Genomic Selection – potential for pig breeding

Brian Kinghorn

The Institute for Genetics and Bioinformatics, University of New England, Armidale, NSW 2351

Introduction

Selection programs exploit the observation that ***like begets like***.

Why does this work? It is basically because better animals tend to have better genes[1], and they pass these on to their offspring. However, we do not have to know about the genes to make selection work - animal breeders were selecting animals well before we knew anything about genes.

Understanding of the principles of heredity has helped us to improve the efficiency of selection programs over the last few decades. The performance of the relatives of an animal gives extra evidence about the quality of the genes carried by that animal – because some genes are shared between relatives. This is a key reason that BLUP gives us extra selection response.

But in fact we do not have to know anything about the actual genes affecting a trait to be able to exploit this property of BLUP – we only have to know the theoretical proportion of genes shared in common between animals with different pedigree relationships.

The holy grail of animal breeding is to be able to get a direct handle on the genes that cause performance for key traits, to understand the effects they have, and to use this knowledge to select the best animals for breeding, according to the genes they carry.

There are two key advantages here:

1. Under "like begets like" we choose animals based on their observed trait values or *phenotypes* (and probably the phenotypes of their relatives). This only gives a clue to the quality of their genes – an indirect handle on their genetic merit and ability to yield high-performance offspring. The temptation is to cut out this simple step, and develop a direct handle on the quality of the genes they carry.

2. If we can judge the genetic merit of an animal from knowledge of its genes, then we can access that information soon after a tissue sample can be taken, and run the breeding program as quickly as reproductive constrains allow. This is especially valuable for traits that are difficult to measure, such as carcass traits, or cannot be measured, such as female reproductive traits in boars, or are measured late in life, such as sow longevity.

[1]For convenience the term gene is used synonymously with quantitative trait locus (QTL, a place in the genome where mutations affect traits). In truth, many such mutations are outside genes.

Silver bullets

We have made slow but definite progress over the last 15 or so years in discovering and exploiting individual genes that seem to affect performance traits in pigs. An early example was the "halothane" gene, so-called because of the effect of this anaesthetic gas to induce malignant hyperthermia in pigs with two doses of the susceptible variant of the gene (one from each parent). With favourable effects on lean meat content and unfavourable effects on meat quality, pig breeders used information about genetic status of pigs to influence selection decisions.

Genetic status was known from halothane test results, and mathematical trickery was used to give probabilities for carrying the halothane gene for animals that had not been tested. However, a DNA test was developed in the early 1990's, giving a genetic marker for the susceptible variant of the gene. This effectively launched Marker Assisted Selection (MAS) in pigs. The "halothane gene" test was a "***silver bullet***" to help us improve our breeding programs. A silver bullet means an individual test for a genetic marker that we are quite confident has an important impact on a trait of interest.

Probably most of the well known silver bullet genetic marker tests are based on initial discovery of inheritance patterns that resembled single gene effects – often from eyeballing animals and their data. This is probably most evident in sheep (see Dodds, 2007), but there are examples in pigs (van der Steen et al., 2005). Other markers have been discovered by working directly on genes that we previously suspected might have some impact, examples given by van der Steen et al. (2005) relate to genes in the leptin pathway or other aspects of fat metabolism, and female reproduction (eg. the estrogen receptor locus).

False positives

With the advent of Single Nucleotide Markers (SNPs) we managed to put thousands of "landmarks" on the pig genome, and use this in large studies to identify markers associated with performance. This moved us from a few silver bullets to hundreds of promising markers. Promising indeed – but not without some problems. We look for the needle in the haystack, but many pieces of hay look like a needle. Only a few markers found in this way could be described as silver bullets.

When we find a hundred promising markers chosen out of thousands, many will look promising by random chance, rather than by true effects that they mark. This problem increases as the pool of markers to search through becomes bigger. There are two types of problem here:

> Pure random chance: If all markers had no true effect, some will still be associated with higher performance by random chance. Consider the Melbourne cup with 10 horses running. Take a thousand coins and toss each coin for each horse in the race. There is a better than even chance you will find a coin that is fully associated with the winner (eg Heads for the winner and Tails for the rest). Does this mean you can use that coin to make a betting decision on the next race?
>
> Chance association with pedigree: If a very high EBV boar with many progeny carries a rare variant of a genetic marker that has no merit in itself, that variant will be associated with high performance, and can be seen to be highly

promising. However, this apparent effect will be absent in other populations or lineages, and will soon diminish as its carriers become less related to the high EBV effect.

These sorts of issues have resulted in many "false positives" in marker discovery work. "Eureka! A silver bullet!" followed by "Er, well …" after independent evidence arrives. This is a lesson that has now been well learned, and we use statistical methods to help control this problem, both for marker discovery, and further downstream during marker assisted selection.

Part of these statistical methods effectively tells us to not believe the evidence of our own eyes, but to shrink back the estimates of marker effects – to presume that their effects are smaller than what is immediately apparent.

This is important in marker discovery, but it can also be important when using many markers (say between 10 and 100 markers) to help make selection decisions.

A new paradigm: Genomic Selection

These lessons to shrink marker effect estimates become critical with the advent of SNP chips, and 50,000 or more candidate markers to associate with performance. Luckily, Meuwissen et al. (2001) laid the groundwork for "Genomic Selection" which carries out such shrinking in a very clever manner that also integrates the information to predict the genetic merit of tested animals.

Genomic selection does not chase silver bullets. It does not try to pick individual markers for us to take a chance on. It essentially jumps that step and aims directly at estimating genetic merit of individuals as a function of all the marker information available.

Figure 1 gives information that shows why this should be a strong step forward.

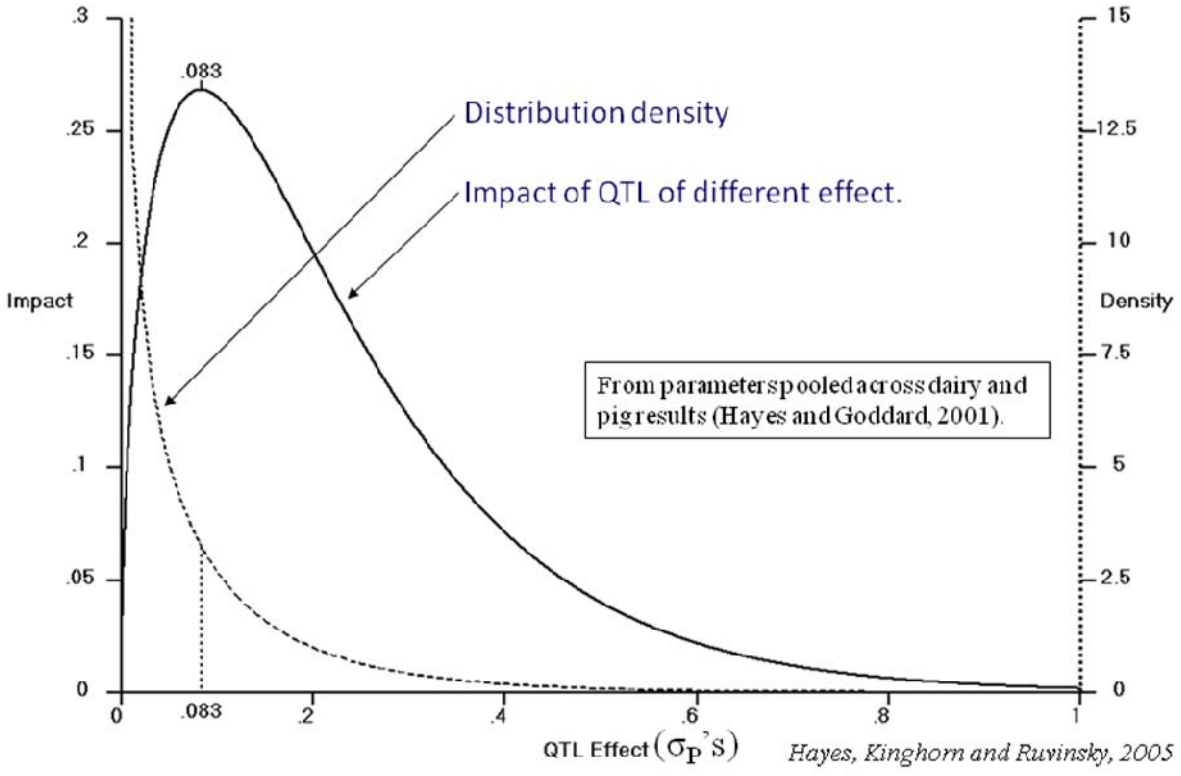

Figure 1. The dotted line shows that there are many genes of small effect and very few genes of large effect - greater than say 0.4 phenotypic standard deviation, or about 5kg in slaughter weight. It is difficult for us to find single markers ("silver bullets") with effects much smaller than this. However, under genomic selection we should be able to exploit these genes reasonably well. This is important, because most impact on genetic variation is given by genes of small effect, as indicated by the solid line, and we need a handle on these if we are going to achieve high selection accuracies. Figure from Hayes et al. (2005)

Part of this process is to effectively shrink the estimated value of most markers back to zero. All the remaining markers are used to evaluate the different segments of DNA that are being passed through the population from parents to offspring (see Figure 2). This DNA makes up the genome, which is organised into chromosomes. Genomic selection evaluates these different segments, whatever individuals are carrying them. It then proceeds to calculate a genomic estimate of breeding value (gEBV) for each animal by adding up the estimated effects of the segments carried by that animal.

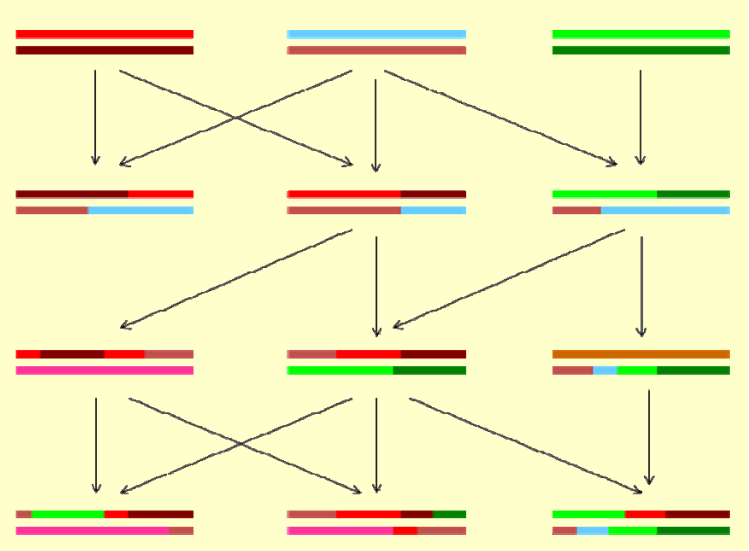

Figure 2. This shows how a chromosome is recombined over generations. Each individual carries two copies of the chromosome – one from each parent. The DNA in the three animals in the last generation can be traced back to the three animals in the first generation (see Figure 3).

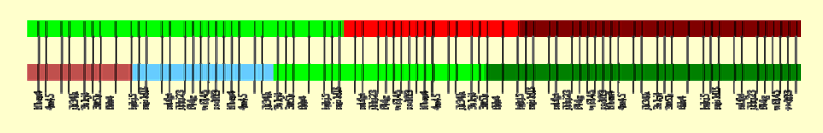

Figure 3. Genetic markers give us "landmarks" on the chromosome. If we have this information on all animals in the pedigree, then tracking where the different chromosome segments have come from is a puzzle we can solve quite well.

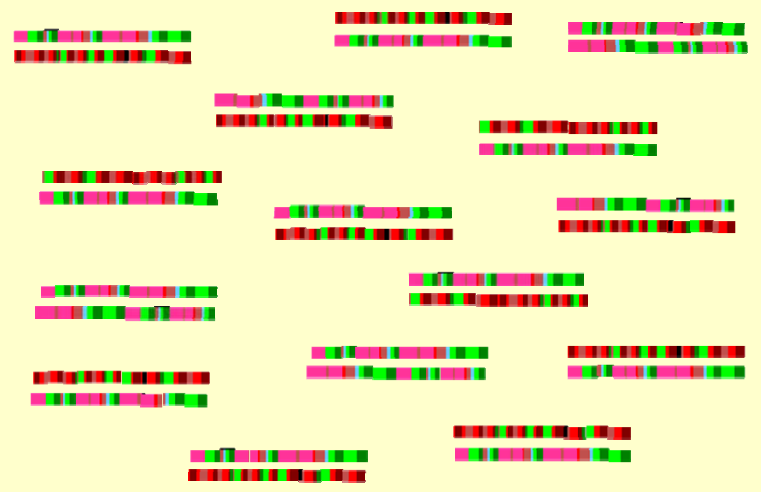

Figure 4. Without pedigree information we cannot track large chunks of DNA, as in Fig. 2, but we can track small chunks. With very many small chunks to track, we need very many genetic markers (say 50,000). This is what we do under genomic selection.

There are two changes of focus when moving to genomic selection:

1. From the viewpoint of exploiting genetic markers: We no longer hunt for specific mutations ("silver bullets" or "needles in haystacks"), a gamble for success, where the risks are largely technical ... we could spend a lot of money and end up with nothing found. Using genomic selection we aim to capture all

that our genomic information can offer, where the risks are largely in costs … we will get a positive result, an accuracy figure, and experience will tell us how good that is likely to be.

2. From the viewpoint of genetic evaluation: The primary focus is to evaluate segments of genome, rather than to evaluate animal EBVs, as in classic BLUP.

Calibration and implementation

These are the two phases of a genomic selection program.

Calibration is used to evaluate the many small segments of genome that are being inherited in the population at large. Genetic inheritance involves mechanisms to cut chromosomes into pieces, assembling them into new combinations for the progeny … variety is the spice of life, and this helps evolution to move along smartly. This means that only small segments remain intact over many generations – and if we look at the DNA of very distantly related population members, only very small segments will be found that are the same by inheritance.

These segments are very small, and so if we want to track them with genetic markers, we need very many genetic markers (say 50,000) to cover the whole genome (all chromosomes).

The calibration exercise uses all that marker information, together with trait records, to evaluate the impact of each segment in comparison to other segments that could be inherited at the same region in the genome. [With very many markers, it is sufficient to treat each marker location as a segment – and evaluate the impact of each marker.]

As indicated in Figure 4, with very many markers we do not need knowledge of pedigree for genomic selection. This means that we can calibrate on commercial animals that have highly relevant trait measurements under commercial conditions. [There may be some technical problems in calibration exercises on crossbred animals, and this has yet to be fully understood.]

Implementation is simple by comparison. As indicated already, for each selection candidate we use marker tests to find which genome segments that animal carries, and add up the estimated values of all these segments to give the gEBV for an animal. The calibration step will highlight which genome segments carry the most useful genes for each trait, so we do not need the full complement of genetic markers to predict the status of selection candidates for just these genome segments. And so we can make a separate DNA test system for candidates, with perhaps only a few thousand markers to be genotyped. Time will tell if this saves enough costs to warrant setting up a separate smaller test (say 3,000 markers rather than say 50,000 markers), bearing in mind that the choice of which marker to include in the reduced test will depend on traits covered, and most probably breed.

A key feature of implementation will be selection among juvenile animals, as discussed below.

Selection accuracies under genomic selection

Early information about the likely value of genomic selection came from computer simulations, and was rather startling. Most work was targeted at dairy breeding structures, and accuracies of more than 80% were predicted – similar to those for a bull with a dozen daughter records. These predictions have moderated with use of real data, and 50% to 80% seems to be a reasonable target for traits of moderate heritability. This might be similar to selection accuracy for growth traits immediately off test in pigs.

50% to 80% accuracy would compare very favourably to accuracies we currently achieve for reproductive traits, except for older widely used boars. Unfortunately we would likely need very large calibration datasets (probably more than 20,000 quality records and genotypings) to get genomic selection accuracies up to these levels for such low heritability traits.

The jury is still out on the likely selection accuracies that we will achieve in pigs.

The need for ongoing calibration

If a calibration exercise were to involve very many markers and a dataset(s) that is widely representative of the target breed/population to be improved, selection accuracies would probably diminish only slightly over 10 or more years. However, these conditions are unlikely to be sufficiently well met for this. Re-calibration every few years, or possibly an ongoing low level of updating, is likely to be required in practice.

Using fewer markers to save costs

One issue with pigs is the potentially high cost of genotyping in relation to the value of each individual animal, especially if genomic selection is to be used to select among young littermates. David Habier (2008) has put forward a solution to this. The young candidates are genotyped for a relatively small number of markers (say 500), and these are "tagged" to the corresponding markers in their parents, which are fully genotyped (say 50,000 markers). The inherited segments of DNA can then be tracked with reasonable confidence, and only a small compromise in selection accuracy is involved.

An alternative is to use only about 3000 markers on each animal, parents and candidates (Kinghorn, 2007). The disadvantage here is that we would need genotyping of all parents, for at least a few generations of pedigree depth, and that the gEBVs would only be available for the breeding line concerned, as reflected in Fig 1. It may turn out that genotyping an animal for 3000 markers is not much cheaper than genotyping it for 50,000 markers, in which case it is probably better to go for the bigger number and have results that can be used across the population/breed.

Patterns of implementation

If genomic selection comes to be well used in pig breeding programs, its biggest impact will probably be made through selection of animals at a young age, with consequent reduction in the generation interval. This is most promising for reproductive traits, health and robustness traits, and carcass traits. With high costs of genotyping, it is

likely that we will attempt to preselect among littermates for genotyping, on whatever information is available at decision time. This may involve delaying of decisions until animals are old enough for any economically relevant recordings to be made.

The state of play for pigs

The International Porcine SNP Chip Consortium should soon have a preliminary 60K SNP panel available (http://www.animalgenome.org/pigs/newsletter/No.92.html). Pricing seems likely to be between US$100 and US$200 per sample for registered users, putting the likely genotyping cost of a calibration exercise at somewhere between US$100,000 and US$1 million.

An alternative for the large breeding companies is to continue to develop their own smaller SNP panel resources as long as that is deemed to be more cost efficient. This will tend to use careful application of MAS to include many markers in otherwise standard EBV calculations (eg. McLaren, 2007), but may move to use a pedigree approach to genomic selection that requires relatively few markers (Kinghorn, 2007). However, the latter approach may be circumvented if the cost of large SNP panels comes down sufficiently in the next year or so.

Revolution or evolution?

Classic BLUP and genomic selection are technically very different. Does that mean that we should simply replace PigBLUP with Genomic Selection? Such a revolution is most unlikely. There may eventually be a replacement, but that is not on the horizon.

It will be somewhat difficult to technically integrate the two approaches within a single statistical analysis. However, there is a relatively straightforward way to make them co-exist in the pipeline to deliver information that breeders can use for selection. This involves treating the BLUP EBVs and the gEBVs as different "traits" to be included in a selection index, so that a single integrated EBV figure results for each animal/trait (Goddard, 1999; Amer 2007).

Conclusion

Genomic selection is highly fashionable right now. There is special interest in the dairy industry, driven largely by the prospect of evaluating bulls quite accurately without the need for a progeny test. Evaluate juveniles rather than grandfathers – it certainly speeds things up.

The impact on pig breeding is unlikely to be so dramatic, and yet if genomic selection proves to be as reliable as we hope it will be, we should be able to select among young animals on traits such as performance under commercial conditions, meat quality and sow longevity. However, the investment required to achieve this will be high, especially for difficult and lowly heritable traits. The need for large quality datasets and strong financial resources may restrict activity to large companies and cooperative efforts.

References

Amer, P.R. (2007). "Wide ranging strategies for incorporating genomic information in selection." Seminar delivered at *Assoc. Advmt. Anim. Breed. Genet.* 17th meeting.

Dodds, K.G. (2007). "Use of gene markers in the New Zealand sheep industry." *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17**: 418-425.

Goddard, M.E. (1999). "New technology to enhance genetic improvement of pigs." *Manipulating pig production*. **7**: 44-52.

Habier, D. (2008). "Genomic Selection using low-density marker panels." Talk at AGBU, 23 September 2008.

Hayes, B.J. and Goddard, M.E. (2001). "The distribution of the effects of genes affecting quantitative traits in livestock." *Genetics Selection Evolution* **33**: 209–229.

Hayes, B.J., Kinghorn, B.P. and A. Ruvinsky (2005). "Genome Scanning for Quantitative Trait Loci." In *Mammalian Genomics*, CAB Press, pp 507-537.

Kinghorn, B.P. (2007). "Where to from here? The technical landscape ahead." *Proc. Assoc. Advmt. Anim. Breed. Genet.* **17**: 509-514.

McLaren, D.G. (2007). "Recent developments in genetic improvement of pigs." *Manotoba Swine Seminar*. 1 Feb 2007. [https://www.gov.mb.ca/agriculture/livestock/pork/pdf/bab21s15.pdf]

Meuwissen, T. H.E., Hayes, B. J. and Goddard, M.E. (2001). "Prediction of total genetic value using genome-wide dense marker maps." *Genetics* **157**: 1819–1829.

van der Steen, H.A.M., Prall, G.F.W. and Plastow, G.S. (2005). "Application of genomics to the pork industry." *J. Anim. Sci.* **83**(E. Suppl.):E1–E8.