QTL discovery and applying results to industry

Richard Kerr

Animal Genetics and Breeding Unit

Introduction

Pig breeders deal with traits such as growth, meat quality, and reproduction and disease resistance. The analytical model used in PIGBLUP assumes many genes of small effect influence these traits. This model is referred to as the infinitesimal model. Though these traits may be easy to record, we say they have complex inheritance. In contrast, a single gene for example, determines blood type, and we say it is a trait with simple inheritance. Human geneticists will often abbreviate traits with complex inheritance to just "complex traits". Animal breeders will more often refer to them as performance or quantitative traits. Molecular biological techniques are now providing new understandings of the number and effect of genes influencing complex or performance traits.

Foremost of these techniques is the rise of the genetic linkage map. The genetic map is similar to a road map in that it provides the structure for the eventual location of genes. The location of the gene on the map is called the locus. Any locus, which influences a performance trait, is called a quantitative trait locus (QTL). The reference points on a genetic map are called genetic markers. The goal of this presentation is to discuss known associations between genetic markers on the porcine map and performance traits, and furthermore, discuss how these known associations can be used in practical breeding programs.

Genetic markers

Before discussing genetic markers it maybe helpful to recall some essential facts on genes. The genome is organized into structures called chromosomes. A chromosome is made up of a large molecule of deoxyribonucleic acid (DNA). DNA has two strands that wrap around each other to resemble a twisted "ladder" whose sides are connected by "rungs" of chemicals called bases (see Figure 1. below). Four different bases are present in DNA - adenine (A), thymine (T), cytosine (C) and guanine (G). The particular order of the bases arranged along the "ladder" is called the DNA sequence. Because A binding with T and C binding with G forms the "rungs", base pairs (*bp*) is the unit of measurement for a DNA sequence. The pig genome has been estimated as ~ 2.7 χ 109 *bp*. That is 27,000,000,000 *bp*.

One of the greatest anomalies in genetics is that the majority of the DNA sequence has no apparent function. Such sequence is called intron sequence. A gene is a region of the chromosome where the sequence actually specifies information necessary to build a particular protein, which in turn could influence a performance trait. Genes have start and stop signals, between which there are introns and protein coding sequences (exons). While some of the intron sequence within a gene has a regulatory function, there is no apparent reason why genes are interspersed with so much non-functional

DNA sequence. Consider the skeletal muscle ryanodine receptor (ryr1) gene for example. Recent research (Wen et al. 1996) has shown this gene is at least 120,000 *bp* and contains as many as 110 exons. It is estimated that only 12% of the gene sequence is coding protein. Some of the introns are as large as 5000 *bp*. (Fujii et al., 1991 have shown the association of the substitution of T for C at base 1843 of this gene with malignant hyperthermia susceptibility.)

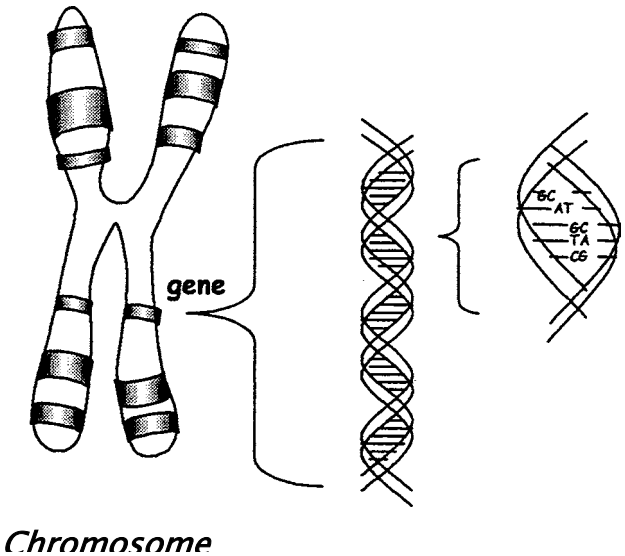


Figure 1. Chromosomes are made up of DNA. DNA resembles a "twisted ladder". A gene is a sequence of *bp* that has the information necessary to build a protein.

A genetic marker is a physical location on a chromosome and has two main features: A. it can easily be identified; and B. its inheritance can be monitored. A marker should also have variants. Most markers used in genetic linkage maps of livestock are intron sequences outside a gene. A fewer number of markers are intron sequences within a gene. Geneticists call markers within genes type I or direct markers and markers that are located outside of genes type II or indirect markers (see Figure 2.). Type I markers can also include sites within exons. Marker assisted selection (MAS) simply refers to the use of markers to enable more accurate selection, especially for traits that had previously been difficult to breed for.

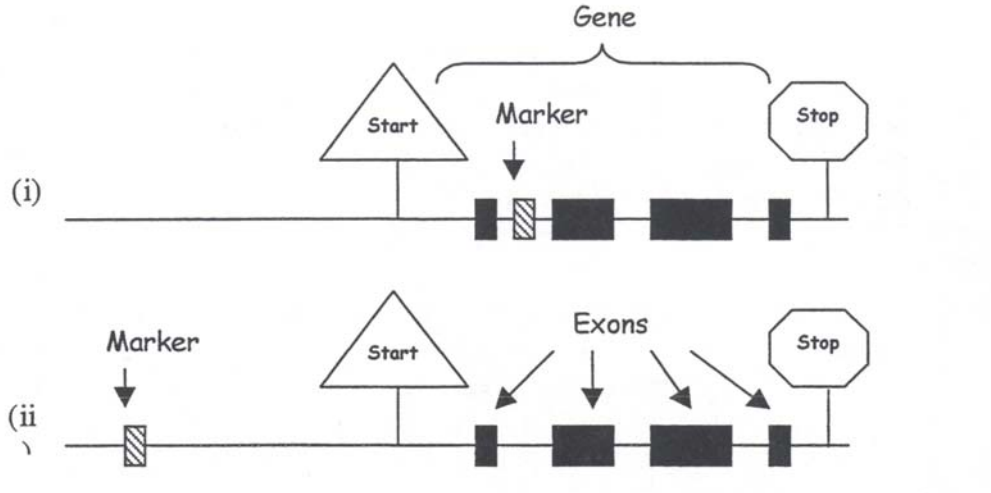


Figure 2: Types of gene markers: (i) type I or direct marker within a gene; (ii) type II or indirect marker, close to a gene

Dr Moran elsewhere in these proceedings has covered how meiotic recombination can result in the separation of two gene markers originally on the same chromosome. The closer the markers are to each other - the more "tightly linked" - the less likely a recombination event will fall between them and separate them. Recombination frequency thus provides an estimate of the distance between two markers on a linkage map.

QTL mapping

QTL mapping is one of the more "in vogue" statistical problems in the world today. The problem is catching the attention of many statisticians, eager to contribute better algorithms and methods. These proceedings are certainly not the forum to explain any statistical theory. Instead a crude analogy may suffice. Suppose you were asked to uncover an unspecified "treasure" in a piece of countryside. The only aid you were given is a topographical land map. Some clues are given such as the treasure is located on a specific type of hill with certain types of boulders. The method in finding the treasure amounts to seeing how well the data at hand fits with all the references supplied in the map. Obviously the better the quality of the map the easier it is to find. There is also the potential for there not to be any treasure at all!

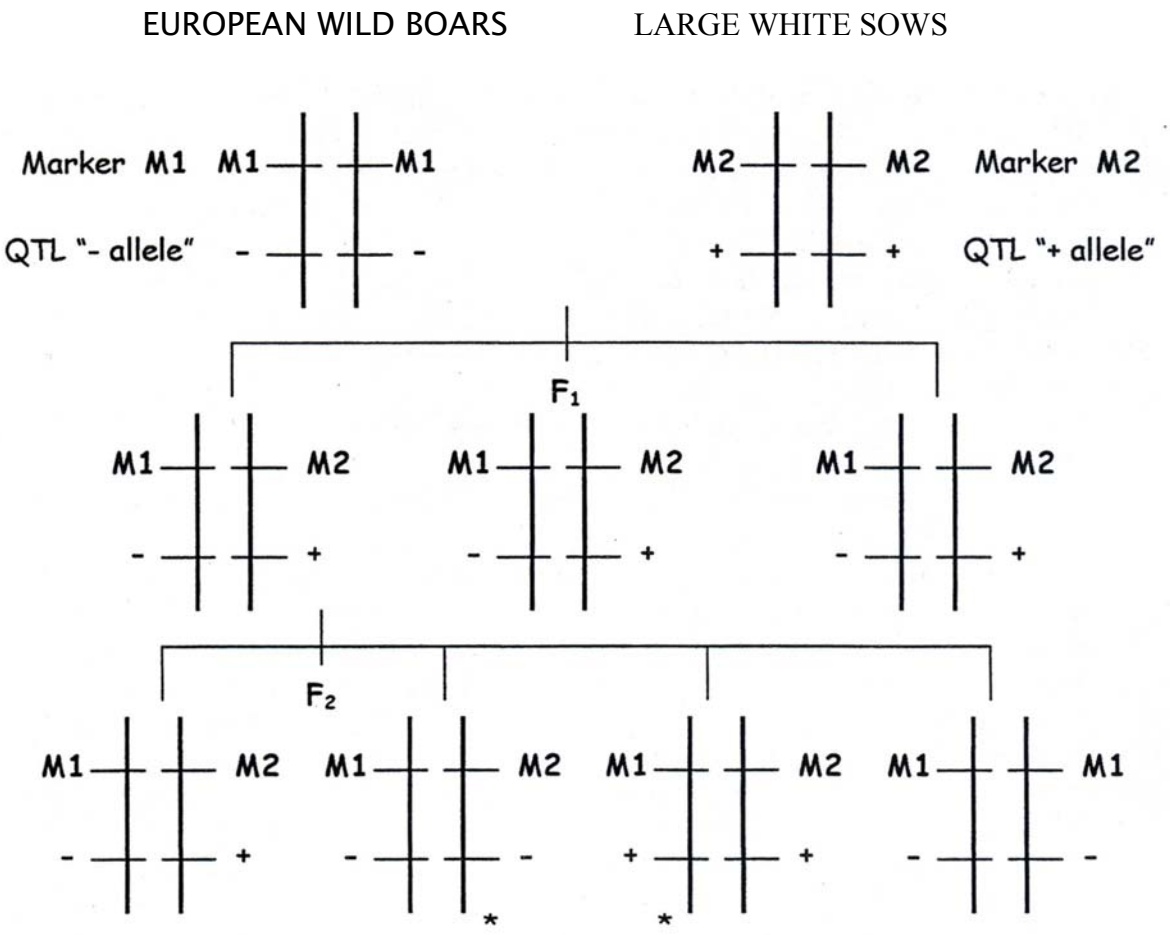
Similar principles apply in QTL mapping. The researcher has to see how well the data at hand, being marker genotypes and performance data, fits with each of the reference points on a linkage map, in terms of uncovering a QTL. The greatest problem in QTL mapping is that the genotype status of the QTL is unknown. The clue to uncovering this missing data comes from the simple fact that half the progeny of a parent will on average receive one of the QTL alleles and the other half will on average receive the other allele. Given that the QTL influences a trait, a contrast in phenotypic performance in the progeny due to the different alleles can be expected. This contrast is very important because without it the statistical algorithms used to uncover the missing information on the QTL would not work. Specific breeding designs will help maximize the chances that the parent is heterozygous for the QTL and that the two QTL alleles have opposing effects. An intercross between two divergent breeds is one such design. To illustrate, consider one of the first livestock QTL mapping experiments to be published (Andersson et al. 1994). Researchers at the Swedish Agricultural University used an intercross between European wild boars and Large White sows. A schematic explaining the design is shown in Figure 3. The case of a single marker linked to a hypothetical QTL which influences fat deposition will be used.

Both alleles at the QTL in the wild population are " - alleles", in that they cause greater fat deposition, relative to the Large White population. Evolutionary history has meant that one particular variant of the marker (Ml) is consistently linked to the QTL in the wild population. Another variant of the marker (M2) is consistently linked to the "+ alleles" in the Large White population. When crossed they create F_1 offspring, which become the ideal parents for a QTL mapping study - they are heterozygous at both the marker locus and the QTL and the QTL alleles have opposing effects.

On inspecting the F_2 progeny we see that meiotic recombination has created new marker - QTL associations. For example, the M2 marker allele is now associated with the "- allele" in some of the progeny. The frequency of these events tells us about the distance between the marker and the QTL. All F_2 progeny are performance tested for

the trait. We calculate trait means for progeny that have been inferred as homozygous for the "- allele", and for those that have been inferred as homozygous for the "+ allele". The difference in mean values for these two groups indicates the size of the effect of the QTL.

Figure 3. Design used by Andersson et al. (1994) to map QTL for growth and fatness.



*Recombinant: frequency of this event reflects the distance between genes for the marker and QTL

The main result of the Swedish study was the identification of a QTL on chromosome 4 with large effects on growth, length of small intestine and fat deposition. F_2 animals homozygous for the wild boar "allele" had 10 kg less in weight at 6 months of age than those homozygous for the Large White "allele".

While using an intercross between divergent breeds has its advantages, there are also some disadvantages. It is a fact that most populations of Large White will most likely be fixed for the favorable " + allele". Thus it is difficult to see the results having any immediate implications for breeding in commercial pig populations. The Swedish

group has more recently published findings on QTL for carcass composition and meat quality, using the same population (Andersson-Eklund, 1997). Generally the wild boar

alleles at these QTL give a shorter and less meaty carcass at equal carcass weight. However, the wild boar allele of one of the QTL on chromosome 3 increased the longissimus muscle area by 1.5cm^2. Thus there are instances where QTL, identified using divergent populations, could have commercial relevance.

A "genome scan" refers to when all markers on a linkage map are tested for their proximity to putative QTL. Typically between 100 and 300 markers are needed to cover the whole genome. Geneticists often talk about dense and sparse coverage. A dense coverage would usually mean markers are spaced approximately 5cM (centiMorgans) apart. This means that any two consecutive markers are separated by recombination 5% of the time. Sparse coverage would usually mean markers are spaced approximately 20-30 cM apart. (A genetic distance of 1 cM is roughly equal to a physical distance of 1 million bp.) Research groups that complete genome scans will often present their results in tables showing the position, effects and level of confidence of identified QTL. Sometimes they will display lod score or F-ratio "profiles". An F-ratio profile for a hypothetical quantitative trait is shown in Figure 4. It is not relevant here what a profile means in statistical terms but peaks in the profile indicate a possible QTL.

Consider the results of the Swedish group's most recent publication (Edfors-Lilja et al., 1998) on mapping QTL for immune capacity in the pig. A F_2 design was again used, developed by crossing European wild boars with Swedish Yorkshire sows. Two hundred F_2 progeny had been typed for 236 markers, using a mix of type I and type II markers (the reason for using both types will be explained below). A genome scan revealed possible QTL for the following traits: the total number of white blood cells (WBC), mitogen induced lymphocyte proliferation, interleukin-2 production (IL-2) and antibody (namely IgG) response to the *Escherichia coli* antigens K88 and O149. These results are very encouraging and the research community is a step closer to identifying the actual genes involved.

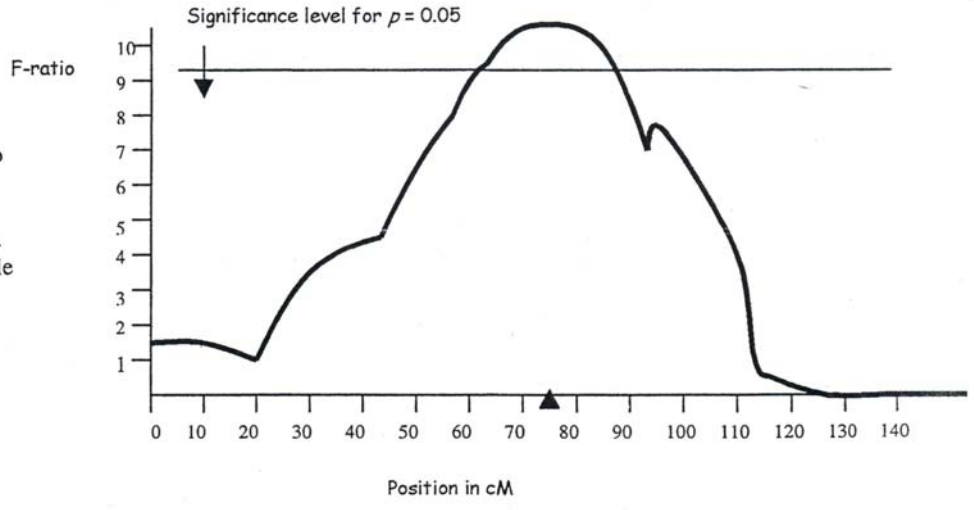


Figure 4. F-ratio profile for a hypothetical quantitative trait. The black triangle indicates the genetic position

Gene identification

1. (Comparative) Positional candidate gene identification

Gene identification in humans and mice is a lot more advanced than in livestock species. Many thousands of genes have been documented in publicly accessible libraries. Pig geneticists use information on their function and the biological systems they affect in trying to identify the genes responsible for QTL mapped in the pig. Any genes they consider applicable for testing in pigs are candidate genes.

The positional candidate gene approach assumes that the swine linkage map can be aligned with the human and mouse maps. Having both maps well populated with type I markers (markers within known genes) aids their alignment, assuming that the known genes are common to both species. The strategy is to position a type I marker that is common to the mouse or human, close (within 5 cM) to the QTL mapped in the pig. It is hoped that amongst the expected plethora of genes surrounding the marker in the human or mouse, there is one that has a function, which is applicable to the trait under study in the pig.

As far as performance traits in the pig are concerned, there have been no cases yet reported of genes being discovered using this approach. However the gene for the dominant white coat colour in the pig was discovered using positional candidate gene information. The autosomal dominant white gene designated I for inhibition of colour, had been closely linked to the platelet-derived growth factor receptor gene (PDGFRA) on pig chromosome 8 (Johansson et al. 1992). This gene and the region surrounding it share homo logy with regions on mouse chromosome 5. The dominant white spotting locus in the mouse, designated W, is positioned within this region and has been discovered to encode the c-KIT gene. This gene is required for the normal development of melanocytes. A mutation in KIT causes the absence of any melanocytes in the skin. Johansson-Moller et al. (1997) thus tested KIT as a candidate gene for the dominant white mutation in pigs. Their results showed unequivocally that variants of the KIT gene were associated with the white coat colour in pigs.

2. Association studies

Often candidate genes are tested for association with performance traits without any prior QTL mapping of the trait. That is, no positional information exists. The term "association analysis" is sometimes used to distinguish it from the positional candidate gene analysis. Two prominent human geneticists have suggested that the future of the genetics of complex diseases in humans is likely to require large-scale testing by association analysis (Risch and Merikangas, 1996). They argue that even if you have to test every gene in the genome, association studies have far greater power than ascertaining linkage between various loci and complex diseases. In recent years there has been a proliferation of reports of linkage between markers and disease loci. However, Risch and Merikangas state -perhaps an undesirable fact - that very few of these findings have been replicated.

Consider the following association study in swine. Livestock meat production capacity is related to myofibril numbers in the muscle. Dutch researchers (Te Pas et al. 1998) have targeted the M*yo*D gene family as likely candidate genes for meat production

capacity. *Myogenin* induces differentiation of myoblasts, M*yoD1* and *myf-5* regulate myoblast and satellite cell proliferation and *myf-6* is involved in maintenance of mature muscle fibers. Because these genes have been characterized in the human genome, small sections of the human version of the gene can be synthesized. These "probes", as they are termed, will "pull out" the equivalent gene in the pig genome. The pig version of the gene is then sequenced and polymorphic sites within the gene (type I markers) are detected. The polymorphism is generally a single base substitution and gives rise to 2 variants. It is common to denote a polymorphism by a letter, with the variants distinguishable by use of lower and upper case. If the two variants at a polymorphic site are *A* and *a*, then the 3 possible genotypes are denoted *AA, Aa* and *aa*. In the case of *myogenin*, 2 polymorphic sites were found yielding 9 possible genotype combinations. Trait values were regressed onto animals' genotypes using mixed model analysis software. A mixed model is required to fit environmental fixed effects and a random polygenic effect, in addition to the genotype effect. Using this method the Dutch were able to find significant associations between *myogenin* and *myf-5* genotypes with meat production related traits. Tests for the *myogenin* and *myf-5* genotypes are currently under patent.

Table 1. Contrast between *H-FABP* genotypes for IMF(%), backfat thickness and growth (presented by van Erp at the 1997 Pig Breeders Round Table.)

Trait	Genotype	Diff.	s.e.	P-value
IMF				
	HH-hh	.40	19	.04
	Hh-hh	.05	.15	.75
IMF*				
	HH-hh	.32	.18	.08
	Hh-hh	.01	.14	.97
STD-BFT (mm)				
	HH-hh	.38	.21	.07
	Hh-hh	.09	.18	.61
STD-BW(kg)				
	HH-Hh	.40	1.30	.76
	Hh-hh	-.62	1.12	.58

IMF* model including backfat thickness as a covariable

Another Dutch group (van Erp, unpublished) has also found significant associations between variants of the heart fatty acid binding protein (H-FABP) gene and variation in intramuscular fat (IMF) deposition in pigs. Table 1 shows the results of mixed model analyses involving various fatness measurements. Of three polymorphic sites found within the pig H-FABP gene only one was associated with changes in fatness. Table 1 shows the contrasts between the genotypes. Pigs homozygous for the H variant had an increase of .32 percentage units in IMF at equal backfat thickness, relative to animals homozygous for the h variant. It has been reported that the frequency of the H variant in Duroc and Dutch Landrace populations is 70%. This result gives further proof that IMF reduction is not completely correlated with backfat reduction. Tests for the

variants of the *H-FABP* gene, and the use of the tests in breeding programs have also been patented.

Results of QTL mapping and gene identification: applying to industry

Identifying the variant at a polymorphic site within a gene, as described above for the *H-FABP* and *myogenin* genes, is an example of a **direct test**. An **indirect test** is where only linkages between marker loci and QTL have been discovered and the breeder has to test for the status of a QTL on the basis of marker information. The same considerations apply to both types of tests when deciding to use them in a commercial breeding program.

One of the first considerations is the cost. It is probable that direct tests will have been patented. In the case of an indirect test, the location of the QTL and the identity of the linked markers will in many instances be concealed by the research group. There are perhaps two avenues to access this intellectual property (IP). One is to pay directly for the information. The other is to purchase DNA samples of the resource population used in the creation of the IP, and the phenotypic measurements, and to repeat the experiment with a self obtained panel of markers. Obviously paying for a direct test would have "better value for money", since a result is almost assured. In regard to an indirect test, there is always the risk that the claims about linkages prove inconclusive or contradictory when a different population is used. Thus an important point when considering an indirect test is whether the findings have been successfully replicated.

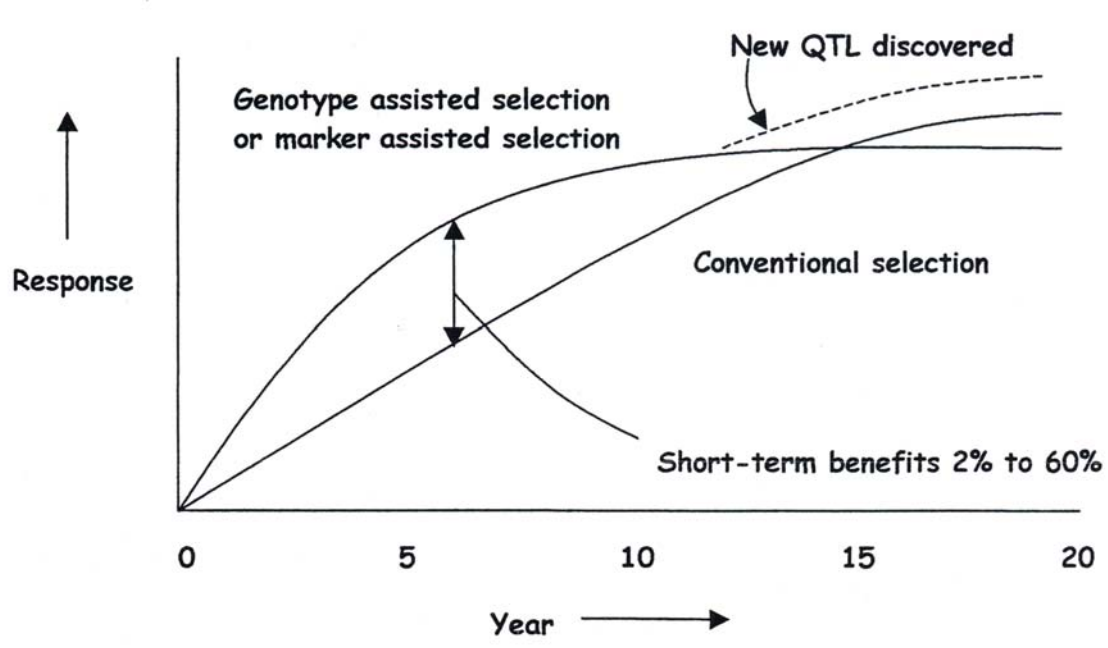


Figure 5. Long term response compared for MAS and conventional selection.

The decision to implement a DNA test on a large scale is also dependent on the current status of the gene or QTL in the population. A preliminary survey of animals will reveal whether the "good" variant of a gene, or the "good" allele of a QTL is low or high in frequency. It is possible that a future, enhanced PIGBLUP will assist in determining what gene variants or QTL alleles are likely to be present in your herd.

This information would come about through links between your herd and other genotyped herds. The breeder must also consider that conventional selection (using polygenic EBVs) could increase the frequency of good alleles with only slightly less efficiency than DNA tests. The graph in Figure 5 is a typical result of many simulation studies to compare MAS and conventional selection over many generations.

The magnitude of the short-term benefit is primarily a function of whether direct or indirect tests are used and whether the trait(s) in question can be easily measured. A direct marker for a hard or costly to measure trait can extend the short-term benefit to as much as 60%. However, regardless of the short-term benefits conventional selection in the long term can "catch up" and even exceed the long-term response achieved through genotype or marker assisted selection. The phenomenon of depletion of polygenic variance is often seen when there is too much selection emphasis on major genes. However, this graph is misleading in many ways. It doesn't emphasize the cumulative return from breeding genotypically superior breeding stock over many generations prior to the "cross over" point. Response from conventional selection would have to exceed the response achieved through MAS for a considerable period beyond this point before cumulative return is greater. The financial benefit from incorporating DNA tests into a breeding program has also to consider the net present value of MAS and conventional breeding. One final aspect of this graph is that discovery of new QTL will help prevent the "cross over" point from ever being realized.

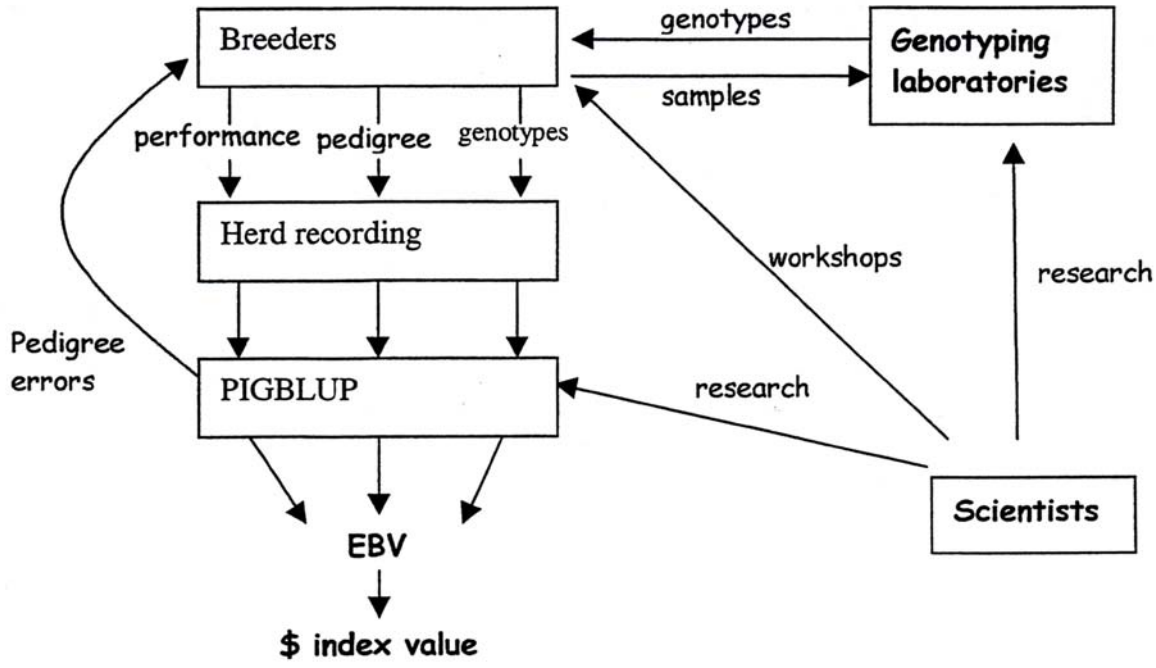


Figure 6. Likely pathways required for an Australian pig breeder to implement MAS

Once the decision to implement DNA tests has been made, there are many logistical issues to consider. It is important to remember the breeding objective usually entails improving more than one trait at a time. Thus the results of direct and indirect DNA tests have to be linked into the selection index just as any other criteria such as PIGBLUP EBVs would be. Successful use of DNA tests will depend on careful and meticulous performance and pedigree recording. Figure 6 is a flow diagram showing the likely pathways required for a breeder to implement MAS.

On farm performance recording and breeding value prediction via PIGBLUP already exist. To implement MAS the breeder is required to send a blood or tissue sample to a genotyping laboratory, who in turn send back genotype information. Herd recording systems will need to be enhanced to store such information. Future releases of PIGBLUP will be able to combine genotype information with performance data, within a multi-trait analysis, to predict breeding values for both the QTL and polygenic components of each trait. Previously, breeders had only to consider one EBV for each trait, for each animal. This EBV was the sum merit of all the many thousands of genes, each of small effect, which contribute to the trait's variation. The EBV would have contained the hidden effects of any QTL. Now, markers enable the partitioning of the QTL and polygenic effects and breeders now have to consider two EBVs for each trait. It is useful to apply the EBV concept to a QTL because it implies variation within the population. This variation stems from the fact that marker-QTL associations are not consistent. There will be some degree of co variance between members of a family, but generally associations differ between families. In pig breeding the dominant structure is the paternal half-sib family. Thus the focus of acquiring EBVs for QTL centers on the boar. It is likely that due to cost only the elite family lines will have EBVs for QTL. PIGBLUP will then provide the tools for the breeder to combine all available information.

In pig breeding, the links to and from genotyping laboratories are especially important. The period between DNA sample collection and selection of parents is considerably less than in other species. In this time the genotypes have to be assayed, and if using indirect tests, marker genotypes have to be combined with phenotypic information to determine EBVs for the QTL. This information then has to be combined with other trait information to yield an EBV for total genetic merit. Figure 7 shows a hypothetical case scenario for acquiring a QTL EBV.

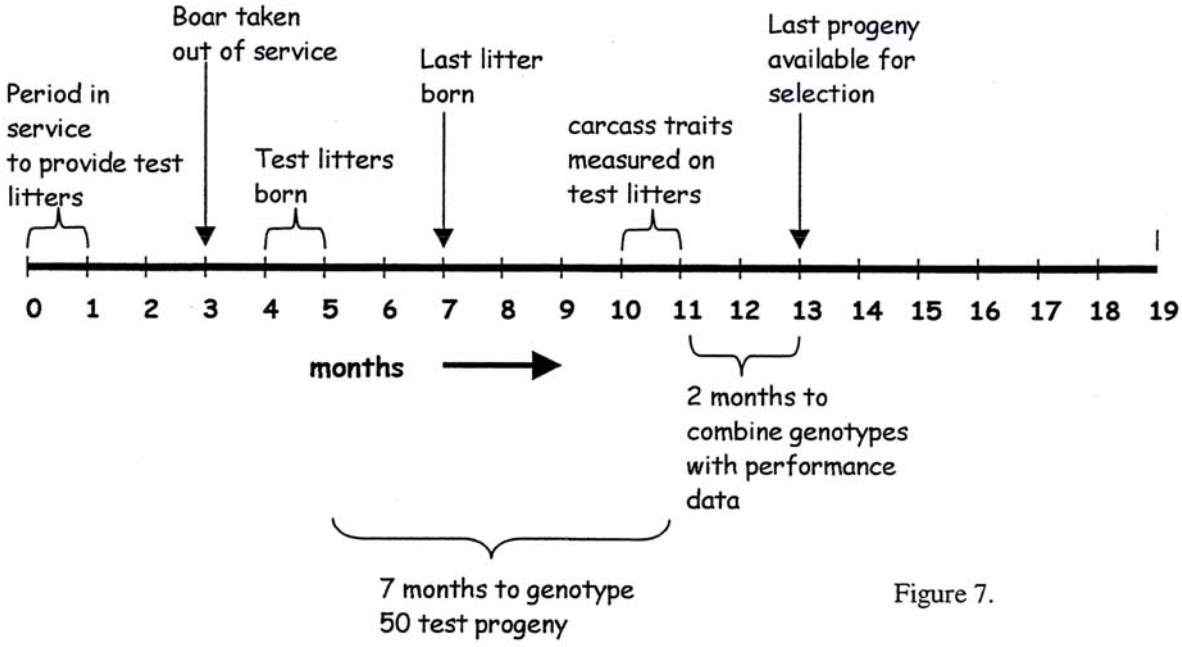


Figure 7.

The case scenario can be explained as follows. It has been decided that QTL information is needed for a boar used in natural service and his descendents. The breeder has decided that it is only worthwhile to DNA test for difficult to measure traits, such as a meat quality trait. There are no direct markers available, thus an indirect test is needed. To obtain a reliable EBV for the QTL approximately 50 progeny with performance data are needed. This translates to 6 litters assuming 8 piglets per litter. Assuming 1.5 litters born per week, the boar's first month of service is needed to produce the test progeny. The boar stays in service for another 2 months. Performance data for the meat quality trait is measured when the test progeny are slaughtered at approximately 6 months of age. At most, 7 months is all the time available to genotype the 50 progeny for the 2 to 3 markers per QTL. And perhaps more limiting is the 2-month time span between data collection and when the boar's last progeny are available for selection. Data analysis must be completed within this time, otherwise all DNA testing has been futile. It is likely that a boar targeted for DNA tests should be left in service longer, enabling greater exploitation of the results. A key point is that genotyping does not end with the test progeny. For MAS to work, all progeny that are candidates for selection must also be genotyped. This information will determine the value of the progeny EBV for the QTL, much the same way polygenic EBVs can be calculated without performance data, but from the pedigree alone.

Learning from experience: happenings here and overseas

Though simulation studies are useful tools for predicting outcomes, they are heavily dependent on the assumptions we use to model the data. Only real experience can tell us whether response curves behave as depicted in Figure 5. There are most likely many logistical issues that we haven't yet considered.

The Animal Science department at Sydney University, in collaboration with Bunge Meat Industries and AGBU has nearly completed a project involving mapping QTL for meat quality, growth and carcass traits in commercial lines. The results should tell us for the first time whether we can identify QTL in mostly purebred populations. Nearly every other QTL project to date has involved crossing wide divergent breeds. A follow up study, to be initiated shortly, will attempt to replicate the findings in other Bunge lines.

The most significant development in the US is the large collaborative project funded by the USDA Cooperative State Research Scheme. NC-210 "Mapping the Pig Genome" involves 10 research stations. In a recent review of Swine Genetics in the U.S.[1] Larry Young describes the contribution of each station. Together the resource populations at each station represent a wide variety of domestic and imported breeds and generations of controlled selection. Linkage relationships between marker loci and QTL alleles, and the relative effects of QTL alleles, are bound to differ by population and even by family. One of the aims of the project is to test marker-QTL linkages in as many resource families as possible. Experiments will be designed and initiated at Iowa, Nebraska, Ohio, Oklahoma and Indiana to measure response from MAS. As Larry Young describes it "regional collaborations will provide the QTL information

[1] http://mark.asci.ncsu.edu/nsijy95proc/review.html

necessary for, and facilitate the start of, the first comprehensive experimental test of marker-assisted selection."

A similar scheme has been proposed in Europe. A cooperative project titled "PigQTech" has been initiated to transfer QTL technology to industry. There are seven participating organisations: Swedish University of Agricultural Sciences; Scan Avel HB; Roslin Institute; PIC; IRTA; COPAGA; and Universitat Autonoma de Barcelona. The primary goal of the project is to examine whether experimental results apply to commercial populations. In other words to answer the question how universal is a QTL effect? Other questions they are examining include how best to sample outbred populations and how to analyse the data.

Conclusions

Given the developments in North America and Europe it would seem cooperation is the key to successful application of QTL technology. No single breeder can hope to place all pieces of the jigsaw together on their own. AGBU are hoping to play their role in this cooperation. Research programs are underway to develop the analytical software required for MAS to become a common industry practice.

Acknowledgements

This work was funded in part by the Pig Research Development Council Project US36. I am grateful to Dr Susanne Hermesch and Dr Ron Crump for helpful comments.

References

Andersson, L., Haley, C.H., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K., Adders son-Ecklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J. and Lundstrom, K. (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* 263:177 I'm 4.

Andersson-Ecklund, L, Marklund, L., Lundstrom, K., Haley, C.S., Andersson, K., Hansson, I., Moller, M. and Andersson, L. (1997) Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross. *J. Anim. Sci.* 76:694-700.

Edfors-Lilja, I., Wattrang, E., Markland., L., Moller, M., Andersson-Ecklund, L., Andersson, L. and Fossum, C. (1998) Mapping quantitative trait loci for immune capacity in the pig. *J. Immunology* 160:829-835.

Fujii, J,. Otsu, K., Zorzato, P., de Leon, S., Khanna.V.K., Weiler, J.E., O'Brien, P.J., Maclennan, D.H. (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science 253:448-451.

Johansson, M., Ellegren, H., Marklund, L., Gustavsson, U., Ringmar-Cederberg, E., Andersson, K., Edfors-Lilja, I. and 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.

Johansson Moller, M., Chaudhary, R., Hellmen, E., Hoyheim, B., Chowdhary, B. and Andersson, L. (1996) Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. *Mamm. Genome* 7:822-830.

Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Nature Genetics* 13:1516-1517.

Te Pas, M.F.W., Harders, F.H., Soumillion, A., Verburg, F.J. and Meuwissen, T.H.E. The *MyoD* gene family influences birth weight, postnatal growth rate and carcass meat content of pigs. In proceedings *16th Int. Conf. Anim. Genet. ISAG*, Auckland, August 1998.

Wen, G., Leeb, T., Reinhart, B., Schmoelzl, S. and Brenig, B. (1996) The porcine skeletal muscle ryanodine receptor gene structure coding region 1 to 10614 harbouring 71 exons.