NEW TOOLS TO MAKE GENETIC PROGRESS

Jack Dekkers and Max Rothschild
Department of Animal Science, Center of Integrated Animal Genomics
Iowa State University
239 Kildee Hall, Ames, Iowa 50011
E-mail: jdekkers@iastate.edu or mfrothsc@iastate.edu

ABSTRACT

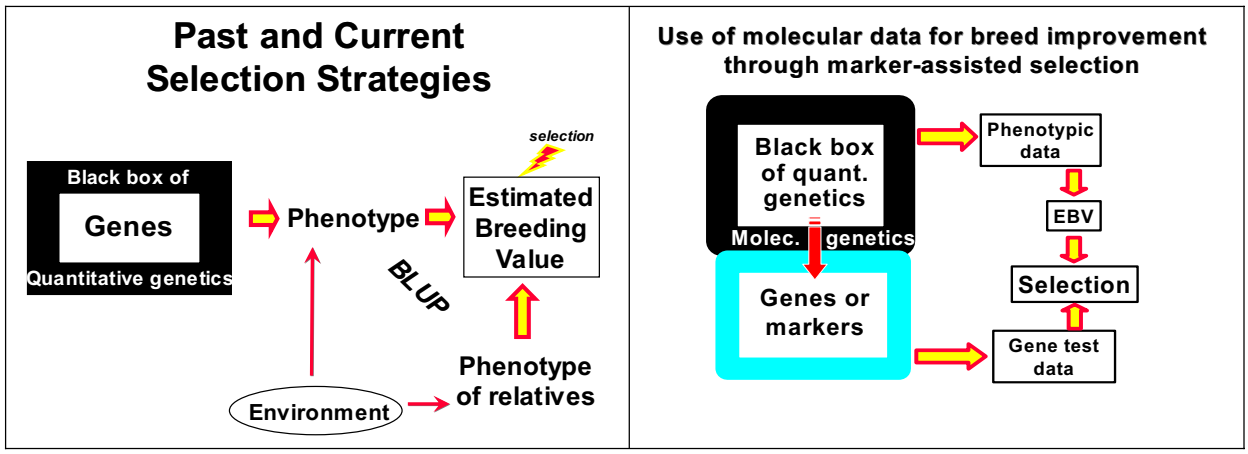
Advances in molecular genetics have opened opportunities to enhance strategies for genetic improvement of pigs by directly selecting on genes or chromosomal regions that harbor genes that affect traits of interest. In this paper, we review molecular technologies that have become available, the current state of the use of gene- or marker tests in pig breeding programs, and future prospects. The main conclusion is that, while current applications of molecular technology in selection are limited, recent developments in molecular genotyping technology will greatly accelerate the rate of implementation of molecular methods for pig breeding in the fore-seeable future. These developments include ongoing efforts to sequence the pig genome, availability of high-density genetic marker maps, and cost-effective high-throughput genotyping of large number of markers across the genome. These opportunities open great opportunities for more effective selection to enhance performance under commercial conditions.

INTRODUCTION

To date, most genetic progress for quantitative traits in pigs has been made by selection on phenotype or on estimates of breeding values (EBV) derived from phenotype, without knowledge of the number of genes that affect the trait or the effects of each gene. In this quantitative genetic approach to genetic improvement, the underlying genetic basis of traits has essentially been treated as a 'black box' (Figure 1a). Despite this, the substantial rates of genetic improvement that have been and continue to be achieved are clear evidence of the power of quantitative genetic approaches to selection. This success does, however, not mean that genetic progress could not be enhanced if we *could* gain insight into the black box of quantitative traits, in particular for traits that are currently difficult to improve. The latter include traits with low heritability (litter size, disease resistance), traits that are difficult to measure (disease resistance), traits that can only be measured on one sex (litter size), traits that are measured late in life (longevity), or traits that require the animal to be slaughtered (meat quality). By being able to study and assess the genetic make-up of individuals at the DNA level through genetic tests, molecular genetics has given us the tools to make those opportunities a reality (Figure 1b). Molecular data is of interest for use in genetic selection because gene tests have heritability equal to 1 (assuming no genotyping errors), can be done on both sexes and on all animals, can be done early in life, and may require the recording of less phenotypic data. The purpose of this paper is to review the current status and future prospects for the use of molecular genetic tools for genetic improvement. Although molecular

genetic data is useful for other purposes, including parentage verification and traceability, the focus of this paper will be on the use of molecular genetics to enhance within-breed improvement.

Figure 1a. Quantitative genetic selection. **Figure 1b. Use of molecular data in selection.**



CURRENT STATUS

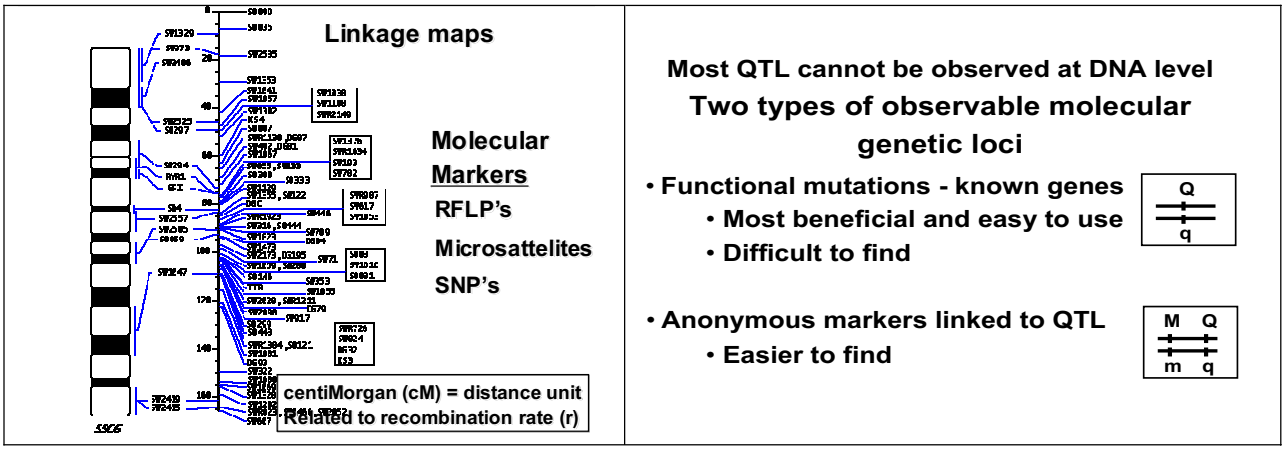
Through the use of molecular genetic technology, a large number of genes have been mapped over the past 10 years in the main livestock species (Figure 2). Although some of these genes have a functional role in the animal's physiology (i.e. they contain the genetic code for a protein), most are non-functional or 'neutral' genes (Figure 3). The latter are referred to as 'genetic markers'. The fact that genetic markers are non-functional does, however, not mean that they are not useful. In particular, genetic markers can be used to identify genes that affect the quantitative traits we are interested in (so-called quantitative trait loci or QTL). The important difference between genetic markers and their linked QTL is that we can determine what genotype an animal has for a genetic marker but not directly for the QTL. However, if the observable genetic marker is linked to the QTL, we can use a genetic marker to indirectly select for the QTL, which is the concept behind marker-assisted selection (MAS).

A marker that is linked to the QTL and, therefore, associated with phenotype, can be detected by comparing the mean phenotypes of individuals that have alternate marker genotypes (Figure 4). A difference in mean phenotype indicates that the marker is linked to a QTL.

Over the past decades, tremendous advances have been made in the use of molecular genetics to find genes or markers linked to genes that affect traits of economic importance in livestock. The main strategies that have been used to find such genes include genome-scans in breed crosses and candidate gene association studies. The breed-cross genome scan approach to QTL detection uses genetic markers spread over the genome to identify genomic regions that harbor QTL. In pigs, the main populations used in these studies have been F2 crosses between breeds or lines. An example of such a cross is the three-generation F2 population that was developed at Iowa State University (Malek et al. 2001a,b) (Figure 5). These studies have

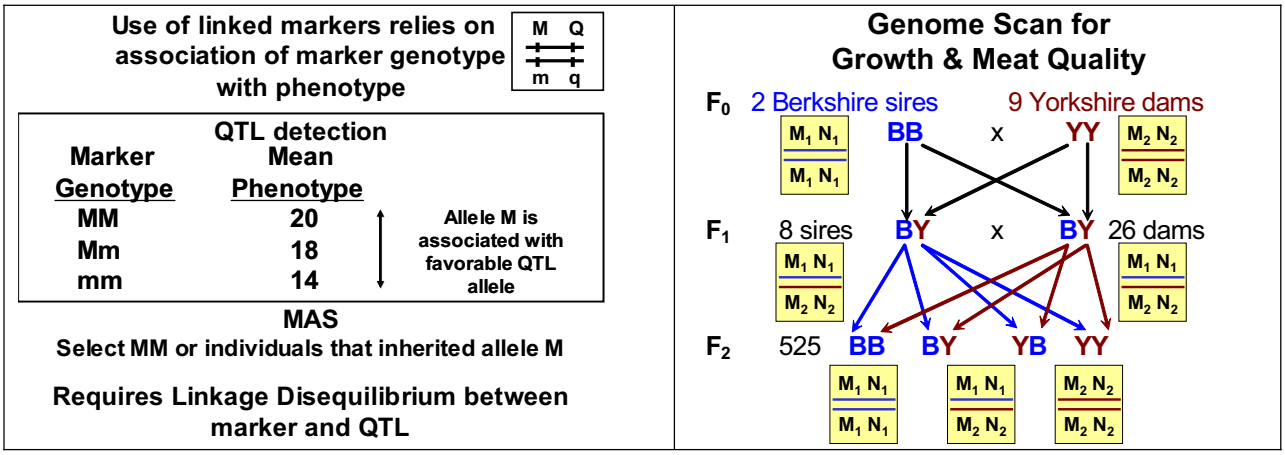
identified many regions of the genome that are associated with economic traits. A database that summarizes the results from most studies is available on the web at: http://www.animalgenome.org/QTLdb/pig.html.

Figure 2. Example linkage map (Rohrer et al.). **Figure 3. Types of molecular genetic loci.**



http://www.genome.iastate.edu/maps/marcmap.html

Figure 4. Principle of marker-QTL associations. **Figure 5. Breed-cross genome scan design.**



Although breed crosses are very powerful to detect QTL, a problem with the breed-cross genome scan approach is that the markers that are found to be associated with the trait in these crosses may actually be quite some distance from the gene that causes the effect. In addition, these approaches detect genes that differ between the breeds that are used in the cross and these genes may not show variation within a breed, which is what is required for within-breed selection. Both these factors limit the direct utility of results from breed-cross studies for within-breed selection.

The candidate gene approach utilizes knowledge from species that are rich in genome information (e.g., human, mouse), effects of mutations in other species, previously identified QTL regions, and/or knowledge of the physiological basis of traits to identify genes that are

thought to play a role in the physiology of the trait. Following mapping and identification of polymorphisms within the gene, the association of genotype at the candidate gene with phenotype can be estimated in a closed pig breeding population. In contrast to the breed-cross genome scan approach, the candidate gene approach identifies markers that are at or close to the causative gene and that segregate within the breeds. These markers can, therefore, be more directly used for within-breed selection.

To date, these techniques for finding genes and QTL, in particular the candidate gene approach, have resulted in the discovery of several genes or markers that are used in the industry. Prime examples are the ryanodine receptor gene (halothane gene) for meat quality, the estrogen receptor gene for litter size, and genetic markers for QTL for growth, backfat, litter size and disease on several chromosomes. These and others are summarized in Table 1.

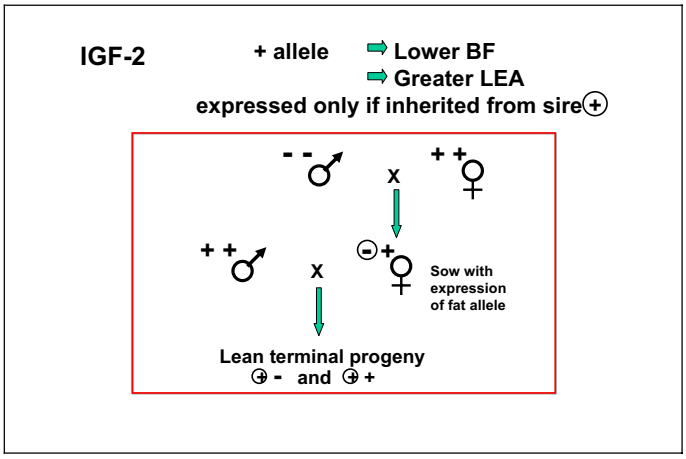
Table 1. Candidate genes and gene tests identified and used in the industry.

Candidate genes	Traits	Industry use
HAL	meat quality/stress	yes
KIT	white color	yes
MC1R	red/black color	yes
MC4R	growth and fatness	yes
RN, PRKAG3	meat quality	yes
AFABP, HFABP	intramuscular fat	??
CAST	tenderness	yes
IGF2	carcass composition	yes
ESR, PRLR, RBP4	litter size	yes
FSHB	reproduction	unknown
NRAMP, SLA	disease susceptibility	unknown
FUT1	disease susceptibility	yes
Trade secret tests	several traits	yes

Recent gene and QTL mapping studies have also revealed that the effect of some genes or QTL depends on whether it was inherited through the sow or the boar. For example the IGF2 gene, which affects carcass composition, has been found to be 'paternally expressed', which means that only the copy that is inherited from the boar is expressed in the offspring (Van Laere et al. 2003). This opens opportunities for the strategic use of genes in crossbreeding programs, as illustrated in Figure 6.

By producing sows from a cross between a boar that is homozygous for the fat (-) allele for IGF2, and mating this sow to a terminal sire that is homozygous for the lean (+) allele, all market pigs will be lean because their sire allele is the lean allele. But, by having inherited the fat allele from their sire, the sows will have the reserves that may help them through gestation and lactation (Buys et al. 2006), but will not pass this on to their progeny.

Figure 6. Tactical use of imprinted genes in cross breeding (Buys et al. 2006).



CURRENT AND FUTURE EFFORTS

At present, we have the ability to select pigs on the basis of individual gene tests for improved reproductive performance, growth rate, leanness and meat quality. Already this has meant benefited genetic improvement of several of these traits. Despite these advances, it is clear that the genes and QTL that have been identified to date only represent the tip of the iceberg and that the majority of genes for the traits of interest have not (yet?) been detected. These genes, therefore, continue to reside in the 'black box' domain (Figure 1b). But imagine for the moment, using not just 5 or 10 genes to select for a trait but 100s or 1000s of genes to improve pig production and create specialized pork products. This has already begun in the US where the two largest swine breeding companies, PIC USA and Monsanto Choice Genetics, are using in some cases 100 to 400 markers in selection in some of their lines. Several additional efforts are currently underway that could further increase these numbers, as described in the following.

Genome Sequencing

The completion of the human genome sequence in the beginning of 2001 has catapulted our understanding of our genetic complexity as human beings. Furthermore, mining this wealth of information allows biologists to understand human diversity including traits like height and weight or eye and hair color, and even more complex traits like susceptibility to various diseases. This means that in the next 10-20 years a whole new form of medicine, called genomic medicine, may make it possible to develop individualized diagnoses, treatments and cures for each person based on their individual and unique genotype. This will revolutionize medicine. Around the world, scientists are spending billions of dollars to learn more about the human genome and these results may also be used to better understand pig health, reproduction, growth, and behavior by comparing the pig genome sequence to the human genome sequence. However, given that our competitors in the chicken and beef industries already have the chicken and cattle genomes sequenced, it is crucial that we also move forward with sequencing the pig if we are to remain competitive.

What is sequencing? Sequencing is the unraveling of the DNA to understand the genetic code (Figure 7). It is equivalent to breaking down books into individual sentences and even specific letters in these sentences and words. The letters in the genetic code (A, T, G, C) are combined into "words" and these words are the genes that control traits or contribute to phenotypes of the animal like rate of growth, level of fat, reproductive performance and disease susceptibility.

Figure 7. Unraveling of chromosomal information to the individual genes.

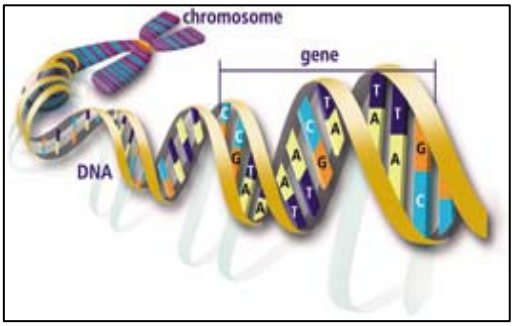


Figure adapted from DOE human genome figure.

Knowing the genetic code requires that we apply modern molecular biology or laboratory methods to break up the code into smaller pieces and then "read" the code.

Progress of the sequencing efforts. Pig genome sequencing began in part when a Danish-Chinese project was initiated several years ago. This project produced a 0.6 X sequence coverage but to have excellent sequence, a 6X copy of sequence is needed. The new effort initiated recently by the US, UK and other country partners has as its goal a 3X -4X coverage, with additional sequencing coverage being obtained from foreign lab contributions, including Canada. Funding to sequence the pig genome is an international effort provided by the USDA, National Pork Board, Iowa Pork Producers Association, University of Illinois, Iowa State University, North Carolina Pork Council, North Carolina State University, the Wellcome Trust Sanger Institute, UK and a number of research institutions from around the world including those from China, Denmark, France, Japan, Korea, Scotland and the U.K. Already this new effort is progressing nicely. Updates can be seen daily at http://www.animalgenome.org/pigs/genomesequence/. These updates are provided as part of the USDA Bioinformatic Coordinator's team effort. Other information about the sequencing can be seen at that page and web pages at the Sanger Institute and the University of Illinois (see http://piggenome.org/index.php). Additional details about the sequencing efforts can be read from the Pig Genome Update also at http://www.animalgenome.org/pigs/newsletter/index.html or at the International Genome Consortium Sequencing Newsletter (http://piggenome.org/newsletter.php).

How does sequencing help? At present we have good but not complete maps of the pig genome. Sequencing will provide not only the "ultimate genetic map" but will allow us to have the tools to hunt down mutations of interest in our own specialized herds and families. This genome sequence of the pig serves as a template to look into the sequence differences in pigs of interest for traits that are economically important (see next section). Sequencing the

swine genome is an investment in basic research with both long- and short-term goals. The potential usefulness of genes in selection for improved pig performance will be determined more quickly if the pig genome sequence is available. Discovery and elimination of undesirable forms or alleles of these genes will be accelerated. Past examples include removal of mutant or negative alleles of the stress gene (HAL) and the Rendement Napole (RN) gene. In the last 10 years, several genes have been identified which improve performance and leanness (IGF2, MC4R), meat quality (CAST, PRKAG3) and reproduction (ESR, PRLR). Sequencing of the pig genome offers the ability to multiply these discoveries into the 1000s and speed the rate of these discoveries. Greater federal funding for pig genomic research can be leveraged to provide more rapid application in these areas. The pig genome sequence can also be used to provide insights into how genes work together. This will allow better genetic planning to allow pig breeders and producers to select animals possessing certain sets of genes that interact in a favorable manner for a particular production system or niche market. Sequencing the pig genome will dramatically accelerate identification of determining the genetic basis of economic traits and their interaction with the environment, which could revolutionize pork production.

For the average pork producer, the many benefits include improved growth and litter size performance due to identification of genes affecting these traits. The genome sequence is a powerful tool, which will enable discoveries for improving traits of interest for producers regardless of their operational size. However, producers and companies associated with more advanced research groups or breeding companies may have the opportunity to leap frog with new genomic strategies. For these better positioned producers and early adopters, more advanced opportunities are likely to include in the next 5-20 years the ability to produce pigs with improved immune response abilities (vaccine ready pigs), growth primed sire lines and development of increased niche and branded products representing unique or special attributes that one producer or one company wishes to use to increase market share and profits. It is likely that producers will have the ability to select certain genetic lines in the future that will require specialized feeds but that could outperform existing lines.

High-Density SNP Genotyping

Genome sequencing typically uses the DNA from a single individual. Genetic selection, however, requires us to identify locations in genome where individuals differ in sequence. These so-called single nucleotide polymorphisms (SNPs) can be identified by comparing the detailed sequence of the single individual to the sequence of other individuals, e.g. from other breeds. For example, in the chicken, over 2.8 million SNPs were identified by comparing the sequence of the Red Jungle Fowl to that of three domesticated breeds (International Chicken Polymorphism Map Consortium, 2004). Efforts to identify large numbers of SNPs have also been initiated through the Danish-Chinese project and in-house by some pig breeding companies. This large number of SNPs enables sufficient numbers of markers to be placed along the genome (e.g. 6 to 50 thousand) such that most QTL will have one or more SNPs located close enough that they can be detected by within-breed association studies. Note that this is similar to candidate gene studies, except that every region of the genome is evaluated, rather than only the candidate gene regions. Studying this many markers is now also possible because of the development of less expensive high-throughput genotyping technology, which

allows large numbers of individuals to be genotyped for a large number of markers at a reasonable costs (estimates as low of $300 for genotyping an individual for 40,000 SNPs have been quoted). This will greatly accelerate the discovery of genes associated with traits and will allow analysis to be conducted directly within a breed and even on commercial pigs.

Genomic Selection

When only a limited number of markers or genes are available, a large proportion of genes that affect the trait will remain in the 'black box' of quantitative genetics (Figure 1b). In this case, selection on marker data alone will not result in great advances in genetic improvement but marker data must be used in combination with regular EBV estimated from phenotypic data on the individual itself and/or its relatives, to ensure that balanced genetic progress is achieved for all genes that affect the trait. This, however, changes if animals can be genotyped for a large number (5,000 or more) of markers across the genome, as is now possible at much reduced costs using high density SNP genotyping. With such technology, Meuwissen et al. (2001) showed that an individual's EBV could be estimated with accuracies as high as can be achieved by progeny testing based only on the individual's genotypes for the markers across the genome. In this strategy, which Meuwissen et al. (2001) called genomic selection, estimates of marker effects are obtained using phenotypes and marker genotypes from a previous generation, which are then used to estimate the breeding value in new generations without the need for additional phenotypes. Although the practical feasibility of genomic selection has yet to be demonstrated, applications of genomic selection are near or underway in several livestock breeding programs.

Genomic selection does not require the actual location of genes that affect the trait to be known. Instead, statistical methods similar to animal model BLUP EBV are used to estimate breeding values of each of many regions across the genome based on associations of phenotype with alternate marker genotypes that exist in the population in each region. Then, the breeding value of an individual can be estimated by simply summing the EBV of the marker genotypes that the individual has for each region.

Marker-Assisted Selection for Commercial Crossbred Performance

A major limitation of today's pig breeding programs is that most selection is in purebred herds, where pigs are raised under high biosecurity. Several studies have, however, shown that purebred performance under nucleus conditions can be a poor predictor of performance of crossbreds raised under commercial circumstances, with genetic correlations as low as 0.4 to 0.7. These limitations can be overcome by collecting phenotypic data on crossbred progeny raised under commercial conditions and using this data to estimate breeding values of purebred pigs, but this is difficult and expensive to implement. These limitations can, however, be overcome by selecting on effects of markers estimated on commercial crossbreds, as illustrated in Figure 8. Results in Table 2 suggest that this cannot only improve response to selection for commercial crossbred performance, but also reduce rates of inbreeding.

Figure 8. **Diagram of a pyramid breeding program, with selection among purebreds in a purebred environment and illustrating the sources of phenotypic and marker data that can be used for selection among purebreds.**

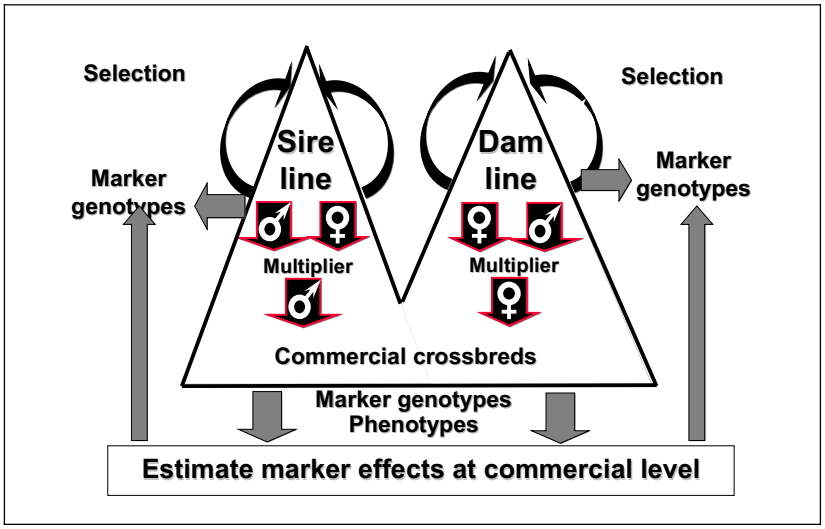


Table 2. **Potential benefit of using marker data to improve commercial crossbred performance.**[1]

Data used for selection	% of genetic variance explained by markers							
	9		25		49		64	
	Resp.	Inbr.	**Resp.**	Inbr.	**Resp.**	Inbr.	**Resp.**	Inbr.
Purebred phenotype	**100**	2.09	**100**	2.09	**100**	2.09	**100**	2.09
Purebred phenotype + Crossbred phenotype	**137**	3.02	**137**	3.02	**137**	3.02	**137**	3.02
Purebred phenotype + X-bred marker data	**108**	1.90	**124**	1.56	**145**	1.25	**158**	1.12

[1] Selection was for commercial crossbred performance for a trait with heritability 0.4 and a genetic correlation of 0.7 between purebred nucleus and commercial crossbred performance, mimicking selection for growth in pigs. Resp = response relative to selection on purebred phenotype (=100%); inbr = rate of inbreeding per generation

CONCLUSIONS

In the past decade, several genes and many genomic regions affecting economic traits have been identified and several of these have been incorporated in selection programs. The impact of molecular genetics on pig breeding programs and pig production is, however, expected to dramatically accelerate in the future through complete sequencing of the pig genome and

availability of large numbers of markers. Sequencing efforts have started and are moving along nicely. Results of these efforts are already being used to help select markers for improved growth and meat quality. Given the funding available, about $15 million presently, it is likely we will have a draft sequence of the pig genome by late 2007 or early 2008. Will companies and seedstock breeders be ready to take advantage of these discoveries? Producers must ask the difficult questions. Are they ready to use the new genetics and genomics information? Are they positioned to 1) understand the information and 2) to use it effectively? Are there genetic systems in which this information can be used more effectively to improve pig production? Are there niche markets for new products that can be produced using these technologies? Team work and partnerships with the right seedstock breeders or breeding companies and university or government research faculty are likely to be keys in transforming this public information from a useful resource to a real payoff. Only then will producers, companies and geneticists help members of the pig industry really bring home the bacon.

ACKNOWLEDGEMENTS

Collaborative research efforts of Drs. G. Plastow, A. Mileham, A.M. Marcos, R. Fernando and members of the Rothschild and Dekkers labs are appreciated. A number of researchers world wide have participated in the sequencing efforts and in particular Dr. Larry Schook, UI and Jane Rodgers and Sean Humphray, Sanger Center are noted. The author wishes to thank financial support received from the USDA NRSP8 which supports the National Pig Genome Coordination project. Support for the pig genome sequencing comes from the USDA, National Pork Board, Iowa Pork Producers Association, University of Illinois, Iowa State University, North Carolina Pork Council, North Carolina State University, the Wellcome Trust Sanger Institute, UK and a number of research institutions from around the world including those from China, Denmark, France, Japan, Korea, Scotland and the U.K. Funding for individual research has been provided in part by Sygen and PIC USA, Monsanto Co., Hy-Line Int., and by Hatch, Iowa Agricultural Experiment Station and State of Iowa funds.

LITERATURE CITED

Buys, N., G. Van den Abeele, A. Stinckens, J. Deley, and M. Georges. 2006. Effect of the IGF2-intron3-G3072A mutation on prolificacy in sows. 8th World Congress Genet. Appl. Livest. Prod. Comm. No. 06-22.

Malek, M., Dekkers, J.C.M., Lee, H.K., Baas, T.J., Rothschild, M.F. 2001a. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. Mamm. Genome 12: 630-636.

Malek, M., Dekkers, J.C.M., Lee, H.K., Baas, T.J., Rothschild. 2001b. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition. Mamm. Genome 12: 637-645.

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

Van Laere, A.-S., M. Nguyen, M. Braunschweig, C. Nezer, C. Colette, L. Moreau, A. Archibald, C. Haley, N. Buys, M. Tally, G. Andersson, M. Georges, and L. Andersson. 2003. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. Nature 424: 832-836.