RESEARCH REPORT
ANIMAL SCIENCE

Title: Increased Accuracy of Selection for Nutrient Utilization in Duroc Pigs by Application of Genomic Tools – NPB #11-076

Revised

Investigator: Joseph P Cassady

Institution: North Carolina State University

Date Submitted: 4-1-2013

Industry Summary

The pig industry continues to face many challenges with regard to costs of production. Feed costs continue to be the largest variable cost. One way of reducing feed costs is by increasing the efficiency with which pigs convert feed into saleable product. This can be done by altering the composition of the diet or selecting pigs with superior genetic potential for efficient use of nutrients. Here we report results from a project designed to use the latest advances in genomic technology to improve the accuracy with which we identify pigs having superior genetic ability to utilize nutrients. Recent advances in molecular genetics have made it possible for geneticists to improve the accuracy of genetic prediction. Utilization of genetic markers to analyze economically important traits has provided opportunities to identify chromosomal regions harboring genes influencing those traits. In some cases this has led to the discovery of the causal mutations within genes which are responsible for the change in performance. The ProcineSNP60 BeadChip allows scientists to genotype a pig for 60,000 genetic markers in a single assay for less than $100 per animal. This provides a wealth of information regarding the genetic makeup of each pig. The objective of this study was to use genomic information to identify genomic regions associated with feed efficiency and production traits, including average daily gain (ADG), average daily feed intake (ADFI), feed conversion ratio (FCR), residual feed intake (RFI), ultrasound back fat thickness (BF), muscle depth (MD), inner muscular fat percentage (IMF), birth weight (BW), and weaning weight (WW) in purebred Duroc boars. Individual feed intake records on more than 1000 boars were collected from electronic feeders. While the PorcineSNP60 BeadChip does include 60,000 markers, not all markers will be useful in every experiment. For our project a total of 40,008 markers proved useful and were evaluated. Three different statistical strategies were utilized to identify the chromosomal regions of interest. The reason for using multiple approaches is to obtain the best information available. Each approach has different strengths and weaknesses.

Several important genomic regions were identified in this study, associated with the analyzed traits. A region on pig chromosome 1 was found which explained a large proportion of the genetic variance for both ADG and ADFI. Additional regions with smaller effects were identified on chromosomes 4, 6, 8, 14 and X. The genetic architecture of the analyzed traits was characterized by a limited number of genes or genomic regions with large effects and many regions with smaller effects. The region on chromosome 1 might be used to improve ADG and decrease ADFI in pigs.

To add confidence to our results, we obtained feed intake data on an additional 504 Duroc boars from the same population. Those boars were genotyped with GGP-Porcine BeadChip (8,826 markers) which has recently become commercially available. A procedure known as "imputation" was used to predict the 60k genotypes based on the GGP-Porcine BeadChip genotypes. Imputation has been shown to be highly accurate when done correctly. The advantage of this approach is that pigs can be genotyped using the GGP-Porcine BeadChip for less than one-half the cost of the Porcine60K BeadChip. After imputation, 35,871 genotypes were obtained for each of the 504 boars typed by the GGP-Porcine BeadChip. Inclusion of the data from the additional 504 boars supported the conclusions from the original analysis that chromosomal regions affecting nutrient utilization can be identified.

In conclusion, genomic regions influencing economically important traits associated with nutrient utilization were identified. Inclusion of genomic marker data in the estimation of breeding values for average daily feed intake and average daily gain is expected to improve the accuracy of genetic prediction for these traits. Improved accuracy of prediction should lead to faster genetic improvement.

Learning outcomes:

- Regions within chromosomes influencing nutrient utilization were identified.
- Genetic markers can be used to improve the accuracy with which we predict a pigs genetic potential for average daily feed intake and average daily gain which are critical components of nutrient utilization.
- Improved accuracy of prediction would be expected to result in more accurate genetic selection and faster genetic improvement.

Funding, wholly or in part, was provided by The National Pork Board, the National Swine Genome Coordinators Office, Smithfield Premium Genetics, and the North Carolina Agriculture Research Service.
Keywords: Nutrient Utilization, Genomics, Pigs, Accuracy, Selection

Scientific Abstract

Efficiency of producing saleable products in the pork industry is largely determined by feed costs and by the amount and quality of lean meat produced. The objective of this study was to identify genomic regions associated with variation in feed efficiency and several related production traits in a Duroc terminal sire population. Traits studied were ADFI (Average daily feed intake), ADG (Average daily gain), FCR (Feed conversion ratio), RFI (Residual feed intake), ultrasound back fat (BF), muscle depth (MD), inner muscle fat (IMF), birth weight (BW) and weaning weight (WW). Individual feed intake and serial pig body weights were recorded using the FIRE system on 1047 individuals and edited by linear mixed model to adjust errors of individual visits to the feeders and subsequent body weight measures were estimated from both linear and robust regression. Heritabilities were 0.50 ± 0.102, 0.32 ± 0.096, 0.08 ± 0.057, 0.10 ± 0.054, 0.59 ± 0.086, 0.39 ± 0.087, 0.54 ± 0.011, 0.65 ± 0.10 and 0.64 ± 0.11 for ADG, ADFI, FCR, RFI, BF, MD, IMF, BW and WW estimated using animal model by ASReml3.0 and were used as prior for following analysis. All phenotypes were pre-adjusted for fixed effects using linear models before GWAS. Genotyping was performed using Illumina PorcineSNP60K BeadChip. After quality control 40,008 SNPs remained for analysis. Single trait association analyses were performed using a Bayes-B model. The proportion of phenotypic variance explained by markers was 0.182 for ADFI, 0.278 for ADG, 0.187 for FCR, 0.054 for RFI, 0.506 for BF, 0.348 for MD, 0.267 for IMF, 0.241 for BW and 0.370 for WW. Significant regions were

identified by using three different significance tests; posterior windows variance, Bayes factor and bootstrapping. Significance was declared for regions where test's significance overlapped. Regions associated with ADFI were mapped to SSC 1 and 10; with ADG mapped to SSC 1, 4, 11, and 14; FCR on SSC 4, BF on SSC 1, 14 and 18, MD on SSC 13 and 23, WW on SSC 8, 16, 17 while no QTL were identified for RFI, IMF and BW. Genes associated traits within the informative QTL regions were annotated. In conclusion, we have identified several genomic regions associated with traits affecting nutrient utilization that could be considered for future genomic prediction to improve feed utilization.

Introduction

Feed costs represent the largest variable cost in pork production. Therefore, nutrient utilization is a trait of primary economic importance to the swine industry. Substantial differences in feed intake exist and ~30% of these are explained by genetics (Lo, McLaren et al. 1992; Cai, Casey et al. 2008; Hoque and Suzuki 2008; Hoque, Katoh et al. 2009). Feed efficiency of the growing pigs has been selected and improved through indirect selection of pigs with faster growth rate and relatively less feed intake to achieve maximum profit. Thus, average daily gain (ADG) and average daily feed intake are vital components of nutrient utilization traits. Residual feed intake (RFI) and feed conversion ratio (FCR) are anther two component traits of feed efficiency. RFI represents the difference between the observed feed intake and predicted feed intake based on requirements for maintenance and growth. High RFI corresponds to higher feed intake for a given production level. While FCR is a measure of feed efficiency, kg of product/ kg of feed (equals to ADG /ADFI). It has been previously demonstrated that traits contributing to feed efficiency are heritable and genetically related to economically important traits such as back fat. Therefore, genetic selection for improved feed utilization is expected to significantly increase efficiency of production.

Although feed efficiency has improved to some degree by selection for growth and reduced backfat, further improvements require direct selection on feed intake. This is, however, complicated by the difficulty and expense of recording feed intake on large numbers of animals, but possible if the genes that are responsible for differences in feed intake and feed efficiency are known.

Nearly two decades of advances in the QTL mapping of complex economic important traits, including feed utilization traits, growth traits, production traits, etc., based on porcine genetic linkage map and discovery of useful candidate genes have led to new breeding programs which incorporate molecular technologies. Leading up to 2007, 1675 QTLs were identified from 110 publications (Rothschild, Hu et al. 2007) for economically important traits in pigs. Genomic information enables the swine industry to utilize marker information to conduct marker-assisted selection along with traditional performance information to improve traits of economic values.

The recent availability of high-density single nucleotide polymorphism (SNP) panels has advanced such studies by capitalizing on population-wide linkage disequilibrium at positions across the genome. The availability and accessibility of the commercially released Porcine60K BeadChip has greatly facilitated whole-genome association studies, contributing to increased accuracy of selection by application of Marker assisted selection or genomic selection.

Objectives:

The objective of this project was to identify opportunities for improving growth rate, nutrient utilization, and survival in a Duroc terminal sire population. The objective was accomplished by doing a genome wide association study (GWAS) utilizing the ProcineSNP60 BeadChip and established statistical methodologies. The ProcineSNP60 BeadChip allows scientists to document a pig's genotype for more than 60,000 genetic markers spread across the entire pig genome. Combining the genetic marker information with performance data and applying the appropriate statistical techniques allows researchers to identify regions of the pig genome associated with genetic variation in the performance trait of interest. Specific questions addressed were the following:

1. Where are the genomic regions associated with nutrient utilization and its component traits, including feed intake and growth rate?

2. How are other economically important traits, such as pig birth weight (related to pig viability), backfat, loin depth, and IMF, impacted by the genomic regions identified?

3. What proportion of genetic variance in the traits of interest is explained by the genomic regions of interest?

4. Do results from Question #1 lead to "genomic enhanced" EPDs with increased accuracy?

5. Would a smaller SNP Chip (generically referred to as a "reduced" Chip) be equally as effective as the PIG SNP60 Chip? This would be important if the "reduced" Chip resulted in lower genotyping costs.

Materials and Methods

Animal and phenotype data collection

The data came from one Duroc nuclear farm of Smithfield Premium Genetics (Roanoke Rapids, NC), including 1047 Duroc boars from the mating of 64 sires and 421 sows. Individual piglet birth weight and litter information were recorded based on SPG's protocol and body weight were recorded when they were weaned at the mean age of 25 days. Growth and feed intake were measured on each pig, starting at an average age of 85 days and ending after approximately 45 or 90 days. Boars with similar age were grouped into contemporary groups and fed in the FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, KS, USA)electronic feeders. Individual feed intake and body weight were recorded when the pig visited the feeder, in total there are 323,639 individual visits recorded. When the boars reached about 120 Kg, ultrasound back fat, muscle depth and inner muscular fat percentage and test weight were collected for each individual.

Phenotype data editing

Average daily feed intake (ADFI) was obtained by editing 323,639 records collected from the FIRE system. Data from electronic feeders have been found to contain substantial amounts of errors (De Haer, Merks et al. 1992; McDonald and Nienaber 1994; Eissen, Kanis et al. 1998) resulting from feeder malfunctions and animal-feeder interactions. To obtain an accurate prediction of individual feed intake, editing methods are required that efficiently identify and correct errors in data from electronic feeders. In this study, feed intake records were edited based on the method proposed by (Casey, Stern et al. 2005). Database and edit systems developed by Casey et al. (2005) were used.

The main steps in the edit procedures were to **(i)** identify errors in each visit (a feeding event from a pig's entrance into the feeder to its exit) by 16 criteria and count the number of errors of each type for each day. Error frequency (5%) was higher than in previous reports (Casey, Stern et al. 2005; Cai, Casey et al. 2008); **(ii)** compute error-free FI for each pig and day by summing feed consumed in visits without identified errors; **(iii)** estimate the effect of error counts on error-free daily FI by fitting a linear mixed model to error-free daily FI observations with batch as fixed effect, 32 variables created from the 16 error counts, and ADG and BW as covariates, and pig as a random effect;

$$y_{imnp} = B_i + b_1 BW_{mn} + b_2 ADG_m + \sum_{p=1}^{16} b_{3p} ETP_{pmn} + \sum_{p=1}^{2} b_{4p} OTD_{pmn} + \sum_{p=6}^{14} b_{4p} OTD_{pmn} + \sum_{p=4}^{5} b_{5p} FID_{pmn} + \sum_{p=15}^{16} b_{5p} FID_{pmn} + e_{imnp}$$

where

y_{imnp} = DFI_{EF} from the m^{th} pig on the n^{th} day. The m^{th} pig was of the i^{th} batch group.

B_i = fixed effect of the i^{th} batch

P_m = effect of the m^{th} pig, which was assumed random with $P_m \sim N(0, \sigma_P^2)$

$b_1 BW_{mn}$ = linear effect of body weight from the m^{th} pig on the n^{th} day in the test period

$b_2 ADG_m$ = linear effect of ADG over the test period from the m^{th} pig

$b_{3p} ETP_{pmn}$ = linear effect of the percentage of all visits from the m^{th} pig on the n^{th} day that contained error type p

$b_{4p} OTD_{pmn}$ = linear effect of recorded occupation time summed over visits from the m^{th} pig on the n^{th} day that contained the error type p

$b_{5p} FID_{pmn}$ = linear effect of recorded feed intake summed over visits from the m^{th} pig on the n^{th} day that contained the error type p

e_{imnp} = residual with $e_{imnp} \sim N(0, \sigma_e^2)$

(iv) adjust error-free daily FI for each pig and day for feed consumed in error visits by adding estimates of covariates from step; **(v)** compute ADFI for each pig by averaging daily FI during the test period.

Average Daily Gain was calculated using two methods (1) Simple linear regression using weaning weight (~26 days, ~17 kg) and ultrasound test weight (~103 days, ~260kg) of 889 boars, assuming linear growth; (2) Robust regression using 272,248 single pig body weight records from the FIRE system. Body weight records from the FIRE system contained outliers which showed abnormal growth patterns when plotting it against age, previous studies showed that robust regression could be applied to edit this serial pig body weight data (Zumbach et al., 2008; Chen et al., 2010). Two steps were used to edit the body weight data in robust regression. (1) A quadratic growth curve was estimated for each pig assuming small or 0 weights for points far away from the curve using robust regression with bisquare weight function in R. By fitting a robust regression, each data point was assigned a weight (from 0 to 1) to minimize the influence of abnormal data point. (2) Data with weight less than 0.5 were treated as outliers. Animals with less than 20 body weight records or with less than 30 days of information were excluded. The retained body weight records were averaged to create average daily weight and then used to calculate ADG. Average daily gain for 927 boars was obtained by combining the two data sets from the above analysis.

FCR was calculated based on ADFI and ADG (FCR = ADG / ADFI) and RFI was calculated using a linear mixed model (Cai et al., 2008):

$$y_{ij} = TA_{ij} + ADG_{ij} + BF_{ij} + batch_i + e_{ij},$$

Where y_{ij} is ADFI; TA_{ij} is fixed regression covariate of on test age for each boar entered to the FIRE system; ADG_{ij} is fixed regression covariate of ADG; BF_{ij} is fixed regression covariate of BF; $batch_i$ is random effect of contemporary group effect. Measures of RFI for individual pigs were obtained as the residuals from the above mode.

Genetic parameters estimation

Pedigree information on each genotyped animal was traced back 3 generations, a total of 2593 individuals were included in the pedigree file. Genetic and residual (co)variances for the 9 traits were estimated using animal models in ASReml (Gilmour et al., 2009).

$$Y_{ijkl} = batch_i + litter_j + parity_k + pig_{ijkl} + e_{ijkl}$$

Where Y_{ijkl} is BW; $batch_i$, $litter_j$, and $parity_k$ are fixed effects for batch, litter size, and sow parity; pig_{ijkl} is random pig effects; and e_{ijkl} is model residual. Weaning weight was analyzed with the above model with weaning age included as an additional covariate. ADG, ADFI and FCR were analyzed with above model with litter and parity removed from the model while BF, MD and IMF were analyzed adding ultrasound test weight as covariate. RFI was analyzed only including mean as fixed effect since RFI was obtained from the model with all fixed effects included.

Genotyping

Genomic DNA samples of 1037 boars were used for the Illumina PorcineSNP60K BeadChip genotyping. The SNP call rates ≤ .90, MAF (minor allele frequency) ≤ .002, individual call rate ≤ .90, and p-value $<$ .0001 for a chi-square test for Hardy-Weinberg equilibrium were excluded from the genotype data set. After quality control was completed 40,008 SNPs remained. Missing SNP genotypes were imputed for all available boars (n=1022) with pedigree information from a Duroc terminal sire population using AlphaImpute1.1.0. A total of 40,008 out of 64,232 SNP were qualified for association analysis, including 3540 SNPs on autosomes and X chromosome.

Genome-wide association analysis

Four feed efficiency traits and 5 production traits including ultrasound BF, MD, and IMF were pre-adjusted for fixed effects and covariates before conducting GWAS. The association analysis was implemented separately for each feed efficiency trait or production trait with the Bayes B model averaging approach described by (Kizilkaya, Fernando et al. 2010) using GenSel v4.0 software (http://bigs.ansci.iastate.edu). The following statistical model was used for single marker regression:

$$y = X\beta + Zu + e,$$

where, y is the vector of pre-corrected phenotypes, X is an incidence matrix of fixed effects (β), Z is a matrix of SNP genotypes that were fitted as random effects (u), and e is the vector of random residual effects assumed to be normally distributed, $N(0, \sigma_e^2)$. The fixed factors used in this statistical model were population mean. Individual SNP effects were estimated from a mixed model with a probability of 0.995 that any SNP would have a zero effect such that approximately ~400 SNPs were fitted per iteration of each Markov chain. A total of 100,000 Markov iterations, with a burn-in of 30,000 iterations, were run for the analyses. The results from these analyses included posterior distributions for the effects of each of 40,008 markers. Each single marker effect with genetic variance explained by the marker was estimated through the genome.

Windows of 1 Mb length of non-overlapping adjacent SNPs on autosomes and X chromosome were

constructed based on the physical map (Build 10.2), a total of 1780 windows was obtained. For each window and each 50th iteration of the chain, sampled values for the effects of the SNPs in the window were then used to compute a sample of the posterior distribution of the true breeding value for that window for each individual by multiplying the sampled SNP effects with the individual's SNP genotypes and summing across all SNPs in the window. The variance across individuals of the resulting sample breeding values ascribed to each window (window variance) was then used to provide a sample of its posterior distribution (Wolc et al., 2012). Resulting sample window variances were divided by the variance explained by all SNPs across the genome in that iteration of the MCMC chain to convert window variances to proportions of genetic variance explained by the window. The proportion of genetic variance by each window was ordered by size and windows with larger effects (exceed 0.1%) were selected as putative QTL regions to conduct the following bootstrapping analysis.

Significant testing

Marker significance was tested using 3 approaches, (1) windows variance and probability of inclusion from windows analysis. (2) Bayes factor for both single marker and windows. (3) Bootstrapping hypothesis testing.

Windows that captured more than 0.1% of genetic variance in over 50% of the samples were declared to explain significantly more variance than expected.
The Bayes Factor (BF) was calculated for each SNP by:

$$BF = \frac{\Pr(H1|y)\,(1-\pi)}{[1-\Pr(H1|y)]\pi}$$

Where H1 is the hypothesis that the marker is linked to a QTL, Pr(H1|y) is the posterior probability of the hypothesis and π is the prior probability of the hypothesis. (1 - Pr(H1|y)) and (1 - π) represent, the posterior and prior probability for the alternative hypothesis, respectively. Bayes factor for single markers from the association analysis and 1 Mb non-overlapping windows of adjacent markers were calculated and markers and windows with extremely strong evidences of association were defined as significant markers and windows.

The significance level of the putative candidate genomic regions was also estimated using bootstrap analysis (only 4 traits have been done). Bootstrap samples were produced using the posterior means of the 40,008 SNPs to construct the distribution of the test statistic (genetic variance of 1Mb window) for each putative QTL. A bootstrap sample of the vector y for replicate j was created using the posterior means of the mean and SNP effects, except that all those SNP contained in the window that formed the QTL were excluded, and a vector of simulated residuals were added, formed by sampling a vector of independent standard normal deviations, one deviation for each animal, scaled by the posterior mean of the residual standard deviation, according to the model below:

$$y_j = \hat{\mu} + \sum_{i=1, i \notin QTL}^{i=4008} Z_i \hat{u}_i + \hat{\sigma}_e e_j$$

These bootstrap samples are constructed according to the null hypothesis of no QTL in the identified SNP window. Each bootstrap sample was reanalyzed using the Bayes B model used for the real data, and the genetic variance of the SNP window corresponding to the QTL were accumulated across all the bootstrap samples, for comparison to the test statistic represented by the genetic variance of the SNP window identified in the analysis of the real data. If just 1 bootstrap statistic from the 1,000 simulated exceeded the test statistic from the real data, the comparison-wise P-value was determined to be $0.001 < P < 0.002$.

Validation using an independent sample

A total of 504 genotyped and phenotyped boars as an independent sample, were used to validate the results from GWAS. Phenotypes of those animals were collected, edited and pre-adjusted of corresponding fixed effects in same procedures described above. Genotypes of the additional 504 boars were typed using 10K chip instead of 60K, which included 8826 SNPs chosen from the 60K BeadChip. Quality control for the marker genotypes were performed using the same criterion described above and 6028 SNPs remained. Imputation was performed using MaCH1.0 and Minimac at each chromosome. After imputation, 35,871 genotypes were obtained for each boar typed by 10K chip as validation genotype data. Conformation of the estimates of 35,871 SNPs increased the confidence in the previous association in training data.

Results

Phenotype records editing and statistics

Sixteen errors associated with individual feed intake were identified in this study. The error type, error rate, and coefficient for covariate associated with each error type estimated from the linear mixed model were summarized below in table 1. At least 1 error type was identified in 35% of the records which is much higher than the reported error rate ~ 5% (Casey et al., 2005; Cai et al., 2008). Error free daily feed intake is 1567 g/d before any adjustment, lower than the adjusted daily feed intake 1996.9 g/d, which indicates that only deleting the records with errors would underestimate the actual feed intake. Thus, proper adjustment is needed for the phenotype editing to ensure accurate prediction.

Table 1. Error type and coefficient from the linear mixed model for adjustment

Error index	Error type	Error rate	Coefficient of PTE	SD	Coefficient of OTD	SD	Coefficient of FID	SD
1	FIV-low	0.1439	-7.23E+02	-35.15	0.05044	4.31	-	-
2	FIV-high	0.0377	6.14E+02	6.92	-0.1843	-8.3	-	-
3	FIV-0	0.0053	-1.68E+03	-63.07	-	-	-	-
4	OTV-low	0.00056	-1.81E+03	-16.84	-	-	1.27E-01	3.89
5	OTV-high	0.1439	1.75E+03	4	-	-	-1.86E-01	-1.28
6	FRV-high-FIV-low	0	0.00E+00	0	-	-	-	-
7	FRV-high-strict	0.0031	-1.35E+03	-12.53	0.2341	5.45	-	-
8	FRV-high	0.037	-7.12E+02	-10.79	0.1729	9.5	-	-
9	FRV-0	0.0024	-6.46E+02	-4.57	-0.1208	-3.79	-	-
10	FRV-low	0.1594	-6.71E+02	-30.52	-0.1402	-23.37	-	-
11	LWD-low	0.0278	-1.00E+03	-22.2	-0.05053	-3.77	-	-
12	FWD-high	0.0396	-7.99E+02	-17.61	0.04453	3.15	-	-
13	FWD-low	0.0278	-5.57E+02	-13.38	0.004975	0.24	-	-
14	FWD-high	0.0396	-1.17E+03	-15.9	-0.009948	-0.52	-	-
15	LTD-low	0.0227	-1.10E+03	-25.27	-	-	-6.06E-02	-3.99
16	FTD-high	0.0227	-1.39E+03	-34.35	-	-	3.71E-02	2.32

Note: SD columns were the slandered error estimated for the coefficient in the previous column.

Robust regression for single pig analysis has been performed using serials of pig body weight data from the FIRE system. Outliers were identified (approximately 12% percent of total data points) and removed from the data set due to lower weight (< .4). 488 pigs were excluded from the analysis because of missing records and ADG more than 3SD from the mean, resulting ADG obtained for the rest of 599 pigs. Figure 1 showed the body weights on age before and after editing. Clearly, data points far from growth curve were removed based on the analysis. From simple linear regression, ADG for 889 boars were obtained. The ADG for 516 pigs were obtained from both robust regression and simple linear regression, with correlation of 0.89. An additional 83 boars, obtained from simple linear regression, were added to 889 boars, giving a total of 972 boars with edited ADG. Statistics were listed in Table 2 for ADG from linear regression and robust regression. ADG from robust regression have higher mean and larger SD than the ADG from simple linear regression.

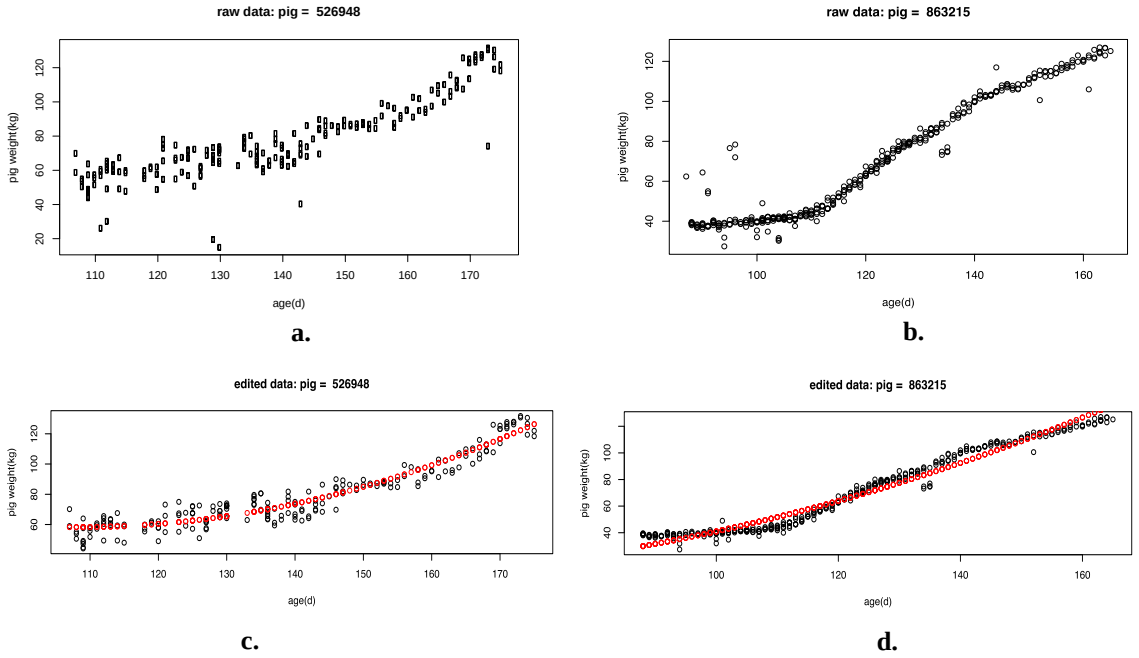


Figure 1. Pig weight editing via robust regression. a and b are row pig weights (kg/d) for two randomly selected pigs and c and d are edited pig weight with age (d) on x axis. Red points are estimated pig weights.

Table 2. Statistics for ADG from linear regression (LR) and robust regression (RR).

ADG(g/d)	LR	RR
N	899	599
Min	0.5193	0.4043
1st Quintile	0.7417	0.6966
Median	0.7915	0.8406
Mean	0.7887	0.8284
3rd Quintile	0.8399	0.9822
Max	1.0802	1.0996
SD	0.0784	0.2378

Summary statistics of the 9 traits to conduct GWAS were listed in Table 3. Number of animals available to perform GWAS was from 730 to 1047 boars. Mean ADFI and ADG are shown in Table 3 below, are 2 kg and 0.776 kg respectively. FCR was calculated from the ratio of ADG and ADFI, with mean 39.4%, indicating about 40% of the energy obtained from ADFI was transmitted to weight gain. Residual feed intake was the model residual for ADFI, with mean 0 g and SD 261g. The average individual birth weight and weaning weight is 2 kg and 7.75 kg, respectively. The ultrasound traits, including BF, MD and IMF were measured and summarized, with mean 1 cm, 4.15 cm, and 0.0388% respectively.

Table 3. Characteristics for nine traits analyzed

Statistics	ADFI(g)	ADG(g)	FCR	RFI(g)	BW(kg)	WW(kg)	BF(cm)	MD(cm)	IMF(%)
N	972	972	972	972	1047	889	1047	1047	730
min	1272	449.3	0.208	-843.1	1.025	2.948	0.5063	2.377	2.423
1stQ	1818	725.3	0.359	-139.7	1.796	6.396	0.8432	3.818	3.435
mean	2003	776.9	0.394	0	2.028	7.752	1.007	4.149	3.688
3rdQ	2179	828.6	0.419	144.2	2.259	9.072	1.1352	4.43	3.959
max	3551	1092.9	0.647	1457	3.847	15.785	1.9845	5.782	4.81
SD	288.71	86.698	0.059	260.953	0.344	1.934	0.221	0.468	0.398

Genetic parameter estimation

Estimates of additive genetic variance, residual variance and heritability were listed in Table 4, estimated from the edited data set containing the 9 traits of interest from animal models. Moderate heritability for ADG, ADFI, BF, MD, IMF, BW and WW were obtained from present analysis, with heritability 0.5, 0.3, 0.58, 0.39, 0.54, 0.65, and 0.64, respectively. However, heritability estimates for FCR and RFI were lower, less than 10%. Because of the structure of the dataset it was not possible to include maternal effects in the model. As a result the heritability estimates for BW and WW are much higher than estimates reported by other investigators.

Table 4. Genetic parameter estimation

Trait	Additive variance	Residual variance	Heritability
ADG(g/d)	3909(950.3)	3873(689.6)	0.5023(0.1026)
ADFI(g/d)	24256(8064.2)	52878(6503.6)	0.3145(0.0955)
FCR	2.358E-3(0.180E-3)	2.889E-3(0.204E-3)	0.0755(0.0568)
RFI(g/d)	5476.8(3087.5)	51088(3521.3)	0.0968(0.0537)
BF(inch)	2.985E-3(0.547E-3)	2.107E-3(0.379E-3)	0.5862(0.0854)
MD(inch)	1.110E-2(0.279E-2)	1.720E-2(0.215E-2)	0.3903(0.0872)
IMF(%)	0.0693(0.0172)	0.0587(0.0126)	0.5415(0.1119)
BW(kg)	0.0801(0.0164)	0.0435(0.0109)	0.6480(0.1019)
WW(kg)	1.5992(0.3465)	0.8878(0.2349)	0.6430(0.1081)

Association analysis

Single marker association

Bayes B model was implicated in this study on each trait of interest separately. Posterior means for residual variance and genetic variance, and marker heritability were shown in Table 5. QTLs were detected for each trait, except for RFI, with markers explaining a small amount of genetic variance. For most traits, the marker heritability obtained from the association study is about half of the heritability obtained from the genetic parameter estimation, indicating that the markers could not explain all the genetic variation and that missing heritability existed.

Manhattan plots for genetic variance explained by single SNP makers against each chromosome (Fig. 2) show potential candidate genes associated with the strong evidence of QTL regions for each trait, except for RFI and BW. Obvious QTL regions for ADFI were SSC 1, 8, 10, 14, while for ADG were SSC 1, 4, 14 with larger effect SNPs. A large putative QTL was mapped in SSC4 for FCR. QTL was identified for RFI on SSC2. QTL for ultrasound BF, MD and IMF were identified on SSC1 and 18, SSC 5, 11, 13 and SSC9, 10,

15, respectively. QTL regions on SSC3, 8, 10, 16 and SSC4, 9 were identified for WW if considered single locus contribution to genetic variance.

Table 5. Posterior means of variance explained by whole-genome markers for each trait

Trait of interest	Posterior Mean of residual variance	Posterior Mean of genetic variance	Estimated Total Variance	Marker heritability
ADG(g)	4604.09	1718.19	6385.28	0.2789
ADFI (g)	59696.3	13142	72416.6	0.1815
FCR (%)	0.003029	0.000696848	0.00372585	0.187031
RFI(g)	542.496	31.0123	573.508	0.0540747
BF (cm)	0.0262772	0.026938	0.0532152	0.506208
MD (cm)	0.13538	0.0723993	0.207779	0.348444
IMF (100%)	0.544172	0.197766	0.741938	0.266554
BW(kg)	0.084821	0.0269415	0.111763	0.241061
WW(kg)	2.2515	1.32427	3.57577	0.370346

Note: a. Marker heritability defined here is the proportion genetic variance explained by markers.

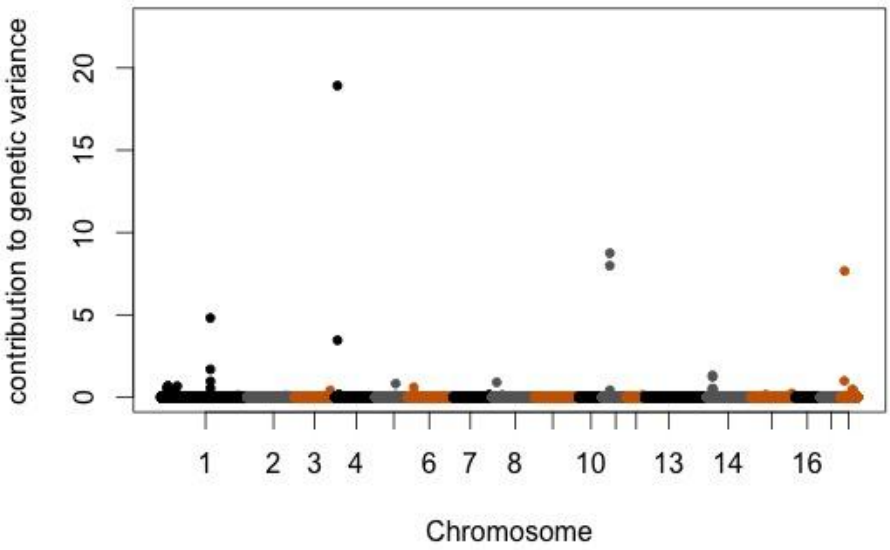
Prior sensitivity was assessed using ADFI (kg/day), by using different variance and pi hyper-parameters (Table 6). From the results, the posterior mean of each scenario was not affected greatly with different prior set up.

Table 6. Posterior means of parameters from different prior set up.

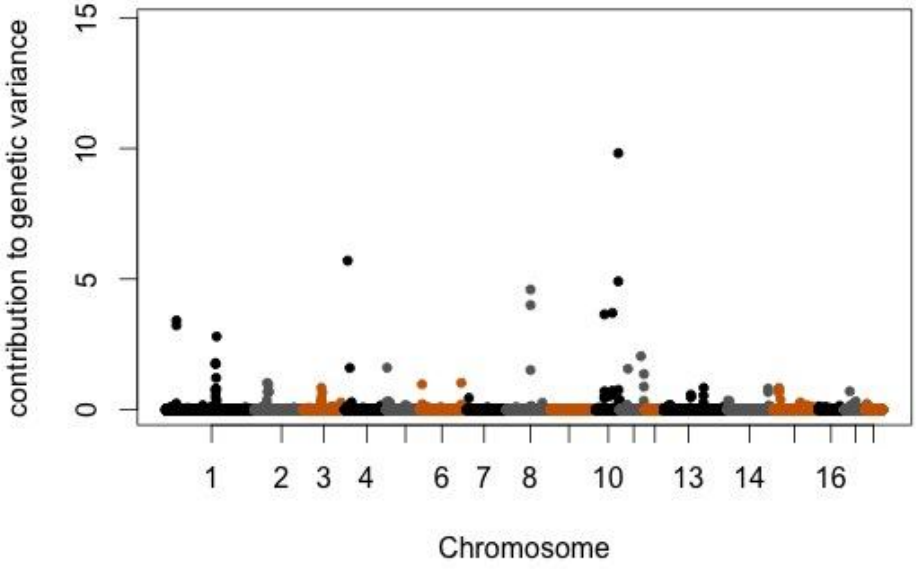
Bayes type	prior π	h^2	gVar	rVar	totalV.	prop.V.	marker V.	posterior π	time (s)
B	0.995	0.05	0.021	0.082	0.102	0.202	0.202	-	17815
	0.995	0.35	0.037	0.070	0.108	0.346	0.346	-	18310
	0.995	0.49	0.041	0.067	0.107	0.377	0.377	-	19524
	0.995	0.75	0.043	0.063	0.106	0.408	0.408	-	19019
	0.995	0.9	0.044	0.060	0.105	0.423	0.423	-	17661
B	0.5	0.49	0.086	0.042	0.128	0.674	0.674	-	22737
	0.8	0.49	0.084	0.043	0.126	0.664	0.664	-	20164
	0.95	0.49	0.076	0.046	0.122	0.622	0.622	-	18236
	0.99	0.49	0.054	0.058	0.112	0.479	0.479	-	17565
	0.995	0.49	0.041	0.067	0.107	0.377	0.377	-	19524
Cpi	0.5	0.49	0.015	0.081	0.096	0.156	0.156	0.635	7217
	0.8	0.49	0.016	0.080	0.097	0.170	0.169	0.790	8835
	0.9	0.49	0.017	0.080	0.097	0.177	0.177	0.958	5830
	0.95	0.49	0.017	0.080	0.097	0.178	0.178	0.949	6784
	0.995	0.49	0.017	0.080	0.097	0.176	0.176	0.725	9499

Note: a. pMean of Rvar stands for posterior mean of residual variance; b. pMean of GVar for posterior mean of genetic variance; c. TVar for total variance, which is sum of a and b; d. for proportion genetic variance explained by markers.

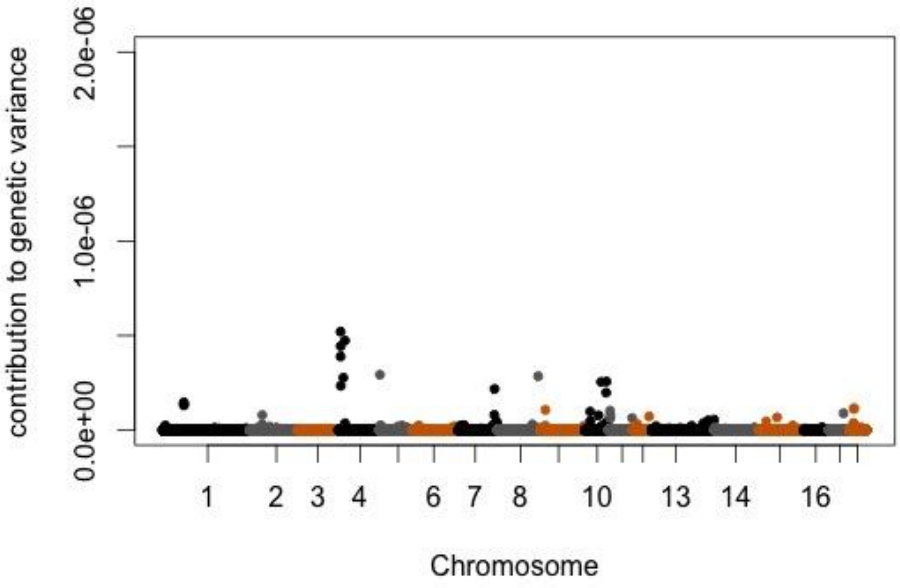
The cumulative distribution of proportion of single marker variance when ranked by size for each traits show that a limited number of markers were sufficient to explain more than 50% of the genetic variance for ADFI (500 SNP markers) and ADG (300 SNP markers) (shown in Fig 3), respectively.


contribution to genetic variance
0
5
10
15
20
1
2
3
4
6
7
8
10
13
14
16
Chromosome

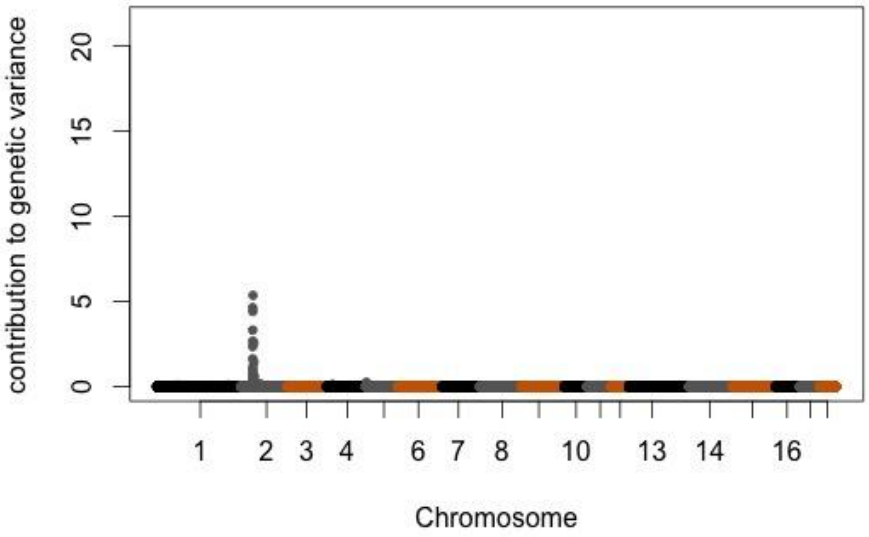
a. ADG


contribution to genetic variance
0
5
10
15
1
2
3
4
6
7
8
10
13
14
16
Chromosome

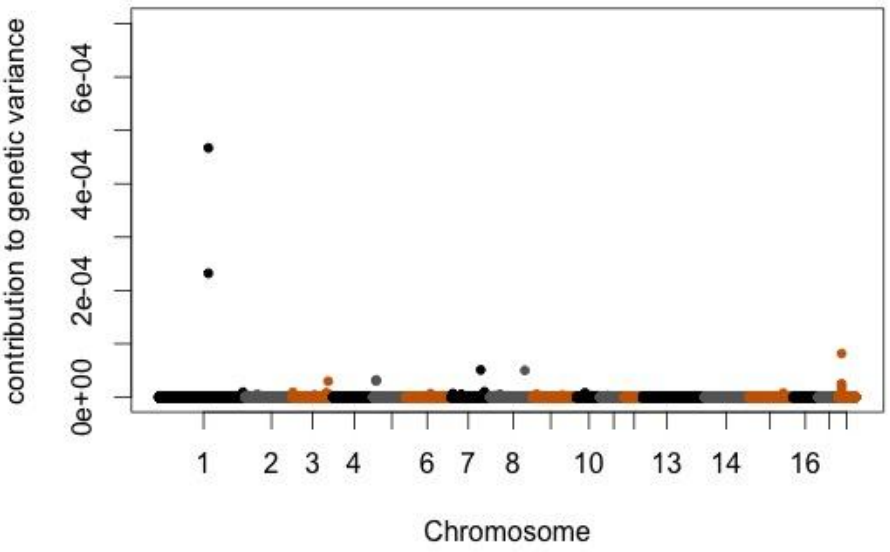
b. ADFI


contribution to genetic variance
0.0e+00
1.0e-06
2.0e-06
1
2
3
4
6
7
8
10
13
14
16
Chromosome

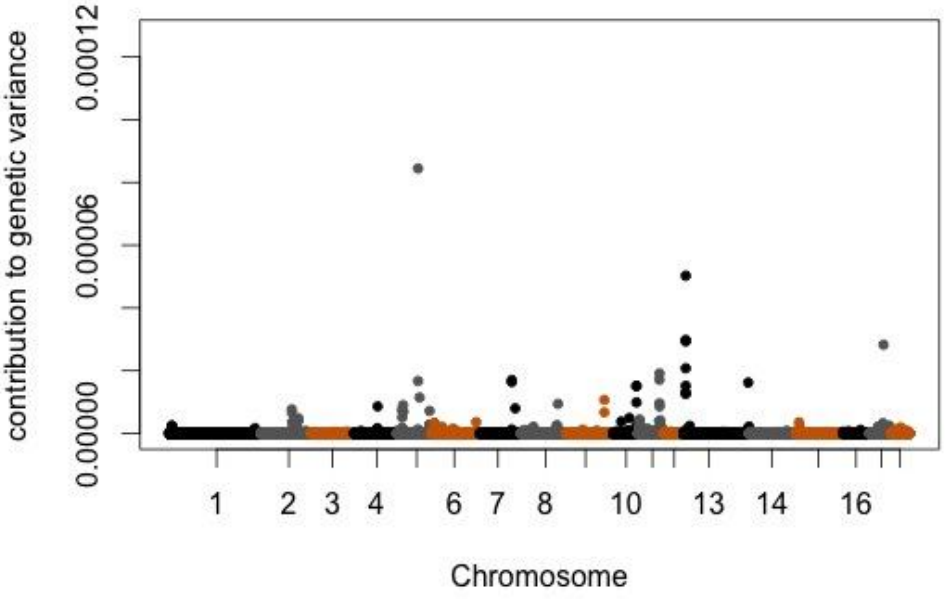
c. FCR


contribution to genetic variance
0
5
10
15
20
1
2
3
4
6
7
8
10
13
14
16
Chromosome

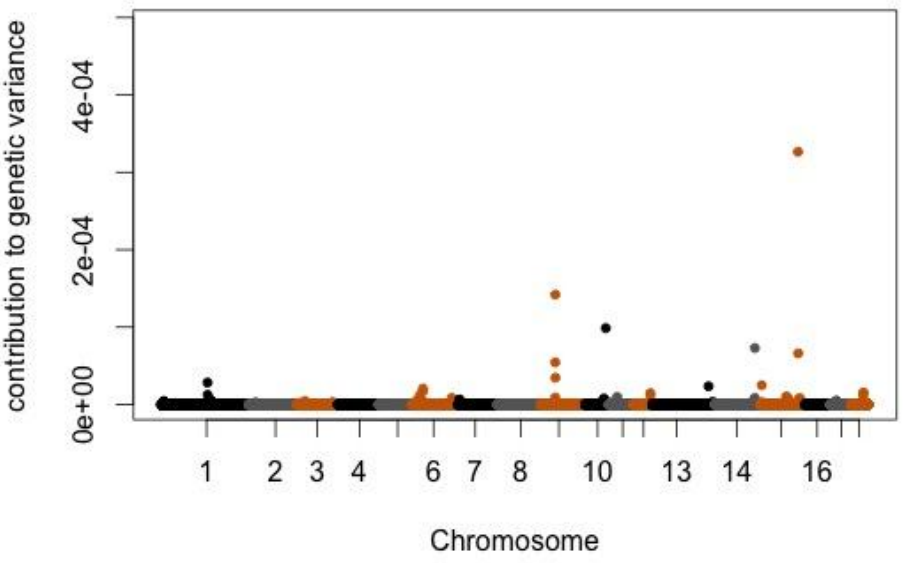
d. RFI


contribution to genetic variance
0e+00
2e-04
4e-04
6e-04
1
2
3
4
6
7
8
10
13
14
16
Chromosome

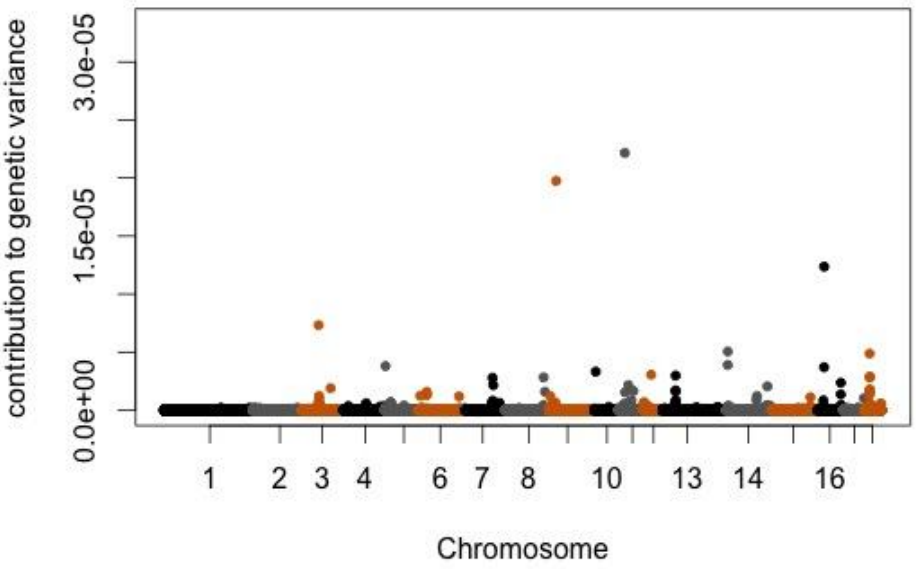
e. BF


contribution to genetic variance
0.00000
0.00006
0.00012
1
2
3
4
6
7
8
10
13
14
16
Chromosome

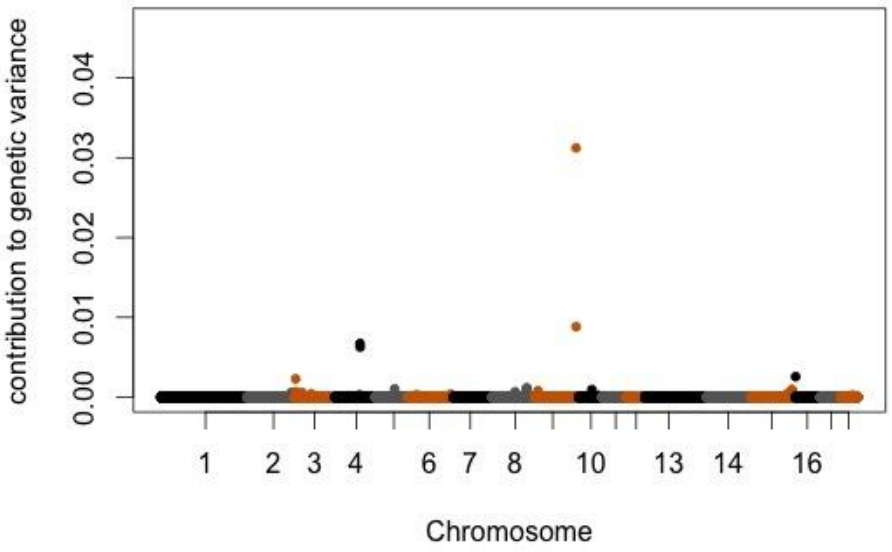
f. MD


contribution to genetic variance
0e+00
2e-04
4e-04
1
2
3
4
6
7
8
10
13
14
16
Chromosome

g. IMF


contribution to genetic variance
0.0e+00
1.5e-05
3.0e-05
1
2
3
4
6
7
8
10
13
14
16
Chromosome

h. BW



g. WW

Fig. 2 Manhattan plots of traits of interest. Note: chromosome 20 is for the unmapped SNPs.

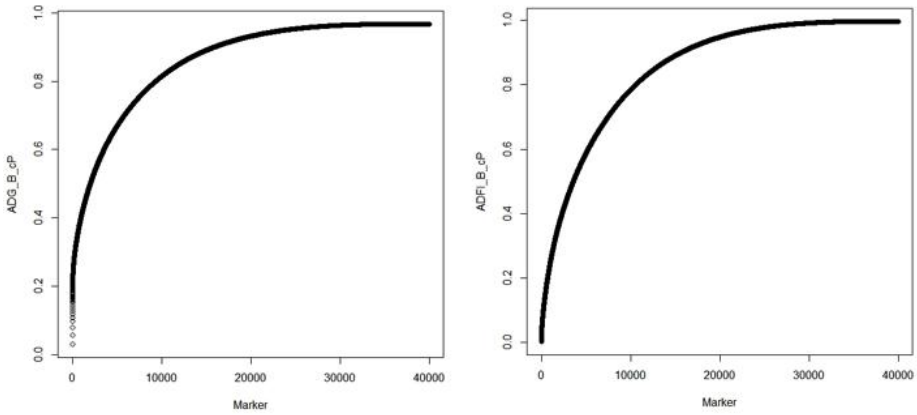


Fig. 3 Cumulative distribution of variance explained by single marker. Left: ADG, right ADFI.

Windows variance analysis

Window variances for each trait were computed, represented in Table 7. For each trait, significant windows were defined as windows with variance account for no less than 1% of genetic variance with probability of inclusion (PPI) larger than 0.5. Significant windows with large proportion of variance explained were identified for several traits.

Four 1-Mb windows on SSC 1, 4, 11, 14 were identified as significant windows for ADG, with proportion of genetic variance explained 4.8, 5.04, 2.69 and 2.39%, respectively. The four significant windows identified for ADG cumulatively explained 15% of genetic variance. For FCR, 1 window on SSC4 was detected in the present study, which could explain 2.65% of genetic variance for FCR. Two significant windows were identified for BF, on SSC 1, 14, 18, with 5.73 and 2.14 percent of genetic variance explained. Windows on SSC 13 and SSC X were identified for MD, explaining 2.67 and 2.62 percent of genetic variance.

Several 1-Mb windows were detected for IMF and BW with large proportion of variance explained; however, none of them are significant. Four significant windows were identified for weaning weight on SSC 4, 8, 9 and 15, with 3.48, 2.32, 7.02 and 2.41 percent of genetic variance explained and a total of 15.22%. Although significant windows for ADFI, IMF and BW were not identified in this study, windows with large proportion of genetic variance explained were detected. No QTL was identified for RFI in this study.

After significant QTLs, less significant QTLs and marker information were analyzed. It is interesting to find that windows with large variance identified for ADFI and ADG on SSC1, at 166Mb and 169Mb, were adjacent regions, indicating that there may be the same gene having effects on both of the traits. Two significant windows were identified for both BF and MD, indicating that they might be the same genes associated with the two production traits.

Table 7. Window variance and inclusion probability.

Trait	Window	#SNPs	Chr	#Mb	%Var	Cum%Var	PPI
ADFI	1497	16	10	74	1.84	1.84	0.42539
	164	11	1	166	1.72	3.57	0.363029
	1207	11	8	79	1.71	5.27	0.378619
	1477	24	10	54	1.61	6.88	0.36971
	341	20	2	46	1.56	8.45	0.314031
ADG	599	22	4	6	5.04	5.04	0.821826
	167	14	1	169	4.8	9.84	0.703786
	1530	22	11	27	2.69	12.52	0.561247
	1885	15	14	20	2.39	14.91	0.503341
	1529	18	11	26	2.2	17.12	0.463252
FCR	598	30	4	5	2.65	2.65	0.570156
	1477	24	10	54	1.31	3.96	0.358575
	2127	7	15	115	0.9	4.86	0.244989
	1124	9	7	128	0.86	5.71	0.273942
	1509	11	11	6	0.83	6.54	0.256125
RFI	340	19	2	45	0.68	0.68	0.124722
	337	28	2	42	0.5	1.18	0.140312
	336	18	2	41	0.46	1.64	0.093541
	599	22	4	6	0.34	1.98	0.100223
	341	20	2	46	0.33	2.3	0.097996
BF	167	14	1	169	5.73	5.73	0.942094
	2331	11	18	11	2.14	7.87	0.64588
	1881	28	14	16	0.87	8.74	0.503341
MD	1659	26	13	6	2.67	2.67	0.786192
	2415	17	23	36	2.62	5.29	0.688196
IMF	1325	21	9	49	1.73	1.73	0.452116
	2140	25	15	128	1.33	3.06	0.407572
	1644	21	12	56	1.19	5.48	0.396437
BW	2339	7	18	19	1.43	3.38	0.342984
	1696	22	13	43	1.21	4.59	0.42539
WW	1412	18	9	141	7.02	7.02	0.939866
	677	8	4	84	3.48	10.5	0.654788
	2152	17	15	140	2.41	12.91	0.650334
	1247	16	8	119	2.32	15.22	0.621381

Bayes factor analysis

Across all nine traits investigated, 104 SNPs had a Bayes factor >20 for at least 3 traits (Fig.4 and Table 8), while 22, 27, 12, 32, 24, 8, 3, 33 SNPs had a Bayes factor > 20, for ADFI, ADG, FCR, BF, MD, IMF, BW, and WW, respectively. The number of SNPs with a Bayes factor > 3 was 510, 505, 443 and 517, 046, 361, 269, 545 for ADFI, ADG, FCR, BF, MD, IMF, BW, and WW, respectively. From the results, ADFI and ADG have markers in the same region, SSC1 169Mb with large variance explained. BF and MD have same markers significantly associated with the traits on SSC 4, 6 and 9.

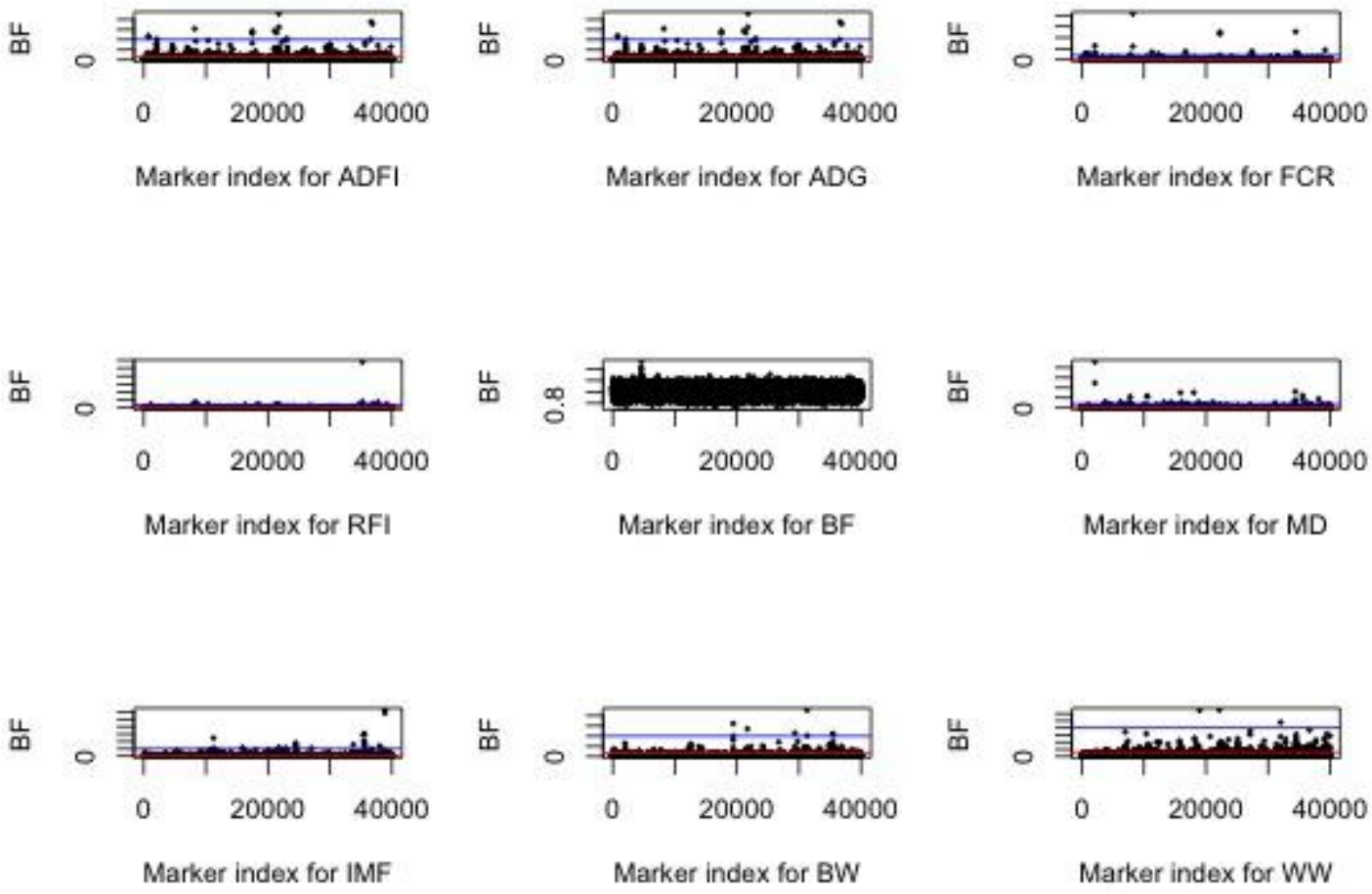


Fig. 4 Bayes Factor analysis for 9 traits investigated

Table 8. Results from Bayes Factor analysis.

Trait	Marker	Marker Effect	PPI	Marker Variance	Chr_bp	%Varaince	Bayes Factor
ADFI	773	-0.0071	0.094	1.52E-05	1_34930033	1.605	20.744
	774	-0.0073	0.097	1.63E-05	1_34940995	1.723	21.45
	2090	-0.0068	0.107	2.28E-05	1_169437312	2.405	23.794
	8219	0.0083	0.136	3.37E-05	4_7491116	3.547	31.271
	10375	-0.0057	0.092	1.04E-05	5_1103371	1.097	20.235
	12036	0.0142	0.104	1.37E-05	6_7478333	1.439	23.197
	13901	-0.0069	0.122	2.05E-05	6_139182650	2.165	27.781
	17484	0.0106	0.122	2.47E-05	8_79987974	2.608	27.729
	17485	0.0112	0.127	2.76E-05	8_79993046	2.91	28.871
	21075	0.0076	0.111	1.63E-05	10_28367534	1.723	24.746
	21406	0.0116	0.151	3.59E-05	10_54856762	3.787	35.421
	21769	0.0092	0.165	4.23E-05	10_74699721	4.454	39.323
	21772	-0.0108	0.189	5.88E-05	10_74794196	6.199	46.316
	22257	0.0064	0.098	1.41E-05	11_27565041	1.482	21.499
	23143	-0.0103	0.099	9.51E-06	11_81062647	1.003	21.792
ADG	472	0.0022	0.111	1.23E-06	1_19628038	0.672	24.746
	528	-0.0014	0.107	9.64E-07	1_22586549	0.526	23.745
	544	0.0027	0.162	2.46E-06	1_23113704	1.341	38.527
	1002	-0.0044	0.16	2.07E-06	1_55429289	1.128	37.877
	2089	-0.002	0.116	2.07E-06	1_169294975	1.127	26.011
	2090	-0.0042	0.23	8.81E-06	1_169437312	4.806	59.542
	2093	0.0027	0.151	3.58E-06	1_169691789	1.953	35.283
	2095	0.0028	0.155	3.76E-06	1_169732994	2.052	36.587
	7779	-0.0012	0.102	6.83E-07	3_123609384	0.373	22.702
	8199	0.0097	0.531	4.13E-05	4_6741671	22.55	225.126
	8205	0.0045	0.255	8.59E-06	4_6948977	4.686	68.186
	11273	-0.0036	0.188	2.80E-06	5_68920051	1.528	46.014
	12293	0.0018	0.123	1.22E-06	6_20873704	0.664	27.884
	16694	0.0028	0.181	2.42E-06	8_15607164	1.322	43.89
	22240	-0.0057	0.367	1.58E-05	11_26777802	8.645	115.177
	22246	-0.007	0.416	2.09E-05	11_27092241	11.392	141.753
	22249	-0.0013	0.092	8.69E-07	11_27188841	0.474	20.163
	27387	-0.0017	0.109	1.26E-06	14_19804468	0.689	24.244
	27393	-0.0018	0.116	1.51E-06	14_20176710	0.822	26.241
	27398	-0.0025	0.153	2.72E-06	14_20514852	1.485	35.919
	27400	-0.0026	0.16	3.01E-06	14_20560420	1.645	37.989
	27420	-0.0014	0.093	8.99E-07	14_21858383	0.491	20.332
	34434	0.0057	0.179	2.93E-06	18_13573617	1.599	43.358
	34462	0.0128	0.398	1.63E-05	18_14581968	8.888	131.565
	34873	-0.0015	0.107	8.51E-07	18_41861687	0.464	23.77
	34874	-0.0016	0.113	9.51E-07	18_41891138	0.519	25.428
	34875	-0.0015	0.106	8.22E-07	18_41916528	0.449	23.645

FCR	8160	0.001	0.13	5.36E-07	4_5226550	3.62	29.841
	8161	0.001	0.125	4.69E-07	4_5254496	3.169	28.377
	8162	9.00E-04	0.115	4.01E-07	4_5268656	2.708	25.96
	8178	-9.00E-04	0.117	4.09E-07	4_5815515	2.761	26.419
	8380	-0.001	0.12	3.76E-07	4_14726013	2.539	27.034
	8506	-0.0018	0.146	5.30E-07	4_20413793	3.583	33.912
	10375	0.0012	0.128	4.29E-07	5_1103371	2.897	29.342
	16306	8.00E-04	0.099	3.07E-07	7_124403181	2.073	21.915
	18332	8.00E-04	0.116	3.43E-07	8_139226228	2.317	26.241
	21406	-0.0021	0.196	1.18E-06	10_54856762	7.962	48.666
	21720	0.0011	0.118	3.67E-07	10_71993874	2.48	26.521
	21748	-9.00E-04	0.102	2.94E-07	10_73205880	1.985	22.604
BF	389	-5.00E-04	0.129	1.09E-07	1_15325380	1.436	29.447
	406	3.00E-04	0.095	5.46E-08	1_16364763	0.722	20.987
	8289	6.00E-04	0.14	1.07E-07	4_11316413	1.421	32.288
	8301	8.00E-04	0.198	3.08E-07	4_11776710	4.07	49.006
	8309	4.00E-04	0.095	6.05E-08	4_11992812	0.8	20.914
	9243	-4.00E-04	0.101	6.42E-08	4_74053481	0.85	22.283
	11974	-7.00E-04	0.194	2.44E-07	6_4454671	3.229	47.868
	12216	-5.00E-04	0.146	1.35E-07	6_16885226	1.79	33.912
	12520	6.00E-04	0.149	1.45E-07	6_35894949	1.913	34.898
	14298	6.00E-04	0.17	1.82E-07	7_5533512	2.412	40.644
	19297	0.001	0.212	2.93E-07	9_44151170	3.88	53.666
	19299	0.0011	0.224	3.41E-07	9_44182184	4.505	57.443
	20212	-8.00E-04	0.167	1.44E-07	9_133405291	1.901	39.953
	21037	6.00E-04	0.147	1.10E-07	10_25637980	1.457	34.294
	21039	7.00E-04	0.151	1.18E-07	10_25691510	1.556	35.311
	24258	-4.00E-04	0.105	6.33E-08	12_60893153	0.837	23.247
	24259	-4.00E-04	0.1	5.68E-08	12_60899098	0.752	22.234
	24260	-4.00E-04	0.102	6.67E-08	12_60915610	0.883	22.653
	24266	-3.00E-04	0.096	5.74E-08	12_61064389	0.759	21.133
	24384	-5.00E-04	0.118	8.93E-08	13_2113025	1.181	26.675
	26125	-5.00E-04	0.128	1.22E-07	13_131469409	1.61	29.185
	30097	4.00E-04	0.111	7.80E-08	15_26329613	1.032	24.923
	32931	-0.0011	0.284	6.10E-07	16_81442583	8.07	78.739
	33602	4.00E-04	0.096	5.24E-08	17_36307204	0.693	21.035
	33913	-4.00E-04	0.103	6.30E-08	17_54487116	0.833	22.826
	34693	-4.00E-04	0.11	8.05E-08	18_30768879	1.064	24.596
MD	389	-5.00E-04	0.129	1.09E-07	1_15325380	1.436	29.447
	406	3.00E-04	0.095	5.46E-08	1_16364763	0.722	20.987
	8289	6.00E-04	0.14	1.07E-07	4_11316413	1.421	32.288
	8301	8.00E-04	0.198	3.08E-07	4_11776710	4.07	49.006
	8309	4.00E-04	0.095	6.05E-08	4_11992812	0.8	20.914
	9243	-4.00E-04	0.101	6.42E-08	4_74053481	0.85	22.283

	11974	-7.00E-04	0.194	2.44E-07	6_4454671	3.229	47.868
	12216	-5.00E-04	0.146	1.35E-07	6_16885226	1.79	33.912
	12520	6.00E-04	0.149	1.45E-07	6_35894949	1.913	34.898
	14298	6.00E-04	0.17	1.82E-07	7_5533512	2.412	40.644
	19297	0.001	0.212	2.93E-07	9_44151170	3.88	53.666
	19299	0.0011	0.224	3.41E-07	9_44182184	4.505	57.443
	20212	-8.00E-04	0.167	1.44E-07	9_133405291	1.901	39.953
	21037	6.00E-04	0.147	1.10E-07	10_25637980	1.457	34.294
	21039	7.00E-04	0.151	1.18E-07	10_25691510	1.556	35.311
	24258	-4.00E-04	0.105	6.33E-08	12_60893153	0.837	23.247
	24259	-4.00E-04	0.1	5.68E-08	12_60899098	0.752	22.234
	24260	-4.00E-04	0.102	6.67E-08	12_60915610	0.883	22.653
	24266	-3.00E-04	0.096	5.74E-08	12_61064389	0.759	21.133
	24384	-5.00E-04	0.118	8.93E-08	13_2113025	1.181	26.675
	26125	-5.00E-04	0.128	1.22E-07	13_131469409	1.61	29.185
	30097	4.00E-04	0.111	7.80E-08	15_26329613	1.032	24.923
	32931	-0.0011	0.284	6.10E-07	16_81442583	8.07	78.739
	33602	4.00E-04	0.096	5.24E-08	17_36307204	0.693	21.035
	33913	-4.00E-04	0.103	6.30E-08	17_54487116	0.833	22.826
	34693	-4.00E-04	0.11	8.05E-08	18_30768879	1.064	24.596
IMF	7080	-0.0095	0.124	1.16E-05	3_64987375	1.893	28.299
	8304	0.0077	0.162	2.92E-05	4_11911872	4.754	38.357
	23928	-0.0085	0.107	7.98E-06	12_43223831	1.298	23.894
	24083	-0.0071	0.14	1.64E-05	12_51945052	2.674	32.288
	33006	0.0225	0.416	0.0002447	16_84730003	39.781	141.929
	33961	-0.007	0.137	1.90E-05	17_57374863	3.093	31.591
	33969	-0.0057	0.115	1.27E-05	17_57717887	2.073	25.859
BW	27111	-0.0038	0.102	6.94E-06	14_5475231	5.316	22.653
	27112	-0.0034	0.095	5.83E-06	14_5564968	4.468	20.914
	32021	0.004	0.102	8.02E-06	16_21945031	6.15	22.727
WW	6269	-0.0572	0.197	0.001337918	3_4652086	2.302	48.728
	6276	-0.0443	0.16	0.000831411	3_4858648	1.43	38.018
	6975	-0.023	0.092	0.000248235	3_57634758	0.427	20.091
	9401	-0.0883	0.249	0.003750799	4_84435521	6.453	66.086
	9404	-0.0882	0.252	0.003744418	4_84811655	6.442	67.221
	14026	-0.0326	0.128	0.000467837	6_147232271	0.805	29.316
	18020	0.0238	0.093	0.000267279	8_119116910	0.46	20.453
	18021	0.025	0.098	0.000295115	8_119154287	0.508	21.572
	18710	0.0303	0.098	0.000319653	9_10290534	0.55	21.596
	18720	0.072	0.202	0.00178095	9_10623910	3.064	50.373
	20391	0.2228	0.562	0.02390132	9_141912772	41.123	255.027
	20393	0.1038	0.278	0.005237628	9_141988713	9.011	76.509
	21233	-0.028	0.118	0.000385637	10_44461025	0.663	26.547

31550	0.0538	0.155	0.001148203	15_140309646	1.976	36.392
31551	0.0551	0.156	0.001205242	15_140332793	2.074	36.754
31552	0.0518	0.148	0.001067154	15_140347111	1.836	34.431
31553	0.0586	0.164	0.001360677	15_140398200	2.341	39.038
31555	0.0307	0.096	0.000375915	15_140454469	0.647	21.255
31695	0.0481	0.174	0.001105167	16_595287	1.901	41.949

Bootstrapping analysis

For the nine traits investigated in this study, 500 bootstrap samples were created and reran by Bayes B in order to construct the distribution of the 40,008 SNPs of the test statistic (genetic variance explaining by each markers) for each putative QTL (There are 1000 bootstrap samples for ADFI). P-value related to each single marker was computed based on the counts of bootstrap samples exceeding test statistics. Negative log P-values against markers were plotted to investigate significant markers for 9 traits analyzed (show in Fig. 5). Significance was determined by the largest –log(p) value 6.216 (9.21 for ADFI, and summary of each trait was listed in Table 9. Compared with results from Bayes factor analysis, more markers were determined to be significant from bootstrapping analysis.

Table 9. Results summary for bootstrapping analysis.

Trait	No. of bootstrap samples	No. of markers exceeding threshold
ADFI	1000	99
ADG	500	51
FCR	500	3
RFI	500	12
BF	500	95
MD	500	54
IMF	500	109
WW	500	43
BW	500	170

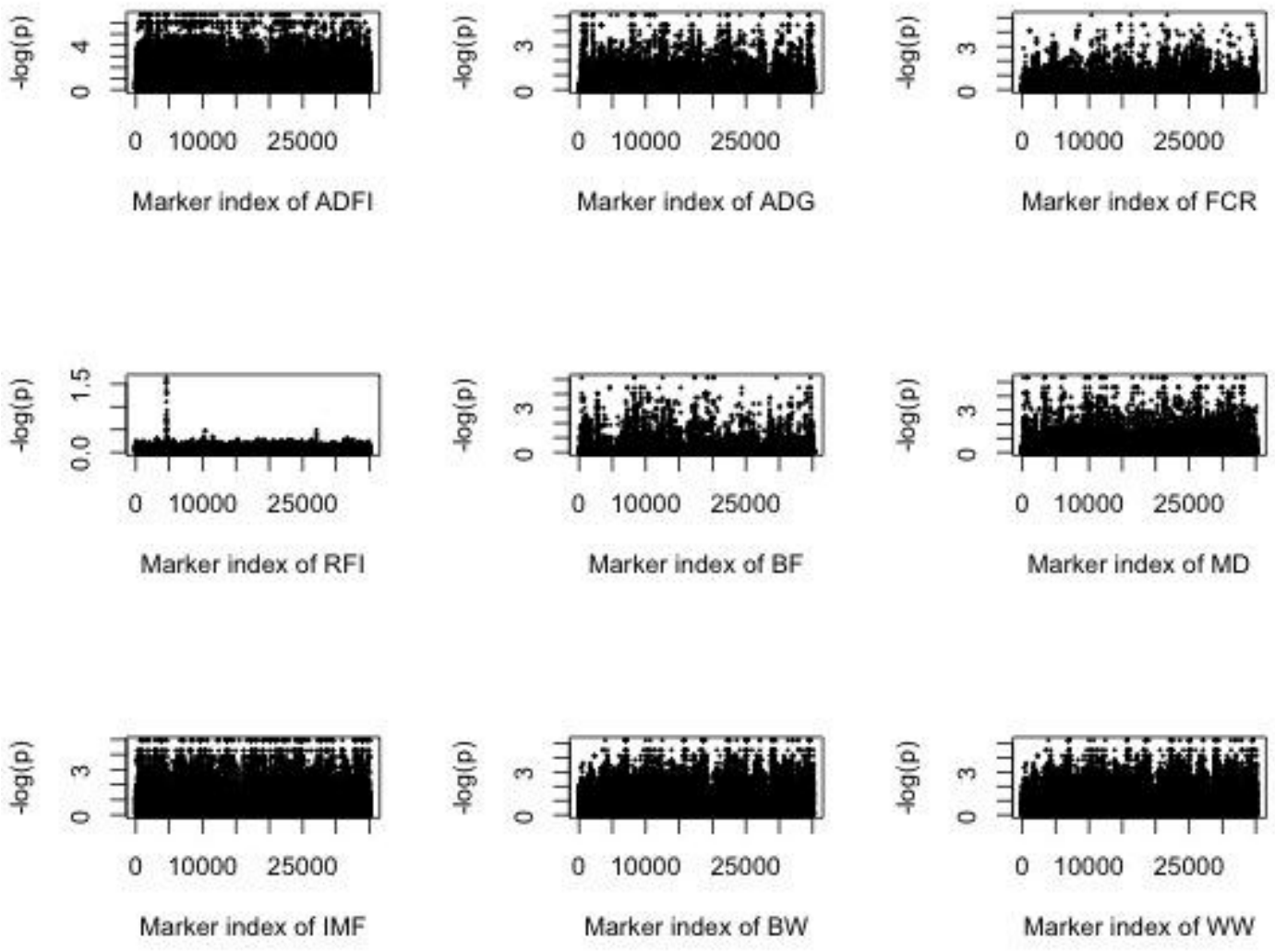


Fig. 5 Negative log p value against marker index for 9 traits

Candidate genes for ADG and ADFI

Significant regions were identified by using three different significance tests; posterior windows variance, Bayes factor and bootstrapping. Significance was declared for regions where tests significance overlapped. Regions associated with ADFI were mapped to chromosome 1 and 10, shown in Fig. 6. Similarly, regions associated with ADG mapped to chromosome 1, 4, 11, and 14, shown in Fig. 7 below. Summary of common significant markers and nearby genes (within 2 Mb from significant marker) for ADG and ADFI were summarized in Table 10. Candidate genes for all the traits analyzed were chosen based on posterior windows analysis (Table 11). Those genes were in the significant windows declared above.

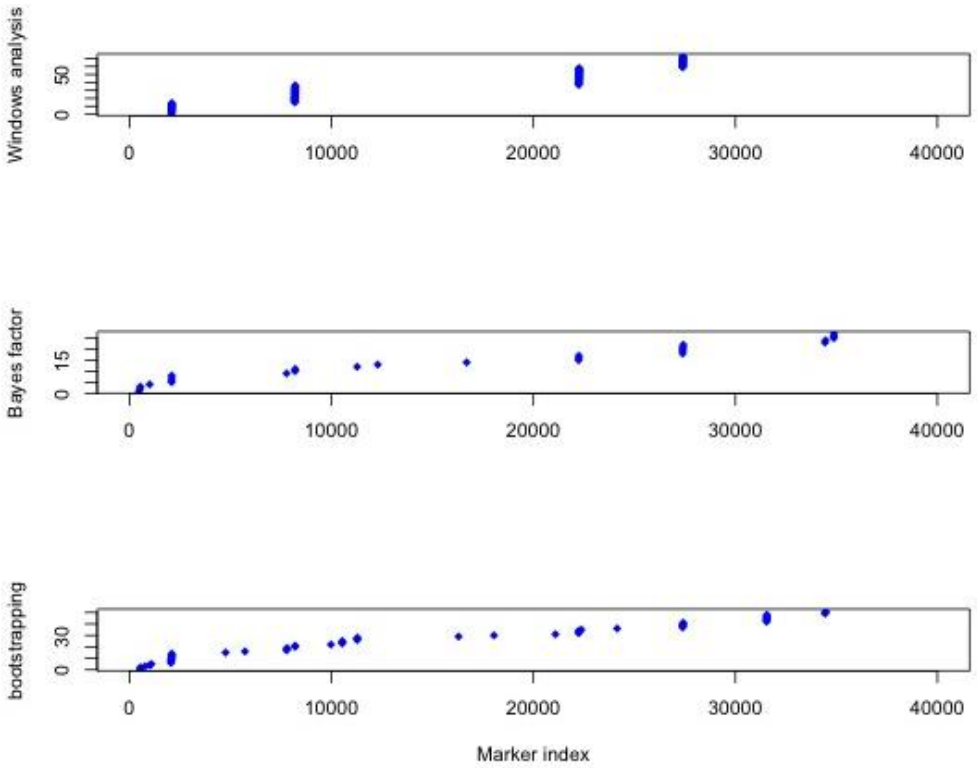


Fig. 6 Significant makers for ADG

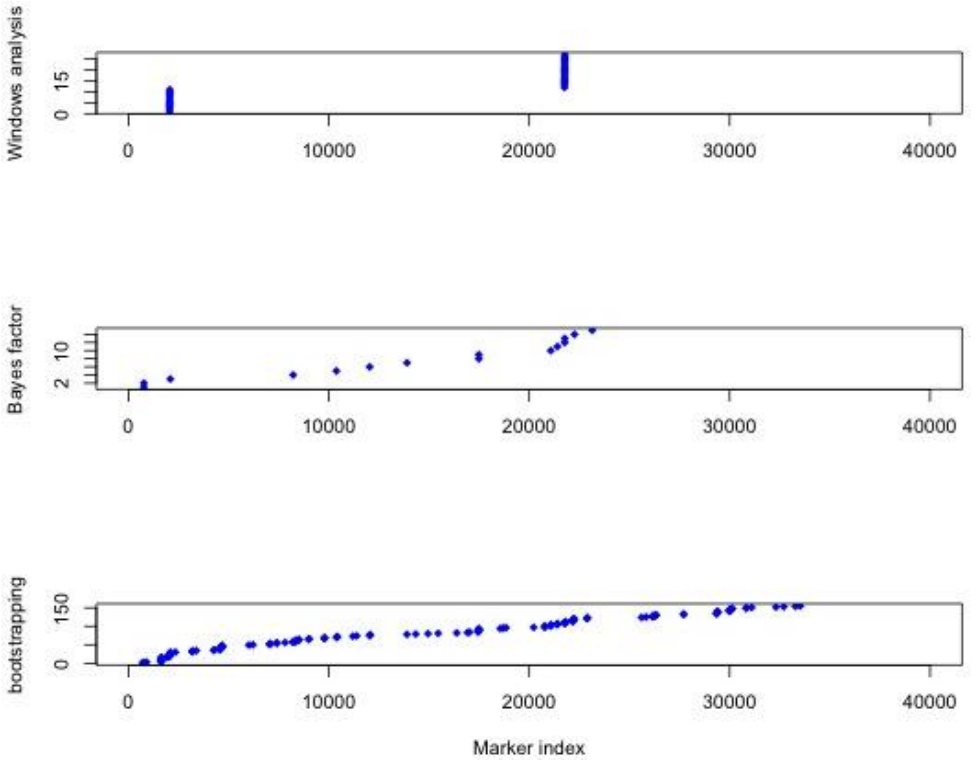


Fig. 7 Significant makers for ADG

Table 10. Results summary for common significant markers associated with both ADG and ADFI.

Trait	Marker Index	Chromosome	Position (bp)	Genes nearby
ADG	2089	1	169294975	LOC100517320,
	2090	1	169437312	SOCS6, RTTN,
	2093	1	169691789	LOC100738041,
	2095	1	169732994	DOK6, etc.
	8199	4	6741671	LOC100738458,
	8205	4	6948977	KHDRBS3, ZFAT
	22246	11	27092241	None
	22249	11	27188841	None
	27393	14	20176710	LOC10051875,
	27398	14	20514852	LOC100518936, etc.
	27400	14	20560420	
ADFI	2080	1	168733437	LOC100517320, SOCS6, RTTN, DOK6, etc.
	21769	10	74699721	PFKP, LOC100521215, etc.
	21772	10	74794196	

Validation of an independent sample

To validate the marker estimates in association for 9 traits analyzed in this study, 500 additional samples were investigated and Bayes factors were re-computed to declare significant markers in validation dataset (Fig. 8). Markers with Bayes factor exceeding 3 were claimed to be significant and extremely significant with Bayes factor 20. Significant markers were identified in validation data for all traits except RFI. Significant markers were identified on SSC6 and SSC7 for ADFI, and for most all chromosomes for ADG, FCR, BF, MD, IMF, BW and WW.

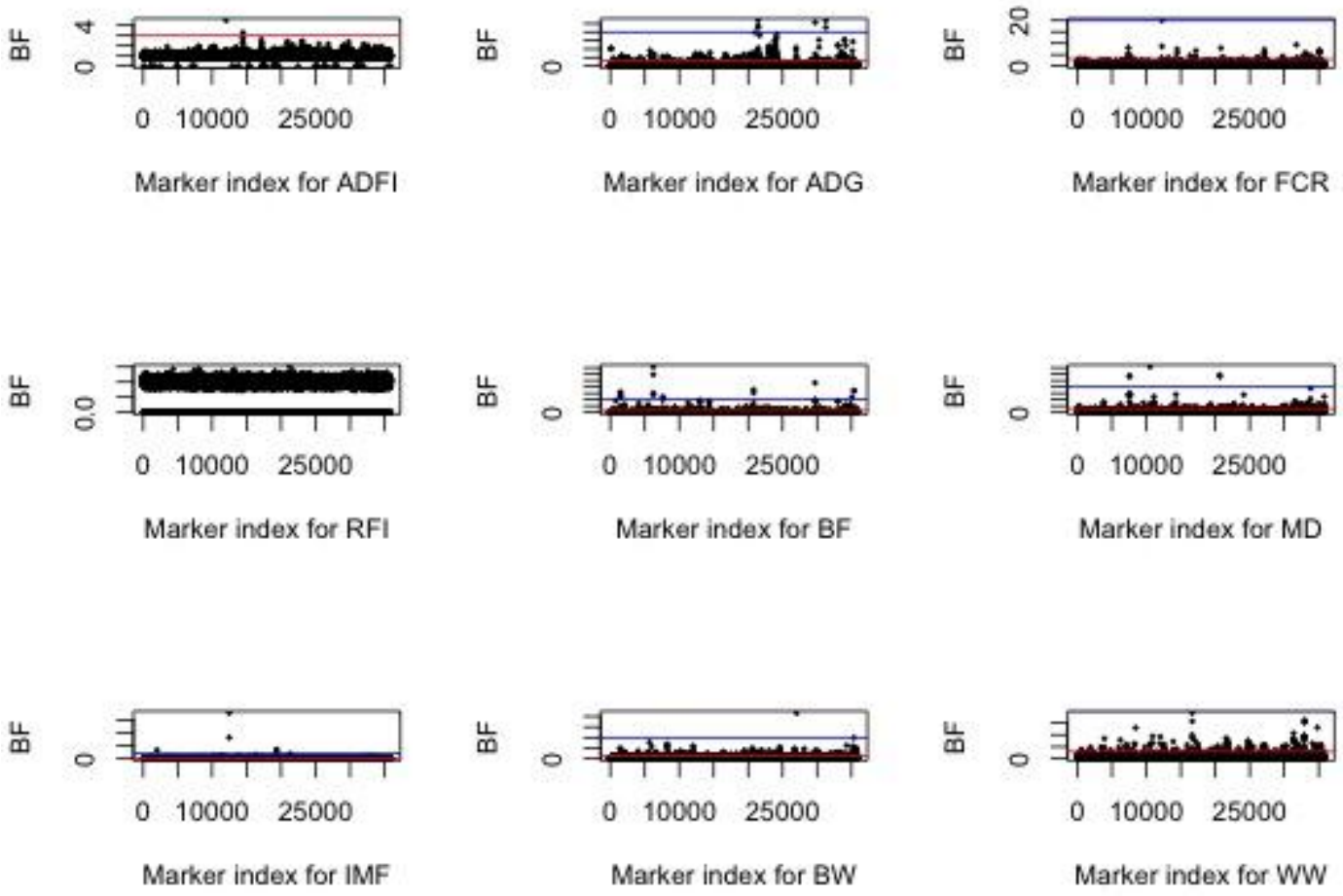


Fig. 8 Bayes Factor analysis for 9 traits in validation data

Correlation of proportion of genetic variance explained by each marker between training and testing data were investigated for 35,870 markers on SSC1 to SSC18 and X chromosome. Absolute value for correlations were less than 0.001 for ADFI, ADG, FCR, RFI, BF, MD, IMF, BW and WW, respectively.

Table 11. Candidate genes for all traits analyzed.

SSC	genes	Biological process and KEGG pathways
1	SOCS6	cytokine-inducible negative regulators of cytokine signaling, involved in regulation of growth
1	DOK6	play a role in the RET (MIM 164761) signaling cascade, insulin receptor binding
1	CNDP2	may involve in Metabolism of amino acids and derivatives, variants found in diabetes patients
1	HPGD	metabolism of prostaglandins, lipoxygenase pathway
1	OPRM1	negative regulation of cell proliferation, positive regulation of ERK1 and ERK2 cascade(growth factor signaling)
1	CYB5A	reduces ferric hemoglobin (methemoglobin) to ferrous hemoglobin, Metabolism, Vitamin Metabolism
6	DTNA	associated with various forms of muscular dystrophy
6	TMCO4	neural regulation
8	TECR*	may involve in process of fatty acid biosynthesis, Metabolism of lipids and lipoproteins
8	PHF17	histone acetylation, negative regulation of cell growth, response to stress, regulation of transcription, DNA-dependent
10	ACO1	involves in cellular iron homeostasis, aconitase activity, citrate metabolic process, intestinal absorption
10	B4GALT1	carbohydrate metabolic process, metabolism of proteins,
10	KIF5B	play roles in Diabetes pathways, insulin Signaling, insulin Synthesis and Processing, Adaptive Immune System
10	CFH	innate immune response, complement activation, alternative pathway
18	SLPI	immune response, negative regulation of endopeptidase activity
x	PNPLA4	may be involved in adipocyte triglyceride homeostasis ,lipid catabolic process, obesity
x	STS	catalyzes the conversion of sulfated steroid precursors to estrogens during pregnancy, sphingolipid metabolism
x	PNPLA4	lipid catabolic process, triacylglycerol degradation

Discussion

In the present study, a GWAS using the pocineSNP60 BeadChip was performed using Bayes B model averaging with random SNP effects for pig feed efficiency and production traits, including ADG, ADFI, FCR, RFI and BF, MD, IMF, BW, WW in purebred Duroc population. The analysis of feed efficiency traits and production traits revealed promising results for marker-assisted selection to improve feed utilization and production traits of interest. All traits analyzed were moderately heritable at ~40%, which has been previously demonstrated by breeders that feed efficiency and production traits could be improved through selection.

Feed intake visit records from electronic feeders

Electronic FIRE systems are frequently used by pig companies, predominantly in the United States, to measure individual feed intake on group housed growing pigs. Due to multiple factors including rodent activity, moisture and dust in the environment as well as the behavior of the pigs, the electronic feeders are prone to malfunctions and may produce erroneous feed intake and body weights records (Chen, Misztal et al. 2010; Zumbach, Misztal et al. 2010). Individual records were used to compute average daily feed intake for each pig. The mean of ADFI for this age of Duroc boars was ~ 2 kg, similar results have been published by (Cai, Casey et al. 2008) for Yorkshire pigs. Due to the high error rate detected in the body weight of pigs using the FIRE system data weaning weights and weights taken when ultrasound data was recorded were used to estimate ADG. Weight gain from weaning until the ultrasound weight was assumed to be linear. The estimated ADG was included in the models for DFI adjustment as well as RFI computation. Developing of

methodologies in further studies to overcome these limitations would improve the quality of phenotypic data and increase the power of genome wide association study.

Genetic parameter estimation

Heritability for each trait in Table 4 was estimated using data in the present study. All traits were moderate to highly heritable. Genetic Variance obtained from this study was used as prior for the Bayes B association study.

Genomic regions associated with feed efficiency and production traits

In this study, Bayes B method was used to preform Genome wide association studies on 8 traits of interest, including 4 feed efficiency traits and 4 production traits. QTL regions associated with each trait were identified and potential candidate genes were found in these regions. Based only on the single marker analysis, QTL associated with ADG were found on SSC1, 4, 14; QTL associated with ADFI were on SSC1, 8, 10, 14, X; QTL associated with larger effects were identified on the X chromosome associated with FCR; QTL associated on BF were found on 1, 4, 6, 9, 10, 18, and QTL associated with MD were on chromosomes 1, 6, 10, X. Markers with large effect on IMF were identified on SSC17 and SSCX, explaining approximately 0.01% and 0.6% genetic variance by two single markers.

Potential candidate genes were found in these regions, the biology functions and pathways were summarized in Table 3, most of them were involved in metabolism regulation. A common QTL region affecting both ADG and ADFI were identified on SSC1, which previous studies have reported (Kim, Larsen et al. 2000; Fan, Lkhagvadorj et al. 2010). Two potential candidate genes *HPGD* and *GALNT7* identified on SSC14 may associated with ADG. GALNT7 codes for a protein that function as an enzyme in initiation step of O-glycosylation, which is in the O-gycan biosynthesis pathway and carbohydrate metabolic process (Bennett, Mandel et al. 2012). HPGD, involved in the lipoxygenase pathway (Cho, Hamza et al. 2005), may have the function of transforming growth factor beta receptor signaling (Yan, Rerko et al. 2004). Previously identified candidate genes associated with ADG was *TCF7L2* (Fan, Lkhagvadorj et al. 2010) which is located at ~ 135 Mb on SSC14, which is more than 100Mb away. Thus, *TCF7L2* may not be the case in this study. For ADFI, there is another region which may be associated with this traits located at 65 – 80 Mb, where *H-FABP* gene was in this region. H-FABP was previous reported associated with ADG by (C, Oliver et al. 2002; Ovilo, Clop et al. 2002). Therefore our study may provide new evidence for the association.

Common candidate genes for QTLs on for ADG and ADFI

In this study, a region on SSC1 was found to have strong association with both ADG and ADFI, although its effects on ADFI were not significant based on the 10 Mb significant test threshold from previously reported studies (Wolc, Arango et al. 2012). Both ADG and ADFI were mapped to the same region on SSC1 from 164 Mb to 172 Mb, where *SOCS6* (~ 168.99) and *DOK6* (169.78) were found in this region; however, the previously identified candidate gene *MC4R* which has been associated with ADG and (Kim, Larsen et al. 2000; Howard and Flier 2006; Fan, Lkhagvadorj et al. 2010) is located at ~178.5 position on SSC1, about 6.5 Mb away from this QTL region. Howard and Flier (2006) reported SOCS6 was involved in development of leptin and insulin resistance. It has been shown that SOCS6 may impair insulin receptor signaling (Gupta, Mishra et al. 2011). It has been concluded that the SOCS6 protein may be involved in the proteasome mediated degradation (Bayle, Lopez et al. 2006). DOK6 protein is an insulin receptor binding protein, expressed highly at neural and kidney development (Crowder, Enomoto et al. 2004; Kurotsuchi, Murakumo et al. 2010). Both the location and gene function strongly support *SOCS6* and *DOK6* may be involved in ADG and ADFI regulation. Detailed study of the markers around these genes revealed that markers around *SOCS6* and *DOK6* ranked highest, while markers around *MC4R* have smaller effects and explained a small amount of genetic variance. Because 6.5 Mb is not so far, there is no evidence that *MC4R* was not associated with both traits on the other hand.

Markers in the overlapping QTL regions have been investigated for marker effects on both ADG and ADFI in this study (Fig. 6). Markers which have positive effects on ADG and negative effects on ADFI or have strong positive effects on ADG but smaller effects on ADFI may benefit the swine industry by increasing ADG while decreasing ADFI, or increasing both but increasing ADG more rapidly than ADFI.

Markers meeting the two scenarios and explaining a large amount of genetic variance are markers which could be used in further Marker-assisted selection.

Validation using additional 500 boars typed by reduced chip (10K)

Analysis of the validation dataset resulted in QTL regions which overlapped with those previously identified. However, the correlation among the estimated contribution to genetic variance of single markers between the validation data and training data was negligible. This may have resulted from miss imputed marker genotypes, monomorphic markers in the imputed genotypes due to small sample size, or a different number of markers fitted in Bayesian model. As the validation dataset was received a short time before this report was due we continue to work on the analysis of this information. To confirm the association results, further investigations are still needed.

Conclusion

In this study, the genomic prediction model Bayes-B was used to identify genomic regions associated with feed efficiency traits, such as average daily gain (ADG), average daily feed intake (ADFI), feed conversion ratio (FCR), residual feed intake (RFI), swine production traits, including ultrasound back fat thickness (BF), muscle depth (MD), inner muscular fat percentage (IMF), birth weight (BW) in a pure breed Duroc nucleus herd. QTLs were identified for most traits of interest and potential candidate genes were located in QTL regions on SSC1, 4, 8, and 14. QTL regions on SSC1 were found to explain a large proportion of the genetic variance for both ADG and ADFI, markers in this region may be used to improve ADG and ADFI by marker-assisted selection. It was concluded that significant genetic variation in traits influencing nutrient utilization exists. This supports the idea that whole genome selection using molecular markers should improve accuracy of selection.

Acknowledgements

This study was supported by National Pork Board, Smithfield Premium Genetics, National Swine Genome Coordinators Office and North Carolina Agricultural Research Service.

References

Bayle, J., S. Lopez, et al. (2006). "The E3 ubiquitin ligase HOIL-1 induces the polyubiquitination and degradation of SOCS6 associated proteins." FEBS Lett **580**(11): 2609-2614.

Bennett, E. P., U. Mandel, et al. (2012). "Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family." Glycobiology **22**(6): 736-756.

C, O. V., A. Oliver, et al. (2002). "Test for positional candidate genes for body composition on pig chromosome 6." Genet Sel Evol **34**(4): 465-479.

Cai, W., D. S. Casey, et al. (2008). "Selection response and genetic parameters for residual feed intake in Yorkshire swine." J Anim Sci **86**(2): 287-298.

Casey, D. S., H. S. Stern, et al. (2005). "Identification of errors and factors associated with errors in data from electronic swine feeders." J Anim Sci **83**(5): 969-982.

Chen, C. Y., I. Misztal, et al. (2010). "Estimation of genetic parameters of feed intake and daily gain in Durocs using data from electronic swine feeders." J Anim Breed Genet **127**(3): 230-234.

Cho, H., A. Hamza, et al. (2005). "Key NAD+-binding residues in human 15-hydroxyprostaglandin dehydrogenase." Arch Biochem Biophys **433**(2): 447-453.

Crowder, R. J., H. Enomoto, et al. (2004). "Dok-6, a Novel p62 Dok family member, promotes Ret-mediated neurite outgrowth." J Biol Chem **279**(40): 42072-42081.

De Haer, L. C. M., J. W. M. Merks, et al. (1992). "A note on the IVOG®-station: a feeding station to record the individual food intake of group-housed growing pigs." Anim. Prod. **54**: 160-162.

Eissen, J. J., E. Kanis, et al. (1998). "Algorithms for identifying errors in individual feed intake data of growing pigs in group-housing." Appl. Eng. Agric. **14**(6): 667-673.

Fan, B., S. Lkhagvadorj, et al. (2010). "Identification of genetic markers associated with residual feed intake and meat quality traits in the pig." Meat Sci **84**(4): 645-650.

Gupta, S., K. Mishra, et al. (2011). "Suppressor of cytokine signalling-6 promotes neurite outgrowth via JAK2/STAT5-mediated signalling pathway, involving negative feedback inhibition." PLoS One **6**(11): e26674.

Hoque, A. and K. Suzuki (2008). "Genetic parameters for production traits and measures of residual feed intake in Duroc and Landrace pigs." Animal Science Journal **79**(5): 543-549.

Hoque, M. A., K. Katoh, et al. (2009). "Genetic associations of residual feed intake with serum insulin-like growth factor-I and leptin concentrations, meat quality, and carcass cross sectional fat area ratios in Duroc pigs." J Anim Sci **87**(10): 3069-3075.

Howard, J. K. and J. S. Flier (2006). "Attenuation of leptin and insulin signaling by SOCS proteins." Trends Endocrinol Metab **17**(9): 365-371.

Kim, K. S., N. Larsen, et al. (2000). "A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits." Mamm Genome **11**(2): 131-135.

Kizilkaya, K., R. L. Fernando, et al. (2010). "Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes." J Anim Sci **88**(2): 544-551.

Kurotsuchi, A., Y. Murakumo, et al. (2010). "Analysis of DOK-6 function in downstream signaling of RET in human neuroblastoma cells." Cancer Sci **101**(5): 1147-1155.

Lo, L. L., D. G. McLaren, et al. (1992). "Genetic analyses of growth, real-time ultrasound, carcass, and pork quality traits in Duroc and Landrace pigs: I. Breed effects." J Anim Sci **70**(8): 2373-2386.

McDonald, T. P. and J. A. Nienaber (1994). "Modeling feed intake in group-penned growing-finishing swine." Transactions of the ASAE **37**(3): 921-927.

Ovilo, C., A. Clop, et al. (2002). "Quantitative trait locus mapping for meat quality traits in an Iberian x Landrace F2 pig population." J Anim Sci **80**(11): 2801-2808.

Rothschild, M. F., Z. L. Hu, et al. (2007). "Advances in QTL mapping in pigs." Int J Biol Sci **3**(3): 192-197.

Suzuki, K., M. Irie, et al. (2005). "Genetic parameter estimates of meat quality traits in Duroc pigs selected for average daily gain, longissimus muscle area, backfat thickness, and intramuscular fat content." J Anim Sci **83**(9): 2058-2065.

Wolc, A., J. Arango, et al. (2012). "Genetic parameters of egg defects and egg quality in layer chickens." Poult Sci **91**(6): 1292-1298.

Yan, M., R. M. Rerko, et al. (2004). "15-Hydroxyprostaglandin dehydrogenase, a COX-2 oncogene antagonist, is a TGF-beta-induced suppressor of human gastrointestinal cancers." Proc Natl Acad Sci U S A **101**(50): 17468-17473.

Zumbach, B., I. Misztal, et al. (2010). "Use of serial pig body weights for genetic evaluation of daily gain." J Anim Breed Genet **127**(2): 93-99.