers and QTL are available (Zhang and Smith, 1993). Zhang and Smith (1993) showed the loss in efficiency of estimating the marker effects with imprecision. For example, reducing the population size in which marker–QTL associations were estimated from 1000 to 100 reduced the response to selection by about 50%. Similarly, Gimelfarb and Lande (1994) concluded that population size

was the main parameter in determining the relative efficiency of MAS. In contrast, Lande and Thompson (1990) concluded that estimating the marker effects with error did not significantly reduce the response to selection. However, these authors assumed that the appropriate markers were included in the selection index, and that it was only the genetic variance associated with the markers that was subject to error, an unrealistic assumption.

Conclusions and remarks

In an MAI programme, markers can clearly be used to introgress a QTL allele (or other difficult-to-measure gene) which would otherwise be lost or greatly reduced in frequency during the backcrossing programme. However, little work has been done to assess the value of this, particularly in economic terms. For the recovery of the background genotype, as a comparison of the results from Hospital *et al.* (1990) and Groen and Smith (1995) suggests, conclusions from studies on the efficiency of MAI heavily depend on the assumption regarding the distribution of gene effects within and between breeds. The same would also apply to selection from the F_2 cross. It is difficult to define a realistic distribution of gene effects which explains genetic variation between and within lines. One extreme is to assume that the lines are inbred and that alternative alleles are fixed in the breeds (as in Groen and Timmermans, 1992; Visscher *et al.*, 1995). Another extreme is to assume that allele frequencies are nearly the same for different breeds (as in Groen and Smith, 1995). In practice, it would depend on which two breeds we consider. For example, if we compare the Meishan and the commercial European pig, a reasonable assumption may be that the breeds are fixed for alternative alleles (by assuming divergent selection in the past), at least for QTL of major effect. When two commercial pig lines are compared, a reasonable assumption may be that genes of large effects are fixed either for the same or for different alleles, and that other genes have similar frequencies in the two lines. In any event, as the efficiency of MAI and selection from the F_2 will depend on the distribution and frequencies of QTL for economic performance, research is needed to determine these in pig breeds and lines and then to revise theoretical studies in the light of these findings.

In practice, the superiority of selection on markers over phenotypic selection is likely to be greater for sex-limited traits and for traits that are expressed later in life. For example, boars can be selected for markers associated with litter size, and piglets can be selected for markers associated with lean growth when they are one day old (or even *in utero*, if that is desired). Lande and Thompson (1990) derived the relative efficiency of selection for a sex-limited trait using markers (using markers for both sexes, and the phenotypic performance of only one sex) compared to phenotypic selection of a single sex. For example, for a trait with $h^2 = 0.10$, a proportion of genetic variance explained by the markers of 20%, and a selection intensity in males which is twice that in females, the relative benefit of using markers for sex-limited traits is about three times that of using markers for traits that are not sex-limited.

A relevant question for the pig breeding industry is what would happen if marker information were to be ignored and selection from the intercross was performed using phenotypic information alone (e.g. index selection or BLUP). In the short term, theoretically at least, using all available information must lead to more genetic progress. In the long term it is not clear what the best strategy is, because selection response on genes of small effect which is sacrificed by putting selection pressure on a QTL of large effect may not be recovered (Gibson, 1994; Ruane and Colleau, 1994).

SELECTION WITHIN POPULATIONS USING MARKERS

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 marker–QTL associations have to be estimated within families, because we are unlikely to find marker alleles which are associated with certain QTL alleles in all families (unless the marker is also a genetic variant in a gene which is thought to influence the trait directly). In dairy cattle breeding this is the route breeders are taking (Soller, 1990; Weller *et al.*, 1990; Brascamp *et al.*, 1993; Hoeschele and Romano, 1993). In particular, dairy breeders have proposed a granddaughter design (GDD; Weller *et al.*, 1990) to find markers associated with milk yield traits. In the GDD, sons of a particular widely used grandsire are genotyped for markers for which their sire is heterozygous. The sons get their 'phenotypes' from their daughters (the granddaughters of the grandsire), and associations are made between markers and estimated breeding value (EBV; based upon the performance of the daughters) all within one sire family. For nucleus pig breeding programmes, there is an analogous case for litter size. Sons of a widely used (through artificial insemination) boar are genotyped for markers, and get their EBV from observations on their progeny. If useful markers are detected, the grandoffspring of the sons (i.e. the great-grandoffspring of the sire) can be preselected for litter size based upon their marker genotype. Applied to milk traits in dairy cattle, such a scheme gives 10–20% extra genetic progress (Meuwissen and Van Arendonk, 1992), and Brascamp *et al.* (1993) showed that MAS was profitable for the case of competing dairy cattle artificial insemination companies. Although pig breeding nucleus schemes are similar to dairy cattle nucleus schemes, the extra response to selection in pig breeding programmes is likely to be smaller because boars with high EBV for litter size are not as widely used as bulls with high EBV for milk yield, and because the relative economic importance of milk yield in dairy cattle is larger than that of litter size in pigs. Furthermore, a general problem with the GDD is that the population in which to apply the marker information will be three or more generations away from the population in which the marker effects were estimated.

NOVEL APPLICATIONS

The applications addressed above use markers as an adjunct to phenotypic selection for the additive genetic breeding value. However, there are likely to be many other novel applications of using marker information which have not ye been explored in any detail. For example, markers can be used to determine a more exact inbreeding coefficient than calculations solely based on pedigree information. In any mating between related animals the average inbreeding coefficient of the progeny can be predicted; however, because whole segments of chromosomes are inherited, some progeny will be 'more inbred' than others. In a typical mammal, for example, among full-sibs produced by a sire–daughter mating the average inbreeding coefficient will be 25% and the expected standard deviation of the inbreeding coefficient approximately 7% (from Hill, 1993), so that some animals will have an actual level of homozygosity by descent of 10%, while others will be 40% inbred. In a recent study, Christensen *et al.* (1994) looked at 21 informative genetic markers (i.e. markers in progeny that can be unequivocally traced backed to parents) in progeny of a sir daughter mating in pigs, and among 37 animals found that realized inbreeding, as assessed using the markers, varied from 6 to 42% with a mean of 25% (i.e. as expected). Furthermore, growth performance was negatively related t the realized inbreeding.

The study by Christensen *et al.* (1994) suggests that markers should allow the actual inbreeding coefficient of animals to be assessed and hence, if desired, animals could be selected from within families on the basis of their realised inbreeding coefficient. However, the extent to which such selection is of value in reducing the long-term accumulation of inbreeding without jeopardising progress under selection is unknown. There is a potential conflict between using markers for selecting the best animals within a family (which are likely to be related) and selecting the least related to minimise inbreeding, and this subject needs further study. However, recent studies on increasing the long-term response to selection using index and BLUP selection have shown that there is scope for decreasing levels of inbreeding substantially without sacrificing response to selection (Brisbane and Gibson, 1994; Wray and Goddard, 1994).

As a complement to the control of inbreeding, markers might be used to improve heterosis in crossbred animals. If there are genes of large effects which are clearly overdominant, i.e. the heterozygote is superior to either homozygote, then markers could be used to genotype animals for the gene and separate (sire and dam) parent lines could be created which are fixed for alternative alleles at the locus. It is not clear whether overdominance for single loci (or for chromosomal segments comprising a number of linked loci) is a common phenomenon in pigs, and hence whether such an approach is worthwhile. Future QTL studies in pigs will clarify this point, but—at least in some plant species—overdominance for chromosomal segments appears widespread (Stuber *et al.*, 1992) and the manipulation of heterotic chromosome segments using markers seems to hold promise.

In theory markers can also be used to characterise different lines or breeds in terms of their allele frequencies for a number of markers. Two possible uses of such characterization would be to help breeders police proprietary rights on their stock or to use genetic distances between lines to predict the overall amount of heterosis when two lines are crossed. Although such possibilities are being pursued to some extent in poultry and in plants, they look less attractive for pigs. It would be difficult to prove that animals come from a certain population on the basis of their marker genotypes alone, and even more difficult to argue such a proof in a court of law! Even if heterosis may be predicted from the allele frequency differences between lines, there are sufficiently few lines of pigs available to make prediction of hybrid vigour as valuable as it is in maize breeding, for example.

TRAIT GENES TO MANIPULATE USING MARKERS

For some of the applications of markers that have been discussed it is necessary to be able to find genes (or sections of the chromosome) which explain a significant proportion of the trait variance (this does not apply to all uses—for example the use of markers to control inbreeding). What genes have been found and what are the prospects of finding more?

Litter size

Some pig breeds, in particular the Chinese breeds, are known for their prolificacy (e.g. Haley and Lee, 1993). The Meishan breed (one of the Chinese breeds) has been crossed with commercial pigs at various experimental pig stations to investigate the nature of the breed differences, in particular those for fertility traits. One outcome of the QTL mapping experiments presently under way may be that QTL are detected for litter size. Indeed, one group in the USA has found a marker (a candidate gene) which is thought to explain about 1.4 piglets born per litter in a (dam) line derived from crosses between the Meishan and US breeds (Rothschild *et al.*, 1994). Sows with one or two copies of the favourable allele for the oestrogen receptor, which originated from the Meishan, had approximately 1 more piglet born alive than sows without any copy of the favourable allele (Rothschild *et al.*, 1994). Markers for such QTL, and in general markers to identify the remainder of the Meishan genome, can then be used through introgression to improve the overall efficiency of pig production.

Lean growth and fatness

Crosses between the Meishan and commercial pigs, or between the wild boar and commercial pigs, may identify QTL of large size segregating in F_2 populations. Andersson *et al.* (1994) analysed approximately 200 F_2 animals from a wild boar × Large White cross and found evidence for QTL on chromosome 4. The two largest QTL, for abdominal fat and average backfat, explained 19 and 18% of the phenotypic variance in the F_2 population, respectively. A QTL for growth, explaining about 12% of the F_2 variance, was also located (Andersson *et al.*, 1994). If these QTL are also segregating in other breeds, these findings can be used in combination with QTL for other traits. For example, if QTL for fatness are segregating in the Meishan population, introgression of an allele for litter size can be combined with extragression of alleles for fatness.

Meat quality

No markers for genes affecting meat quality have yet been mapped in pigs (other than the halothane gene). However, at least two major genes whose effects can be detected at the phenotypic level have been described—a gene affecting the yield of cured meat found in synthetic breeds which contained genes from the Hampshire (Le Roy *et al.*, 1990) and a gene affecting intramuscular fat content, found in a Meishan cross (Janss *et al.*, 1994; Van Arendonk *et al.*, 1994). Both of these should be readily mapped, and once this has been done their selection using linked markers would be significantly easier than selection on the traits directly.

Disease resistance

Disease resistance alleles are the classical example of candidates for transgenic programmes (Smith *et al.*, 1987). Similarly, if some breeds are found to be more resistant for certain diseases of great economic importance to the pig industry, and if markers are available which are associated with these diseases, MAI can be used to introgress the resistance allele in commercial populations. Major genes for diseases may also segregate within the present commercial populations, and the discovery of a marker for malignant hyperthermia (MacLennan *et al.*, 1990) or halothane susceptibility (Eikelenboom and Minkema, 1974) is an example of this. Resistance to the K88 strain of *Escherichia coli* is known to be determined by a recessive gene (Gibbons *et al.*, 1977), and pigs lacking the receptor for K88 are resistant to neonatal diarrhoea caused by this strain of *E. coli*.

In mice more than 20 genes conferring resistance to infectious disease (caused by viruses, bacteria and parasites) have been mapped (Malo and Skamene, 1994). In man, searches for disease genes have also been successful recently. For example, after enormous worldwide efforts, genes causing Huntington's chorea and breast cancer have been found. However, in pigs such diseases with large effects are unlikely to be maintained in the population because affected animals are likely to die or be culled. Diseases of relevance to the pig industry, for example leg problems, are more likely to be controlled by many genes, and have not been characterised well at the phenotypic level, let alone at the genetic level. The example of mice suggests that it may be worthwhile searching for genes

conveying resistance to endemic infectious diseases in the pig, but these studies would not be easy to perform.

Colour genes

Many UK abattoirs penalise slaughter pigs with coloured spots, so there is an economic incentive at present to breed for pure white pigs. For example, one UK breeding company recently launched a commercial line of white Duroc pigs. The colour white is dominant in pigs and the gene responsible has been mapped to chromosome 8 and linked markers are known (Johansson *et al.*, 1992). The advantage of using markers in this case is that heterozygotes for the white allele (which themselves are phenotypically white) can be identified. Using classical selection without any markers will still result in the occasional coloured pig, since heterozygotes will remain in the population for a long time, and so the occasional homozygote for the colour allele will be produced.

Transgenes

Any transgene to be spread through the population will probably be incorporated in the target population using introgression. The difference between introgressing a transgene and a QTL or major gene is that with transgenes we are likely to have few founder transgenic animals (perhaps only one), and that animals carrying the transgene can usually be identified without error, whereas we may only have a linked marker for the QTL or major gene. Nonetheless, markers may be of value in speeding the recovery of the background genotype of the target commercial population.

From the brief discussion above it is clear that there are not many candidate genes available for introgression at present. At various experimental pig stations trials are under way in which divergent pig breeds have been crossed. These experiments may result in the identification of suitable markers for genes of interest to the breeding industry. Evidence for the segregation of alleles for fatness (Andersson *et al.*, 1994) and litter size (Rothschild *et al.*, 1994) in such crosses has already been reported.

MARKER TECHNOLOGY

The prospects of using markers depends to some extent on the state of marker technology (i.e. costs and the number and informativeness of markers). As stated earlier, approaching 1000 microsatellite markers are available; 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 100–150 markers covering the entire genome would be costly. 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 are thus likely to be too optimistic in their predictions. Future studies and actual applications will have to account for this problem and use information from multiple linked markers.

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 variation at the DNA level have already been devised. For example, a large number of random-amplified polymorphic DNA (RAPD) markers can be produced very quickly (Williams *et al.* 1990). Combining such markers with bulked segregant analysis (in which 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 the homozygotes) and have low reproducibility between populations and laboratories, but other types of markers and technologies for exploiting these are being developed and may be more suited to animal populations. Such technological developments could completely alter our perspective on the use of markers in pig breeding programmes.

DISCUSSION

In terms of the potential use of markers in breeding programmes, the main consideration of breeding companies will be that of gain versus risk, i.e. what can they expect to gain by using markers, and what can they lose if things go wrong. The gain versus risk issue is different from the one in traditional phenotype-based selection on a single trait, because estimating genetic parameters (such as h^2) with error for traits in the selection objective usually does not lead to a substantial reduction in response to selection. However, the situation is analogous to the use of indirect indicators (perhaps physiological traits) of genetic merit which themselves have no intrinsic value. In this latter case, Sales and Hill (1976) have shown that poor estimates of genetic parameters can result in less progress being made using indirect indicators than using direct selection. In the case of MAI and MAS, the errors may also have large effects. Potential sources of error are the location of a gene and its effect, i.e. (1) what if we think a particular marker is linked to a gene of interest whereas it isn't? and (2) what if our estimate of a marker effect is far from its real effect? False positives for marker effects are potentially the worst kind of error, because of wasted selection pressure, so testing for significant marker effects should be stringent. At present, the problem in hypothesis testing is one of defining the null hypothesis (i.e. what number and sizes of QTL do we expect *a priori* in a cross between two breeds?), and one of taking multiple testing into account. As an alternative (or complement) to stringent hypothesis testing, estimates of marker-linked effects could be 'regressed back' to some extent (i.e. the weight put on marker-linked effects in the selection index is related to our belief that they are 'real'). One way of doing this would be to treat marker effects as random, and Zhang and Smith (1992) show that this can be advantageous. This is unlikely to be optimal, however, as the regression should be based upon the actual distribution of QTL effects, which is of course presently unknown.

Poor estimates of QTL effects or detection of spurious QTL that do not actually exist is a problem if it leads us to divert selection pressure from worthwhile objectives (e.g.

phenotypic selection). This problem is avoided if selection on markers replaces random selection; then even if the QTL effects are poorly estimated or entirely spurious we should do no worse than random selection. In pigs the opportunity may exist to use markers in this way to select the young piglets that will be performance tested (i.e. in situations where there is reproductive redundancy, so not all animals are performance tested). This might be even more favourable if litter sizes were increased (e.g. using embryo technologies such as multiple ovulation and embryo transfer) to increase the selection opportunities available (Gomez-Raya and Gibson, 1993). Such an approach might be especially favourable if the traits selected using markers were complementary to those recorded on the performance tested individuals, such as litter size in males and disease resistance and meat quality in both sexes.

Poor estimates of the position of QTL which do actually exist can be a problem if selecting on markers thought to be flanking a QTL in an introgression programme, for example. From QTL mapping experiments we can expect a standard deviation of the position of a QTL of approximately 1.5–3 cM even for genes of the largest effect with a dense and informative map (Haley and Knott, 1992). Even for those QTL mapped with reasonable accuracy, in an appreciable number of cases the estimated position of the QTL was not between the markers that actually flanked it.

In introgression programmes the risk is of losing the allele we want to introgress. If only one marker is used for introgression, and if the position of the gene is estimated from a QTL mapping experiment, it is very easy to lose the allele of interest through recombination between the marker and QTL, although using two or more markers could overcome this (Visscher *et al.*, 1995).

Whether or not to embark on a crossbreeding programme with a breed that is inferior, apart from perhaps one superior gene, is a major decision for pig breeding companies. Although the end product may be saleable, animals from the intermediate backcross populations may not be easy to sell. Apart from this cost, there is the cost of genotyping a large number of animals for a large number of markers. Assuming a total map length of 2600 cM and markers approximately 20 cM apart throughout the genome, 130 markers per animal are needed in each backcross generation. As outlined previously, a further cost is that associated with genetic lag: during the time of backcrossing and intercrossing the nucleus population will improve, so that the final product will always lag behind for traits other than the one associated with the introgressed allele. Gama *et al.* (1992) expressed the cost of introgressing a transgene into a nucleus population as genetic lag. They concluded that the transgene should have an effect of approximately 10–20% or more of the mean economic performance to be worthwhile introgressing. Adding the costs of genotyping and reduced sales may result in a higher break-even point for introgressing an allele from one breed into another. Hence, the allele to be introgressed must have a very large effect on economic performance before introgression becomes profitable.

It seems likely that some of the earliest opportunities will be in programmes where there is already a benefit in crossing two different breeds. For example, breeding companies are already crossing the Chinese Meishan with Western breeds to create synthetic maternal lines (Mercer and Hoste, 1994). In such a cross, markers are already available to manipulate colour (Johansson *et al.*, 1992) and litter size (Rothschild *et al.*, 1994), and one would suppose that they could be found for growth rate and subcutaneous fat depths (Andersson *et al.*, 1994) and for the 'intramuscular fat gene' (Janss *et al.*, 1994). Hence there are already opportunities to use markers in current breeding programmes. In future it is likely both that the cost of genotyping animals will come down and that strategies to predict breeding values using markers will improve, and so opportunities for applying markers to breeding programmes are likely to increase.

Finally, we should not forget that the possible extra financial gain to be made by breeding companies by using marker information may well come from the indirect effects of marketing and PR rather from the direct effects of extra genetic progress.

SUMMARY

Given a QTL with an effect worth using in the combining of genes from different breeds or within commercial populations, we have reviewed how to effectively use such information. However, our main point throughout this paper is that we need more information to assess the value of MAS, because different sets of assumptions lead to different conclusions about the value of markers in selection programmes. In particular, we need to know the distribution and effects of QTL between and within breeds. Present experiments being conducted worldwide will hopefully provide the information needed to assess the value of MAS more accurately.

ABBREVIATIONS

BLUP	Best Linear Unbiased Prediction
cM	centimorgan(s)
EBV	estimated breeding value
GDD	granddaughter design
MAI	marker-assisted introgression
MAS	marker-assisted selection
QTL	quantitative trait locus, loci
RAPD	random amplified polymorphic DNA

ACKNOWLEDGEMENTS

This work was partially funded by the Marker Assisted Selection Consortium of the British Pig industry. We thank Derrick Guy, Bill Hill, Sam Hoste, Jon Mercer, Olwen Southwood, Robin Thompson, Hein van der Steen and Annemarie van Heelsum for helpful comments on an earlier version of the manuscript.

REFERENCES

Andersson, L.; Haley, C.S.; Ellegren, H.; Knott, S.A.; Johansson, M.; Andersson, K.; Andersson-Eklund, L.; Edfors-Lilja, I.; Fredholm, M.; Hansson, I.; Håkansson, J.; Lundström, K. (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**, 1771-1774.

Archibald, A.L.; Brown, J.F.; Couperwhite, S.; McQueen, H.A.; Nicholson, D.; Haley, C.S.; Coppieters, W.; Van de Weghe, A.; Stratil, A.; Winterø, A.K.; Fredholm, M.; Larsen, N.J.; Nielsen, V.H.; Milan, D.; Woloszyn, N.; Robic, A.; Dalens, M.; Riquet, J.; Gellin, J.; Caritez, J.C.; Hue, D.; Burgaud, G.; Ollivier, L.; Bidanel, J.P.; Vaiman, M.; Renard, C.; Geldermann, H.; Davoli, R.; Ruyter, D.; Verstege, E.J.M.; Groenen, M.A.M.; Davies, W.; Høyheim, B.; Keiserud, A.; Andersson, L.; Ellegren, H.; Johansson, M.; Marklund, L.; Miller, J.R.; Anderson Dear, D.V.; Signer, E.; Jeffreys, A.J. (1994) The PiGMaP Consortium linkage map of the pig (*Sus scrofa*). *Mammalian Genome*, in press.

Beckmann, J.S.; Soller, M. (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theoretical and Applied Genetics* **67**, 35-43.

Brascamp, E.W.; Van Arendonk, J.A.M.; Groen, A.F. (1993) Economic appraisal of the utilization of genetic markers in dairy cattle breeding. *Journal of Dairy Science* **76**, 1204-1213.

Brisbane, J.R.; Gibson, J.P. (1994) Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* **19**, 135-138.

Christensen, K.; Fredholm, M.; Winterø, A.K.; Andersen, S. (1994) Effect of identical homozygosity on growth in pigs. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* **19**, 155-158.

Eikelenboom, G.; Minkema, D. (1974) Prediction of pale, soft and exudative muscle with a non-lethal test for the halothane induced porcine malignant hyperthermia syndrome. *Netherlands Journal of Veterinary Science* **99**, 421-426.

Ellegren, H.; Chowdhary, B.P.; Johansson, M.; Marklund, L.; Fredholm, M.; Gustavsson, I.; Andersson, L. (1994) A primary linkage map of the porcine genome reveals a low rate of genetic recombination. *Genetics* **137**, 1089-1100.

Gama, L.T.; Smith, C.; Gibson, J.P. (1992) Transgene effects, introgression strategies and testing schemes in pigs. *Animal Production* **54**, 427-440.

Gibbons, R.A.; Selwood, R.; Burrows, M.; Hunter, P.A. (1977) Inheritance of resistance to neonatal *E. coli* diarrhoea in the pig: examination of the genetic system. *Theoretical and Applied Genetics* **51**, 65-70.

Gibson, J.P. (1994) Short-term gain at the expense of long-term response with selection of identified loci. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* **21**, 201-204.

Gimelfarb, A.; Lande, R. (1994) Simulation of marker assisted selection in hybrid populations. *Genetical Research* **63**, 39-47.

Gomez-Raya, L.; Gibson, J.P. (1993) Within-family selection at an otherwise unselected locus in dairy cattle. *Genome* **36**, 433-439.

Groen, A.F.; Smith, C. (1995) A stochastic simulation study on the efficiency of marker-assisted introgression in livestock. In preparation.

Groen, A.F.; Timmermans, M.M.J. (1992) The use of genetic markers to increase the efficiency of introgression—a simulation study. *Proceedings of the XIX World's Poultry Congress* **1**, 523-527.

Haldane, J.B.S. (1919) The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics* **8**, 299-309.

Haley, C.S. (1991) Considerations in the development of future pig breeding programs. *Australasian Journal of Animal Science* **4**, 305-328.

Haley, C.S.; Knott, S.A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**,315-324.

Haley, C.S.; Lee, G.J. (1993) Genetic basis of prolificacy in Meishan pigs. *Journal of Reproduction and Fertility*, **48**, Supplement, 247-259.

Hill, W.G. (1993) Variation in genetic composition in backcrossing programs. *Journal of Heredity* **84 (3)**, 212-213.

Hillel, J.; Schaap, T.; Haberfield, A.; Jeffreys, A.J.; Plotzky, Y.; Cahaner, A.; Lavi, U. (1990) DNA fingerprints applied to gene introgression in breeding programs. *Genetics* **124**, 783-789.

Hillel, J.; Verrinder Gibbins, M.; Etches, R.J.; Shaver, D. McQ. (1993) Strategies for the rapid introgression of a specific gene modification into a commercial poultry flock from a single carrier. *Poultry Science* **72**, 1197-1211.

Hoeschele, I.; Romano, E.O. (1993) On the use of marker information from granddaughter designs. *Journal of Animal Breeding and Genetics* **110**, 429-449.

Hospital, F.; Chevalet, C.; Mulsant, P. (1992) Using markers in gene introgression breeding programs. *Genetics* **132**, 1199-1210.

Janss, L.L.G.; Van Arendonk, J.A.M.; Brascamp, E.W. (1994) Identification of a single gene affecting intramuscular fat in Meishan crossbreds using Gibbs sampling. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* **18**, 361-364.

Johansson, M.; Ellegren, H.; Marklund, L.; Gustavsson, U.; Ringmar-Cederberg, E.; Andersson, K.; Edfors-Lilja, I.; Andersson, L. (1992) The gene for dominant white color in the pig is closely linked to *ALB* and *PDGFRA* on chromosome 8. *Genomics* **14**, 965-969.

Lande, R.; Thompson, R. (1990) Efficiency of marker assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743-756.

Le Roy, P.; Naveau, J.; Elsen, J.M.; Sellier, P. (1990) Evidence for a new major gene affecting meat quality in pigs. *Genetical Research* **55**, 33-40.

MacLennan, D.H.; Duff, C.; Zorzato, J.; Fujii, J.; Phillips, M.; Korneluk, R.G.; Frodis, W.; Britt, B.A.; Worton, R.G. (1990) Ryanodine receptor gene is a candidate for predisposition to malignant hyperthermia. *Nature* **343**, 559.

Malo, D.; Skamene, E. (1994) Genetic control of host resistance to infection. *Trends in Genetics* **10**, 365-371.

Mercer, J.T.; Hoste, S. (1994) Prospects for the commercial use of Chinese pigs. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* **17**, 327-334.

Meuwissen, T.H.E.; Van Arendonk, J.A.M. (1992) Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. *Journal of Dairy Science* **75**, 1651-1659.

Michelmore, R.W.; Paran, I.; Kesseli, R.V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 9828-9832.

Rohrer, G.A.; Alexander, L.J.; Keele, J.W.; Smith, T.P.; Beattie, C.W. (1994) A microsatellite linkage map of the porcine genome. *Genetics* **136**, 231-245.

Rothschild, M.F.; Jacobson, C.; Vaske, D.A.; Tuggle, C.K.; Short, T.H.; Sasaki, S.; Eckardt, G.R.; McLaren, D.G. (1994) A major gene for litter size in pigs. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* **21**, 225-228.

Ruane, J.; Colleau, J.J. (1994) The value of marker assisted selection when one QTL is marked. *Proceedings of the 45th Annual Meeting of the European Association of Animal Production, Edinburgh*, p. 80 (abstract).

Sales, J.; Hill, W.G. (1976) Effect of sampling errors on efficiency of selection indices. 2. Use of information on associated traits for improvement of a single important trait. *Animal Production* **23**, 1-14.

Smith, C.; Meuwissen, T.H.E.; Gibson, J.P. (1987) On the use of transgenes in livestock improvement. *Animal Breeding Abstracts* **55**, 1-10.

Smith, C.; Simpson, S.P. (1986) The use of polymorphisms in livestock improvement. *Journal of Animal Breeding and Genetics* **103**, 205-217.

Smith, C.; Smith, D.B. (1993) The need for close linkages in marker-assisted selection for economic merit in livestock. *Animal Breeding Abstracts* **61**, 197-204.

Soller, M. (1990) Genetic mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. *Journal of Dairy Science* **73**, 2628-2646.

Stuber, C.W.; Lincoln, S.E.; Wolff, D.W.; Helentjaris, T.; Lander, E.S. (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* **132**, 823-839.

Van Arendonk, J.A.M.; Janss, L.L.G.; Brascamp, E.W. (1994) Advances in methods for detection and exploitation of quantitative trait loci in pigs. *Proceedings of the 45th Annual Meeting of the European Association of Animal Production, Edinburgh*, p. 77 (abstract).

Visscher, P.M. (1994) Bias in genetic R^2 from halfsib designs. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* **18**, 394-397.

Visscher, P.M.; Haley, C.S.; Thompson, R. (1995) Introgression of genes in backcross programmes using genetic markers and phenotypes. In preparation.

Weller, J.I.; Kashi, Y.; Soller, M. (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science* **73**, 2525-2537.

Williams, J.G.K.; Kubelik, A.R.; Livak, J.; Rafalski, J.A.; Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531-6535.

Wishart, J. (1931) The mean and second moment coefficient of the multiple correlation coefficient in samples from a normal population. *Biometrika* **22**, 353-361.

Wray, N.R.; Goddard, M.E. (1994) Increasing long-term response to selection. *Genetics Selection Evolution* **26**, 431-451.

Zhang, W.; Smith, C. (1992) The use of marker assisted selection with linkage disequilibrium. *Theoretical and Applied Genetics* **83**, 813-820.

Zhang, W.; Smith, C. (1993) The use of marker assisted selection with linkage disequilibrium: the effects of several additional factors. *Theoretical and Applied Genetics* **86**, 492-496.