Genome-wide selection for reproductive traits in swine

David G McLaren[1], Matthew A Cleveland[2], Nader Deeb[2], Selma Forni[2], Alan J Mileham[1], Scott Newman[2], Olwen I Southwood[3] and Lizhen Wang[2]

[1]Genus R&D, 1525 River Rd., DeForest, WI 53532, USA; [2]Genus R&D, 100 Bluegrass Commons Blvd., Suite 2200, Hendersonville, TN 37075, USA; [3]Genus Breeding, Alpha Building, London Road, Nantwich, CW5 7JW, UK

The introduction of high-density SNP arrays in livestock species has enabled genomic evaluations on a scale not possible just a few years ago. Faster genetic gains are realized from application of genomics in pigs by increasing the accuracy of selection. This is especially important for lowly heritable reproductive traits, where female selection candidates have not yet expressed a phenotype at the point of selection and male candidates have no phenotype at all. Litter size was one of the first traits for which the breeding company PIC implemented genome assisted selection, using a 196 SNP panel in 2010. Significant improvements in accuracy for all selection index traits are now achieved using single step genomic evaluation incorporating genomic information into multivariate mixed model evaluations. Genotyping costs have been a barrier to full-scale implementation, but this challenge has been overcome by imputing high density genotypes from low density panels on selection candidates. Next generation sequencing technology is revolutionizing genomics research. The recently published draft pig genome sequence, along with the availability of phenotypes and tissue samples on thousands of animals, often with complete pedigree, facilitate discovery of genomic regions, genes and causative mutations for traits such as disease resistance. In the future, genome editing has the potential to introgress beneficial alleles from rare or indigenous breeds not present in improved commercial lines, as well as increasing the expression of beneficial genes through directed gene duplication. Additionally, nonlinear and nonparametric methods may be applied to further improve statistical predictions of genetic merit and performance.

Introduction

Genomic information has been used to increase accuracy of selection for a range of commercially important traits in commercial pig breeding, beginning with use of the HAL-1843® 'stress' gene test in 1991 (Fujii et al., 1991, MacLennan & Phillips, 1992). This review traces the history of genetic improvement for reproductive traits in swine from application of multivariate best linear unbiased prediction (BLUP) beginning in the early 1990s, through marker assisted selection (MAS), genome assisted selection (GAS), and the recent application of genomic selection (GS). The review concludes with a brief consideration of new statistical and genomic technologies,

currently the subject of research, expected to further improve accuracy and rate of genetic improvement in reproductive and other quantitative traits of economic importance to pork production.

Quantitative genetic selection to improve sow reproduction

Reproductive performance is a crucial component of any pig breeding program as efficient pork production requires highly productive sows (Dekkers *et al.*, 2011). Pig breeding companies began using quantitative genetics-based multi-trait BLUP evaluation systems (Henderson, 1975, Harris & Newman, 1994) to improve pig reproduction traits starting in the early 1990s, and good genetic progress as evidenced by both genetic and phenotypic trends has been achieved (Figure 1).
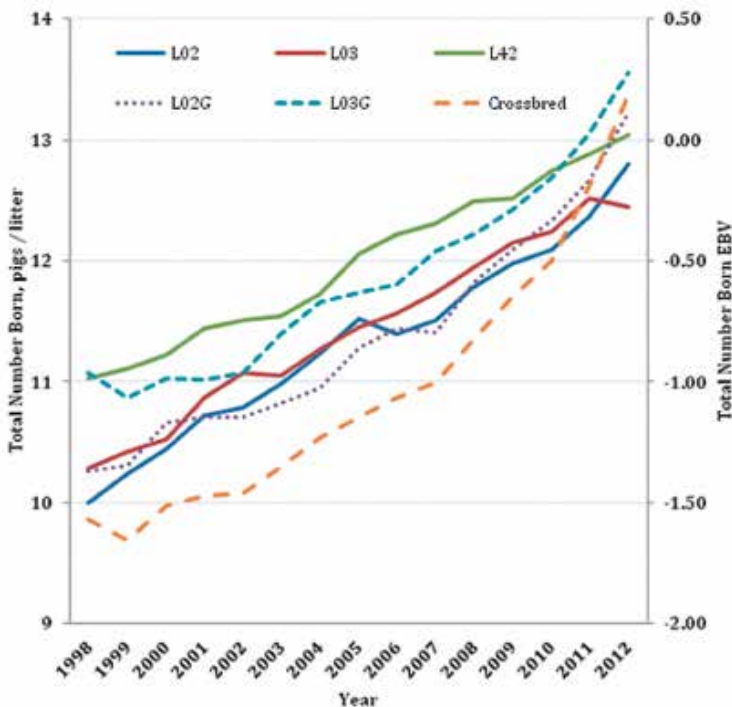
Fig. 1 Fifteen year genetic (dotted/dashed lines) and phenotypic (solid lines) trends in PIC purebred (L02, L03) and crossbred (L42 = L02 × L03) sow litter size (total number born per litter).

Quantitative genetics applies to traits which are affected by the action of many genes, as distinct from qualitative traits controlled by just a few major genes. The assumed genetic model behind the past success of genetic selection for reproductive (and other quantitative) traits is that such complex traits are genetically controlled by an infinite number of genes, each with an infinitely small effect (the infinitesimal gene model, Lynch & Walsh, 1998). Based therefore on unrealistic biological assumptions, gene frequencies are changed by selection without any knowledge of how many genes are involved, where they are located, the additive effects they each contribute to the trait selected for, or the frequencies of favorable alleles. That this statistical-genetic "black box" method works, which it does (McLaren, 2007), is not surprising given genome wide association studies (GWAS) using tens or hundreds of thousands of single

nucleotide polymorphisms (SNPs) have confirmed the existence of hundreds or thousands of genes throughout the genome contributing small effects to the overall additive genetic variance for most, if not all, quantitative traits (McCarthy *et al.*, 2008, Flint & Mackay, 2009).

Complex statistical models adjust for multiple sources of non-genetic (environmental) variation and yield the estimated sum of the additive effects of alleles an individual carries, the estimated breeding value (EBV), required to select genetically superior parents of the next generation. This process depends on the phenotypic performance of individuals and their relatives, the genetic relationship among these individuals, and known sources of environmental variation.

Phenotypes and computations

The structure of porcine genetic improvement programs involves a high biosecurity nucleus tier where selection occurs in purebred populations (Dekkers *et al.*, 2011). Commercial production takes place in different environments and is based on crossbred sows (Figure 2). Thus, success of selection depends on how much genetic progress is realized in crossbred pigs. The genetic correlation between purebreds and crossbreds is a measure of efficiency (accuracy) of genetic selection in purebred animals for crossbred performance. The smaller the correlation, the more important crossbred information becomes. Genetic correlations between purebred nucleus and commercial crossbred performance for economically important traits in the pig can deviate significantly from unity (Merks, 2001, Nakavisut *et al.*, 2005), indicating not all the improvement
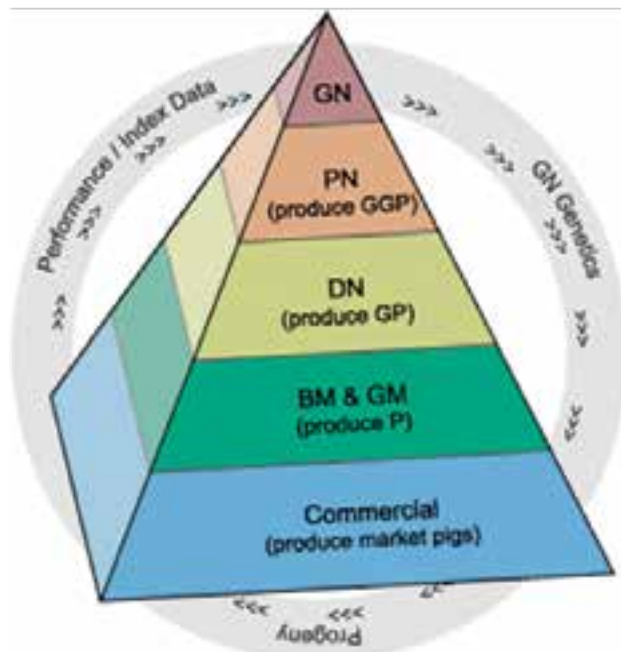
Fig. 2. The classic pig production pyramid. Pure line genetic improvement made at the genetic nucleus (GN) level is multiplied in production nucleus (PN) farms where female great grandparents of the commercial generation are produced and supplied to daughter nucleus farms producing grandparents (GPs) for gilt and boar multipliers (GM and BM) which produce the commercial crossbred parent sows. For example, dam lines A, B, C and sire lines D and E might be improved in the GN, PNs multiply CxC and ExE, DNs produce BxC sows, GMs produce Ax(BxC) parent sows and BMs DxE parent boars. Increasingly PN and DN functions are merged and pure line terminal sires used in a Ex(BxC) commercial crossbreeding system. The ratio of sows in GN/PN/DN:GM:Commercial is approximately 1:4:40.

predicted based on purebred traits in the nucleus environment will be realized by crossbreds in a commercial environment.

In PIC's genetic data recording system, semen from maternal line boars used at the nucleus level is distributed to cooperating multiplier herds to produce F1 parent sows whose lifetime performance is recorded in commercial herds (Casey *et al.*, 2006, Perez *et al.*, 2006, Knap & Su, 2008). Measurements are collected from both crossbred and purebred sows for total number born and percentage of pigs stillborn. In addition, preweaning survival, litter weaning weight and weaning to first mating interval are measured on purebred litters or sows in the nucleus farms. Repeated records (when available) from each sow are used in estimating breeding values. Supporting information from each litter record is incorporated into genetic evaluation models in the form of fixed non-genetic effects and (or) linear covariates. Examples include farm, line, parity, number of services (returns), farrowing date, mating type (purebred or crossbred), age at first farrow, start litter weight after cross fostering, number of nursing days and number of pigs after cross-fostering. Multiple trait repeatability animal models are used for genetic parameter estimation for all these traits. Genetic parameters are obtained using the REMLF90 program (Misztal, 1999). Using a high performance computing environment all seven traits can be run together, but this can be time-consuming. Alternatively, two-trait models are used to speed the estimation process and matrix bending performed to obtain a positive-definite matrix for use in multivariate genetic evaluation. Retrospective analyses are performed to estimate genetic trends in both pure and crossbred populations. As illustrated in Figure 1, selection based upon mixed statistical model analysis of swine reproduction traits has been successful despite the traits being lowly heritable, sex limited and expressed late in a crossbred female's life.

Early use of molecular genetics in pig breeding

The search for "major" genes associated with economic traits in pigs began in the late 1980s and early 1990s when researchers started to look for quantitative trait loci (QTL) using microsatellite DNA markers in divergent crosses (Beuzen *et al.*, 2000, Bidanel & Rothschild, 2002, Dekkers, 2004, Bidanel, 2011, Dekkers *et al.*, 2011). The concept being that, particularly for important traits that are lowly heritable, sex limited and / or not expressed in selection candidates, if there were QTL of relatively large effect their detection could be exploited by marker assisted selection (MAS) schemes to accelerate rates of genetic improvement by improving the accuracy of selection (Figure 3, van der Steen *et al.,* 2005).

In particular, diverse crosses between highly prolific Chinese breeds (e.g., Meishan) and western pigs gave hope that genes involved in traits associated with reproduction could be identified and useful variation exploited in pig breeding. Major drawbacks to this approach, however, were that associations found in such divergent crosses did not easily translate into commercial populations; the distances between markers were very large; and there was a significant chance that beneficial Chinese alleles did not exist in commercial populations.

An alternative was the candidate gene approach where DNA sequence variation, commonly SNPs, would be identified in genes thought to be involved in traits of interest and investigated for statistically significant association with that trait, either in a divergent cross or a commercial population. A significant advantage of this approach was that, provided a SNP had a reasonable minor allele frequency (>0.1) in a commercial population and had an economically significant effect on a trait of interest, it could be immediately used in a MAS process to improve traits such as litter size. This was taking place in the 1990s, which was followed by a decade that saw phenomenal growth in high throughput genotyping and leaps in the large data base computing capacity required to apply the technology outside the research laboratory.
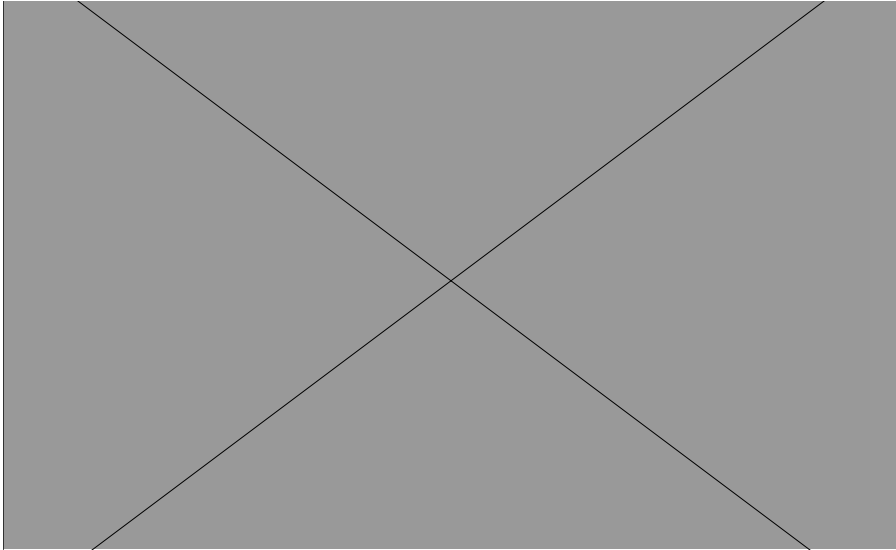
Fig. 3 Relationship of accuracy of selection to rate of genetic gain, Delta G.

Collaboration between PIC and Max Rothschild's group at Iowa State University was particularly successful in identifying SNPs in candidate genes for litter size in Chinese Meishan and crossbred Meishan populations, and then validating associations in commercial populations. Early successes from this collaboration resulted in markers in the estrogen and prolactin receptor genes as well as the retinol binding protein gene, all of which were used in MAS by PIC (Messer *et al.*, 1996, Rothschild *et al.*, 1996, Vincent *et al.*, 1997). The estrogen receptor gene marker was used to move the favorable allele from a frequency of around 0.7 to fixation in PIC's Large White origin maternal line (*unpublished data*).

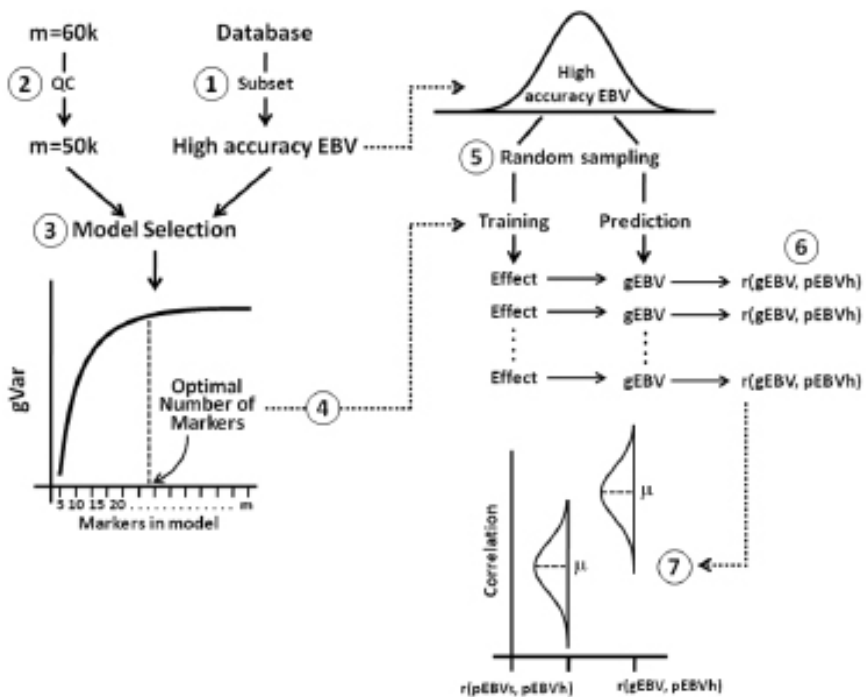## Genome Assisted Selection (GAS) for single traits using small SNP panels

In 2008 the International Swine Genome Sequencing Consortium collaborated with Illumina to develop a nominal 60,000 SNP genotyping array (PorcineSNP60 BeadChip) which became commercially available in early 2009 (Ramos *et al.*, 2009). PIC used this new resource to identify multiple SNPs significantly associated with litter size in commercial crossbred females and selected the best 196 SNPs to form a genotyping panel used in female line GAS between 2010 and 2012 (Deeb *et al.*, 2011). Other researchers have also reported a number of informative QTL and genes within QTL for reproductive traits, including lifetime productivity (Onteru *et al.*, 2011, 2012), based upon GWAS using the PorcineSNP60 BeadChip.

The effectiveness of genome scans depends largely on the level of linkage disequilibrium (LD) within the target population that can be captured by markers. The higher the level of LD the fewer markers are needed to capture the genomic regions contributing to the trait of interest. In a survey of several PIC pure lines Deeb *et al.* (2010) and Cleveland & Deeb (2012) found a few thousand equally-spaced SNPs achieved average LD of 0.2 or higher in most pure lines and crosses. Moreover, of the 64k markers on the commercially available PorcineSNP60 BeadChip, over 50k SNPs had allele frequencies of 5% or higher in most PIC lines and crosses, indicating this chip provides enough markers to be effectively used in any PIC population.

As stated previously, reproduction is one of the most difficult traits to improve through selection in pigs due to low heritability, sex limited expression, and expression late in life in

commercial crossbred animals. This, and its high economic value, made number born alive per litter (NBA) one of the first target traits for GAS at PIC.

Deeb *et al.* (2011) used a training set of 2,160 PIC dam line animals with high accuracy (r > 0.8) EBVs for NBA (pEBVh – high accuracy polygenic estimated breeding values) to estimate SNP effects on NBA. Trimming the pedigree and eliminating performance records to emulate information available when animals would have completed performance test and been selected off-test, they calculated the animals' EBVs at point of selection (pEBVs). All animals were genotyped using the Illumina PorcineSNP60 BeadChip. Then a subset of 196 markers was selected that best explained the genetic variation for NBA in this population (Figure 4). The predictive ability of the marker panel was evaluated in a series of cross-validations where, in each iteration, the correlation between the marker predicted or genomic EBV (gEBV) and the pEBVh was calculated. The final implementation of the NBA panel involved blending the gEBV with pEBVs using methodology similar to VanRaden *et al.* (2009) to produce blended EBVs (bEBV) which further improved accuracy (Table 1).

Fig. 4 Process used to develop a small SNP panel to use in genome assisted selection (GAS).

1. Subset of the population is selected based on accuracy of their breeding values. 2. SNPs preselected based on minor allele frequency and Hardy-Weinberg equilibrium. 3. SNPs with the smallest effect are trimmed out from the model. 4. The optimal number of SNPs is selected balancing cost of genotyping (number of SNPs) and the genetic variance explained by the model. 5. Cross validation – randomly selecting a subset of the population for training and predicting on the remaining subset. 6. Calculate the correlation between gEBV and pEBVh in the prediction subset. 7. Calculate the average correlation between gEBV and pEBVh from all subsets in the cross validation and compare to the average correlation between pEBVs and pEBVh.

Table 1. Correlation (r) between litter size (NBA) EBVs[a] at selection with proxy "true" breeding values[b].

Correlation (r)		Improvement in accuracy
Pedigree only r(pEBVs, pEBVh)	0.560	
Genomic only r(gEBV, pEBVh)	0.766	+ 36.8%
Blended pedigree plus genomic r(bEBV, pEBVh)	0.787	+ 40.5%

[a]NBA EBV is the estimated breeding value for number of pigs born alive per litter.
[b]pEBVh are high (>0.8) accuracy polygenic estimated breeding values.

Genomic selection (GS)

GS is a relatively new technology that incorporates large-scale "genome wide" DNA information in genetic evaluations, with the aim of increasing the accuracy of selection compared to GAS and thus further accelerating rates of genetic improvement (Figure 3). GS is characterized by selection for tens or hundreds of thousands of markers simultaneously, covering the entire genome in a dense manner so that all genes are expected to be in linkage disequilibrium with at least some of the markers and thus, potentially, all of the additive genetic variation can be accounted for by markers (Goddard & Hayes, 2007, 2009). Proposed by Meuwissen *et al.* in 2001, it has only recently become practical to consider for implementation in pig breeding schemes (e.g., Lillehammer *et al.*, 2011) as costs of genotyping have decreased.

GS increases the precision of predicting genetic differences between animals and decreases the age at which accurate ranking of animals can be obtained. In practice, GS refers to selection decisions based on genomic estimated breeding values (GEBVs). The GEBVs are estimates of heritable genetic differences calculated using DNA molecular information, pedigree and records. Requirements for successful implementation of GS include a large number of SNPs across the genome, a large number of samples from pedigreed animals with phenotypic data (Goddard, 2009, Meuwissen, 2009), advanced statistical genetics methodologies and algorithms, sufficient computing power for analysis and real-time implementation (Cole *et al.*, 2012), and affordable genotyping technology – all of which are available to the pig breeding industry today. Of these factors, affordability of SNP chips is most limiting. Benefits of GS are population and trait specific, depending on the number of genotyped animals, the extent of relationships between genotyped animals and the rest of the population, trait heritability, and the number of records available for the trait.

Multi trait Genomic Selection (GS) using a genomic relationship matrix

The concept of identity by descent (IBD) is classically quantified as coefficients of relationship determined by homologous alleles descended from a common ancestor (Powell *et al.*, 2010). This fundamental concept has many uses in genetics, including predicting genotype frequencies, mapping genes, estimating genetic variance and predicting inbreeding depression. In livestock genetic evaluation, relationship coefficients have been extensively used as estimates of genetic covariance between individuals. Traditionally, relationships were estimated using only pedigree information. Recently, molecular markers for many loci across the genome have been used to estimate relationships and proved to be more precise than pedigree information (Van Raden, 2008, Forni *et al.*, 2011).

Marker-based relationships can better estimate the proportion of chromosomes segments shared by individuals because high-density genotyping can identify genes identical in state

D.G. McLaren et al.

that may be shared through common ancestors not recorded in the pedigree. A genomic relationship matrix (G) can be calculated by different methods (Gianola & van Kaam, 2008; Van Raden, 2008). Because an entire population is unlikely to be genotyped, Misztal *et al.* (2009), Aguilar *et al.*, (2010) and Christensen & Lund (2010) proposed the integration of a pedigree-based relationship matrix (A) and a genomic relationship matrix (G) into a single matrix (H). Genetic evaluation using H as the genetic covariance matrix between individuals is called single step genomic evaluation (SSE). Besides the computation of H, no further modifications in models and software commonly used in livestock genetic evaluation are required in the SSE. The method has become popular for computational straightforwardness; especially in swine breeding programs where information is accumulated and genetic evaluation is often computed weekly (Forni *et al.*, 2011, Christensen *et al.*, 2012).

Pedigree-based relationships are calculated with respect to a base population in which an arbitrary relationship across individuals is defined, usually equal to zero. The individuals in the base population are called founders, and a typical assumption is that they do not share genes from older ancestors. Relationship and inbreeding coefficients from later generations are estimated as deviations from the founders' relatedness. In the SSE, genomic relationship coefficients should be calculated with respect to the same base population as pedigree coefficients. Methods to calculate genomic relationships analogous to the pedigree-based coefficients have been proposed (Van Raden, 2008). However, compatibility is difficult to achieve if there was genetic drift, artificial selection or genotype sampling in the population. Ideally, genomic relationships should be estimated using the allele frequencies from the same base population chosen to calculate the pedigree-based coefficients. This information can rarely be extracted from historical data and approximations must to be used. Forni *et al.* (2011) showed that observed allele frequencies in genotyped samples of swine commercial populations could be used as approximations regardless of the effects of selection if additional scaling of the G matrix was performed. Other methods for scaling G according to the pedigree-based matrix were proposed by Christensen *et al.* (2012). Both authors showed that errors in allele frequency estimates and incorrect scaling could result in biased estimates of relationships and genetic effects.

Molecular markers allow more accurate identification of genetic differences between genotyped individuals in the absence of their own or progeny phenotypes, as is the case for sow reproduction traits. Several studies in different species have shown that incorporating high-density molecular markers in genetic evaluations produced more accurate predictions of genetic values for genotyped animals than traditional pedigree-based evaluations (Van Raden, 2008; Hayes *et al.*, 2009, Forni *et al.*, 2011). In addition, SSE can produce more accurate predictions for non-genotyped animals (Christensen *et al.*, 2012). Improvement in accuracy of breeding values for reproductive traits with genomic information obtained by PIC is shown in Table 2.

Table 2. Average accuracy of breeding values estimated based on pedigree (EBV) and based on combined pedigree and PorcineSNP60 BeadChip genomic information (SSE GEBV) of 2,023 young pigs.

| | Accuracy | | |
Trait	EBV	GEBV	Diff*
Total number of piglets born	0.25	0.42	68%
Stillborn in the litter (%)	0.26	0.43	65%
Piglet survival until weaning (%)	0.17	0.26	53%
Litter weaning weight (kg)	0.23	0.35	52%
Interval between weaning and next mating (days)	0.17	0.30	76%

*Diff = relative increase in accuracy with genomic information

Implementing Genomic Selection (GS)

All existing genomic information should be available for inclusion in the breeding value at the time of selection for genome-wide selection to be successful. This includes the genotypic data of the selected pig and its parents, if collected. If the young pig is to be genotyped and these data used for selection the time required for all procedures, from sample collection to final breeding value estimation, needs to be known so all have time to take place. The time commitments for several aspects of this process are known and relatively fixed. These include the time required for DNA extraction; the actual genotyping process; the subsequent manipulation of the resulting genomic data, including quality control evaluations to weed out low accuracy calls and minimize the introduction of incorrect genotypes into the evaluation. Estimation of these time commitments allows for the timely identification of animals and collection of their DNA for analyses which reduces genotyping costs associated with animals not on test or those removed during testing. However, collection of samples is usually simpler when performed at a point where there are facilities and time to collect and record the DNA samples accurately. In PIC nucleus farms tissue sample collection is done on all pigs typically within about 24 hours of birth, at the same time the young pigs are recorded and individually tagged.

Samples are collected and placed into uniquely bar-coded tubes. The tube label's code is linked with the piglet tag and recorded directly into the database using hand held computers. On arrival at the tissue storage facility, labels are scanned together with their box and freezer location and recorded directly into the database. In this way, an individual pig's tissue for DNA extraction can be quickly located whenever required.

As the quantity of DNA information collected increases, and especially the depth of the data, i.e., parents and offspring with genotypic data, these data are also used for quality control of the pedigree. If young pigs are to be genotyped prior to selection, then advancing the genotyping procedures by a few weeks will allow these genotypes to be used to check for Mendelian inconsistencies.

Although as many procedures as possible are put into place to avoid errors – such as bar coding of tissue samples, scanning data directly into databases, individual electronic tags for pigs - mislabeling can still occur. Pigs may be cross-fostered before the sampling procedures take place and not recorded correctly, samples may be placed in the wrong tubes, a tag may be read incorrectly, etc. Although these errors occur at a very low level, any misidentification of the pedigree or genomic data can lead to a reduction in the overall accuracy of the breeding value.

A low incidence of Mendelian conflicts may indicate genotyping errors, but a high incidence is an indicator that there is an incompatibility between the genotypes of parent and offspring. What is does not say is whether the error is due to the parent or the offspring, or whether it is a pedigree or sample issue. This requires further investigation and possibly some extra genotyping to identify the cause and suggest possible solutions. The ideal situation is that data are corrected, e.g., samples are re-assigned to the correct pigs, or pigs are re-assigned to the correct parents, resulting in usable data and pigs. In the event that the correct information cannot be verified, then the data (pedigree and/or genotypes) are removed and not used in the genetic evaluation. In PIC's genetic evaluations over the past 12 months $< 0.2\%$ of individuals have been excluded due to irresolvable Mendelian conflicts. The overall result aimed for is to maximize the accuracy of the breeding values.

Imputation for cost-effective genomic analysis

The introduction of high-density SNP arrays in livestock species has enabled genomic evaluations on a scale not possible just a few years ago. Implementation of these genome-wide approaches

has been challenging, however, due to the size of the available datasets. Large amounts of data, both phenotypic and genotypic, are needed to make meaningful use of genomic information, but, as described earlier, the cost of genotyping has been a barrier to full-scale implementation for many in the livestock industry. Genotype imputation has the potential to overcome this challenge and increase the amount of genotype data for a fraction of the cost of high-density genotyping.

Imputation is the process of using heuristic or statistical rules to fill in genotypes that are missing because either the marker was not on the genotyping panel or was not called. Imputation has been a focus in human research for a number of years, particularly as researchers have attempted to create larger datasets by combining genotypes from different arrays, and thus many genotypes are missing (see Marchini & Howie 2010 for a comprehensive overview). A number of approaches have been developed to impute genotypes and generally fall into two categories: population-based methods using linkage disequilibrium information (e.g., Scheet & Stephens, 2006, Browning & Browning, 2007, Howie *et al.*, 2009) and pedigree-based methods that use linkage information (e.g., Druet & Georges, 2010, Daetwyler *et al.*, 2011, VanRaden *et al.*, 2011, Hickey *et al.*, 2012). Livestock populations differ in structure from human populations. Approaches have been developed that take advantage of the information available in most livestock breeding programs (e.g., a complete pedigree) and to address the requirement of imputing from very low-density panels (e.g., Hickey *et al.*, 2012).

Faster genetic gains are realized from genomics by increasing the accuracy of EBVs at the time of selection, usually when the animal is young. This is especially important for traits related to reproduction where female selection candidates have not yet expressed a phenotype and male candidates will not have a phenotype at all. Increased EBV accuracies have been observed when using dense genotypes (Table 2) but these gains are constrained when dense genotypes are not available for selection candidates. A large number of candidates are produced in each litter and it is generally cost-prohibitive to exhaustively genotype at high density. Genotyping strategies combined with imputation are needed for a cost-effective implementation of GS.

There are a number of issues to consider when determining an appropriate genotyping strategy, including extent of pedigree recording, access to an imputation approach that is accurate when using a density that is affordable, availability of affordable low-density genotyping platforms (in addition to a high-density platform) and determination of target animals to be genotyped. A general strategy is to densely genotype a relatively small number of individuals (e.g., some parents) and sparsely genotype a relatively larger group of other individuals (e.g., selection candidates and other parents). Huang *et al.* (2012) evaluated a number of genotyping strategies in pigs in terms of impact on imputation accuracy and cost. They found that the cost of genotyping could be greatly reduced when genotyping selection candidates for a very small panel (384 SNPs) and sires and grandsires for the full 60k SNPs, with little reduction in the content of the genomic information (as characterized by imputation accuracy). Adding low-density genotypes for female parents offered small additional increases at reduced cost.

PIC has developed a low-density panel that could be used to impute 60k genotypes (Cleveland *et al.*, 2012, Deeb *et al.*, 2011, 2012). In contrast to other small panels that have been developed for GAS, the ideal panel for imputation is one that is usable across all populations of interest and contains SNPs with particular properties that are evenly distributed across the genome, rather than specific for any trait. Genotypes for all SNPs on the high-density platform can then be used for genomic evaluation. To select the SNPs for inclusion in the panel, each region of the genome was assigned to one of a large number of bins (~ 1,000). A small number of SNPs within each bin were identified based on minor allele frequency, the Chi-square test statistic for Hardy-Weinberg equilibrium and percent called genotypes across all populations of interest.

Additionally, SNPs with low confidence in genome map position were excluded. The final panel for imputation contained ~450 SNPs.

A pipeline has been developed to impute high-density genotypes on a large-scale in a production environment using a high-performance computing cluster (Figure 5). The basis for the pipeline is the software package *AlphaImpute* (Hickey *et al.*, 2012), which combines simple phasing rules, long-range phasing, haplotype libraries, segregation analysis and recombination modeling. *AlphaImpute* will impute genotypes for any number of genotype densities in the data, to the highest density, for all animals and loci. The imputed genotypes are the sum of the imputed alleles or allele probabilities and therefore no genotypes are left un-imputed, which is a requirement for some genomic evaluation software.
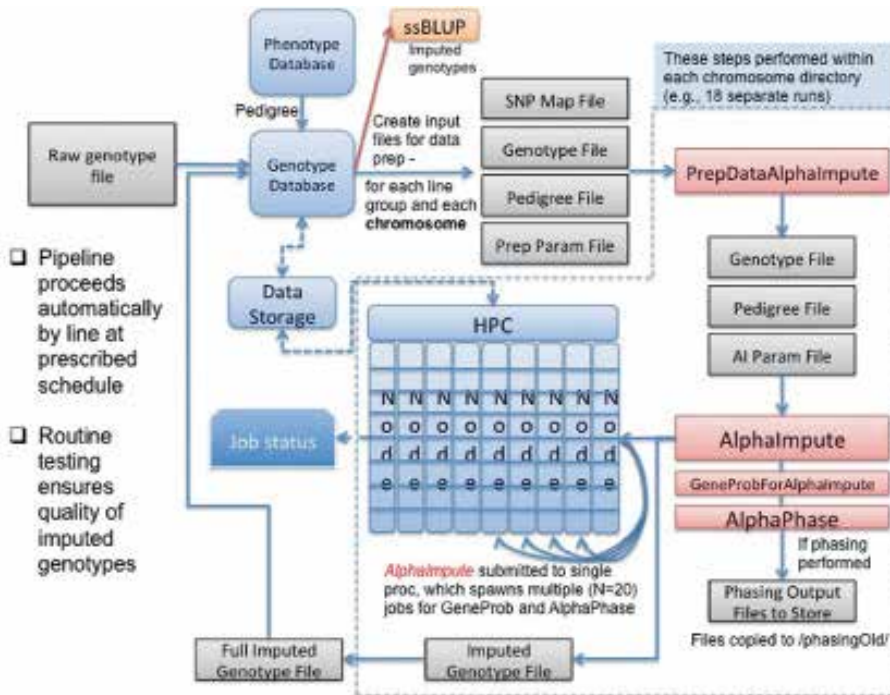
Fig. 5 Imputation pipeline. ssBLUP is single step genomic evaluation (SSE); Data storage is 114TB; HPC is the high performance computing cluster (21 nodes, 216 cores, 1,664 GB RAM); AlphaImpute is the imputation software.

The accuracy of imputation is dependent on the density of the low-density genotyping panel and the level of genotyping in close relatives. Hickey *et al.* (2012) showed that accuracy of imputation, measured as the correlation between true and imputed genotypes, was 0.98 for individuals genotyped at 1% of the total SNPs (representing a panel of ~600 SNP), where both parents were genotyped at high density. Pigs with only one parent and one grandparent genotyped had accuracies that still exceeded 0.93. Huang *et al.* (2012) tested a subset of the same population and reported imputation accuracies exceeding 0.96 for a panel of 384 SNPs when parents and grandparents were genotyped at high density. Accuracies decreased with declining levels of genotyping of close relatives but remained above 0.90 when at least sires and grandsires were genotyped. In production, imputation accuracies generally exceed 0.95 when both parents are genotyped and 0.90 when sires and grandsires are genotyped (Table 3).

Table 3. Accuracy of imputation measured as the correlation between true and imputed genotypes, for animals in three PIC pig lines genotyped on a low-density 500 SNP panel, given alternative levels of high-density genotyping (40k SNPs filtered from the PorcineSNP60 BeadChip) in parents and grandparents.

	Accuracy		
Line	Both Parents*	Sire/MGS	Sire
A	0.95	0.90	0.88
B	0.97	0.91	0.91
C	0.92	0.89	0.91

*Both Parents = both parents were genotyped at high density; Sire/MGS = sire and maternal grandsire were genotyped at high density (but not dam); Sire = sire was genotyped at high density (but not grandsire or dam)

The next stage of innovation will involve imputing sequence variants from a mixture of high- and low-density information. The current imputation approaches will need to be modified to properly use the available information and to address the computational challenges inherent in handling much larger datasets than are routine today. Additionally, few full sequences are available for livestock species to perform the testing that will be needed to put these new imputation approaches into practice. With decreasing sequencing costs and increasing computing power, however, these challenges will be overcome and low-cost (imputed) sequence will be available for large numbers of animals to implement into genomic evaluation programs.

New methods for Genomic Selection (GS)

Every methodology used for genetic evaluation is based on assumptions about how genes affect different traits in a population. These assumptions are established based on our knowledge of how genes interact with each other and influence important traits. The availability of large-scale DNA information in the last few years has been modifying our knowledge about the genes and their mode of action at a very fast pace. The more we use genomics information, the more we learn about genes' behavior and, consequently, we can update the assumptions implicit in our methodology for genetic evaluation.

The expectation is that in the next few years we will probably update and improve the statistical methods used to incorporate genomic information into the calculation of breeding values based on new knowledge acquired through research with genomic data. A potential improvement in statistical methods may come with the use of nonparametric statistics. Nonparametric statistics comprise a set of techniques that do not assume a fixed structure for a model. Therefore, it allows the building of statistical models with very flexible assumptions about how genes affect different traits. Nonparametric methods are commonly used in several fields of engineering and computer science, and are the basis of artificial intelligence.

In a partnership with the Statistical Genetics group of the University of Wisconsin – Madison, Genus researchers are currently investigating the potential of several nonparametric methods to improve the prediction of genetic differences based on genomic data. These procedures have the potential to capture more complex effects of multiple genes, effects of gene-by-gene interactions, and effects of gene-by-environment interactions that are not considered in current SSE but are present in the crossbred populations we select to improve. Initial results of the project suggested that nonparametric methods might be particularly helpful predicting genetic differences leading to better performance of crossbred sows. Over the next two years we will learn how to use these statistical techniques to sharpen our ability to exploit genomic information, and to further focus on the improvement of prediction of genetic differences that have a high impact on commercial performance.

Breeding has relied heavily on only additive inheritance for many years and thus linear models including only genomic additive effects have usually been used to analyze quantitative traits. However, they may not be the most appropriate models when interactions or non-linearity between genotypes and phenotypes are present. Nonlinear models have been proposed as an alternative to enhance the prediction of complex traits (Gianola & van Kaam, 2008). Nonlinear methods can capture complex signals from data and can improve predictive accuracy, even in linear systems. Tussell *et al*. (2012) evaluated the performance of several models including genomic information to predict litter size phenotypes in three pig populations and concluded that nonlinear neural networks were the best method to predict future performance of crossbred populations (Table 4).

Table 4. Average correlations (standard deviation in parentheses) obtained between predicted and observed litter size phenotypes in a 10-fold cross-validation in two purebred and crossbred pig populations (Tussell *et al*., 2012).

| Model	Line 1	Line 2	Crossbred	Average of the 3 datasets
Ridge Regression	0.187 (0.058)	0.234 (0.072)	0.272 (0.058)	0.231
Bayesian LASSO	0.189* (0.059)	0.239* (0.074)	0.274 (0.055)	0.234
Genomic BLUP	0.186 (0.058)	0.231 (0.072)	0.263 (0.068)	0.227
Kernel-Hilbert Spaces**	0.188 (0.051)	0.231 (0.071)	0.276 (0.058)	0.232
Neural Networks**	0.174 (0.074)	0.222 (0.074)	0.291 (0.068)	0.229
Regularized Neural Networks**	0.032 (0.112)	0.234 (0.064)	0.312*† (0.068)	0.193
Pedigree BLUP	0.111 (0.048)	0.132 (0.061)	0.250 (0.065)	0.164

*Best predictability within line data set
†Best predictability among models and lines
**Nonparametric models

Reproduction under PRRS and Genomic Selection (GS)

Porcine Reproductive and Respiratory Syndrome (PRRS) is an infectious viral disease in pigs characterized by reproductive failure in sows and respiratory distress in growing pigs. A 2011 US Pork Checkoff study (http://www.pork.org/ResearchDetail/1499/AssessmentoftheEcono. aspx 2012) estimated the disease costs the U.S. industry $664M annually, and PRRS has similar economic impacts on pig health and performance in all major pork producing countries.

The effect of PRRS virus (PRRSv) infection usually is very acute, causing a dramatic increase in abortions, mummified piglets, still-births and reduction in number born alive (Deeb *et al.*, 2012). A recent study (Boddicker *et al.*, 2012a), where 600 nursery aged pigs were challenged with PRRSv, revealed moderate heritability for both viremia and growth following infection, and led to the discovery of genomic regions associated with the host response to PRRS virus infection (Figure 6). Since then, these regions have been validated in other unrelated populations (Boddicker *et al.*, 2012b).
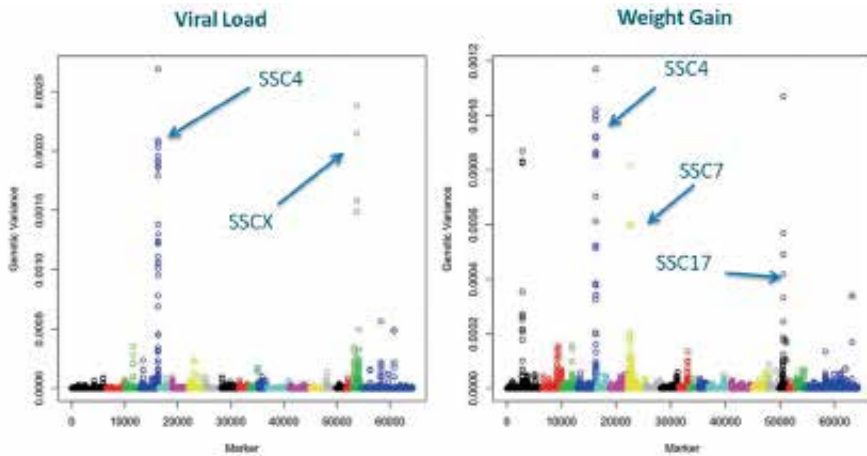
Fig. 6 A 33 SNP region of chromosome 4 (SSC4) explained 15.7% of genetic variance for viral load and 11.2% for weight gain in 600 PIC pigs (after Boddicker *et al.*, 2012a).

These findings are the first steps to implement GS to reduce the impact of PRRSv on sow reproduction and growing pigs. The distinction in this case will be that routine phenotypes from performance testing will not be available and combined with GEBVs for resilience. GEBVs developed from GWAS in discovery (training) data sets will have to be evaluated under field conditions and continually fine tuned with ongoing challenge and field data set collection.

Next Generation Sequencing (NGS) technology has revolutionized the field of genomics research and introduced new tools to animal breeders that up to this point were too expensive and too technologically challenging. This combined with the recently published draft of the pig genome sequence (Groenen *et al.*, 2012) and availability of phenotypes and tissue samples on thousands of animals, often with complete pedigree, form a very powerful tool to discover genomic regions, genes and causative mutations that can lead to a swift and effective selection against PRRSv infection. PIC is involved in several projects using RNAseq technology and targeted resequencing within known genomic regions and candidate genes associated with immune response in general or with PRRS response. The discovery of functional genes and networks of genes can also be the target for a more biologically direct approach through the use of genome-editing tools such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs, Marx, 2012). All in all, we believe the next five years will see many technological advances in the battle against PRRS which will lead to a significant improvement in animals' resilience to PRRSv infection, lead to a significant improvement in sows' reproductive performance, and eventually lead to the eradication of PRRS.

Genetic modification in pig breeding

Transgenic technology was first applied to livestock species in the late 1980s and was mainly focused on aspects of growth (see Pursel *et al.*, 1987). Guthrie *et al.*, (2005) tried to improve ovulation rate in pigs by expressing the human BCL-2 gene in ovaries, but no improvement was seen. The major drawbacks of transgenesis in livestock were that DNA sequences from other organisms were used; transgene insertion was random, which made transgene expression unpredictable; integration often involved the insertion of multiple copies of the transgene and selectable markers, usually drug resistance, were also standard. These factors were viewed negatively by consumers.

Modern techniques in genetic modification offer the ability to make precise changes in the genomes of livestock species. In particular genome editing using either zinc finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs) allows precisely targeted DNA sequences to be altered in livestock without the addition of any foreign DNA, or the use of selectable markers. Whyte *et al.*, (2011) demonstrated ZFNs could be used in pigs by knocking out an eGFP transgene. Carlson *et al.*, (2012) used TALENs to create minipigs containing mono- and biallelic mutations of the LDL receptor gene as medical models for familial hypercholesterolemia. A collaboration ultimately aimed at producing ASFv resistant pigs involving PIC, The Roslin Institute, and Recombinetics Inc., has recently produced live pigs with deletions in the RELA gene following direct injection of ZFN and TALEN mRNA into pig zygotes (B. Whitelaw, 2012, personal communication).

Genome editing has the potential to introgress beneficial alleles from rare or indigenous breeds which are not present in improved commercial lines, as well as increasing the expression of beneficial genes through directed gene duplication. Recent work in a mouse model shows that expression of porcine FSH, in addition to the mouse's own FSH genes, led to increased ovulation rate and litter size (Bi *et al.*, 2012). This group is now looking to see if adding an extra copy of the pig FSH genes, using the same construct, would increase litter size in pigs (Ning Li, 2012, personal communication).

Conclusion

Genomic selection (GS) in pigs has evolved from simple *ad hoc* utilization of single gene and quantitative trait loci (QTL) tests in the 1990s through initial genome wide applications enabled by availability of the PorcineSNP60 BeadChip in 2009 through full integration into single step multivariate breeding value estimation today (Table 5). As such, it has become something of a unifying theory for animal improvement, increasingly blurring past distinctions between performance testing *vs.* progeny testing. A genomic relationship matrix that can distinguish among animals with the same parents at birth by their genetic code, now coupled with the necessary computing tools and datasets, allows breeders to rapidly predict a specific animal's genetic potential (Forni *et al.*, 2011).

It is important to understand GS is not a case of "out with the old, in with the new". There was perhaps a naive notion within the animal improvement industry a decade ago that GS might eliminate the need for performance and progeny testing in livestock improvement. The reality is the genomic information is most beneficially used alongside continually updated abundant, accurate, 'real-world' performance data. Livestock industries continue to invest in recording animal performance, in pig breeding more so than ever today, as phenotypes become increasingly valuable for genomic "training" (estimation of SNP effects).

The full benefits to the pork supply chain of the latest genomics technology breakthroughs are still two to three years away, due to genetic lag. New genotyping technologies and competition among service providers will continue to reduce the cost of genotype data collection, with the result that GS will become an increasingly cost effective and powerful tool with which to obtain productivity improvement in animals of all species.

Agriculture and society have entered a critical phase as the global population grows while availability of land and fresh water for agriculture diminishes (Zering *et al.*, 2012). By the year 2050 it is predicted the world population will grow from 7 billion people today to 9.1 billion, the per capita income will rise by 150%, and global consumption of meat, milk and eggs will double (*FAO, 2006*).

GS will make a significant contribution, but application of GS alone will not meet growing world population needs for animal protein. Rapidly developing genome sequencing and editing

Table 5. Incorporation of genomic research into swine breeding practice.

Years	Technology	Contribution	References
1991-present	Single gene tests; candidate genes; QTL (quantitative trait loci); MAS (marker assisted selection)	Targeted *ad hoc* applications, e.g., elimination of the deleterious HAL-1843® stress allele; fixing coat color and F18 E. coli resistance alleles; quantitative trait (litter size, meat quality, feed intake, growth) applications	Dekkers, 2004
2009-2012	Genome wide association studies (GWAS); Genome assisted selection (GAS)	Population-trait specific genomic estimated breeding values (EBVs) blended with classical pedigree EBVs to form a single EBV for selection	Deeb *et al.*, 2011
2012-present	Genomic selection (GS); single step genomic evaluation (SSE); imputation (Im)	Simultaneous multi-trait genetic evaluations using both the additive relationship (pedigree) and genomic relationship matrices to achieve a 20-40% increase in overall rates of selection index improvement	Forni *et al.*, 2011; Cleveland *et al.*, 2012; Hickey *et al.*, 2012
Future	Non parametric statistical methods; Genomic breeding values for disease, e.g., PRRS tolerance; Genome editing to create disease resistance	To be determined	Boddicker *et al.*, 2012a,b; Carlson *et al.*, 2012; Deeb *et al.*, 2012; Tussell *et al.*, 2012

technologies, combined with greater understanding of the functional genomics of reproductive and other traits of economic importance, offer significant promise for further advances.

References

Aguilar I, Misztal I, Johnson D, Legarra A, Tsuruta S & Lawlor T 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **93** 743-752.

Beuzen ND, Stear MJ & Chang KC 2000 Molecular markers and their use in animal breeding. *The Veterinary Journal* **160** 42-52.

Bi M, Tong J, Chang F, Wang J, Wei H, Dai Y, Chu M, Zhao Y & Li N 2012 Pituitary-specific overexpression of porcine follicle-stimulating hormone leads to improvement of female fecundity in BAC transgenic mice. *PLoS One* **7(7)** e42335.

Bidanel JP 2011 Biology and genetics of reproduction. In *The Genetics of the Pig*, edn 2, pp 218-241. Eds MF Rothschild and A Ruvinsky. CABI Publishing.

Bidanel JP & Rothschild MF 2002 Current status of quantitative trait locus mapping in pigs. *Pig News and Information* **23(2)** 39N-53N.

Boddicker N, Waide EH, Rowland RR, Lunney JK, Garrick DJ, Reecy JM & Dekkers JC 2012a Evidence for a major QTL associated with host response to porcine reproductive and respiratory syndrome virus challenge. *J Anim Sci.* **90(6)** 1733-46.

Boddicker NJ, Lunney JK, Rowland RRR, Garrick DJ, Reecy JM, & Dekkers JCM 2012b Genetic basis of host response to PRRSV infection. *Proc 2012 International PRRS Symposium/NSIF Conference:*20 (Abstr.)

Browning SR & Browning BL 2007 Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81** 1084-1097.

Carlson DF, Tan W, Lillico SG, Stverakova D, Proudfoot C, Christian M, Voytas DF, Long CR, Whitelaw CB & Fahrenkrug SC 2012 Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. USA* **109(43)** 17382-7.

Casey D, Perez M, McLaren D & Short T 2006 Crossbred Breeding Values: Selecting for Commercial Performance in Pigs. *Proc. 8th World Congr. Genet. Appl. Livestock Prod* Belo Horizonte, Minas Gerais, Brazil. **06**-26

Christensen OF & Lund MS 2010 Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* **42** 2.

Christensen OF, Madsen P, Nielsen B, Ostersen T & Su G 2012 Single-step methods for genomic evaluation in pigs. *Animal* **10** 1-7.

Cleveland MA & Deeb N 2012 Selecting markers and evaluating coverage. In *Quantitative Trait Loci: Methods and Protocols*, pp 55-71. Ed SA Rifkin**.** New York: Humana Press**.**

Cleveland MA, Yu N, Foertter F, Deeb N, & Hickey JM 2012 An imputation pipeline for cost-effective genomic selection in commercial pig breeding. *63rd Annual Meeting of the European Association for Animal Production, Bratislava, Slovakia* **18** 232 (Abstr.).

Cole JB, Newman S, Foertter F, Aguilar I, & M. Coffey M 2012 Really big data: Processing and analysis of very large data sets *J Anim Sci.* **90**:723-733.

Daetwyler HD, Wiggans GR, Hayes BJ, Wooliams JA & Goddard ME 2011 Imputation of missing genotypes from sparse to high density using long-range phasing. *Genet.* **189** 317-327.

Deeb N, Cleveland MA & Forni S 2010 Linkage disequilibrium decay in commercial pigs. *Plant and Animal Genome XVIII Conference, San Diego, CA USA* **P602** (Abstr.).

Deeb N, Cleveland MA, Mileham AJ, Gladney C, Yu N, Sebbana S, Wang L & Southwood OI 2011 Genomic selection using small panels. *Plant and Animal Genome XIX Conference, San Diego, CA, USA* **P606** (Abstr.).

Deeb N, Cleveland MA & Forni S 2012 Genomic selection in pigs: are we there yet? *International Plant and Animal Genome XX Meeting, San Diego, CA, USA* **W698** (Abstr.).

Dekkers JCM 2004 Commercial application of marker-and gene-assisted selection in livestock: Strategies and lessons. *J. Anim. Sci.* **82** E313-E328.

Dekkers JCM, Mathur PK & Knol EF 2011 Genetic improvement of the pig. In *The Genetics of the Pig*, edn 2, pp 390-425. Eds MF Rothschild and A Ruvinsky. CABI Publishing.

Druet T & Georges M 2010 A Hidden Markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genet.* **184** 789-798.

Food and Agriculture Organization of the United Nations (FAO) 2006 Livestock's Long Shadow. FAO, Rome. http://www.fao.org/docrep/010/a0701e/a0701e00.htm

Flint J & Mackay TF 2009 Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19(5)** 723-33.

Forni S, Aguilar I & Misztal I 2011 Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* **43** 1.

Fujii J, Otsu K, Zorzato F, De leon S, Khanna VK, Weiler JE, O'Brien PJ & MacLennan DH 1991 Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **25** 448-451.

Gianola D & van Kaam BCHM 2008 Reproducing kernel Hilbert spaces regression methods for genomic prediction of quantitative traits. *Genetics* **178** 2289-2303.

Goddard ME 2009 Genomic selection: prediction of accuracy and maximisation of long term response. *Genetics* **136** 245-257.

Goddard ME & Hayes BJ 2007 Genomic selection. *J. Anim. Breed. Genet.* **124(6)** 323-330.

Goddard ME & Hayes BJ 2009 Mapping genes for complex traits in domestic animals and their use in

breeding programmes. *Nature Reviews Genetics* **10** 381-391.

Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ *et al.* 2012 Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491** 393–398.

Guthrie HD, Wall RJ, Pursel VG, Foster-Frey JA, Donovan DM, Dawson HD, Welch GR & Garrett WG 2005 Follicular expression of a human beta-cell leukaemia/lymphoma-2 (Bcl-2) transgene does not decrease atresia or increase ovulation rate in swine. *Reprod. Fertil. Dev.* **17(4)** 457-66.

Harris DL & Newman S 1994 Breeding for profit: Synergism between genetic improvement and livestock production (A review). *J. Anim. Sci.* **72** 2178-2200.

Hayes BJ, Bowman PJ, Chamberlain AJ & Goddard ME 2009 Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* **92** 433-443.

Henderson CR 1975 "Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics* **31**(2) 423–448.

Hickey JM, BP Kinghorn, B Tier, JHJ van der Werf & MA Cleveland 2012 A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genetics Selection Evolution*. **44**:9.

Howie BN, Donnelly P & Marchini J 2009 A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5(6)** e1000529.

Huang Y, Hickey JM, Cleveland MA & Maltecca C 2012 Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genetics Selection Evolution* **44** 25.

Knap PW & Su G 2008 Genotype by environment interaction for litter size in pigs as quantified by reaction norms analysis. *Animal* **2** 1742-1747.

Lillehammer M, Meuwissen THE & Sonesson AK 2011 Genomic selection for maternal traits in pigs. *J. Anim. Sci.* **89** 3908–3916.

Lynch M & Walsh B 1998 Genetics and Analysis of Quantitative Traits. Edn 1. Sinauer Associates Inc.

Marchini J & Howe B 2010 Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11** 499-511.

Marx V 2012 Genome-editing tools storm ahead. *Nature Methods* **9** 1055–1059.

MacLennan DH & Phillips MS 1992 Malignant hyperthermia. *Science* **256** 789-794.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP & Hirschhorn JN 2008 Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* **9(5)** 356-69.

McLaren D 2007 Recent developments in genetic improvement of pigs. *Proc. Manitoba Swine Seminar* **21** 131-149.

Merks JWM 2001 Genetic improvement at the commercial level compared to genetic progress at the nucleus level. Proc.National Swine Improvement Federation Conference. NSIF **26.**

Messer LA, Wang L, Yelich J, Pomp D, Geisert RD & Rothschild MF 1996 Linkage mapping of the retinol-binding protein 4 (RBP4) gene to porcine chromosome 14. *Mammalian Genome* **7(5)**:396.

Meuwissen THE 2009 Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* **41** 35.

Meuwissen THE, Hayes BJ & Goddard ME 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157** 1819–1829.

Misztal I 1999 REMLF90 Manual http://nce.ads.uga.edu/~ignacy/numpub/ blupf90/docs/remlf90.pdf

Misztal I, Legarra A & Aguilar I 2009 Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science* **92** 4648-4655.

Nakavisut S, Crump R, Suarez M & Graser HU 2005 Genetic correlations between the performance of purebred and crossbred pigs. *Proc. Assoc. Advmt. Anim. Breed. Genet.* **16** 99-102.

Onteru SK, Fan B, Du Z-Q, Garrick DJ, Stalder KJ & Rothschild MF 2012 A whole genome association study for pig reproductive traits. *Anim. Genet.* **43(1)** 18-26.

Onteru SK, Fan B, Nikkila MT, Garrick DJ, Stalder KJ & Rothschild MF 2012 Whole genome association analyses for lifetime reproductive traits in the pig. *J. Anim. Sci.* **89** 988-995.

Perez M, Casey D & McLaren D 2006 Crossbred breeding values: Selecting for commercial performance. *Proc. 37th Ann. Meet. Amer. Assoc. Swine Vets* 83-84.

Powell JE, Visscher PM & Goddard ME 2010 Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews: Genetics* **11** 800–805.

Pursel VG, Rexroad CE Jr, Bolt DJ, Miller KF, Wall RJ, Hammer RE, Pinkert CA, Palmiter RD & Brinster RL 1987 Progress on gene transfer in farm animals. *Vet Immunol Immunopathol.* **17(1-4)**:303-12.

Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P *et al* 2009 Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE* **4(8)** e6524.

Rothschild M, Jacobson C, Vaske D, Tuggle C, Wang L, Short T, Eckardt G, Sasaki S, Vincent A, McLaren D, Southwood O, van der Steen H, Mileham A & Plastow G 1996 The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proc Natl Acad Sci USA.* **93(1)**:201-5.

Scheet P & Stephens M 2006 A fast and flexible statistical model for large-scale population genotype data: application to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78**:629-644.

Tussel L, Perez P, Forni S, Wu XL & Gianola D 2012 A comparison of methods for predicting litter size in

commercial pig lines. *J. Anim. Sci.* **90 Suppl. 3** 721 (Abstr).

van der Steen HAM, Prall GFW & Plastow GS 2005 Application of genomics to the pork industry. *J. Anim. Sci.* **83** E1-E8.

VanRaden PM 2008 Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91** 4414-4423.

VanRaden PM, Van Tassell, GR Wiggans, TS Sonstegard, RD Schnabel, JF Taylor & FS Schenkel 2009 Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. **92** 16-24.

Van Raden PM, O'Connell JR, Wiggans GR & Weigel KA 2011 Genomic evaluations with many more genotypes. *Genetics Selection Evolution* **43** 10.

Vincent AL, Wang L, Tuggle CK, Robic A, Rothschild MF 1997 Prolactin receptor maps to pig chromosome 16. *Mammalian Genome* **8(10)** 793-4.

Whyte JJ, Zhao J, Wells KD, Samuel MS, Whitworth KM, Walters EM, Laughlin MH & Prather RS 2011 Gene targeting with zinc finger nucleases to produce cloned eGFP knockout pigs. *Mol Reprod Dev.* **78(1)** 2.

Zering KD, Centner TJ, Meyer D, Newton GL, Sweeten JM & Wooruff S 2012 Water and Land Issues Associated with Animal Agriculture: A U.S. Perspective. *Council for Agricultural Science and Technology (CAST) Issue Paper* 50. http://www.cast-science.org/publications/?water_and_land_issues_associated_with_animal_agriculture_a_us_perspective&show=product&productID=261302