ABSTRACT

JIAO, SHIHUI. The Genetics of Feed Efficiency and Feeding Behavior in Pigs. (Under the direction of Dr. Christian Maltecca.)

The overall objective of this research was to study the genetics of feed efficiency and feeding behavior with the use of feed intake measures recorded by electronic feeding system. The first study aimed at determining if two alternative implementations (termed MI and MICE) of multiple imputation were more effective to adjust errors occurring in feed intake collected by electronic feeders than the well-established linear mixed model (LMM) approach. In our results, multiple imputation outperformed the LMM approach in all simulated scenarios with mean accuracies of 96.71%, 93.45% and 90.24% obtained with MI and 96.84%, 94.42% and 90.13% with MICE, compared to 91.0%, 82.63% and 68.69% using LMM, for daily feed intake. In the second study we investigated measures of intake and growth, to determine the potential of genomic information in improving the efficiency of swine production. Magnitudes of heritability from pedigree analysis were moderate for growth, feed intake and ultrasound traits; heritability estimates were 0.32 ± 0.09 for FCR (feed conversion ratio) but only 0.10 ± 0.05 for RFI (residual feed intake). Comparatively, heritability estimates using marker information by Bayes-A models were about half of that from pedigree analysis, suggesting “missing heritability”. Moderate positive genetic correlations between growth and feed intake (0.32 ± 0.05), backfat (0.22 ± 0.04), as well as, negative genetic correlations between growth and feed efficiency indicate selection solely on growth may lead to undesirable increases in feed intake, backfat and reduced feed efficiency. Accuracies of genomic prediction ranged from 9.4% for RFI to 36.5% for backfat, providing

new insight into pig breeding and future selection programs using genomic information. The use of molecular information to dissect the genetic architecture underlying efficient growth in pigs was the subject of the third study. A region (166-140 Mb) on SSC 1, approximately 8 Mb upstream *MC4R* gene, was significantly associated with ADFI, ADG and backfat, where *SOCS6* and *DOK6* are proposed as the most likely candidate genes. Another region affecting weaning weight was identified on SSC 4 (84-85 Mb), harboring genes previously found to influence both human and cattle height: *PLAG1*, *CHCHD7*, *RDHE2*, *MOS*, *RPS20*, *LYN* and *PENK*. The fourth study aimed at dissecting different measures of feed efficiency in their relationship with feeding behavior and growth, exploring the use of efficient methods for the prediction of genomic breeding values, and accounting for the social interactions among individuals. Non-heritable social interaction has been observed for traits associated with measures recorded by electronic feeders and it is suggested that there is a need to include those effects to reduce bias for genetic parameter estimation when the variance explained by social interaction has been found significant. After accounting for social interaction, RG (residual growth) and RIG (combined measure of RFI and RG) have been found as two good measures of feed efficiency due to their moderate heritability and strong genetic correlation with other production traits. Feeding behavior traits were found moderately heritable and some were highly correlated with feed efficiency, which are worth further investigation. Increased accuracies have been shown when apply single-step GBLUP over BLUP for feeding behavior, feed efficiency, growth and off-test traits in a validation setting. Further research has been carried out to identify the underlying genetic regions affecting measures of feed efficiency and feeding behavior. Genomic regions have been identified for ANVD

(average daily number of visits), AOTV (average occupation time per day) and FCR and candidate genes of those significant regions and regions approaching significant threshold for RG and RIG have been annotated. Finally, gene networks for all candidate genes located were built to investigate the relationships among genes in common pathways.

The Genetics of Feed Efficiency and Feeding Behavior in Pigs.

by
Shihui Jiao

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Animal Science and Poultry Science

Raleigh, North Carolina

2015

APPROVED BY:

_____	_____
Dr. Christian Maltecca Chair of Advisory Committee	Dr. Joseph P. Cassady
_____	_____
Dr. Todd See	Dr. Mark Knauer
_____	_____
Dr. Alison A. Motsinger-Reif	Dr. Kevien Potter

DEDICATION

To my parents.

BIOGRAPHY

Shihui Jiao was born in Zhengding of city Shijiazhuang, Hebei Province, China in1986, to Jinhua Jiao and Zhengxia Wang. She grew up with a younger brother, Shigeng Jiao and a large family of relatives in the small town full of historical temples and towers.

The love for animals drove her to make the first big decision in her life: to take animal science as her major in the top ranking agricultural university in China. As a college student, she thoroughly enjoyed taking diverse subjects as an animal science major, such as biology, genetics, nutrition, chemistry, physiology, microbiology, and of course biostatistics.

The strong academic atmosphere and her eagerness to explore the new led her to attend graduate school in China Agricultual University and worked on indentifying the causative gene for Arachnomelia Syndrome in cattle, a congenital, lethal genetic disorder. It was the desire to work as an animal breeder led her to pursue a doctoral degree in quantitatve genetics in animal science with the research project on improving feed efficiency traits in pigs with the incorporation of genomic information. During the 4 years of her Ph.D., she also obtained her master's degree in statistics at NC State with a concentration in statistical genetics.

ACKNOWLEDGMENTS

I would first like to acknowledge Dr. Christain Maltecca, as a great mentor and advisor, for taking me as a Ph.D. student and guiding me in my course studies and research at NC State University, especially his dedication to research and wisdom he has shared with me throughout the past 4 years.

I would also to thank my committee for their helps and inputs in my resarch. I enjoyed working with each member of the committee. I would also like to say a special thanks Dr. Joe Cassady for accepting me as a Ph.D. student at NC State, his professional, finantial and personal support during the first 3 years of my Ph.D. program, his willingness to be on my committee and for traveling so far for my final oral defense.

This work cannot be completed without the helps and supports from Dr. Kent Gray, Dr. Yijian Huang and all others in Smithfield Premimun Genetics. I would also like to take the chance to thank Dr. Kent Gray for offering me the opportunity to work at SPG as an student intern during the summer 2014.

A special thanks goes out to Dr. Gene Eisen and Dr. O. W. Robison for leading the genetics journal club and their willingness to share their wisdom and insigts in animal breeding, and also to my fellow graduate students and Dr. Francesco Tiezzi for many simulating discussions, many helps and many laughs.

Finally, I would like to thank my parents and my fiancé Zhaohui for all the love and support. Many thanks to my parents for setting exmaples for me through out my life, for being kind, being positive and being hard-working.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ADFI: average daily feed intake

$ADFI_a$: adjusted average daily feed intake

ADG: average daily gain

AFIV: average feed intake per visit

AFRD: average daily feeding rate

AFRV: average feeding rate per visit

ANVD: average number of visit per day

AOTD: average daily occupation time

AOTV: average occupation time per visit

BF: ultrasound backfat thickness

BLUP: best linear unbiased prediction

bp: base pairs

BV: breeding value

BW: body weight

BW_Off: off-test body weight

DFI: daily feed intake

DFI_a: adjusted daily feed intake

DFI_e: daily feed intake in error visits

EBV: Estimated breeding value

FCR: feed conversion ratio

FDR: false discover rate

FIRE: feed intake recording equipment (Osborne Industries, Inc., Osborne, Kansas, USA)

FIV: feed intake per visit

FRV: feeding rate per visit

GBLUP: genomic best linear unbiased prediction

GEBV: genomic estimated breeding value

GS: genomic selection

GWAS: genome-wide association study

HD: high-density

IBD: identical by decent

IBS: identical by state

IMF: ultrasound intramuscular fat percentage

IOVG: individual feed intake recording in group housing ()

LASSO: least absolute shrinkage and selection operator

LD: linkage disequilibrium

LMM: linear mixed model

MAF: minor allele frequency

MAS: Marker-assisted selection

Mb: Mage base pairs

MD: ultrasound muscle depth

OTD_e: daily occupation time in error visit

OTV: occupation time per visit

PPA: posterior probability of association

RFI: residual feed intake

RG: residual growth

RIG: combined measure of residual feed intake and residual growth

SNP: single nucleotide polymorphism

SEP: standard error of prediction

SNPBLUP: single nucleotide polymorphism best linear unbiased prediction

SSC: *Sus scrofa* chromosome

ssGBLUP: single-step genomic BLUP

QTL: quantitative trait loci

CHAPTER 1

Literature review

INTRODUCTION

The ultimate goal of most livestock industry is to improve profitability. In the past, this has been achieved effectively by the genetic improvement of the output traits (production traits) of economic importance, such as lean meat growth rate in pigs, milk yield in dairy cattle or egg production in poultry. Success has been witnessed in the past several decades in livestock production with an emphasis on selection for increasing production traits (Rauw et al., 1998). Since the profit is a function of both inputs and outputs, traits related to inputs of production such as feed efficiency traits have also long been recognized and received substantial attentions especially by the swine and poultry industries, for which cost of feed is easier to quantify.

The pork industry faces a daunting challenge when it comes to utilizing feed resource more efficiently for saleable pork production, since feed cost remains the largest variable cost in pork production, approximately 50 to 85% of total production cost (McLone and Pond, 2003; Hoque et al., 2009; Henman, 2003). Sustained price increases for feed and the growing demand for pork force the pork producers to use strategies that maximize feed efficiency in order to 'produce more with less'. Feed efficiency is an aggregate measure of the general individual's ability of converting energy from feed to growth and body composition while reducing the maintenance energy cost. Therefore, pigs with high feed efficiency are ideal to select in order to improve feed utilization in the future generations. However, individual feed intake is expensive to measure and feeding behavior may be potentially used as indicator traits due to less equipment, less costly and easer maintaining to measure than the use of conventional electronic feeding systems (Maselyne et al., 2015). Over the past decades there

has been an abundance of data collected by electronic feeding systems, which have been used and continue to be used to understand the basic physiology and genetics of feed efficiency and feeding behavior. This information is advantageous for developing selection programs that can be targeted to improve the efficiency of feed utilization in swine herds.

Individual feed intake

Individual feed intake is difficult and expensive to measure. For several decades, pigs from nucleus herds have been evaluated for feed intake in central test stations, where the pigs were penned individually. However, pigs in commercial herds are normally housed in groups. Potential problem of genotype x environment interaction have been found between nucleus and commercial herds (Merks, 1989), as pigs housed individually in central test stations eat more, grow faster, and are fatter than pigs housed in groups. This genotype x environment may hamper the selection since it is generally recommended that selection should be carried out under the environmental conditions in which the improved breed is destined to live in (Falconer, 1952) to achieve maximum response. To avoid this, efforts have been made to develop automatic feeding systems to collect individual feed intake on pigs housed in groups, with a single-space feeder equipped in each pen where pigs as pen mates share the same environment.

The development and availability of those feeding systems has greatly facilitated the feed intake collection process and these systems (Maselyne et al., 2015) are now widely used in research and breeding herds. Among the different feeding systems two are the most popular choices for growth-finishing pigs: FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, Kansas, USA) and IOVG (individual feed intake

recording in group housing, Insentec B.V., Marknesse, the Netherland), while the electronic sow feeder (Nedap, Groenlo, the Netherlands) is normally utilized for sows. The design of these feeders is similar in principle (Maselyne et al., 2015). The pig visiting the feeder is identified, and then feed consumption and duration of feeding time are recorded. Take FIRE system as an example, each pen is equipped with a feeder that allows pigs access to feed 24 h per day, but only one pig in the pen can eat at each time because of a protective crate. When a pig enteres the feeder, weight of the feed eaten, pig ID, body weight of the pig, duration time of feeder visiting, and feeder number are recorded. The second advantage other than the ability to measure individual feed intake for group-housed pigs is that some behavior measures can also be recorded (Maselyne et al., 2015; Maselyne et al., 2014). Additionally, traits are often measured over an entire test period, but electronic feeders measure each visit to the feeder so curves for the traits related to feed intake can be obtained. As a result, selection can be applied to curve parameters to change the shape of the curve (Schnyder et al., 2001; Schulze et al., 2002). More importantly, with the longitudinal measures recorded for feed intake, the underlying genetic mechanisms of the dynamics of feed intake and growth can be studied (Howard et al., 2015).

Despite the numerous benefits recognized in the use of automatic feeding systems, there were several issues reported regarding their use. First, it has been known that data collected by those systems contain substantial errors and outliers that are due to feeder malfunction and animal feeder interaction (De Haer et al., 1992; Eissen et al., 1998; Casey, 2003; Chen et al., 2010). Different types of errors contained in such records have been categorized by Eissen et al. (1998) and Casey et al. (2005) and the overall error rates found in

several recorded datasets ranged from 6% to up to 35% (Eissen et al., 1998; Casey, 2003; Jiao et al., 2014a). Secondly, one single-space feeder is normally placed in a pen in the performance testing nucleus herds whereas in commercial herds multiple-space feeder are normally employed, resulting in potential occurrence of genotype x environment. The single feeding space and the protection the feeder provides for the pig by the electronic feeding systems can change the feeding behavior compared to normal farm conditions with other type of feeders (Maselyne et al., 2015). From the results of the study by Nielsen et al. (1996), pigs visited the feeder much more frequently with shorter durations, eating significantly less per visit accessed to four-spaced feeder compared to pigs kept in the groups given access to a automatic single-space feeder. Lastly, the electronic feeding system is expensive, thus in order to maximize the use of the single-space feeder the number of pigs distributed in the same pen is also maximized. This could negatively impact the accuracy of feed intake measures (Brisbane, 2002; Eissen et al., 1999) and may cause social interactions among pen mates (Bergsma et al., 2008; Chen et al., 2010).

Automatic feeding systems are now used in genetic nucleus and research herds to improve or study traits related to feed intake. Although there are many benefits using these feeders as stated, the issues associated with those electronic feeding systems must be addressed in order to optimize the use of electronic feeders.

Inaccurate measures of feed intake and related measures may hamper the use of those records for further analysis or evaluation and different types of errors in such datasets were categorized, and then identified by several authors (Eissen et al., 1998; Casey et al., 2005; Chen et al., 2010). Data cleaning or editing is necessary and cannot be overlooked.

Previously reported error rates in feed intake collected by electronic feeders varied among different datasets. In a similar population of Duroc, Jiao et al. (2014a) found the overall error rates ranging between 14% and 35%. Eissen and colleagues (1998) reported error visits representing 6% of the total 385,329 feeding visits for 250 pigs. Similarly, Casey (2003) reported percentages of identified error visits of 4.33%, 5.62% and 18.74% for three different datasets with 863,590 total visits for 893 pigs, 290,073 total visits for 591 pigs and 162,638 visits for 237 pigs, respectively.

Due to the fact that simply deleting unreliable records would underestimate the true daily feed intake (Casey, 2003; Eissen et al., 1999), editing or adjustment methods have been established in order to obtain more accurate measures. A linear mixed model (LMM) has been proposed by Casey (2003) to adjust records containing errors after removing visits with missing values. This approach has proven especially useful when strategies that alternate the use of an electronic feeder between two pens of pigs every other week (Grignola et al., 2002) to maximize the number of pigs tested per feeder is employed. However, negative impacts using those strategies have been reported by several authors (Schulze et al., 2001; Von Felde et al., 1996), mainly due to stress in the tested pigs. In spite of this, accuracy of this adjustment model were reported higher than the method used by Eissen et al. (1999). However, applying LMM adjustments to a dataset with extreme values remains a challenging task since those values tend to severely bias the estimates (Osborne and Overbay, 2004).

The LMM approach for error adjustment presents additional limitations. First, the model cannot handle missing data in the feeding visits, which need to be deleted when fitting the mixed model equation. Secondly, the model is relatively cumbersome and data need to be

processed to construct covariates associated with each error types before the actual mixed model application. And lastly, subjective constraints for daily feed intake or occupation time with errors need to be used to limit the bias arising from influential or extreme values in the estimation and prediction.

In order to utilize serials body weight measures collected using FIRE feeders to obtain growth rate, regression models established using robust regression have applied with better performance compared to least squares (Zumbach et al., 2010). Growth rate obtained by robust regression, have been used in genetic analyses by several authors such as Chen et al. (2010), Jiao et al. (2014a) and Howard et al. (2015).

Given that adjustment methods based on LMM are problematic, alternative editing methods are needed to best use the data collected by electronic feeding systems. Multiple imputation might be in this case a viable option. Multiple imputation was introduced by Rubin (1976), as a method with the very general task of 'filling in' missing values. This approach has gained increasing popularity and the past several decades have seen implementations spanning many areas of statistical analysis (Rubin, 1996; Allison, 2002). The key concept of this technique is the use of the distribution of the observed data to estimate a set of plausible values for the missing data. Its generality and recent software development makes it a potentially advantageous approach in feed intake data edits since it treats errors as unobserved missing values. In contrast to LMM, multiple imputation is very general and can be easily implemented in a variety of settings with minimal data preprocessing or ad hoc adjustments since the development of software packages (Su et al., 2009; van Buuren and Groothuis-Oudshoorn, 2011).

Measures of feed efficiency

Feed intake and its utilization span a series complex biological processes and physiological pathways, along with interactions and response to the environment conditions (Arthur and Herd, 2005). As a reflection of this complexity, feed efficiency does not refer to one trait, but rather encompasses all traits associated with the efficiency of feed utilization. Furthermore, differences in the formulae used to compute a particular feed efficiency trait may lead to differences in estimates or even selection outcomes (Herd and Arthur, 2005; Hoque et al., 2009). Different measures of feed efficiency have been used in literatures, such as feed conversion rate (FCR) or its reciprocal (feed : gain ratio), residual feed intake (RFI) (Koch et al., 1963), residual growth (RG) and combined measure of residual feed intake and residual growth RIG (Crowley et al., 2010; Berry and Crowley, 2012). Feed conversion ratio was often defined as the ratio of growth rate in kg over feed intake in kg or the reverse of the ratio (termed feed gain ratio). The measure of residual feed intake was introduced by Koch et al. (1963) and defined as the difference of actual feed intake and the predicted feed intake. The predicted feed intake can be obtained by considering the size of the individual (body weight or metabolic mid body weight) and its production, such as growth rate, backfat and muscle compositions. Therefore, residual feed intake was thought to capturing the energy expenditure in maintenance. However, residual feed intake has its drawback that the slow growing animal with less feed intake may appear to be efficient. To overcome this problem, residual growth was invented as the difference of the actual growth and predicted growth rate, the predicted growth was computed based on the feed intake and tissue composition of this animal (Crowley et al., 2010). Thus, RG can be used as an indicator of efficiency given a

certain level of feed intake and production performance. Berry and Crowley (2012) introduced RIG as a combined measure of residual feed intake and residual growth in order to combine both advantages of the two measures of efficiency. They defined RIG as the difference of residual growth and residual feed intake, both scaled to mean 0 and standard deviation 1 to remove the potential calling problems.

The two most common measures of feed efficiency are FCR and RFI, while RG and RIG are seldom used as efficiency measures in swine research (MacNeil and Kanp, 2015) albeit they have been employed both in poultry (Willems et al., 2013) and beef cattle (Crowley et al., 2010). Residual feed intake is commonly used as standard measure of feed efficiency at a given level of production, so that the predicted feed intake can be defined by different sets of production traits, as suggested by several authors (Johnson et al., 1999; Hoque et al., 2009; Do et al., 2013). However, the relationships between feed efficiency with other traits may not consistent across different of stages of maturity (Archer et al., 2002) or different populations or breeds (Do et al., 2013).

The two commonly used measure of feed efficiency FCR and RFI suffer from several problem. For example, FCR (or Feed : Gain ratio) has a close correlation with both feed intake and rate of gain (Carstens et al., 2003); this may lead to animals with heavier mature weights and greater maintenance requirements as animals with similar FCR may differ greatly in their rate of gain and feed intake (Smith et al., 2010). Additionally selection based on ratio traits may result in different response in the component traits and cannot be accurately predicted in future generations (Gunsett, 1984). In contrast, RFI is independent of level of production phenotypically (such as body size, growth and backfat thickness). In a

study carried out by Campo and Turrado (1998) RFI was employed as trait to decrease the FCR and was shown to be better than direct selection on the ratio in chicken. Nonetheless, residual feed intake (RFI) is still genetically correlated with the production traits (Kennedy et al., 1993). Additionally, this trait may lack acceptance by producers because slow growing animals eating relatively less of feed may actually have good RFI. Crowley et al. (2010) and Berry and Crowley (2012) argued that RG and RIG may be better measures of feed efficiency because improved RG is associated with faster growth rate on average, given same level of feed intake and RIG by combining RFI and RG measures should be able to retain advantages of both. Indications exist that RG and RIG might be two good measures of efficiency and feed utilization in beef cattle (Crowley et al., 2010) and turkey (Willems et al., 2013). However the genetics of the two traits, the genetics relationships between these and other traits in pigs remain to be determined.

Assessing various measures of feed efficiency in pigs may shed lights on how to choose a particular useful feed efficiency measure in the future selection program. The genetics of different measures of feed efficiency and interrelationships with other important economic traits may provide new information and new understanding of the various measures.

Selection for feed efficiency

It is has been long recognized that feed efficiency is of great economic importance to the swine industry. In the past decades most improvement in feed efficiency has been made indirectly by selection on lean growth rate (Cleveland et al., 1983; McPhee et al., 1988). However undesirable side effects have surfaced with intense selection for lean growth rate in

pigs (Rauw et al., 1998; Lonergan et al., 2001), resulting in reduced pork quality and increased reproduction and health problems. On the other hand, only approximately 65% of phenotypic variation in feed intake can be accounted for by growth rate and other performance traits, such as backfat thickness or muscle depth (Cai et al., 2008). The remaining phenotypic differences can be evaluated by feed efficiency (as based on different definitions). As leanness approaches optimum levels, direct selection for feed efficiency can create more incentives for producers in swine production, in which case individual feed intake is needed on a large number of animals in the breeding population.

Moderate heritability estimates for two main measures of feed efficiency FCR and RFI have been reported in previous studies. Heritability estimated for FCR have been shown to moderate to high, ranging from 0.12 to 0.58 with an average of 0.30, whereas reported estimates for RFI have been slightly lower, from 0.1 to 0.42 with mean 0.24 (Johnson et al., 1999; Rothschild and Ruvinsky, 2010; Cai et al., 2008; Gilbert et al., 2008). Recently Do et al. (2013) reported estimates for FCR of approximately 0.30 for Danish Duroc, Landrace and Yorkshire with two measures of RFI heritabilities ranging from 0.34 to 0.40 across breeds. Likewise, MacNeil and Kanp (2015) reported an estimate for RFI with 0.22 for Canadian Duroc boars. Little has been known regarding to the estimates of heritability for RG and RIG in pigs, since the two measures of feed efficiency were seldom used in previous studies, but in beef cattle (Crowley et al., 2010) and poultry (Willems et al., 2013). MacNeil and Kanp (2015) reported the heritability for RG was 0.21 in a Canadian Duroc population with 3291 growing-finishing boars phenotyped using IOVG feeder (individual feed intake recording in group housing, Insentec B.V., Marknesse, the Netherland). No estimates for RIG have been

reported in pigs. However, Willems et al. (2013) reported the estimates of RG and RIG in turkey were 0.19 and 0.23 and Crowley et al. (2010) found that RG was moderately heritable in beef cattle.

Genetic correlations between FCR and RFI and other production traits have been reported in different breeds/populations. Genetic correlations with RFI and growth rate have been generally small. Hoque et al. (2008) reported genetic correlation between daily gain (DG) and RFI of -0.05, DG and FCR of -0.22 and FCR and RFI of 0.95. Negative correlations (-0.37 and -0.54) have been also found in two feeding regimes between average daily gain (ADG) and FCR by Schulze et al. (2001). However, Mrode and Kennedy (1993) found weak positive genetic correlation between growth measures and RFI, ranging from -0.18 to -0.34. Genetic correlation between efficiency and backfat has been show to be low when residual feed intake is adjusted for backfat (-0.15) but intermediate when adjusted for average daily gain (-0.34), and high when adjusted for lean growth (-0.61) in a study from Mrode and Kennedy (1993). Do et al. (2013) reported moderate to negligible genetic correlation between FCR and BF (-0.36 and -0.03) in Danish Duroc and Yorkshire population but not in Landrace where a moderate genetic correlation between FCR and growth (-0.38 and -0.31) was found in Danish Duroc and Yorkshire and of opposite direction in Danish Landrace (0.26).

The effectiveness of direct selection for feed efficiency has been demonstrated by several selection experiments for feed conversion rate and residual feed intake (Bernard and Fahmy, 1970; Cameron and Curran, 1994; Cai et al., 2008). In pig breeding program, feed efficiency or its inverse has been incorporated in selection objectives and selection criteria

(de Vries and Kanis, 1992). Cai et al. (2008) reported how selection for reduced RFI significantly decreased the feed requirement in pigs at a given rate of growth and backfat thickness in comparison to a random control line in a 5-generation selection experiment in Yorkshire pigs. Another divergent selection experiment against high RFI in Large White pigs found higher carcass percentage and weight of loin in low RFI line whereas the line divergence in growth, body weight and backfat thickness was close to zero after 3 generation of divergent selection for RFI (Gilbert et al., 2006).

Feeding behavior

Phenotypes of feeding behavior have been studied in the past, focusing on the mechanisms of free feeding (Young and Lawrence, 1994) in pigs and the relationship between feeding behavior and growth or feed intake. Analysis of feeding behavior were mainly concentrated on testing hypothesis of controlling food or diet selection, including the roles of hunger and satisfaction (Tolkamp et al., 2002), on predicting illness (Cornou et al., 2008; Sowell et al., 1998), or on identifying certain behavior traits for incorporation into selection index (Rauw et al., 2006; Do et al., 2013) in order to get indirect selection response on feed efficiency.

Feeding behavior is often referred to ‘putting the head into the through’ (or a feeding visit) instead of ‘chewing the food’, because the later one is a more difficult measure to record without interfering with the animal (Maselyne et al., 2015). The feeding event can be measured on animals individually housed or group housed. The use of electronic feeding system on farms makes individual feed intake easier to measure for group housed pigs (Maselyne et al., 2015) allowing also to collect feeding behavior measures, such as duration

of each feeding visit and the body weight of the animal. One potential problem with the use of electronic feeders is that to maximize the usage of feeder, one single-space electronic feeder often is equipped in a pen with a large group of pigs. This may introduce social interactions like competition or cooperation among pigs sharing the same pen. Social interactions among pen mates have been reported by several authors in feed intake and growth of pigs (Arango et al., 2005; Bergsma et al., 2008; Chen et al., 2010) and in feed consumption of dairy cattle (Huzzey et al., 2014).

Previous studies have shown that group size affects feeding behavior. Walker et al. (1991) reported significant difference in occupation time per visit and percentage of days the feeder were occupied by a pig with different group size of 10, 20 and 30 pigs per pen. The results indicated that as the group size increase, occupation time per visit by a pig decrease but the time the feeder were occupied by the same pig increased. Hyun and Ellis (2001) evaluated the influence of group size on different feeding behaviors using single-space FIRE feeders in growing pigs and found that number of visit per day and occupation time per day decreased whereas feed intake per visit and feeding rate increased as group size increased. Similar results were also found by follow up experiments (Hyun and Ellis, 2002). Altered feeding behavior were also found with different group size but same space allowance per pig by Neilson et al. (1995), and the results indicated that pigs in large groups had fewer visits to feeders, longer occupation time per visits and faster feeding rate. The results from those studies suggest that competition exists in pigs sharing a single-space feeder and pigs may adjust their feeding behavior based on group size.

The relationship between feeding behavior (summarized from feeding events) and feed efficiency has been investigated in previous publications (Do et al., 2013; Rauw et al., 2006; Rohrer et al., 2013; von Felde et al., 1996; Haer et al., 1993; Labroue et al., 1997), but limited research has been done on the relationships between behavior measures and different measures of feed efficiency, especially when social interaction is considered.

Little was known regarding the genetics of behavior traits and the relationship with other economically important traits in large populations. This lack of information might reflect the fact that collecting individual feed intake is expensive. Feeding behavior traits have been found to be moderate to high heritable. Do et al. (2013) reported that heritabilities of six feeding behavior traits were all moderate to high and consistent across Danish Duroc, Yorkshire and Landrace breeds. Similar results have been seen in other studies (Rauw et al., 2006; Labraue et al., 1997). However, Rohrer et al. (2013) found the estimates of heritability for feeding behavior traits were variable (ranging from 0.16 to 0.60) in a pedigreed Landarace-Yorkshire-Duroc composite population. Is should be noted though that in that case a small number of animals were measured (n = 1162). Some of the feeding behavior measures have been shown to be strongly genetically correlated with growth and feed efficiency traits, although the strength of the correlations depends on breeds and population (Do et al., 2013; Rauw et al., 2006; Labraue et al., 1997).

It would be possible to use feeding behavior measures as selection criterion to improve feed efficiency, if feeding behavior traits were strongly genetically correlated with feed efficiency in the population investigated. Devices have been developed to record individual feeding behaviors without measuring individual feed intake (Maselyne et al.,

2015; Maselyne et al., 2014). Those devices would be less costly and easier to maintain than electronic feeding systems (Brown-Brandl et al., 2013).

Genomic evaluations of feed efficiency and feeding behavior

Nearly all traits of great economic importance in domestic animals are for the most part quantitative and complex in nature, such as milk yield in dairy cattle, growth rate in pigs and egg production in chicken (Falconer and Mackay, 1996). The classical model of quantitative traits, introduced by Fisher (1930), states that a phenotypic value is controlled by an infinite number of genes each with an infinitesimal small effect as well as by non-genetic or environmental factors. Under these premises a great effort has been dedicated to predict the total additive genetic value of individuals, or termed as their estimated breeding value (EBV) through the used of BLUP mixed models (Henderson, 1953) and selection index theory (Hazel, 1943). Exceptional genetic improvement has been made for most livestock production traits under this framework. Nonetheless the genetic architecture of the trait investigated is treated as a black box in BLUP models, with no knowledge of the number of genes that affect the trait, their individual additive and dominance effects, or the degree of epistasis (Dekkers and Hospital, 2001). The traditional EBV predictions have relied heavily on the animals' phenotypes and their relationships derived from the pedigree information. However, traditional selection becomes less efficient for traits that are sex specific (milk yield, number of born), measured late in life (longevity), expensive to measure (disease resistance, feed efficiency), can only be measured after sacrifice (meat quality, carcass traits), or have a low heritability (fertility) (Dekkers and Hospital, 2001). To date individual feed

intake is still a character expensive to measure, thus traditional selection for feed efficiency becomes less efficient.

In the recent past, researcher have tried to identify genes or quantitative trait loci (QTL) on the constructed linkage maps of major livestock, after the discovery of a few number of genes of very large effect, such as the effect of the gene affecting halothane sensitivity in pigs (Lundstrom et al., 1995). Between the end of the 20th and beginning of 21th century, thousand of QTL have been located in pigs, associated with economically important traits such as growth rate, leanness, feed intake, meat quality and disease resistance (Rothschild et al., 2007; Ernst et al., 2013). Although a lot of effort has been dedicated to mapping QTL in livestock (Rothschild et al., 2007), the discovery of the underlying mutations affecting certain traits and implementation of discovered QTL through marker assistant selection (MAS) has been limited (Dekkers, 2004) because of rare replication of QTL studies, the limited variation explained by the identified QTL, and difficulty in determining economic weight for a single marker.

The discovery of a many thousands, and even millions of single-nucleotide polymorphisms (SNP) has greatly improved our ability to use information on DNA polymorphisms to select livestock (Dekkers, 2012). These SNPs can be genotyped in a cost-effective way by modern SNP-chip genotyping technologies. For most of the major livestock species (pigs, chickens, sheep, and cattle), low-cost SNP chips with approximately 50,000 genome-wide SNPs are currently available. Although these SNPs are unlikely to directly cause functional genetic differences between animals, they can serve as genetic markers, flagging the genome to track the surrounding regions, which may contain functional

mutations controlling the phenotype of interest. The availability of these dense marker information, provide us more room to investigate the underlying genetic architectures of traits of interests and the ability to conduct genome-wide selection (Meuwissen, 2007).

The landmark paper from Meuwissen et al. (2001) marks the start of the genomic selection (GS) era. After this publication, there has been a rapid development of genomic selection tools and genomic selection to implement GS in animal breeding programs of various species, including dairy, beef, pig, layer chicken, etc (Meuwissen et al., 2013; Hayes et al., 2009; van Eenennaam et al., 2013; Wolc et al., 2012). In contrast to traditional selection (BLUP), GS relies on a large number of individual in a genotyped and phenotyped training population from which a genomic prediction equation is derived that can then be used to estimate the genomic estimated breeding value (GEBV) of unphenotyped individuals as a selection candidate based solely on their marker genotypes. Thus, GS is more efficient for traits that are sex specific, measured late in life, expensive to measure, or have a low heritability. The use of GS has been fully applied to dairy industry (VanRaden et al., 2008) and is implemented to beef (Garrick, 2011), swine (Abell et al., 2014) and poultry (Wolc et al., 2012) industries.

Unlike dairy industry, implementing the genomic selection tool is hindered by the ratio of the value of an individual to the cost of its genotyping in other species including pigs (Abell et al., 2014; Habier et al., 2009). To reduce genotyping costs low-density genotyping strategies have been implemented (Huang et al., 2012; Habier et al., 2009; Cleveland and Hickey, 2013).

There are many methods developed for the estimation of the genomic breeding

values. A direct method to incorporate genomic information into the traditional BLUP framework is to replace the numerator relationship matrix (A matrix) with a realized or genomic relationship matrix (G matrix) (VanRaden, 2008), termed GBLUP. The A matrix is based on the expected additive variance-covariance relationship matrix based on the pedigree, which is twice the coancestry between individuals. The coancestry or coefficient of kinship is the probability that the alleles/genes carried by the two individuals descend from the same ancestor (IBD) (Nejati-Javaremi et al., 1997). The expected relationship does not account for Mendelian sampling, thus full siblings are assumed to have the same relationship with their parents. Additionally, the founders of a pedigree are assumed to be unrelated with zero inbreeding. In contrast, the G matrix is derived from alleles being identical by state (IBS), thus estimating the realized proportion of the genome that two individuals share including the base animals (Goddard et al., 2011; Hayes et al., 2009). The G matrix picks up distance relationships prior to recording of the pedigree and Mendelian sampling is accounted for, which allows for variation in progeny genotypes to be determined, in particular full or half siblings (Nejati-Javaremi et al., 1997). In field study, the GBLUP approach is attractive because the implementation is straightforward using existing BLUP routines and small computation demands (VanRaden et al., 2008).

Alternative methods rely on the estimation of specific marker effects. Genomic selection assumes that every QTL (large or small) can be explained by several nearby markers, which together explain all the genetic variation caused by QTL. The idea of genomic selection is thus to fit all markers simultaneously in the model, irrespective of whether they are statistically significant (Meuwissen et al., 2001). In the easiest

implementation (SNPBLUP) all markers are fitted so that each has the same variance. Habier et al. (2007) and VanRaden (2008) noticed that the this model and GBLUP model are equivalent. The SNPBLUP model makes the assumption that all SNP effects are independent and identically distributed, coming from the same normal distribution with small variance. Thus the SNP effects estimated from the model is typically very small and are very sensitive to the prior. Furthermore, SNP effects are normally distributed with constant variance may be unrealistic: some SNP in LD with causal mutations may capture the QTL effects thus have large effects while other SNP may have no effects. Therefore, various genomic prediction models have been developed with different prior assumptions for SNP effects (Meuwissen et al., 2001; Habier et al., 2010; Tibshirani, 1996, Whittaker et al., 2002, Park and Casella, 2008) and those approaches have been called the Bayesian alphabets (Gianola et al., 2009). In the BayesA model, the SNP effects are assumed to have a *t* distribution (Meuwissen et al., 2001). This allows some SNP to have larger effects than they do under an assumption of normality, because a *t*-distribution has fatter tails than a normal distribution. In the BayesB model (Meuwissen et al., 2001), a fraction (p) of the SNPs is assumed to have no effect, and a fraction ($1 - p$) has an effect drawn from a *t*-distribution. BayesC (Habier et al., 2011) is similar to BayesB except that the SNPs with an effect are assumed to have a normally distributed effect instead of *t*-distribution. BayesC is computationally simpler to solve than BayesB and yields almost similar accuracies. BayesR is an extension of BayesC with multiple normal distributions, which allows good prediction under a range of genetic architectures, including for traits where there are genetic variants of moderate effect (Erbe et al., 2012). Other commonly used prior distribution for the estimation of SNP effects are

double-exponential distribution (also called the Laplace distribution), which result in Bayesian least absolute angle and selection operator (Bayesian LASSO) (Park and Casella, 2008). de los Campos et al. (2009) further extended Bayesian LASSO to accommodate pedigree in formation as well as covariates other than markers. In summary, the main difference between the methods is their prior distributions and their assumptions about the distribution of marker effects. de Los Campos et al. (2013) reviewed the relationships among the Bayesian models and some models can be considered special case or limiting case of others.

The various models (GBLUP, BayesA, B, C, R and Bayesian LASSO) have been compared with simulation and real data analysis. In general, the method with assumptions about the distributions of SNP effects that most closely agree with the true distribution gives the highest accuracy of GEBVs (Meuwissen et al., 2014). Higher prediction accuracy often observed when SNP effects have prior distributions with spike at 0 than distributed with equal variance, however, GBLUP yield similar accuracy in real data analysis (Hayes et al., 2009; Meuwissen et al., 2014).

Usually the number of animals genotyped is far less than the number of animals that would be in the pedigree. However, the above motioned genomic prediction methods can only incorporate genotyped individuals in the mixed model equations. To overcome this limitation, a covariance matrix that has both A matrix and G matrix would be needed. Misztal et al. (2009) and Legarra et al. (2009) augmented the A and G matrix and this resulted in a joint distribution of genotyped and ungenotyped genetic values, with a pedigree-genomic relationship matrix H. In this matrix, genomic information is transmitted to the

covariances among all ungenotyped individuals. This extension of genomic relationship matrix as combination of A and G matrix can be used in BLUP procedure, and is generally termed single-step genomic BLUP (ssGBLUP). Single-step GBLUP (or ssGBLUP) has been used widely in different species for routinely genetic evaluations (Misztal et al., 2013) with its advantages being easy implementation and the use of standard BLUP routines with genotyped or non-genotyped animals. The increased accuracies of predictions by using ssGBLUP incorporated with genotypes over pedigree-based BLUP have also been reported by Christensen et al. (2012) for growth and feed intake and by Forni et al. (2011) for sow litter size in analyses using swine data. Similar conclusion has also seen from Guo et al. (2015) in genomic evaluation of litter size and piglet mortality in Danish pigs using ssGBLUP.

One of the advantages of utilizing estimated marker effects derived from a training population is that GEBV can be predicted on individuals without phenotypes for multiple generations, which results in decreased cost and labor during phenotype collection and the generation interval is reduced because GEBV can be predicted as soon as they are born or in-utero under extreme situations. The inability of accurate marker effects to persist across multiple generations (Goddard 2009; Habier et al, 2007) and divergent populations or prediction of crossbred individuals (Kizilkaya et al. 2010; Kachman et al. 2013; Ibanez-Escriche et al. 2009) is still an active research topic. The primary reason for the decrease in marker effect accuracy is LD changes across time and population, thus leading to a lower correlation between the marker and QTL.

Genome-wide association study

The accuracy of prediction for future phenotypes or genetic merit using high-density SNP makers also depends on the genetic architecture of the trait of interest (Hayes et al., 2010). With more dense genomic information, the genetic architecture of a trait can be better understood, such as the number of genes/loci that affect a trait and the additive effects of individual genes/loci. Identifying genes for complex traits would greatly enhance our understanding of the traits. For livestock, there would be a practical benefit in locating genes causing the phenotypes to further facilitate breeding and selection (Hayes and Goddard, 2010).

Genome-wide association study also known as LD mapping is used to identify trait-marker association relationships based on linkage disequilibrium (LD). The association study needs both phenotypes and genotypes of a sample of animals. Since most of traits of importance in livestock is polygenic in nature, that the phenotypic variance can explained by a lot of variants of tiny effects, the association analysis unavoidably encounter the 'small n big p' problem (de los Campos et al., 2013). Several methods have been adopted to overcome this problem in order to identify genes associated with the trait of interest. The simple linear model (Kang et al., 2010) was commonly used to carry out the genome wide association analysis with single SNP effect fitted as fixed effect in the mixed model with a polygenic breeding value of each animal, accounting for several forms of confounding including the population structure and familial relatedness or the genetic background other than this SNP marker. Alternatively, practical and theoretical advantages of using Bayesian approaches for the assessment of whole genome-wide association have been shown in recent papers (Lunn et

al., 2006; Wakefield et al., 2007; Hosking et al., 2008; Fridley et al., 2009; Wakefield et al., 2008). Benefits of exploring Bayesian models in the contexts of genome-wide association studies were reviewed thoroughly by Stephens and Balding (2009): Bayesian approach considers the background knowledge of relationship between genotype and phenotypes (biological information); incorporates uncertainty in the model (flexibility of the model) and results in probability statement about the parameters; allows variable and model selection in a natural way.

Identifying genes or markers in domestic animals for quantitative traits with economical values is growing in importance with the advent of high throughput genotyping methods. Attention on methods by which the number of false positives can be controlled without missing too many scientifically interesting associations is drawn increasingly. Thus over the last few years several more or less conservative measures of evidence in GWAS under Bayesian models have been employed (Fan et al., 2011; Wolc et al., 2012; Veerkamp et al., 2012) for single SNP or genomic segments covering multiple loci taking account of LD among SNPs, however often with little justification on the choice of method, such as Bayes factor (Veerkamp et al., 2012), Posterior probability of association and variances explained by SNP winodws (Wolc et al., 2012) and non-parametric bootstrapping methods (Fan et al., 2011). In human, the use of Bayes factor as summary of the evidence given the observed data has been advocated in both medical (Goodman et al., 1999) and epidemiology study (Balding et al., 2006). Wakefield (2009) described approximate Bayes factors that gave same rankings of p-values and suggested that Bayes factor is an alternative of p-value. However, Stephens and Balding (2009) argued posterior probability of association is a more

appropriate measure of evidence compared to Bayes factor based on Bayesian contexts. Oppose to parametric methods, permutation and bootstrapping use data itself to obtain empirical estimate of the maximal distribution. Both permutation and bootstrapping methods proceed by resampling data under the null hypothesis (Nichols and Hayasaka, 2003).

Success in genome wide association study also relies on the ability to account for population structures (Pritchard, 2000) as failure to do so can cause spurious associations (false positives) in the genome wide association study. Population structure can strongly influence patterns of LD within a genome. Unlike human population, in livestock population structure can be created by multiple offspring per sire, selection for specific breeding goals and breeds or strains within the population, population mating patterns and admixture (Falconer and Mackay, 1996). Consequently, the highly structure nature of livestock population can reduce the number of animals used to conduct association mapping since relative information can be used in addition to its own records (Hayes and Goddard, 2010). As another advantage in contrast to human, there is no need to use SNP panels as dense as human's to conduct association mappings using samples from one breed in those livestock species. However, on the other hand, a long extent of LD may be problematic for fine mapping of a QTL, because association may be detected at a distance far from the causative locus (Hayes and Goddard, 2010).

To fully understand complex quantitative traits and accurately predict breeding values using whole genome molecular markers, it is essential to reveal the genetic architecture underlying these phenotypes. Many genome wide association analyses have been conducted due to the availability of dense marker panels and the development of statistical tools,

including a few studies tried to locate the genes for feed efficiency in pigs. Onteru et al. (2010) reported genomic regions containing insulin release genes (eg., *GLP1R*, *CDKAL*, *SGMS1*) with RFI and ADFI. Later on, a total of 79 significant SNP associations with FCR in pigs on 6 chromosomes were identified in Denmark Duroc pigs, of which 10 SNP reach the genome-wide significance threshold on SSC 4 and 14 (Sahana et al., 2013). Noval regions affecting two different measures of RFI were reported in a genome-wide association study in Danish Duroc pigs on chromosome 1, 8, 9, 13 and 18 and indicated the candidate genes in those genomic regions involve in biological processes and metabolic pathways (Do et al., 2014a), while the gene detection in Yorkshire boars revealed different regions (Do et al., 2014b). The results reported by Do et al. (2014a, 2014b) may indicate that the genomic regions influencing feed efficiency in different breeds were slightly different and it is interesting to see if those genes were functioned in the same pathways. Little is known for the underlying genetic architecture of feeding behavior in pigs. The only association study conducted by Do et al. (2013) reported several genomic regions on SSC 1, 8, 11, and 12 associated with several feeding behavior measures.

CONCLUSION

Feed efficiency traits are of substantial importance in swine breeding. Moderate heritability of feed efficiency traits, mainly feed conversion ratio (FCR) and residual feed intake (RFI) have been founded in selection experiments or field data analysis. Both indirect and direct selections for feed efficiency improvement in pigs have been recognized to be effective. Unfortunately due to the fact that individual feed intake is expensive to measure even with the use of automatic feeding systems, some breeding companies have not

incorporate feed efficiency in the breeding program yet. Data collected by electronic feeding system has long known to contain errors and new method need to be established to better utilize those data with higher accuracy. Feeding behavior may be good indicator traits for feed efficiency because they are much more easier to collect with less equipment, but still need further investigations. The availability of dense SNP marker information has greatly facilitate identifying the underlying genes influencing feed efficiency and feeding behavior traits and genomic selection seems promising as increased accuracy of genomic selection has been reported for other traits in pigs and important production traits in other species.

The subsequent 5 chapters will address those questions through the following objectives;

Chapter 2

Obtaining accurate individual feed intake records is the key first step in achieving genetic progress toward a more efficient pig in nutrient utilization. Feed intake records collected by electronic feeding systems contain errors (extreme values or outliers exceeding certain cut-off criterions), which are due to feeder malfunction or animal feeder interaction. The objective of this study was therefore to evaluate the performance of two alternative implementations of multiple imputation, denoted as MI and MICE, in replacing errors and missing observations occurring in feed intake data, compared with the well-established LMM approach, under different simulated scenarios.

Chapter 3

The availability of the Porcine60K BeadChip has greatly facilitated whole-genome association studies, contributing to increased accuracy of selection by application of marker-

assisted or genomic selection. This project aimed at identifying genomic regions associated with feed efficiency and production traits. As part of the work, the objectives of this chapter were (1) to explore heritabilities and genetic correlations for ADFI, ADG, feed efficiency traits (FCR and RFI), ultrasound traits (BF, MD and IMF), BW at birth and weaning, using either pedigree or marker information; and (2) to access accuracy of genomic prediction of the nine traits in a Duroc terminal sire population.

Chapter 4

With more dense genomic information, the genetic architecture of a trait can be better understood, such as the number of genes/loci that affect a trait and the additive effects of individual genes/loci. The objective of this study was to identify genomic regions associated with variations in feed intake, average daily gain, the two commonly used feed efficiency measures (feed conversion ratio and residual feed intake) and real-time ultrasound traits in a Duroc terminal sire population.

Chapter 5

The present chapter had three objectives. The first was to estimate genetic parameters for different measures of feed efficiency and other traits (including growth, off-test measurements and feeding behavior traits) with pedigree or marker information. The second objective was to quantify the effect of social interaction among pen mates for traits collected using FIRE system. The last objective was to compare the accuracy of prediction for all those traits of interests using traditional BLUP or single-step genomic BLUP (ssGBLUP) approach.

Chapter 6

The aim of this study was to identify genomic regions and candidate genes associated with different measures of feed efficiency including feed conversion ratio (FCR), residual feed intake (RFI), residual growth (RG) and the combined measure of residual feed intake and residual growth (RIG) and five feeding behavior traits including daily occupation time (AOTD), number of visits to feeders (ANVD), daily feeding rate (ADFR), occupation time per visit (AOTV) and feed intake per visit (AFIV) across the testing period in a Purebred Duroc population.

LITERATURE CITED

Abell, C. E., J. C. M. Dekkers, M. F. Rothschild, J. W. Mabry, and K. J. Stalder. 2014. Total cost estimation for implementing genome-enabled selection in a multi-level swine production system. Genet. Sel. Evol. 46(1):32.

Allison, P. D. 2002. Missing data: Quantitative applications in the social sciences. Br. J. Math. Stat. Psychol. 55:193-196.

Arango J., I. Misztal, S. Tsuruta, M. Culbertson, and W. Herring. 2005. Estimation of variance components including competitive effects of Large White growing gilts. J. Anim. Sci. 83:1241-1246.

Archer, J. A., P. F. Arthur, R. M. Herd, P. F. Parnell, and W. S. Pitchford. 1997. Optimum postweaning test for measurement of growth rate, feed intake and feed efficiency in British breed cattle. J. Anim. Sci. 75:2024–2032.

Arthur, P. F., and R. M. Herd. 2005. Efficiency of feed utilisation by livestock-Implications and benefits of genetic improvement. Can. J. Anim. Sci. 85(3):281-290.

Balding, D. J. 2006. A tutorial on statistical methods for population association studies. Nat. Rev. Genet. 7:781-791.

Bergsma, R., E. Kanis, E. F. Knol, P. Bijma. 2008. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genet.178:1559-1570.

Berry, D. P., J. J. Crowley JJ. 2009. Residual intake and body weight gain: a new measure of efficiency in growing cattle. J Anim. Sci. 90:109-115.

Bernard C., and M. H. Fahmy. 1970. Effect of selection on feed utilization and carcass score in swine. Canadian J. Anim. Sci. 50(3): 575-584.

Brisbane, J. R. 2002. Prediction of within-herd differences in total feed intake between growing pigs. Can. J. Anim. Sci. 82:283-293.

Brown-Brandl, T.M., G. A. Rohrer, and R. A. Eigenberg. 2013. Analysis of feeding behavior of group housed growing-finishing pigs. Comput. Electron. Agric. 96:246-252.

Cai, W., D. S. Casey, and J. C. M. Dekkers. 2008. Selection response and genetic parameters for residual feed intake in Yorkshire swine. J. Anim. Sci. 86:287-298.

Cameron, N. D., and M. K. Curran. 1994. Selection for components of efficient lean growth rate in pigs 4. Genetic and phenotypic parameter estimates and correlated responses in performance test traits with ad-libitum feeding. Anim. Prod. 59: 281-291

Carstens, G. E., C. M. Theis, M. D. White, T. H. Jr. Welsh, B. G. Warrington, R. D. Randel, T. D. A. Forbes, H. Lippke, L. W. Greene and D. K. Lunt. 2002. Residual feed intake in beef steers: I. Correlations with performance traits and ultrasound measures of body composition. J. Anim. Sci. 80 (Suppl. 2): 135.

Chen, C. Y., I. Misztal, S. Tsuruta, B. Zumbach, W. O. Herring, J. Holl, and M. Culbertson. 2010. Estimation of genetic parameters of feed intake and daily gain in Durocs using data from electronic swine feeders. J. Anim. Breed. Genet. 127:230-234.

Casey, D. S. 2003. The use of electronic feeders in genetic improvement programs for swine. PhD Diss. Iowa State Univer., Ames.

Casey, D. S., H. S. Stern, and J. C. M. Dekkers. 2005. Identification of errors and factors

associated with errors in data from electronic swine feeders. J. Anim. Sci. 83:969-982.

Christensen, O. F. and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Gen. Sel. Evol. 42:2.

Christensen, O. F. P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal. 6:10:1565-1571.

Cleveland, E. R., R. K. Johnson, R. W. Mandigo, and E. R. Peo, Jr. 1983. Index selection and feed intake restriction in swine. II. Effect on energy utilization. J. Anim. Sci. 56:570-578.

Cleveland, M. A., and J. M. Hickey. 2013. Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation. J. Anim. Sci. 91(8): 3583-3592.

Crowley, J. J., M. McGee, D. A. Kenny, D. H. Crews, R. D. Evans, and D. P. Berry. 2010. Phenotypic and genetic parameters for different measures of feed efficiency in different breeds of Irish performance-tested beef bulls. J. Anim. Sci. 88(3):885-894.

Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics. 185:1021-1031.

De Haer, L. C. M., J. W. M. Merks, H. G. Kooper, G. A. J. Buiting, and J. A. van Hattum. 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.

de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, and J. M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genet. 182(1):375-385.

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 193:327-345.

Dekkers, J. C. M. and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3:22-32.

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

Dekkers, J. C. M. 2012. Application of genomic tools to animal breeding. Current Genomics. 13:207-212.

De Vries, A. G., R. Kerr, B. Tier, and T. Long. 1994. Gametic imprinting effects on rate and composition of pig growth. Theo. Appl. Genet. 88 (8):1037-1042.

Do, D. N., A. B. Strathe, J. Jensen, T. Mark, and H. N. Kadarmideen. 2013a. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. J. Anim. Sci. 91(9):4069-4079.

Do, D. N., A. B. Strathe, T. Ostersen, J. Jensen, T. Mark, and H. N. Kadarmideen. 2013b. Genome-wide association study reveals genetic architecture of eating behavior in pigs and its implications for humans obesity by comparative mapping. Plos One, 8(8):14102-14109.

Do, D. N., A. B. Strathe, T. Ostersen, S. D. Pant, and H. N. Kadarmideen. 2014a. Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. Front. Genet. 5:307.

Do, D. N., T. Ostersen, A. B. Strathe, T. Mark, J. Jensen, and H. N. Kadarmideen. 2014b. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. BMC Genet. 15(1): 27.

Eissen, J. J., E. Kanis, and J. W. M. Merks. 1998. Algorithms for identifying errors in individual feed intake data of growing pigs in group-housing. Appl. Eng. Agric. 14:667-673.

Eissen, J. J., A. G. De Haan, and E. Kanis. 1999. Effect of missing data on the estimate of average daily feed intake of growing pigs. J. Anim. Sci. 77(6):1372-1378.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114-29.

Ernst, C. W. and J. P. Steibel. 2013. Molecular advances in QTL discovery and application in pig breeding. Trends Genet. 29:215-225.

Falconer, D. S. 1952. The problem of environment and selection. Am. Naturalist. 86: 293-298.

Falconer D. S., and T. F. C. Mackay.1996. Introduction to Quantitative Genetics. Longman.

Fan, B., S. Lkhagvadorj, W. Cai, J. Young, R. M. Smith, J. C. M. Dekkers, E. Huff-Lonergan, S. M. Lonergan, and M. F. Rothschild. 2010. Identification of genetic markers associated with residual feed intake and meat quality traits in the pig. Meat Sci. 84:645-650.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendilian inheritance. Trans. Roy. Soc. Edinburgh 52:399-433.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol. 43:1.

Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genet. Sel. Evol, 43(17):10-1186.

Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genet. Sel. Evol, 43(17):10-1186.

Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. Genetics. 183:347-363.

Gilbert, H., J.P. Bidanel, J. Gruand, J. C. Caritez, Y. Billon, P. Guillouet, H. Lagant, J. Noblet, P. Sellier. 2008. Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. J. Anim. Sci. 85(12):3182-3188.

Goddard, M. E. and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programs. Nat. Rev. Genet. 10:381-391.

Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2010. Genomic selection in livestock populations. Genet. Res. 92.5-6:413-421.

Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed. Genet. 128:409-421.

Goodman, S. N. 1999. Toward evidence-based medical statistics. 2: The Bayes factor. A. Int. Med. 130(12):1005-1013.

Grignola, F., A. C. Clutter, D. S. Casey, J. C. M. Dekkers, and X. Liu. 2002. Effects of feeding strategies using electronic feeders on data quality, feed intake, growth and feed efficiency of swine. Proc. Seventh World Cong. Genet. Appl. Livest. Prod.: 10-12.

Gunsett, F.C. 1984. Linear index selection to improve traits defined as ratios. J. Anim. Sci. 59:1185-1193.

Guo, X., O. F. Christensen, T. Ostersen, Y. Wang, M. S. Lund, and G. Su. 2015. Inproving genetic evaluation of litter size and piglet mortality for both genotyped and nongenotyped individuals using a single-step method. J. Anim. Sci. 93. 2

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 177:2389:2397.

Habier, D., Fernando, R. L. and Dekkers, J. C. M. 2009. Genomic selection using low-density marker panels. Genet.182:343-353.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 12:186.

Henderson, C. R. 1953. Estimation of variance and covariance components. Biometrics. 9(2): 226-252.

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet. 6:e1001139.

Hayes, B., and M. Goddard. 2010. Genome-wide association and genomic selection in animal breeding. Genome. 53(11):876-883.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433-443.

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realised relationship matrix. Genet. Res. Camb. 91:47-60.

Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics. 28:476-490.

Henman, D. 2003. Nutritional management in integrated pig units.p11-132. in J. Wiseman, M. A. Varely, and B. Kemp, eds. Perspectives in pig science. Nottingham University Press, Nottingham, UK.

Hoque, M. A., H. Kadowaki, T. Shibata, T. Oikawa, and K. Suzuki. 2009. Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livest. Sci. 121:45-49.

Hoque, M. A., H. Kadowaki, T. Shibata, and K. Suzuki. 2008. Maternal and direct genetic parameters for production traits and maternal correlations among production and feed efficiency traits in Duroc pigs. Asian-Aust. J. Anim. Sci. 7:961-966.

Hosking, F. J., J. A. C. Sterne, G. D. Smith, and P. J. Green. Inference from genome-wide association studies using a novel Markov model. Genet. Epidem.32: 497-504.

Howard, J. T., S. Jiao, F. Tiezzi, Y. Huang, K. A. Gray, and C. Maltecca. 2015. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. BMC Genet. 16(1):59.

Hu, Z., E. R. Fritz, and J. M. Reecy. 2007. AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. Nucleic. Acids Res. 35 (suppl 1): 604-609.

Huang Y., J. M. Hickey, M. A. Cleveland, and C. Maltecca. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. Genet. Sel. Evol. 44:25.

Huzzey, J. M., D. M. Weary, B. Y. F. Tiau, and M. A. G. von Keyserlingk. 2014. Short communication: Automatic detection of social competition using an electronic feeding system. J. Dairy Sci. 97(5):2953-2958.

Hyun, Y., and M. Ellis. 2001. Effect of group size and feeder type on growth performance and feeding patterns in growing pigs. J. Anim. Sci. 79:803-810.

Hyun, Y., M. Ellis, and R. W. Johnson. 1998. Effects of feeder type, space allowance, and mixing on the growth performance and feed intake pattern of growing pigs. J. Anim. Sci. 76:2771-2778.

Jiao, S., C. Maltecca, K. A. Gray, and J. P. Cassady. 2014a. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. J. Anim. Sci. 92 (6):2377-2386.

Johnson, Z. B., J. J. Chewning, and R. A. Nugent. 1999. Genetic parameters for production traits and measures of residual feed intake in large white swine. J. Anim. Sci. 77:1679-1685.

Kachman, S. 2008. Incorporation of marker scores into national cattle evaluations. Proc. 9th Genetic Prediction Workshop, Kansas City, MO, pp. 88-91.

Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S. Kong, N. B. Freimer, C. Sabatti, and E. Eskin. 2010. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42 (4):348-354.

Kennedy, B. W., J. H. Van der Werf, and T. H. Meuwissen. 1993. Genetic and statistical properties of residual feed intake. J. Anim. Sci. 71:3239-3250.

Kizilkaya, R. L., R. L. Fernando, D. J. Garrick. 2009. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J. Anim. Sci. 88: 544-551.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22: 486-494.

Labroue, F., R. Gueblez, and P. Sellier. 1997. Genetic parameters of feeding behaviour and performance traits in group-housed Large White and French Landrace growing pigs. Genet. Sel. Evol. 29:451-468.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656-4663.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single step, a general approach for genomic selection. Liv. Sci. 166:54-65.

Lonergan, S. M., E. Huff-Lonerga, L. J. Rowe, D. L. Kuhlers, and S. B. Jungst. 2001. Selection for lean growth efficiency in Duroc pigs influences pork quality. J Anim. Sci. 79(8):2075-2085.

Lunn, D. J., J. C. Whittaker, and N. Best. 2006. A Bayesian toolkit for genetic association studies. Genet. Epidem. 30:231-247.

Osborne, J. W., and A. Overbay. 2004. The power of outliers (and why researchers should always check for them). PARE. 9:1-12.

Onteru, S. K., B. Fan, M. T. Nikkilä, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Whole-genome association analyses for lifetime reproductive traits in the pig. J. Anim. Sci. 89:988-995.

Onteru, S. K., D. M. Gorbach, J. M. Young, D. J. Garrick, J. C. M. Dekkers, and M. F. Rothschild. 2013. Whole genome association studies of residual feed intake and related traits in the pig. PloS one. e61756.

MacNeil, M., and R. A. Kemp. 2015. Genetic parameter estimation and evaluation of Duroc boars for feed efficiency and component traits. Can. J. Anim. Sci. 95(2):155-159.

McGlone, J. J., and W. Pond. 2003. Pig Production: Biological Principles and Applications. Delmar Learning, Clifton Park, NY.

Maselyne, J., W. Saeys, and A. Van Nuffel. 2015. Review: Quantifying animal feeding behaviour with a focus on pigs. Physiol. Behav. 138:37-51.

Maselyne, J., W. Saeys, B. De Ketelaere, K. Mertens, J. Vangeyte, E. F. Hessel, B. Sonck, and W. Saeys. 2014. Validation of a High Frequency Radio Frequency Identification (HF RFID) system for registering feeding patterns of growing-finishing pigs. Comput. Electron. Agric.108:209-220.

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

Meiwissen, T. H. E. 2007. Genomic selection: marker-assisted selection on a genome wide scale. J. Anim. Breed. Genet. 124:321-322.

Meuwissen, T.H.E., B. J.Hayes, M. E. Goddard. 2014. Accelerating improvement of livestock with genomic selection. Annu. Rev. Anim. Biosci. 1:221-237.

Merks, J. W. M. 1989. Genotype x environment interactions in pig breeding programmes. VI. Genetic relations between performances in central test, on-farm test and commercial fattening. Livest. Prod. Sci. 22:325-339.

Misztal, I. A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648-4655.

Misztal, I., A. E. Aggrey, and W. M. Muir. 2013. Experiences with a single-step genome evaluation1. Poul. Sci. 92(9):2530-2534.

Mrode, R. A., and B. W. Kennedy. 1993. Genetic variation in measures of food efficiency in pigs and their genetic relationships with growth rate and backfat. Anim. Prod. 56:225-232.

Nejati-Javaremi, A., C. Smith, and J. P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Anim. Sci. 75:1738-1745.

Nielsen, B. L., A. B. Lawrence, and C. T. Whittemore. 1995. Effect of group size on feeding behaviour, social behaviour, and performance of growing pigs using single-space feeders. Livest. Prod. Sci. 44:73-85.

Nguyen, N. H., C. P. McPhee, and C. M. Wade. 2005. Responses in residual feed intake in lines of Large White pigs selected for growth rate on restricted feeding (measured on ad libitum individual feeding). J. Anim. Breed. Genet. 122(4):264-270.

Park, T., and G. Casella. 2008. The Bayesian Lasso. J. Am. Stat. Assoc. 103:681-86

Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. Am. J. Hum. Genet. 67 (1): 170-181.

Rauw, W. M., E. Kanis, E. N. Noordhuizen-Stassen, and F. J. Grommers. 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review.Livest. Prod. Sci. 56(1):15-33.

Rohrer, G. A., T. Brown-Brandl, L. A. Rempel, J. F. Schneider, and J. Holl. 2013. Genetic analysis of behavior traits in swine production. Livest. Sci. 157(1):28-37.

Rothschild, M. F., and A. Ruvinsky. 2010. The genetics of the pig. 2nd ed. CABI, Cambridge, MA.

Rothschild, M. F., Z. Hu, and Z. Jiang. 2007. Advances in QTL mapping in pigs. Intern. J. Biolo. Sci. 3 (3):192.

Rubin, D. B. 1978. Multiple imputation for nonresponse in surveys. John Wiley & Sons. New York.

Rubin, D. B. 1996. Multiple imputation after 18+ years. J. Am. Stat. Assoc. 91:473-489.

Sahana, G., K. Veronika, H. Henrik, N. Bjarne, and O. F. Christensen. 2013. A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. J. Anim. Sci. 91:1041-1050.

Schnyder, U., A. Hofer, F. Labroue, and N. Kunzi. 2001. Genetic parameters of a random regression model for daily feed intake of performance tested French Landrace and Large White growing pigs. Genet. Sel. Evol. 33:635-658.

Schulze, V., R. Roehe, L. B. J, H. Looft, and E. Kalm. 2002. Genetic associations between observed feed intake measurements during growth, feed intake curve parameters and growing-finishing performances of central tested boars. Livest. Prod. Sci. 73:199-211.

Sowell, B.F., J. G. P. Bowman, M. E. Branine, and M. E. Hubbert. 1998. Radio frequency technology to measure feeding behavior and health of feedlot steers. Appl. Anim. Behav. Sci. 59:277-284.

Stephens, M. and D. J. Balding. 2009. Bayesian statistical methods for genetic association studies. Nat. Rev. Genet. 10:681-690.

Su, Y. S., M. Yajima, A. E. Gelman, and J. Hill. 2011. Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. J. stat. softw. 45:1-31.

Suzuki, K., M. Irie, H. Kadowaki, T. Shibata, M. Kumagai, and A. Nishida. 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. 83:2058-2065.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B. 58(1):267-88.

Tolkamp, B.J., D. P. N. Schweitzer, I. Kyriazakis. 2000. The biologically relevant unit for the analysis of short-term feeding behavior of dairy cows. J. Dairy Sci. 83, 2057–2068.

van Buuren, S. and K. Groothuis-Oudshoorn. 2011. MICE: Multivariate imputation by chained equations in R. J. Stat. Softw. 45:1-68.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423.

Veerkamp, R. F., M. P. Coffey, Donagh P. Berry, Y. De Haas, E. Strandberg, H. Bovenhuis, M. P. L. Calus, and E. Wall. 2012. Genome-wide associations for feed utilisation complex in primiparous Holstein–Friesian dairy cows from experimental research herds in four European countries. Animal 6:1738-1749.

von Felde, A., R. Roehe, H. Looft, and E. Kalm. 1996. Genetic association between feed intake and feed intake behaviour at different stages of growth of group-housed boars.

Livest. Prod. Sci. 47: 11-22.

Walker, N. 1990. A comparison of single- and multi-space feeders for growing pigs fed non-pelleted diets ad libitum. Anim. Feed Sci. Technol. 30:169-173.

Walker, N. 1991. The effects on performance and behaviour of number of growing pigs per mono-place feeder. Anim. Feed Sci. Tech. 35:3-13.

Wakefield, J. 2007. A Bayesian measure of the probability of false discovery in genetic epidemiology studies.Am. J. Hum. Genet. 81:208–227.

Wakefield, J. 2008. Reporting and interpretation in genome-wide association studies. Intern. J. Epidem. 37:641-653.

Willems, O. W., S. P. Miller, and B. J. Wood. 2013. Assessment of residual body weight gain and residual intake and body weight gain as feed efficiency traits in the turkey (Meleagris gallopavo). Genet. Sel. Evol. 45(1):26.

Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, W. G. Hill, and J. C. M. Dekkers. 2012. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. Anim. Genet. 43:87-96.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, P. M.

Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565-69.

Zhang, Z, J. Liu, X. Ding, P. Bijma, DJ de Koning, and Q. Zhang. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. PloS one 5 (9):e12648.

Zumbach, B., I. Misztal, C. Y. Chen, S. Tsuruta, M. Łukaszewicz, W. O. Herring, and M. Culbertson. 2010. Use of serial pig body weights for genetic evaluation of daily gain. J. Anim. Breed. Genet. 127: 93-99.

CHAPTER 2

The use of multiple imputation for the accurate measurements of individual feed intake by electronic feeders

Shihui Jiao[*], Francesco Tiezzi[*], Yijian Huang[†], Kent A. Gray[†], and Christian Maltecca[*]

[*] Department of Animal Science, North Carolina State University, Raleigh, NC 27695

[†] Smithfield Premium Genetics, Rose Hill, NC 28458

ABSTRACT: Obtaining accurate individual feed intake records is the key first step in achieving genetic progress toward a more efficient pig in nutrient utilization. Feed intake records collected by electronic feeding systems contain errors (extreme values or outliers exceeding certain cut-off criterions), which are due to feeder malfunction or animal feeder interaction. In this study, we introduced a new feed intake data editing strategy, based on multiple imputation, to replace errors and missing observations occurring in feed intake data. In our work, accuracy of feed intake data adjustment obtained from the conventional linear mixed model (LMM) approach was compared to multiple imputation based on two alternative implementations of multiple imputation by chained equation algorithm, denoted as MI and MICE. The three methods were compared under 3 scenarios with 5%, 10% and 20% simulated feed intake error rates. Each of the scenarios was replicated 5 times. Accuracy of the error-adjustments was measured as the correlation between the true error-free daily feed intake (DFI) or true average daily feed intake (ADFI) and the adjusted DFI (DFI_a) or adjusted ADFI ($ADFI_a$). To investigate the possibility that the error cut-off criterions may affect any of the three methods, the simulation was repeated with two alternative error cut-off criteria. Multiple imputation methods outperformed the LMM approach in all scenarios with mean accuracies of 96.71%, 93.45% and 90.24% obtained with MI and 96.84%, 94.42% and 90.13% obtained with MICE, compared to 91.0%, 82.63% and 68.69% using LMM, for DFI. Similar results were obtained for ADFI. Additionally, multiple imputation methods performed better than LMM regardless of the cut-off criteria chosen when defining errors. In conclusion, multiple imputation is proposed as a more accurate and flexible error-adjustment method for error adjustments in feed intake data collected by electronic feeders.

Key words: feed intake, electronic feeders, multiple imputation

INTRODUCTION

Feed efficiency is a trait of primary economic importance in the swine industry. To achieve genetic progress toward a more efficient pig in nutrients utilization, obtaining accurate individual feed intake is essential. Computerized electronic feeding systems developed to automatically measure feed intake, have greatly facilitated the data collection process. However, it has been known that feed intake data collected by those systems contain errors and outliers due to feeder malfunction and animal feeder interaction (De Haer et al., 1992) and different types of errors contained in such records have been categorized by Eissen et al. (1998) and Casey et al. (2005). Due to the fact that simply discarding feed intake visits containing errors underestimated the true DFI, a linear mixed model (LMM) has been proposed by Casey (2003) to adjust records containing errors after removing visits with missing values. However, applying LMM to
a dataset with extreme values remains challenging since those values tend to severely bias the estimates (Osborne and Overbay, 2004).

Multiple imputation was introduced by Rubin (1976), as a method with the very general task of 'filling in' missing values. This approach has gained increasing popularity and the past several decades has seen implementations spanning many areas of statistical analysis (Rubin, 1996; Allison, 2002). The key concept of this technique is the use of the distribution of the observed data to estimate a set of plausible values for the missing data. Its

generality and recent software development makes it a potentially advantageous approach in feed intake data edits by treating the errors as missing values.

The objective of this study was therefore to evaluate the performance of two alternative implementations of multiple imputation, denoted as MI and MICE, in replacing errors and missing observations occurring in feed intake data, compared with the well-established LMM approach, under different simulated scenarios.

MATERIALS AND METHODS

Animal selection and dataset generation

To mimic realistic error patterns, an error-free dataset and the corresponding simulated datasets with different percentage of error visits were generated from real feed intake records collected by electronic feeders. The flow chart for the data creation and simulation process is summarized in Figure 2.1.

Raw feed intake records employed were collected from 2004-2013 using the FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, KS, USA) system in a Duroc nucleus herd owned by Smithfield Premium Genetics (SPG, Rose Hill, NC). Data included 4,958,077 feed intake visits for 14,901 animals. This dataset containing raw feeding visits was defined as the full-dataset. Briefly, during the testing period, an average of 12 pigs housed together in a pen had 24h access to feed with a single-spaced electronic feeder. When a visit to the feeder occurred, the pig ID, date, entry feed weight (feed weight when a pig entered the feeder), exit feed weight, entry time, exit time and pig body weight were recorded. Quantities measured by the feeding system can be summarized into feed intake per visit (FIV, g), occupation time per visit (OTV, s) and feeding rate per visit (FRV, g/min).

Feed intake per visit was computed as the difference of entry and exit feed weight. Similarly, the occupation time per visit was calculated as the difference between exit and entry time of the visit. Feeding rate per visit was defined as the ratio between FIV and OTV. For each of the parameters previously outlined, errors contained in each visit were defined as values more extreme than a pre-determined cut-off value. Cut-off values frequently employed by the industry for FIV, OTV and FRV are those recommended by Casey (2003), while other cut-off values are generally based on knowledge of the feeders or biology of the pigs (Eissen et al., 1998; Casey, 2003). The eight most commonly used error types occurring in the three variables FIV, OTV and FRV are those based on studies conducted by Eissen et al. (1998) and Casey (2003). The error rates for these error types in the raw feed intake dataset of the present study are displayed in Table 2.1. Two of these error types had zero rates in our data. The following analyses therefore considered only the remaining six error types.

An *"error-free"* complete dataset (EFD) and an *"error"* dataset (ED) containing visits with missing values or errors were generated from the full-dataset by identifying missing feed intake values and each of the error types in each feeding visit event. In order to reduce computation costs while preserving the general validity of the results, the EFD was generated as a subset of the full-dataset in order to include 100 animals born in year 2013. Animals were selected to enter EFD if (1) there were no error visits for the animal; (2) the animal had at least 2 feed intake visits to the feeder a day; (3) the testing period for each individual lasted at least 60 days; (4) each contemporary group (concatenation of birth year, season and house) had at least 15 pigs. After filtering, there were 17,908 feeding visits for the

100 selected pigs belonging to 4 contemporary groups (CG). The data in the EFD were treated as 'true' feed intake records without any errors.

In order to mimic the error occurring patterns, error visits were introduced into the EFD by masking the true values. Error events were randomly assigned so that they were as realistic as possible, including one or several combination of errors in one visit. To achieve this goal, the simulated data were generated by randomly selecting true visits in the EFD then substituting them with random samples of error visits from the ED. To assess the influence of different error rates (ratio of number of visits with errors over total visits) on error adjustment accuracy, three simulation scenarios were considered with error rates of 5%, 10% and 20%, respectively (Figure 2.1). For each scenario, 5 replicates were generated independently to eliminate any sparse randomness of adjustment results. As a result, there were 15 simulated datasets under the 3 simulated scenarios. The error rate for each error type for the 15 simulated datasets is presented in Table 2.2.

The cut-off criteria determining whether a visit contained an error was based on hands-on knowledge of the feeders, the criteria developed early (Eissen et al., 1998) or the distribution of variables such as FIV, OTV and FRV (Casey, 2003). To verify that the choice of cut-off points had no impact on the effectiveness of the error adjustment methods evaluated, the outlined simulation was replicated two additional times, under more and less stringent cut-off criteria for FIV, OTV and FRV. This was done by either doubling or halving the original error thresholds based on their empirical distribution (Figure S2.1). The corresponding error rates are shown in Figure 2.2.

Statistical analysis for error adjustment

Linear mixed model

The justification for adjusting error visits stems from the fact that simply throwing these visits out would severely underestimate DFI (Casey, 2003). A linear mixed model to adjust error records in feed intake collected by FIRE was developed by Casey (2003) and was considered as the conventional approach in this study. Briefly, in this approach, percentage of errors, DFI and daily occupation time (OTD) summarized for visits of a certain error type were regressed on DFI_{ef} (Error-free DFI) as covariates in order to compute an adjusted value. In our data, a matrix representation of the model is as follows,

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{b} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{e}$$

where **y** is a vector of DFI_{ef}; **b** is a vector of estimated effects including the fixed coefficients of CG, regression coefficients for off-test BW, coefficients for PTE (percentage of error) of the 6 error types, regression coefficients for DFI_e (daily feed intake for a certain error type) of error type 4 and 5 and regression coefficients for OTD_e (occupation time per day for a certain error type) of error type 1, 2, 7 and 8 (Table 2.1); $\boldsymbol{u}$ is a vector of animal effect assuming $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2\mathbf{I})$; $\boldsymbol{e}$ is a vector of model residuals assuming $\boldsymbol{e} \sim MVN(\boldsymbol{0}, \sigma_e^2\mathbf{I})$; and **X** and **Z** are the corresponding design matrices. In order to reduce the biases caused by the extreme values in DFI_e and OTD_e (defined as $DFI_e < 0$ g, $DFI_e > 3{,}500$ g, $OTD_e < 0$ s and $OTD_e > 5000$ s) Casey(2003), those extreme values were removed before fitting them into the linear mixed model. Linear coefficient estimates from this linear mixed model were then used to adjust DFI_{ef} (Casey, 2003) by adding an adjustment term, which was computed using those coefficient estimates, PTE, DFI_e and OTD_e for the corresponding error types.

Multiple imputation

Multiple imputation was designed to tackle the general problem of replacing missing values in a dataset. For this specific study, an error was defined as an extreme value of FIV, OTV or RFV departing from its (unobserved) true value. In order to estimate the true value, the extreme value in the error visit can be viewed as a missing value problem and can be 'filled' with the average of a set of plausible imputed values by multiple imputation.

Two multiple imputation methods were employed. The first one employed the R *mi* package (Su et al., 2009) in order to select conditional models for different variable types using regression models. The other method employed the R package *mice* (van Buuren and Groothuis-Oudshoorn, 2011) using multivariate imputation by chained equations.

The fundamental idea of multiple imputation as implemented in both R packages is to use chained equation algorithms to deal with multivariate missing values. The principle of chained equation is based on drawing random samples from the conditional posterior predictive distribution of missing values under a particular Bayesian framework (Rubin, 1978). Using a simplified example, let us denote the response of a univariate sample $Y = (y_1, y_2, \dots, y_n)$ where the first a values $Y_{obs} = (y_1, y_2, \dots, y_a)$ are observed and the remaining values $Y_{mis} = (y_{a+1}, y_{a+2}, \dots, y_n)$ are missing at random. Under an independent normal model $y_i \sim N(\mu, \varphi)$, $i = 1, 2, \dots, n$ and $\theta = (\mu, \varphi)$ is an unknown parameter. The observed-data posterior distribution of θ with the uninformative standard prior $P(\theta) \propto \varphi^{-1}$ is

$$\mu \mid \varphi, Y_{obs} \sim N(\overline{y_{obs}}, a^{-1}\varphi)$$

$$\varphi \mid Y_{obs} \sim (a-1)S^2_{obs}/\chi^2_{a-1}$$

where $\overline{y_{obs}} = a^{-1}\sum_{i=1}^{a} y_i$, $S_{obs}^2 = (a-1)\sum_{i=1}^{a}(y_i - \overline{y_{obs}})^2$, and χ_{a-1}^2 denote a chi-square variate with a-1 degree of freedom. To create an imputation $Y_{obs}^{(l)} = \left(y_{a+1}^{(l)}, \ldots, y_n^{(l)}\right)$, one would draw $y_i^{(l)} \sim N\left(\mu^{(l)}, \varphi^{(l)}\right)$ independently with a random mean $\mu^{(l)} \sim N\left(\overline{y_{obs}}, a^{-1}\varphi^{(l)}\right)$ and a random variance $\varphi^{(l)} \sim (a-1)S_{obs}^2/\chi_{a-1}^2$.

Repeating the procedure for $l = 2, \ldots, m$ results in m proper imputations for Y_{mis}.

More generally, $Y = (Y_{obs}, Y_{mis})$ following some parametric model $P(Y \mid \theta)$ where θ has a prior distribution and $Y_{mis}$is missing. Because

$$P(Y_{mis} \mid Y_{obs}) = \int P(Y_{mis} \mid Y_{obs}, \theta)\, P(\theta \mid Y_{obs})\, d\theta$$

An imputation for Y_{mis} can be created by first simulating a random draw of unknown parameters from their observed-data posterior $\theta^* \sim P(\theta \mid Y_{obs})$ following by a random draw of the missing values from their conditional predictive distribution $Y_{mis}^* \sim P(Y_{mis} \mid Y_{obs}, \theta^*)$.

In both *mi* and *mice* packages (Su et al., 2009; van Buuren and Groothuis-Oudshoorn, 2011), it is assumed that the complete data Y is a partially observed random sample from the p-variate multivariate distribution $P(Y|\theta)$, assuming the multivariate distribution of Y is completely specified by θ (the unknown parameters). To obtain a posterior distribution of θ, the chained equation algorithm proposes to sample iteratively from conditional distributions of the form $P(Y_1|Y_{-1})$, …, $P(Y_p|Y_{-p})$. Thus at the ith iteration, the chained equation is a Gibbs sampler which draws

$$\theta_1^{*(i)} \sim P\left(\theta_1 \middle| Y_1^{obs}, Y_2^{(i-1)}, \ldots, Y_p^{(i-1)}\right)$$

$$Y_1^{*(i)} \sim P\left(Y_1 \middle| Y_1^{obs}, Y_2^{(i-1)}, \ldots, Y_p^{(i-1)}, \theta_1^{*(i)}\right)$$

$$\vdots$$

$$\theta_p^{*(i)} \sim P\left(\theta_p \middle| Y_p^{obs}, Y_1^{(i)}, \ldots, Y_{p-1}^{(i)}\right)$$

$$Y_p^{*(i)} \sim P\left(Y_p \middle| Y_p^{obs}, Y_1^{(i)}, \ldots, Y_p^{(i)}, \theta_p^{*(i)}\right)$$

where $Y_j^{*(i)} = (Y_j^{obs}, Y_j^{*(i)})$ is the j[th] imputed variable at iteration i. Since in this way, the Gibbs sampler can be easily implemented as a concatenation of univariate procedures to fill out the missing data, this algorithm was called chained equation algorithm (van Buuren and Groothuis-Oudshoorn, 2011). As demonstrated in many studies (van Buuren, 1999; Rubin, 2003; Heymans et al., 2007), a low number iteration (say 10 to 20) is often sufficient. Replacing the missing value by a set of imputed plausible values, multiple imputation generates multiple imputed datasets to reflect the uncertainty of imputed values and statistical analysis need to be appropriately applied to combine results obtained from each of them. For simplicity in the current analysis, the mean of the set of plausible values may be viewed as the expectation of the imputed values for the unobserved entry in the dataset.

Although R package *mi* and *mice* are both based on the same chained equation algorithm, they employ different elementary imputation methods to impute numeric missing values: predictive mean matching in package *mi* (Su et al., 2009) and Bayesian linear regression in *mice* (van Buuren and Groothuis-Oudshoorn, 2011), respectively. Predictive mean matching is a semi-parametric imputation method that is restricted to the observed values and can preserve non-linear relations in the conditional model while Bayesian linear regression is faster and more efficient when the residual of the conditional model is normal.

These two slightly different implementations of multiple imputation were used in this study to compare the different implementation performances. For each simulated dataset, the error visits identified and then masked as unobserved values (missing values) were drawn using MCMC (Markov Chain Monte Carlo) as specified above. Convergence of the chains for both MI and MICE were examined by trace plots of the chains and assessed with the use of the CODA package in R (van Buuren and Groothuis-Oudshoorn, 2011; Su et al., 2009).

Measuring method performance

In order to compare the efficiency and accuracy of error adjustment, each method was applied to the same 15 simulated datasets and the results were compared using Pearson correlation of DFI_a or $ADFI_a$ with true DFI or the true ADFI computed using the error-free complete dataset.

RESULTS

Feed intake visits collected by FIRE contained errors and the error rates for the 8 predetermined error types that ranged from 0.0% to 4.46% in the full-dataset. In Table 2.1, FIV-high (4.46%) and FIV-low (3.07%) were the two most common error types. The overall error rate (defined as the ratio of number of visits with at least one error over the total number of visits) was 9.28% in the Full dataset (Table 2.1).

The error rates for the eight error types in the 15 simulated datasets are shown in Table 2.2. The variation of error rates among replicates within the first scenario (containing 5% error visits) was small (SD of error rates ranged from 0.01 to 0.10%) while variability of error rates under scenario 2 (containing 10% error visits) and 3 (containing 20% error visits)

were approximately twice or three times of error rates under scenario 1 (5% error rates), which is expected by design.

Accuracies for adjusted DFI and ADFI are shown in Table 2.3 and 2.4 for all method/scenario combinations. For all scenarios, MI and MICE performed similarly and outperformed LMM with higher accuracy in DFI_a (Table 2.3). For the 5% error rate scenario, all three methods performed well with correlations between true and adjusted DFI of 91.0%, 96.7% and 96.9% for LMM, MI and MICE, respectively. For moderate and high error scenarios (10% and 20%), multiple imputation methods (MI and MICE) were considerably more effective than LMM in terms of accuracy. Average accuracies were 82.6% for LMM and 93.5% and 94.4% for MI and MICE in the 10% error rate scenario. Average accuracy was 68.7% for LMM, versus 90.2% and 90.1% for MI and MICE in the 20% error rate scenario. For adjusted average daily feed intake (ADFI), the trend remained similar although the differences were less marked, with the accuracies of LMM (ranging from 92.8% to 98.9%) consistently lower than the accuracies of MI and MICE (ranging from 98.7% to 99.9% in Table 2.4). We additionally computed the Spearman rank correlation between DFI_a and true DFI (or between $ADFI_a$ and true ADFI). Results can be found in supplemental Table S1 and S2. It should be also pointed out that in all cases LMM underestimated DFIa more as compared to MI or MICE (supplemental Tables S2.3 and S2.4). When the same simulated scenarios were repeated with different cut-off criteria, results did not change. In all cases, MI and MICE outperformed LMM regardless of the cut-off criteria chosen (Figure 2.2).

DISCUSSIONS

There has been considerable interest in feed intake, feed efficiency and feeding behavior in livestock and much of that interest has been centered on the ability of obtaining reliable genetic/genomic predictions for these traits. A sizable body of literature has been produced on the application of genomic information in the prediction of feed intake related traits (Fan et al., 2010; Do et al., 2013; Sahana et al., 2013) and feeding behavior (Onteru et al., 2011). However much less attention has been dedicated to improve the quality of the data that is used in these analyses. Feed intake data collected using electronic feeding system are noisy and this might significantly hinder the quality of any downstream utilization of these data (De Haer et al., 1992; Casey et al., 2005). Phenotypic data cleaning or editing is a necessary and important first step of any genetic improvement program and cannot be overlooked. In our data, the percentage of error visits was moderately high with an average of 9.28% errors visits out of 4,958,077 total feeding visits of 14,901 animals collected from year 2004 to 2013. Previously reported error rates in feed intake collected by electronic feeders varied among different datasets. In a similar population of Duroc, Jiao and colleagues (2014) found the overall error rates ranging between 14% and 35%. Eissen et al. (1998) reported error visits representing 6% of the total 385,329 feeding visits for 250 pigs. Similarly Casey (2003) reported percentages of identified error visits of 4.33%, 5.62% and 18.74% for three different datasets with 863,590 total visits for 893 pigs, 290,073 total visits for 591 pigs and 162,638 visits for 237 pigs, respectively. It is unsurprising that the occurrence of errors in feed intake during electronic recording varies among different

datasets since the electronic feeding systems are placed under varying environmental conditions. This further highlights the need to develop robust error adjustment methods.

The LMM approach developed by Casey (2003) is routinely used to adjust error visits in feed intake records, but presents some limitations. The data processing before the actual mixed model application is cumbersome. For instance, the daily records need to be computed for visits with and without errors. Moreover, subjective constraints (in DFI_e and OTD_e) must be set upon the model to limit the bias arising from influential or extreme values in the predictors, which results in removing a proportion of DFI records before fitting the models. Lastly, the correction for DFI using LMM does not take the unidentified visits (missing values in feed intake) into account because of upfront removal of those visits. Conversely, the multiple imputation approach outlined in the current paper is very general and can be easily implemented in a variety of settings with minimal data preprocessing or ad hoc adjustments. Furthermore, by treating error records as missing, the extreme values have no effects on the MI and MICE models as compared to LMM approach where these have to be removed to ensure unbiased estimates (Casey, 2003). The results of our study provide a practical illustration of the advantages of MI or MICE over LMM in addressing the problem of occurrence of errors or outliers in feed intake data collected by electronic feeding systems.

The use of multiple imputation has benefits also with respect to the final dataset after error adjustment. Its main advantage is that multiple imputation produces a final dataset with individual feeding records, instead of one with individual average daily feed intake across the whole testing period (see for example Eissen et al., 1999; Hebart et al., 2004 for applications in swine and beef cattle). The LMM looses all the information of feed intake and related

measures of individuals on a visit basis. Feed intake is a trait that intensely reflects the day-to-day or hour-to-hour dynamics of an animal's metabolism. To investigate the mechanisms of variation in individual feed intake over testing period (Bermejo et al., 2003; Cai et al., 2011) or daily eating patterns (Young et al., 2011; Rohrer et al., 2013), feed intake and related measures such as feeding time or feeding rate per visit for each individual on test are required. In this situation, multiple imputation is a more natural method of choice since it allows to make use of all the information provided.

Multiple imputation is a mature technique that is continually refined and makes it suitable for handling missing data routinely (Horton et al., 2001). As the fast implementation of this technique in statistical software, multiple imputation become increasing popular to deal with messiness in data, especially in medical and social science research (Royston, 2004; King et al., 2011; Sterne et al., 2009) to avoid bias for population parameter estimation in regression settings and loss of information due to missing values. In addition, multiple imputation is a very general data editing approach and could find a broad applicability in all situations where error prone data are employed. In the field of animal science, missing values and extreme values wrongly recorded occur in all types of data no matter how hard producers or investigators try to prevent them. Multiple imputation might be a proper technique to deal with the error or missing records in such field recorded datasets, such as disease incidence data from dairy producer-recorded health events from on-farm computer systems (Gaddis, 2012).

Like any statistical techniques, it should be used after careful examination. As pointed by Robin (1996) and White et al. (2011) multiple imputation is not free of limitations and

pitfalls. It is for example difficult to impute data points when the dataset contains too many variables with missing values. Furthermore the methodology is sensitive to the error occurrence patterns and is computationally more intensive. It is implied that sensitivity analysis may be needed when applying multiple imputation to a new dataset and parallel computing might serve as a tool to release the computation burden.

CONCLUSIONS

In conclusion, with this work we suggest multiple imputation as an effective alternative to LMM to deal with errors contained in feed intake data collected by electronic feeding systems. Nonetheless, potential application of multiple imputation in field data editing is exciting and encouraging and may need further investigation before use.

LITERATURE CITED

Allison, P. D. 2002. Missing data: Quantitative applications in the social sciences. Br. J. Math. Stat. Psychol. 55:193-196.

Bermejo, J. L., R. Roehe, V. Schulze, G. Rave, H. Looft, and E. Kalm. 2003. Random regression to model genetically the longitudinal data of daily feed intake in growing pigs. Livest. Prod. Sci. 82(2):189-199.

Cai, W., H. Wu, and J. C. M. Dekkers. 2011. Longitudinal analysis of body weight and feed intake in selection lines for residual feed intake in pigs. Asian Australas. J. Anim. Sci. 24(1):17-27.

Casey, D. S. 2003. The use of electronic feeders in genetic improvement programs for swine. PhD Diss. Iowa State Univer., Ames.

Casey, D. S., H. S. Stern, and J. C. M. Dekkers. 2005. Identification of errors and factors associated with errors in data from electronic swine feeders. J. Anim. Sci. 83:969-982.

De Haer, L. C. M., J. W. M. Merks, H. G. Kooper, G. A. J. Buiting, and J. A. van Hattum. 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.

Do, D. N., A. B. Strathe, T. Ostersen, J. Jensen, T. Mark, and H. N. Kadarmideen. 2013. Genome-wide association study reveals genetic architecture of eating behavior in pigs and its implications for humans obesity by comparative mapping. Plos

One, 8(8):14102-14109.

Eissen, J. J., E. Kanis, and J. W. M. Merks. 1998. Algorithms for identifying errors in individual feed intake data of growing pigs in group-housing. Appl. Eng. Agric. 14:667-673.

Eissen, J. J., A. G. De Haan, and E. Kanis. 1999. Effect of missing data on the estimate of average daily feed intake of growing pigs. J. Anim. Sci. 77(6):1372-1378.

Fan, B., S. Lkhagvadorj, W. Cai, J. Young, R. M. Smith, J. C. M. Dekkers, E. Huff-Lonergan, S. M. Lonergan, and M. F. Rothschild. 2010. Identification of genetic markers associated with residual feed intake and meat quality traits in the pig. Meat Sci. 84:645-650.

Gaddis, K. L. P. 2012. Incidence validation and relationship analysis of producer-recorded health event data from on-farm computer systems in the United States. J. Dairy Sci. 95(9):5422-5435.

Hebart, M. L., W. S. Pitchford, P. F. Arthur, J. A. Archer, R. M. Herd, and C. D. K. Bottema. 2004. Effect of missing data on the estimate of average daily feed intake in beef cattle. Anim. Prod. Sci. 44(5):415-421.

Heymans, M. W., S. van Buuren, D. L. Knol, W. V. Mechelen and H. C. D. Vet. 2007. Variable selection under multiple imputation using the bootstrap in a prognostic study. BMC Med. Res. Meth. 7(4):33.

Horton, N. J., and S. R. Lipsitz. 2001. Multiple imputation in practice: comparison of software packages for regression models with missing variables. Am. Stat. 55:244-254.

Jiao, S., C. Maltecca, K. A. Gray, and J. P. Cassady. 2014. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. J. Anim. Sci. 92(6):2377-2386.

King, G., J. Honaker, A. Joseph, and K. Scheve. 2001. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. APAS. 95(01):49-69.

Osborne, J. W., and A. Overbay. 2004. The power of outliers (and why researchers should always check for them). PARE. 9:1-12.

Onteru, S. K., B. Fan, M. T. Nikkilä, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Whole-genome association analyses for lifetime reproductive traits in the pig. J. Anim. Sci. 89:988-995.

Rohrer, G. A., T. Brown-Brandl, L. A. Rempel, J. F. Schneider, and J. Holl. 2013. Genetic analysis of behavior traits in swine production. Livest. Sci. 157(1):28-37.

Royston, P. 2004. Multiple imputation of missing values. Stata J. 4:227-241.

Rubin, D. B. 1978. Multiple imputation for nonresponse in surveys. John Wiley & Sons. New York.

Rubin, D. B. 1996. Multiple imputation after 18+ years. J. Am. Stat. Assoc. 91:473-489.

Rubin, D. B. 2003. Nested multiple imputation of NMES via partially incompatible mcmc. Stat. Neerl. 57(1):3-18.

Sahana, G., K. Veronika, H. Henrik, N. Bjarne, and O. F. Christensen. 2013. A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. J. Anim. Sci. 91:1041-1050.

Sterne, J. A. C., W. R. Ian, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter. 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 338:b2393.

Su, Y. S., M. Yajima, A. E. Gelman, and J. Hill. 2011. Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. J. stat. softw. 45:1-31.

van Buuren, S., H. C. Boshuizen, and D. L. Knook. 1999. Multiple imputation of missing blood pressure covariates in survival analysis. Stat. in Med. 18(6):681-694.

van Buuren, S. and K. Groothuis-Oudshoorn. 2011. MICE: Multivariate imputation by chained equations in R. J. Stat. Softw. 45:1-68.

White, I. R., P. Royston, and A. M. Wood. 2011. Multiple imputation using chained equations: issues and guidance for practice. Stat. Med. 30(4):377-399.

Young, J. M., W. Cai, and J. C. M. Dekkers. 2011. Effect of selection for residual feed intake

on feeding behavior and daily feed intake patterns in Yorkshire swine. J. Anim. Sci. 89(3):639-647.

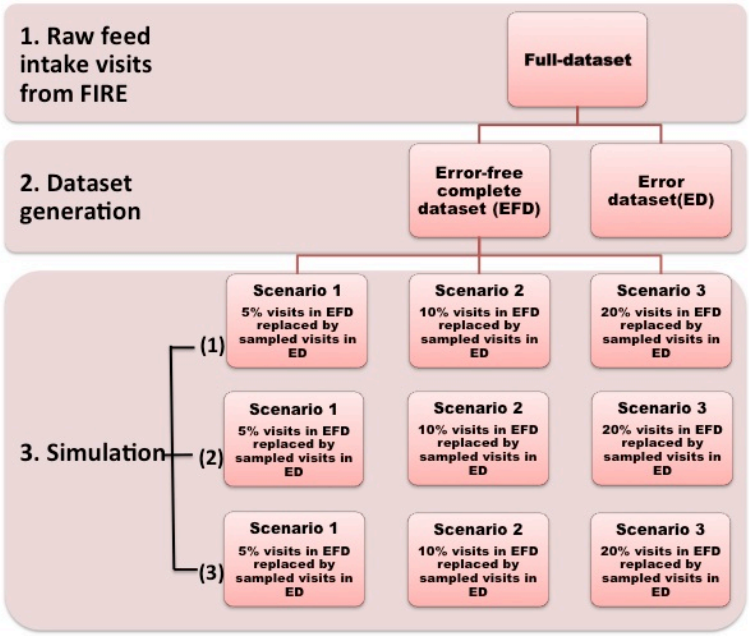

Figure 2.1 Working flow chart for data generation and simulation.

The full-dataset contained 4,958,077 raw feed intake visits collected using FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, KS) from 2004 to 2013 for 14,901 pigs. The error-free complete dataset (EFD) was generated from the full-dataset with animals meet the requirements to enter, if (1) no error visits for the animal; (2) had at least 2 feed intake visits to the feeder a day; (3) the testing period lasted at least 60 days; (4) each contemporary group had at least 15 pigs. After filtering, there were 17,908 feeding visits for 100 selected pigs belonging to 4 contemporary groups. Error dataset (ED) was generated by identifying errors in the full-dataset. The simulation was repeated 3 times (in the above figure (1), (2) and (3)) with different error cut-off criteria (the threshold values defining errors). Within each replication of simulation, three scenarios were simulated with 5%, 10% and 20% visits in EFD substituted by randomly sampled error visits from ED.

Table 2.1 Types of errors in feed intake visits from FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, KS, USA) and rate of error for each error type in full dataset from year 2004 to 2013.

Error index	Error type[1]	Error definition[2]	Error rate (100%)[3]
1	FIV-low	FIV < -20 g for all visits	3.07
2	FIV-high	FIV > 2000 g for all visits	4.46
3	FIV-0	FIV > 20 g or FIV < -20g for visits with OTV = 0 s	0.00
4	OTV-low	OTV < 0 s for all visits	0.53
5	OTV-high	OTV > 3600 s for all visits	0.10
6	FRV-high-FIV-low	FRV > 500 g/min for visits with 0 g < FIV < 50 g	0.00
7	FRV-high	FRV > 350 g/min for visits with FIV > 50 g	2.20
8	FRV-0	FRV = 0 g/min for visits with OTV > 500 s	0.80

[1] Eight error types were proposed by Casey et al. (2005) and Eisen et al. (1998): FIV = feed intake per visit (g); OTV = occupation time per visit (s); FRV = feed intake rate per visit (g/min).

[2] The cut-off criterion were based on Casey et al. (2005) for different error types, which were chose based on the feature of the feeder (error type 1 and 3), the biology of pig for feed intake (error type 4, 5 and 8) or the distribution of the variables FIV, OTV or FRV (error type 2, 5 and 7).

[3] The error rate is computed for the full dataset, where the overall error rate (= number of visits with at least 1 error / total visits) is 9.28 %.

Table 2.2 Error rates in simulated replicated datasets.

Replicate dataset[1]	Error 1[2]	Error 2	Error 4	Error 5	Error 7	Error 8
Error rate 5%						
Rep 1	1.69	2.28	0.30	0.05	1.14	0.48
Rep 2	1.67	2.40	0.27	0.05	1.10	0.49
Rep 3	1.63	2.41	0.26	0.06	1.23	0.47
Rep 4	1.66	2.45	0.26	0.08	1.21	0.40
Rep 5	1.75	2.40	0.25	0.05	1.16	0.36
Mean	1.68	2.39	0.27	0.06	1.17	0.44
SD	0.04	0.06	0.02	0.01	0.05	0.06
Error rate 10%						
Rep 1	3.43	4.83	0.54	0.10	2.34	0.82
Rep 2	3.35	4.75	0.57	0.11	2.35	0.91
Rep 3	3.40	4.85	0.49	0.12	2.45	0.86
Rep 4	3.32	4.86	0.55	0.11	2.35	0.78
Rep 5	3.26	4.78	0.63	0.10	2.45	0.86
Mean	3.35	4.82	0.56	0.11	2.38	0.84
SD	0.07	0.05	0.05	0.01	0.06	0.05
Error rate 20 %						
Rep 1	6.50	9.66	1.17	0.22	4.72	1.74
Rep 2	6.64	9.57	1.20	0.21	4.72	1.72
Rep 3	6.73	9.49	1.10	0.26	4.71	1.70
Rep 4	6.73	9.61	1.13	0.22	4.70	1.67
Rep 5	6.67	9.53	1.17	0.22	4.64	1.73
Mean	6.65	9.57	1.16	0.22	4.70	1.71
SD	0.10	0.07	0.04	0.02	0.03	0.03

[1] Rep represents replicated dataset.

[2] Error is indexed in this table and the unit is 100%. Error 1 is FIV-low; Error 2, FIV-high; Error 4, OTV-low; Error 5, OTV-high; Error 7, FRV-high and Error 8, FRV-0.

Table 2.3 Accuracies[1] of adjusted daily feed intake (DFIa) with three different error adjustment methods[2].

Replication dataset	LMM			MI			MICE		
	5% Rate[3]	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate
1	90.70	82.33	70.63	96.78	92.17	90.48	96.77	94.58	90.40
2	90.50	81.37	66.89	96.37	92.89	89.76	96.83	93.89	90.20
3	90.75	82.89	69.80	97.05	94.59	90.32	96.50	94.79	90.81
4	91.63	83.60	67.69	95.88	94.81	90.81	97.06	93.75	89.62
5	91.38	82.99	68.45	97.49	92.78	89.84	97.03	95.10	89.63
Mean	90.99	82.64	68.69	96.71	93.45	90.24	96.84	94.42	90.13
SD	0.49	0.84	1.52	0.62	1.18	0.44	0.23	0.58	0.52

[1] Accuracies of DFIa with three methods were evaluated with Pearson correlation coefficients of DFIa and true daily feed intake (unit = 100%).

[2] Error adjustment methods include linear mixed model (LMM) approach and multiple imputation with MI and MICE.

[3] To obtain the simulated replication datasets, error visits were introduced to the *'error-free'* complete dataset with rate 5%, 10% and 20%.

Table 2.4 Accuracies[1] of adjusted average daily feed intake (ADFIa) with three different error adjustment methods[2].

Replication dataset	LMM			MI			MICE		
	5% Rate[3]	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate
1	98.89	97.91	91.53	99.87	99.60	98.78	99.88	99.72	98.70
2	98.91	98.19	93.70	99.86	99.55	99.02	99.88	99.58	98.60
3	99.18	97.42	92.49	99.88	99.64	98.81	99.85	99.62	98.95
4	98.37	97.47	92.43	99.33	99.63	98.87	99.88	99.69	98.67
5	99.27	97.24	94.06	99.88	99.52	98.53	99.88	99.67	98.71
Mean	98.93	97.65	92.84	99.76	99.59	98.80	99.88	99.66	98.72
SD	0.35	0.39	1.03	0.24	0.05	0.18	0.01	0.06	0.13

[1] Accuracies of ADFIa with three methods were evaluated with Pearson correlation coefficients of ADFIa and true average daily feed intake (unit = 100%).

[2] Error adjustment methods include linear mixed model (LMM) approach and multiple imputation with MI and MICE.

[3] To obtain the simulated replication datasets, error visits were introduced to the *'error-free'* complete dataset with rate 5%, 10% and 20%.

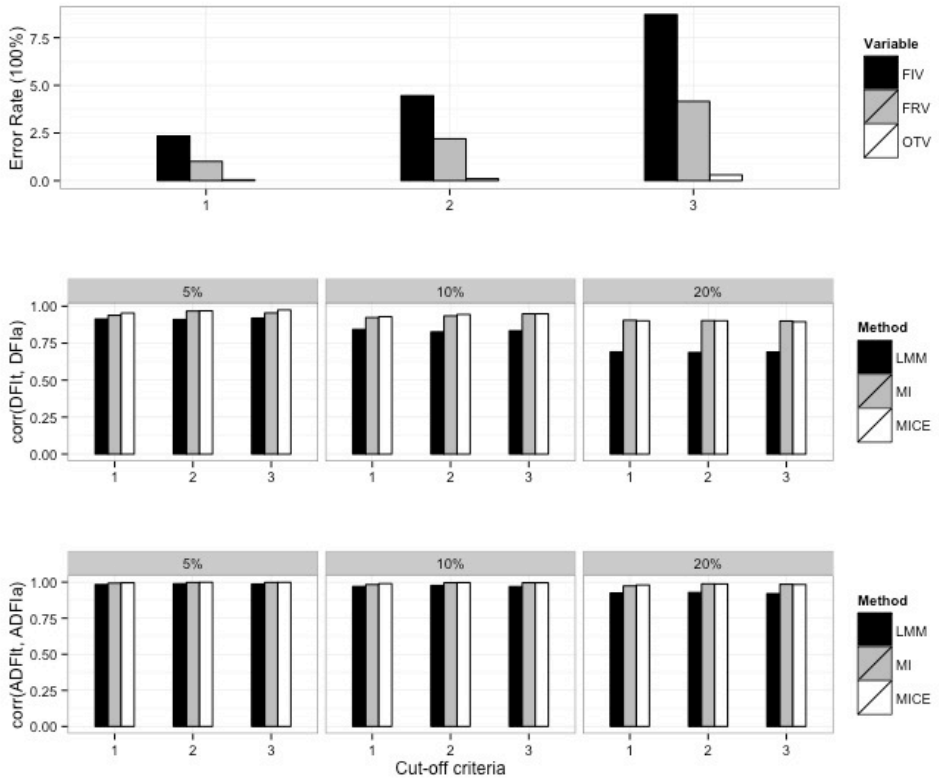

Figure 2.2 The impact of different cut-off criteria (for FIV-high, OTV-high and FRV-high) in full dataset on the performance of error-adjusting methods.

Three different cut-off criteria (denoted as 1, 2 and 3 on x-axis) were used based on distribution of FIV (feed intake per visit), OTV (occupation time per visit) and FRV (feeding rate per visits). The change of cut-off values had impact on the three error types (FIV-high, OTV-high and FRV-high) in the full-dataset (Top figure), where cut-off criteria 2 is used in literature (Casey, 2003) and cut-off criteria 1 is halved that and cut-off criteria 3 is doubled the that in the right tail of their distributions. The middle and bottom figures are shown the change of performance of methods LMM (linear mixed model), MI (multiple imputation by conditional distribution) and MICE (multiple imputation by chained equation).

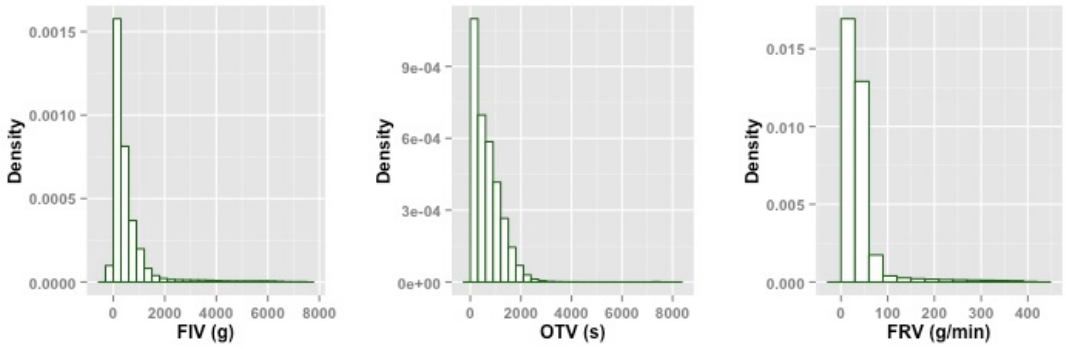

Figure S2.1 Distribution of feed intake per visit (FIV), occupation time per visit (OTV) and feeding rate per visit (FRV) in the full dataset.

Table S2.1 Spearman rank correlation of true daily feed intake (DFI) with adjusted DFI using different kind of error adjustment methods[1].

Replication dataset	LMM			MI			MICE		
	5% Rate[2]	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate
1	0.90	0.80	0.69	0.97	0.93	0.89	0.96	0.93	0.89
2	0.89	0.81	0.67	0.96	0.93	0.89	0.96	0.92	0.89
3	0.90	0.81	0.65	0.97	0.94	0.89	0.96	0.94	0.87
4	0.90	0.81	0.69	0.97	0.94	0.89	0.96	0.92	0.88
5	0.89	0.81	0.67	0.97	0.93	0.88	0.96	0.94	0.88
Mean	0.89	0.81	0.67	0.97	0.93	0.89	0.96	0.93	0.89
SD	0.01	0.00	0.02	0.00	0.01	0.00	0.00	0.01	0.01

[1] Error adjustment methods include linear mixed model approach (LMM) and multiple imputation with MI and MICE.

[2] To obtain the simulated replication datasets, error visits were introduced to the 'error-free' complete dataset with rate 5%, 10% and 20%.

Table S2.2 Spearman rank correlation of true average daily feed intake (ADFI) with adjusted ADFI using different kind of error adjustment methods[1].

Replication dataset	LMM			MI			MICE		
	5% Rate[2]	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate
1	0.94	0.93	0.81	0.99	0.98	0.94	0.99	0.98	0.93
2	0.96	0.93	0.81	0.99	0.98	0.94	0.99	0.97	0.94
3	0.96	0.92	0.82	0.99	0.97	0.94	0.99	0.97	0.95
4	0.93	0.92	0.82	0.99	0.98	0.94	0.99	0.98	0.94
5	0.96	0.91	0.85	0.99	0.97	0.93	0.99	0.98	0.94
Mean	0.95	0.92	0.82	0.99	0.97	0.94	0.99	0.98	0.94
SD	0.02	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.01

[1] Error adjustment methods include linear mixed model approach (LMM) and multiple imputation with MI and MICE.

[2] To obtain the simulated replication datasets, error visits were introduced to the 'error-free' complete dataset with rate 5%, 10% and 20%.

Table S2.3. Regression coefficients[1] of adjusted daily feed intake (DFI) on true DFI for different error adjustment methods[2].

Replication dataset	LMM			MI			MICE		
	5% Rate[c]	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate
1	0.87	0.73	0.61	0.96	0.91	0.93	0.98	0.96	0.93
2	0.85	0.74	0.57	0.96	0.93	0.91	0.97	0.96	0.94
3	0.85	0.75	0.59	0.97	0.95	0.93	0.97	0.96	0.93
4	0.85	0.76	0.58	0.95	0.96	0.93	0.98	0.94	0.93
5	0.86	0.75	0.58	0.98	0.92	0.92	0.98	0.97	0.92
Mean	0.86	0.74	0.59	0.96	0.93	0.92	0.98	0.96	0.93
SD	0.01	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.01

[1] The regression coefficients were obtained from linear regression model of adjusted DFI on true DFI for each scenario.

[2] Error adjustment methods include linear mixed model approach (LMM) and multiple imputation with MI and MICE.

[3] To obtain the simulated replication datasets, error visits were introduced to the *'error-free'* complete dataset with rate 5%, 10% and 20%.

Table S2.4. Regression coefficients[1] of adjusted average daily feed intake (ADFI) on true ADFI for different error adjustment methods[2].

Replication dataset	LMM			MI			MICE		
	5% Rate[3]	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate	5% Rate	10% Rate	20% Rate
1	1.04	1.03	0.92	1.02	1.06	1.10	1.04	1.06	1.02
2	1.02	1.02	1.03	1.03	1.07	1.01	1.03	1.06	1.01
3	0.97	1.01	1.00	1.03	1.05	1.02	1.03	1.05	1.09
4	0.99	1.04	1.03	1.02	1.06	1.10	1.03	1.05	1.01
5	0.99	1.00	1.07	1.02	1.06	1.01	1.03	1.05	1.01
Mean	1.00	1.02	1.01	1.03	1.06	1.01	1.03	1.05	1.02
SD	0.03	0.01	0.05	0.00	0.01	0.01	0.00	0.01	0.02

[1] The regression coefficients were obtained from linear regression model of adjusted ADFI on true ADFI for each scenario.

[2] Error adjustment methods include linear mixed model approach (LMM) and multiple imputation with MI and MICE.

[3] To obtain the simulated replication datasets, error visits were introduced to the *'error-free'* complete dataset with rate 5%, 10% and 20%.

CHAPTER 3

Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc Pigs: I. Genetic parameters estimation and accuracy of genomic prediction

Shihui Jiao[*], Christian Maltecca[*1], Kent A. Gray[†], Joseph P. Cassady[*2]

[*] Department of Animal Science, North Carolina State University, Raleigh, NC 27695

[†] Smithfield Premium Genetics, Rose Hill, NC 28458

ABSTRACT: The efficiency of producing saleable products in the pork industry is largely determined by costs associated with feed and by the amount and quality of lean meat produced. The objectives of this paper were (1) to explore heritability and genetic correlations for growth, feed efficiency and real-time ultrasound traits using both pedigree and marker information; and (2) to assess accuracy of genomic prediction for those traits using Bayes-A prediction models in a Duroc terminal sire population. Body weight at birth (BW at birth) and weaning (BW at weaning), real-time ultrasound traits, including backfat thickness (BF), muscle depth (MD) and intramuscular fat content (IMF), were collected based on farm protocol. Individual feed intake and serial body weight records of 1,563 boars obtained from FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, KS, USA) were edited to obtain growth, feed intake and feed efficiency traits, including ADG, ADFI, feed conversion ratio (FCR) and residual feed intake (RFI). Correspondingly, 1,047 boars were genotyped using the Illumina PorcineSNP60 BeadChip. The remaining 516 boars, as an independent sample, were genotyped with a low-density GGP-Porcine BeadChip and imputed to 60K. Magnitudes of heritability from pedigree analysis were moderate for growth, feed intake and ultrasound traits (ranging from 0.44 ± 0.11 for ADG to 0.58 ± 0.09 for BF); heritability estimates were 0.32 ± 0.09 for FCR but only 0.10 ± 0.05 for RFI. Comparatively, heritability estimates using marker information by Bayes-A models were about half of that from pedigree analysis, suggesting “missing heritability”. Moderate positive genetic correlations between growth and feed intake (0.32 ± 0.05), backfat (0.22 ± 0.04), as well as, negative genetic correlations between growth and feed efficiency traits (-0.21 ± 0.08, -0.05 ± 0.07) indicate selection solely on growth traits may lead to an

undesirable increase in feed intake, backfat and reduced feed efficiency. Genetic correlations among growth, feed intake, and feed conversion ratio assessed by a multiple Bayes-A model resulted in increased genetic correlation between ADG and ADFI, a negative correlation between ADFI and FCR, and a positive correlation between ADG and FCR. Accuracies of genomic prediction for traits investigated, ranging from 9.4% for RFI to 36.5% for BF, were reported that might provide new insight into pig breeding and future selection programs using genomic information.

Key words: Duroc, genetic parameters, genomic prediction, growth and feed efficiency, ultrasound traits

INTRODUCTION

Efficiency of producing saleable pork products is largely determined by costs associated with feed and by the amount of and quality of lean meat produced. Utilizing feed resources more efficiently became a clear challenge that faces the US pork industry (McGlone et al., 2003). In order to produce 'more with less', pork producers need to adopt new technologies to improve feed efficiency of pigs. Ideally, pigs eating less, growing faster, and depositing more lean meat than fat are preferred.

Feed efficiency can be evaluated by feed conversion ratio (FCR) or residual feed intake (RFI). Residual feed intake, first introduced by Koch et al. (1963), was defined as the difference between observed and expected feed intake for an individual (Hoque et al., 2008; Kennedy et al., 1993). Methods for improving feed efficiency through FCR or RFI have been

reported in certain selection programs (Rothschild and Ruvinsky, 2010). However, improving feed efficiency directly by selection is expensive: recording individual feed intake remains labor intensive and time consuming, requiring costive equipment. Additionally, data from FIRE systems often contain errors that require carefully editing before the data can be used (Casey et al., 2005).

The availability of the Porcine60K BeadChip has greatly facilitated whole-genome association studies, contributing to increased accuracy of selection by application of marker-assisted or genomic selection (Fan et al., 2011; Christensen et al., 2012). This project aimed at identifying genomic regions associated with feed efficiency and production traits. As part of the work, the objectives of this paper were (1) to explore heritabilities and genetic correlations for ADFI, ADG, feed efficiency traits (FCR and RFI), ultrasound traits (BF, MD and IMF), BW at birth and weaning, using either pedigree or marker information; and (2) to access accuracy of genomic prediction of the nine traits in a Duroc terminal sire population.

MATERIALS and METHODS

Animal and phenotype data collection

Data used for this study were provided by Smithfield Premium Genetics (Rose Hill, NC). This dataset included 1047 Duroc boars from the mating of 64 sires and 421 sows within a single nucleus farm. These data were also used as reference data to estimate marker effects for genomic prediction. Individual piglet body weight and litter information were recorded within 24 hours of birth and body weight at weaning was recorded at the mean age of 25 days. After weaning, boars were grouped into batches or contemporary groups and fed in FIRE stations, starting at an average age of 85 days. Contemporary groups (or batches)

were defined as boars weaned in the same week. Growth and feed intake were measured on each pig, during a test period of 45 days on average.

Individual feed intake and body weight were recorded when a pig visited the feeder (one in each pen). In total there were 323,639 individual visits recorded. When the boars reached approximately 120 kg, ultrasound backfat, muscle depth and intramuscular fat content and body weight (BW on test) were recorded. The ultrasound images of all animals were captured over the last three ribs via an Aloka 500 ultrasound machine (Corometrics Medical Systems, Wallingford, CT) and analyzed for intramuscular fat using the Swine Image Analysis Software (Designer Genes Technologies, Inc. Harrison, Arkansas).

Additionally, all the traits mentioned above were collected on 516 more boars from the same farm, following the same protocols. Data from the 516 boars were used as validation data in genomic prediction.

Statistical Analysis

Average daily gain was calculated using two methods, in order to maximize information from different sources of data: (1) Simple linear regression using BW at weaning (at an average of 26 days, approximately 7.75 kg) and ultrasound BW on test (at an average of 103 days, approximately 120kg) 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 plotted against age. Previous studies showed that robust regression could be applied to edit this serial pig body weight data (Zumbach et al., 2010; 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 the bisquare weight function in R. By fitting a robust regression with age and squared age as covariates, each data point was assigned a weight (from 0 to 1) to minimize the influence of abnormal data points. (2) Data with weight less than 0.5 were treated as outliers, which is similar to Zumbach et al. (2010). Animals with less than 20 body weight records or with less than 30 days of information were excluded. The predicted body weights from robust regression were used to calculate average daily gain for 599 boars. Average daily gain estimates for 927 boars were obtained by combining the two data sets from the above analysis.

Average daily feed intake was obtained by editing 323,639 records collected from the FIRE system. Data from electronic feeders have been found to contain a substantial number of errors (De Haer et al., 1992; Eissen 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 and coworkers (2005).

The main steps in the edit procedures were to firstly 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 in this study was higher than 5%, which was previous reported (Casey et al., 2005; Cai et al., 2008). Secondly, compute error-free feed intake for each pig and day by summing feed consumed in visits without identified errors. The third step involves estimating the effect of error counts on error-free daily feed

intake by fitting a linear mixed model to error-free daily feed intake observations with batch as fixed effect, 31 variables created from the 16 error counts, and ADG and body weight as covariates, and pig as a random effect,

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

where y_{imnp} = error-free daily feed intake from the m[th] pig on the n[th] day in the i[th] batch; B_i = fixed effect of the i^{th} batch; BW_{mn} = body weight of the m[th] pig on the n[th] day as a covariate; ADG_m = average daily gain of the m[th] pig as a covariate; ETP_{pmn} = percentage of error type p for the m^{th} pig on the n^{th} day as a covariate; OTD_{pmn} = daily occupation time, summing over visits from the the m^{th} pig on the n^{th} day that contained the error type p; FID_{pmn} = daily feed intake summed over visits from the m^{th} pig on the n^{th} day that contained the error type p; P_m = effect of the m^{th} pig, which was assumed random with $P_m \sim N(0, \sigma_p^2)$; e_{imnp} = residual with $e_{imnp} \sim N(0, \sigma_e^2)$; all bs are corresponding coefficients.

For the forth step, adjust error-free daily feed intake for each pig and day for feed consumed in error visits by adding estimates of covariates from last step. Finally, compute ADFI for each pig by averaging daily feed intake during the test period.

FCR was calculated based on ADFI and ADG (ADG/ADFI) and RFI was calculated using an animal model:

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

where y_{ij} = ADFI; TA_{ij} = fixed regression covariate of on test age for each boar that entered the FIRE system; ADG_{ij} = fixed regression covariate of ADG; BF_{ij} = fixed regression

covariate of BF; B_i = random effect of contemporary group effect with $B_i \sim N(0, \sigma_b^2)$; P_{ij} = effect of the j^{th} pig from i^{th} batch, which was assumed random with $P_{ij} \sim N(0, \sigma_p^2)$; e_{ij} = residual with $e_{ij} \sim N(0, \sigma_e^2)$. Measures of RFI for individual pigs were obtained as the residuals from the above model.

Phenotypes of the additional 516 boars were collected and edited in the same procedures described above.

Phenotypes of 1047 animals were used to estimate genetic parameters by incorporating pedigree information. Pedigree for each animal was traced back 3 generations, resulting in a total of 2593 individuals included in the pedigree file. Genetic and residual (co)variances for the 9 traits were estimated using the following animal models in ASReml (Gilmour et al., 2009):

$$Y_{ijkl} = B_i + L_j + P_k + PIG_{ijkl} + D_{ijkl} + e_{ijkl},$$

where Y_{ijkl} = BW at birth; B_i, L_j, and P_k are fixed effects for batch, litter size and sow parity; PIG_{ijkl} = random pig effects with $PIG_{ijkl} \sim N(0, A\sigma_p^2)$, where A is the relationship matrix; D_{ijkl} = maternal effect and e_{ijkl} = model residual with $e_{ijkl} \sim N(0, \sigma_e^2)$. Body weight at weaning was analyzed with the above model with weaning age included as an additional covariate and adding permanent dam effects. ADG, ADFI, and FCR were analyzed with a similar model with litter, parity, and dam effects removed from the model while BF, MD and IMF were analyzed adding ultrasound BW on test as a covariate. Residual feed intake was analyzed only including the mean as a fixed effect since RFI was obtained from the model with all fixed effects included.

Genetic covariance between any two investigated traits were obtained by exploring a bivariate animal model with same fixed effects described for single trait animal models above. However, BW at birth and BW at weaning were excluded due to limited amount of progeny for each sow to accurately estimate maternal effects and permanent environment effects in the two-trait animal models.

Genomic data from 1,047 boars were obtained using the Illumina PorcineSNP60K Bead (Illumina Inc., San Diego, CA). The SNP with call rates $\leqslant 0.90$, MAF (minor allele frequency) $\leqslant 0.002$, and p-value < 0.0001 of a chi-square test for Hardy-Weinberg equilibrium and individual with call rate $\leqslant 0.90$ were excluded from the genotype data set. After quality control was completed, 40,008 SNPs and 1,022 boars remained. Missing SNP genotypes were imputed for all available boars with pedigree information using AlphaImpute1.1.0 (Hickey et al., 2011). A total of 40,008 SNP (including 35,870 SNPs on autosomes and X chromosome) out of 64,232, were qualified for variance components estimation.

A total of 516 additional boars with both genotype and phenotype information were used as a validation dataset for genomic prediction. Genotypes of the 516 boars were obtained using a low-density panel (GGP-Porcine BeadChip, 9K), which included 8,826 SNPs chosen from the 60K bead chip. Quality control for the marker genotypes were performed using the same criterion described above and 6,028 SNPs remained. Imputation was performed using MaCH1.0 (Li et al., 2010) and Minimac (Howie et al., 2012) at each chromosome because imputation using the two packages was faster than AlphaImpute (unpublished study). After imputation, 35,870 SNPs were obtained for each boar typed by 9K

chip as validation set. Imputation accuracies were investigated by masking 30% and 60% SNPs from the 60K chip using non-overlapping sliding windows (Huang et al., 2012) on chromosome 1 and 2.

To simultaneously estimate SNP effects to derive the prediction equation, a single trait Bayesian approach, called Bayes-A by Meuwissen et al. (2001), was used via GenSelv4.0 software (http://bigs.ansci.iastate.edu). The model fitted was,

$$\boldsymbol{y} = \mathbf{1}_{\boldsymbol{n}}^{'}\boldsymbol{\mu} + \boldsymbol{Xu} + \boldsymbol{e},$$

where **y** is a vector of n pre-adjusted phenotypes (taking out fixed effects accordingly); $\boldsymbol{\mu}$ is the general mean; **X** is a design matrix of marker genotypes (m) for n individuals with elements coded as -10, 0, 10 as required by GenSel; **u** is a (n × 1) vector of SNP effects assumed normally distributed with mean 0 and variance σ_u^2; **e** is a vector of residuals assuming $e_i \sim N(0, \sigma_e^2)$. The Bayes-A model has a prior assumption that SNP effects are t distributed with a specific marker variance. We predicted the GEBVs in the validation sets as $\boldsymbol{GEBV} = \boldsymbol{X_2}\hat{\boldsymbol{u}}$ and the predicted phenotypes was then $\hat{\boldsymbol{y}} = \boldsymbol{\mu} + \boldsymbol{X_2}\hat{\boldsymbol{u}}$ where $\boldsymbol{X_2}$ was design matrix of marker genotypes in validation set. Prediction accuracy for each trait was assessed by using the formula derived by Legarra et al. (2008),

$$r(g, \hat{g}) = \frac{r(y, \hat{y})}{H}$$

where r is correlation; g = overall genetic value; $\hat{g}$ = predicted genetic value; y = phenotype; $\hat{y}$ = predicted phenotype value; $H^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$ is the broad-sense heritability.

A three-trait Bayes-A analysis was used to investigate relationships among ADG, ADFI, and FCR adjusting the dimension of the previous model accordingly (Jia et al., 2012). After

solving the model and obtaining SNP effect estimates, breeding values for each window with 50 markers across the entire genome was computed instead of a single marker to better understand genomic regions involved in or associated with two traits.

RESULTS

In this study, growth, feed intake efficiency and real-time ultrasound measurements of 1,047 boars from a Duroc nucleus herd were collected and analyzed. An additional 516 boars were genotyped with a low-density chip and used as validation to evaluate genomic prediction accuracies.

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 the covariate associated with each error type estimated from the mixed linear model for the reference dataset are summarized below in Table 3.1. Error rate ranged from 0% to approximately 16% in the reference dataset. Records in the reference dataset containing at least 1 error type was 35.13%, which is much higher than previously reported error rate of approximately 5% (Casey et al., 2005; Cai et al., 2008). Error rate within the validation set was much lower at 14.92%. Error free daily feed intake was 1567 g/d (1577 g/d in validation set) before any adjustment, which was lower than the adjusted daily feed intake of 1996.9 g/d (2002.3 g/d in validation set). This 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.

Robust regression for single pig analysis was performed using serial measurements 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 value (< 0.5), resulting in ADG computed for 599 pigs. Body weights plotted against age before and after editing of two randomly selected pigs are shown in Figure 3.1. Clearly, data points far from the 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 0.89. Thus ADG of the additional 83 pigs obtained from robust regression were added to the 889 boars, for a total of 972 boars after ADG editing. Descriptive statistics are listed in Table 3.2 for ADG from linear regression and robust regression. Robust regression resulted in a higher mean estimate of ADG and larger standard deviation than estimates from simple linear regression.

Descriptive statistics of the 9 traits investigated in the reference and the validation datasets are listed in Table 3.3.

Genetic parameter estimation

Estimates of additive genetic variance, residual variance and heritability from single trait animal models for the reference dataset are represented in Table 3.4. Moderate heritability estimates for ADG, ADFI, FCR, BF, MD, IMF and BW at birth were obtained from the present analysis, with heritabilities estimated to be 0.44 ± 0.11, 0.66 ± 0.11, 0.32 ± 0.09, 0.58 ± 0.09, 0.39 ± 0.09, 0.54 ± 0.11 and 0.34 ± 0.28, respectively, while heritability estimates for RFI and BW at weaning were lower, less than 10%. Genetic covariance and correlation between any two traits except BW at birth and BW at weaning are shown in 3.5. Moderate positive genetic correlations between growth and feed intake (0.32 ± 0.05), growth and backfat (0.22 ± 0.04) as well as negative genetic correlations between growth and feed

efficiency traits (-0.21± 0.08, -0.05± 0.07) indicate that selection solely on growth traits may lead to undesirable increases of feed intake, backfat, and reduced feed efficiency.

In contrast to the pedigree analysis, genomic variance components and marker heritability from single-trait Bayes-A models are shown in Table 3.6, using 35,870 markers on autosomes and the X chromosome. In comparison of the heritability estimates in Table 3.4, most of the marker heritability estimates were approximately half of the heritability obtained from the classical animal model, except RFI and BW at weaning. Covariance components estimated from a multiple-trait Bayes-A model are displayed in Table 3.7. Genetic correlations among growth, feed intake and feed conversion ratio assessed by a multiple-trait Bayes-A model resulted in increased genetic correlation between ADG and ADFI (0.82), a negative correlation between ADFI and FCR (-0.13) and a positive correlations between ADG and FCR (0.40). Similar to the multiple-trait analysis using traditional animal model, traits borrowed large amounts of information from other traits.

Marker effects estimation and genomic prediction accuracy

Genetic correlations between quantitative traits indicate that measurements of one trait can be informative for other traits. Single trait analysis does not take this information into account. Using single trait models might be a disadvantage not only in variance component estimation, but also in finding common chromosomal regions for traits of interests in a breeding objective. Window BV (breeding values) were computed using $X_i\hat{u}_i$, where i is a genomic region, i.e. we used 50 markers per region i in this study, i = 1, 2, …, 800. Several obvious genomic regions that may affect both ADG and ADFI were identified (Figure 3.2).

To investigate the effect of different loci effects on accuracy of prediction for different traits, we used SNP effects for each trait from a single trait Bayesian approach. Prediction accuracy was computed for each trait and is shown in Table 3.8.

Prediction accuracies were different for each trait, ranging from 0.094 to 0.365. However, compared to genomic prediction projects completed in dairy or beef cattle, accuracy was much lower. The average relationship of animals in the reference and validation set was 0.044. To investigate the effect of relationship between animals in the reference dataset and animals in the validation data, animals in the validation set were divided into 3 groups based on the average numeric relationship of one boar in validation and all other boars in reference set. Group 0 contains boars in the validation dataset having no relationship (within 3 generation) with boars in the reference set (completely unrelated); as well as group 1 represente moderate relationships which were defined as an average relationship less than 0.044; while group 2 included boars in the validation set related more highly than the other groups (with average relationship larger than 0.044). A boxplot (Figure 3.3) was created to illustrate that the closer related animals between the validation and reference datasets, the more accurate the prediction would be.

DISCUSSIONS

Feed intake visit records from electronic feeders

Collection of growth records and real-time ultrasound measurements are relatively simple and has been done on a routine basis in many seed-stock swine units. However, individual feed intake collection is limited due to the fact that it is more difficult and expensive to measure. The FIRE systems are frequently used by pig companies 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 pigs, electronic feeders are prone to malfunctions and may produce erroneous feed intake and body weights records (Chen et al., 2010; Zumbach et al., 2010). A summary of the error rate for each type is shown in Table 3.1. Individual records were used to compute average daily feed intake for each pig. The mean of ADFI for this age of Duroc boars was approximately 2 kg; similar results have been published by (Cai et al., 2008) for Yorkshire pigs. Due to the high error rate detected in body weight measurements of pigs using the FIRE system data, weaning weights and weights taken when ultrasound measurements 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 adjustments as well as RFI computation. Development 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.

Variance components and genetic parameters estimation: animal model vs. whole-genome Bayesian model

Heritability for each trait was estimated using an animal model in the present study (Table 3.4). All the nine traits investigated were moderately to highly heritable except feed efficiency trait RFI as well as BW at weaning. Estimated heritabilities for ADG and ADFI using animal models were 0.44 ± 0.11 and 0.66± 0.11 respectively in the present study and were consistent with estimates of 0.48 and 0.49 reported by Hoque et al. (2008; 2009) in Duroc pigs and the estimates (0.51 and 0.46) published by Schulze and collegues (2001) for

563 boars tested in central stations by electronic feeders. In addition, the heritability estimates for ADG and ADFI in the current study are slightly higher than literature averages (0.29 and 0.29, respectively) (Rothschild and Ruvinsky, 2010). Estimated heritability for real-time ultrasonic traits including BF and IMF were also in agreement with estimates by Lo et al. (1992), who estimated the heritability for BF and IMF to be 0.54 and 0.52, respectively, for crossbreds of Landrace and Duroc pigs and were also similar to the estimates reported by Suzuki et al. (2005). Heritability estimates of two components of feed efficiency traits FCR and RFI in the present study were 0.32 ± 0.09 and 0.10 ± 0.05, respectively. The heritability estimate for FCR was consistent with the estimates in literatures, ranging from 0.12 to 0.58 with an average of 0.30, but heritability for RFI was slightly lower than estimates commonly reported in the literature, from 0.1 to 0.42 with mean 0.24 (Johnson et al., 1999; Rothschild and Ruvinsky, 2010). Kaufmann et al. (2000) indicated that the direct heritability for individual piglet BW at birth and weaning were 0.02 and 0.08 and maternal heritability was 0.21 and 0.16, respectively, in Large White pigs. However, the estimate of direct heritability for BW at birth was much higher than that and had a very high standard error (0.34 ± 0.28), which may result from the smaller sample size.

Results in Table 3.5 represent the genetic correlation estimates for any two traits investigated in the present study except BW at birth and at weaning. Moderate genetic correlations between growth and feed intake (0.32 ± 0.05), backfat (0.22 ± 0.04), as well as negative genetic correlations between growth and feed efficiency traits (-0.21 ± 0.08, -0.05 ± 0.07) were found in the study, indicating that selection solely on growth traits may lead to undesirable increases in feed intake, backfat and a reduced feed efficiency. Lo et al. (1992)

reported genetic correlation between ADG and BF of 0.28, which was lower than the estimate of 0.67 reported by Schulze et al. (2001) and slightly lower than 0.33 suggested by Hoque et al. (2008). Hoque et al. (2008) also reported correlation between DG and FI of 0.34, DG and RFI of -0.05, DG and FCR of -0.22 and FCR and RFI of 0.95, which is higher than the current estimate (0.53). Negative correlations (-0.37 and -0.54) in two feeding regimes have been found between ADG and FCR by Schulze et al. (2001).

In contrast, heritability estimates obtained from genome-wide dense markers were generally lower than estimates obtained from traditional animal models, and in most cases is approximately half of the corresponding estimates for most of the traits (Table 3.6). The reason why the estimates of heritability from whole-genome studies are small may be the result of "missing heritability", which has been a hotly debated issue in human genetics (Eichler et al., 2010), or may be due to different factors of whole genome-wide evaluation. The reason why genome-wide studies only explain a relatively small proportion of heritability and have small effect estimates may be due to the genetic architecture of the traits, epistatic effects, genotype by environments interactions and other similar factors (Makowsky et al., 2011). Genetic correlations among quantitative traits indicate relationships among the traits. Current single trait analyses do not take this information into account and might be disadvantageous in variance component estimation. Jia et al. (2011) reported increased genetic value prediction accuracy with multiple-trait genomic selection, and suggested that low-heritability traits benefit from correlated high-heritability traits in genomic selection. Our results in Table 3.7 strongly agree with the conclusion they made: by borrowing information from other traits, genetic correlation between ADG and ADFI boosted

(See Table 3.5 and Table 3.7). By exploring multiple Bayes-A analysis, genomic regions of improved feed efficiency traits (Less input, more output, which is lower ADFI and higher ADG) were identified on chromosome 1. Further investigation will be needed to uncover the genes or mutations in the identified region (Figure 3.2).

Genomic prediction accuracy

Genomic prediction of future phenotype or genetic merit using dense SNP genotypes can be used for estimating genomic breeding values in selection of livestock, especially in dairy cattle (Hayes et al., 2009) and beef cattle (Garrick et al., 2011) but are rarely reported in swine (Christensen et al., 2012). In the current study, prediction accuracies were different for each trait, ranging from 9.4% to 36.5%, using approximately 1000 animals as reference and 500 animals as validation. The prediction accuracies for growth (ADG), feed intake (ADFI), and feed efficiency (FCR and RFI) were 24%, 15%, 11% and 9.4%, accordingly. Our results are consistent with the studies of Christensen et al. (2012), in which they compared accuracies of predicted breeding values for average gaily gain and feed conversion ratio in Danish Duroc pigs using a single-step method and genomic BLUP, with prediction accuracies of 18%-35% for ADG and 18%-23% for FCR for genotyped animals from different single-trait models. However, compared to genomic prediction studies completed in dairy or beef cattle, the accuracies were low. There are several possible reasons for the low accuracies. Firstly, the sample size of the validation set is small, only approximately 500 boars for each trait, which is half of the sample size of the reference set. Secondly, genotypes for boars in the validation dataset were imputed from 9K panel to 60K; therefore, imputation error can lower the prediction accuracy. The accuracy of imputation was estimated to be

93.12% and 91.69 % when masking 60% markers in 9K for chromosome 1 and 2. When masking only 20% of the makers, imputation accuracies were increased to 98.76% and 99.12%. Furthermore, Hayes et al. (2010) suggested that genomic prediction accuracy may also rely on the underlying genetic architecture of traits, or the heritability.

CONCLUSIONS

In the current study, variance components for nine traits investigated including ADG, ADFI, FCR, RFI, BF, MD, IMF, BW at birth and BW at weaning, as well as genetic correlations among them, were estimated using single-trait and two-trait animal models as well as single-trait Bayes-A and multiple-trait Bayes A approaches. Most traits were moderately to highly heritable with the exception being feed efficiency traits FCR and RFI and BW at weaning. Moderate positive genetic correlations between growth and feed intake, growth and backfat, as well as negative genetic correlations between growth and feed efficiency traits were found in the study. Based on our results, selection solely on growth traits might lead to an undesirable increase in feed intake, backfat and reduced feed efficiency.

Heritability estimates using a whole genome dense marker panel were approximately half of the estimates from a traditional animal model, indicating that “missing heritability” existed in all the traits. Utilizing dense marker genotypes provides a wealth of information regarding the genetic makeup of each pig, providing new insights into estimates of heritability and correlations between traits in a genomic perspective. Accuracies of genomic prediction were investigated for nine traits in pigs, which may provide new insight into pig breeding and future selection programs.

LITERATURE CITED

Cai, W., D. S. Casey, and J. C. M. Dekkers. 2008. Selection response and genetic parameters for residual feed intake in Yorkshire swine. J. Anim. Sci. 86:287-298.

Casey, D. S., H. S. Stern, and J. C. M. Dekkers. 2005. Identification of errors and factors associated with errors in data from electronic swine feeders. J. Anim. Sci. 83:969-982.

Chen, C. Y., I. Misztal, S. Tsuruta, B. Zumbach, W. O. Herring, J. Holl, and M. Culbertson. 2010. Estimation of genetic parameters of feed intake and daily gain in Durocs using data from electronic swine feeders. J. Anim. Breed. Genet. 127:230-234.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6:1565-1751.

De Haer, L. C. M., J. W. M. Merks, H. G. Kooper, G. A. J. Buiting, and J. A. Van Hattum. 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.

Eichler, E. E., J. Flint, G. Gibson, A. Kong, S. M. Leal, J. H. Moore, and J. H. Nadeau. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet.11:446-450.

Eissen, J. J., E. Kanis, and J. W. M. Merks. 1998. Algorithms for identifying errors in individual feed intake data of growing pigs in group-housing. Appl. Eng. Agric. 14:667-673.

Fan, B., S. K. Onteru, Z. Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011.

Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. PLoS ONE 6:e14726.

Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genet. Sel. Evol. 43:17.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433-443.

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet. 6:e1001139.

Hickey, J. M., B. P. Kinghorn, B. Tier, J. F. Wilson, N. Dunstan, and J. H. van der Werf. 2011. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. Genet. Sel. Evol. 43:12.

Hoque, M. A., H. Kadowaki, T. Shibata, T. Oikawa, and K. Suzuki. 2009. Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livest. Sci. 121:45-49.

Hoque, M. A., H. Kadowaki, T. Shibata, and K. Suzuki. 2008. Maternal and direct genetic parameters for production traits and maternal correlations among production and feed efficiency traits in Duroc pigs. Asian-Aust. J. Anim. Sci. 7:961-966.

Howie B., C. Fuchsberger, M. Stephens, J. Marchini, and G.R. Abecasis. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. 44:955-959.

Huang Y., J. M. Hickey, M. A. Cleveland, and C. Maltecca. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. Genet. Sel. Evol. 44:25.

Jia, Y., and J. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192: 1513-1522.

Johnson, Z. B., J. J. Chewning, and R. A. Nugent. 1999. Genetic parameters for production traits and measures of residual feed intake in large white swine. J. Anim. Sci. 77:1679-1685.

Kaufmann, D., A. Hofer, J. P. Bidanel, and N. Künzi. 2000. Genetic parameters for individual birth and weaning weight and for litter size of Large White pigs. J. Anim. Breed. Genet. 117:121-128.

Kennedy, B. W., J. H. Van der Werf, and T. H. Meuwissen. 1993. Genetic and statistical properties of residual feed intake. J. Anim. Sci. 71:3239-3250.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22: 486-494.

Legarra, A., C. Robert-Granié, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611-618.

Li Y., C. J. Willer, J. Ding, P. Scheet, and G. R. Abecasis. 2010. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. 34:816-834.

Lo, L. L., D. G. McLaren, F. K. McKeith, R. L. Fernando, and J. Novakofski.1992. Genetic analyses of growth, real-time ultrasound, carcass, and pork quality traits in Duroc and Landrace pigs: I. Breed effects. J. Anim. Sci. 70:2373-2386.

McGlone, J. J., and W. Pond. 2003. Pig Production: Biological Principles and Applications. Delmar Learning, Clifton Park, NY.

Makowsky, R., N. M. Pajewski, Y. C. Klimentidis, A. I. Vazquez, C. W. Duarte, D. B. Allison, and G. de los Campos. 2011. Beyond missing heritability: prediction of complex traits. PLoS genet. 7:e1002051.

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

Rothschild, M. F., and A. Ruvinsky. 2010. The genetics of the pig. 2nd ed. CABI, Cambridge, MA.

Schulze, V., R. Roehe, H. Looft, and E. Kalm. 2001. Effects of continuous and periodic feeding by electronic feeders on accuracy of measuring feed intake information and their genetic association with growth performances. J. Anim. Breed. Genet. 118:403-416.

Suzuki, K., M. Irie, H. Kadowaki, T. Shibata, M. Kumagai, and A. Nishida. 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. 83:2058-2065.

Zumbach, B., I. Misztal, C. Y. Chen, S. Tsuruta, M. Łukaszewicz, W. O. Herring, and M. Culbertson. 2010. Use of serial pig body weights for genetic evaluation of daily gain. J. Anim. Breed. Genet. 127: 93-99.

Table 3.1 Types of errors in FIRE (Feed Intake Recording Equipment) data, error rates, and coefficients of errors from a linear mixed model for ADFI adjustment.

Error index	Error type[1]	Error rate	Coefficient of PTE[2]	SE	Coefficient of OTD[3]	SE	Coefficient of FID[4]	SE
1	FIV-low	0.14	-7.23E2	35.15	0.05	4.31	-	-
2	FIV-high	0.04	61.40	6.92	-0.18	8.30	-	-
3	FIV-0	0.01	-1.68E3	63.07	-	-	-	-
4	OTV-low	0.60E-3	-0.18E-2	16.84	-	-	0.13	3.89
5	OTV-high	0.14	1.75E3	4.00	-	-	-0.19	1.28
6	FRV-high-FIV-low	0	0	0	-	-	-	-
7	FRV-high-strict	0.30E-2	-1.35E3	12.53	0.23	5.45	-	-
8	FRV-high	3.7E-2	-7.12E2	10.79	0.17	9.50	-	-
9	FRV-0	0.20E-2	-6.46E2	4.57	-0.12	3.79	-	-
10	FRV-low	1.59E-1	-6.71E2	30.52	-0.14	23.37	-	-
11	LWD-low	0.03	-1.00E3	22.20	-0.05	3.77	-	-
12	LWD-high	0.04	-7.99E2	17.61	0.05	3.15	-	-
13	FWD-low	0.03	-5.57E2	13.38	0.50E-2	0.24	-	-
14	FWD-high	0.04	-1.17E3	15.90	-0.01	0.52	-	-
15	LTD-low	0.02	-1.10E3	25.27	-	-	0.06	3.99
16	FTD-high	0.02	-1.39E3	34.35	-	-	0.04	2.32

[1] Sixteen error types were proposed by Casey et al. (2005): FIV = feed intake per visit; FIV-low, FIV-high and FIV-0 referred to FIV < -20, FIV > 2000g and FIV when consumption time = 0 s. OTV = occupation time per visit; OTV-low and OTV-high referred to OTV < 0 s and OTV > 3600 s, respectively. FRV = feed intake rate per visit; FRV-high-FIV-low considered visit with FIV > 500g/min and 0g <FIV < 50g as an error. FRV-high-strict referred to FRV > 110 g/min with FIV≥ 50g and followed by a visit with FIV ≤ -20g. FRV-high considered FRV > 170 g/min with FIV≥ 50g and not followed by a visit with FIV ≤ -20g as an error. FRV-0 referred to FRV = 0 g/s but with OTV > 500s, and FRV-low referred to abs (FRV) = 2 g/min. LWD = leading weight difference = entrance weight of following visit - exit weight of present visit; LWD-low and LWD-high referred to LWD < -20g and LWD > 1800g. FWD = following weight difference = entrance weight of present visit – exit weight of preceding visit; FWD-low and FWD-high considered FWD < -20g or FWD > 1800 g as error visits. LTD = leading time difference; LTD-low referred to LTD < 0s. FTD = following time difference; FTD-high referred to FTD < 0s.

[2] PTE is percentage of visits with error for each error type, as covariate in the linear mixed model.

[3] OTD is daily occupation time summed over visit with error daily, as covariate in the linear mixed model.

[4] FID represents daily feed intake summed over all visits with errors, as covariate in the linear mixed model.

Table 3.2 Statistics for ADG[1] estimates obtained from linear regression and robust regression.

ADG, g/d	Linear Regression	Robust regression
N[2]	899	599
Min	0.52	0.40
1st Quintile	0.74	0.70
Median	0.79	0.84
Mean	0.79	0.83
3rd Quintile	0.84	0.98
Max	1.08	1.10
SD	0.08	0.24

[1] The ADG estimates obtained from linear and robust regression overlapped for 516 pigs; correlation between ADG estimates obtained from linear regression and the ones from robust regression was computed using those boars (r=0.89).

[2] N represents the number of observations.

Table 3.3 Descriptive statistics for nine traits analyzed in the reference and validation dataset.

Statistics	Dataset	ADFI, g	ADG, g	FCR[1]	RFI, g[2]	BW at birth, kg	BW at weaning, kg	BF, cM[3]	MD, cM[4]	IMF(%)[5]
No.	Reference	972	972	972	972	1047	889	1047	1047	730
min		1272.00	449.30	0.21	-843.10	1.03	2.95	0.51	2.38	2.42
1stQ		1818.00	725.30	0.36	-139.70	1.80	6.40	0.84	3.82	3.44
mean		2003.00	776.90	0.39	0.00	2.03	7.75	1.01	4.15	3.69
3rdQ		2179.00	828.60	0.42	144.20	2.26	9.07	1.14	4.43	3.96
max		3551.00	1092.90	0.65	1457.00	3.85	15.79	1.98	5.78	4.81
SD		288.71	86.70	0.06	260.96	0.34	1.93	0.22	0.47	0.40
No.	Validation	506	515	476	473	519	512	516	516	327
min		1000.12	450.20	0.15	-1746.00	0.00	1.05	0.55	2.74	2.90
1stQ		1738.00	732.60	0.29	-362.60	1.71	2.38	0.86	3.68	3.71
mean		2240.00	776.50	0.37	0.00	1.93	2.83	1.03	3.99	3.90
3rdQ		2716.33	833.20	0.41	368.80	2.16	3.24	1.18	4.27	4.06
max		3827.12	1009.00	0.96	3067.00	3.26	5.53	2.19	5.28	5.10
SD		399.21	85.86	0.14	662.81	0.41	0.67	0.24	0.45	0.31

[1] FCR is feed conversion ratio (ADG / ADFI).

[2] RFI is residual feed intake.

[3] BF represents ultrasound backfat thickness.

[4] MD is ultrasound muscle depth.

[5] IMF represents ultrasound intramuscular fat percentage.

Table 3.4 Estimates of variance components and heritability for traits by single-trait animal models.

Trait	Additive variance	Residual variance	Heritability
ADG, g/d	2293.10 (680.60)	2928.70 (485.20)	0.44(0.11)
ADFI, g/d	29806.50 (10411.00)	58838.90 (7888.90)	0.66(0.11)
FCR[1]	4.90E-4 (1.70E-4)	1.06E-4 (1.30E-4)	0.32(0.09)
RFI, g/d[2]	5476.80 (3087.50)	51088.00 (3521.30)	0.10(0.05)
BF, cM[3]	0.30E-2 (0.5E-2)	2.10E-3 (0.4E-3)	0.58(0.09)
MD, cM[4]	0.01 (0.30E-2)	0.02 (0.2E-2)	0.39(0.09)
IMF, %[5]	0.07 (0.02)	0.06 (0.01)	0.54(0.11)
BW at birth, kg	0.08 (0.02)	0.04 (0.01)	0.34(0.28)
BW at weaning, kg	1.60 (0.35)	0.89 (0.23)	0.05(0.09)

[1] FCR is feed conversion ratio (ADG / ADFI).

[2] RFI is residual feed intake.

[3] BF represents ultrasound backfat thickness.

[4] MD is ultrasound muscle depth.

[5] IMF represents ultrasound intramuscular fat percentage.

Table 3.5 Genetic correlations between traits investigated using two-trait animal models. Covariance estimates are shown above the diagonal and correlation estimates are shown below diagonal, SE of estimates presented in parentheses.

Trait	ADG	ADFI	FCR[1]	RFI[2]	BF[3]	MD[4]	IMF[5]
ADG, g/d		3095.91(1110.20)	-0.59 (0.11)	-229.46 (218.41)	0.75 (0.18)	-1.31(1.06)	3.94 (0.98)
ADFI, g/d	0.32 (0.05)		0.99 (1.08)	805.81 (100.28)	3.13 (0.34)	3.10 (0.09)	-24.68 (2.98)
FCR	-0.21 (0.08)	0.13 (0.11)		1.92 (0.27)	-5.36 (0.56)	-8.00E-2(1.40E-2)	-1.8E-3 (0.17)
RFI, g	-0.05 (0.07)	0.07 (0.09)	0.53 (0.31)		-1.81(0.37)	-0.31(0.03)	0.58 (0.12)
BF, cM	0.22 (0.04)	0.36 (0.04)	-0.12 (0.23)	-0.11 (0.19)		2.00E-3 (4.10E-2)	9.50E-4(2.2E-4)
MD, cM	-0.21 (0.05)	0.19 (0.06)	-0.16 (0.28)	-0.04 (0.25)	0.34 (0.12)		0.01(0.09)
IMF, %	0.24 (0.04)	-0.05 (0.06)	-0.14 (0.27)	0.03 (0.23)	0.46 (0.09)	0.44 (0.13)	

[1] FCR is feed conversion ratio (ADG / ADFI).

[2] RFI is residual feed intake.

[3] BF represents ultrasound backfat thickness.

[4] MD is ultrasound muscle depth.

[5] IMF represents ultrasound intramuscular fat percentage.

Table 3.6 Posterior means of variance explained by whole-genome markers for trait investigated.

Trait[1]	Residual variance[2]	Genetic variance[3]	Total variance	Marker heritability[4]
ADG, g	4604.09	1718.19	6385.28	0.28
ADFI, g	59696.30	13142.00	72416.60	0.18
FCR	1.64E-3	0.25E-3	1.89E-3	0.13
RFI, g	542.50	31.01	573.51	0.05
BF, cM	0.03	0.03	0.05	0.51
MD, cM	0.14	0.07	0.21	0.35
IMF, %	0.54	0.20	0.74	0.27
BW at birth, kg	0.09	0.03	0.11	0.24
BW at weaning, kg	2.25	1.32	3.58	0.37

[1] FCR is feed conversion ratio (ADG / ADFI); RFI is residual feed intake; BF is ultrasound back fat thickness; MD represents ultrasound muscle depth; IMF is ultrasound innermuscular fat percentage.

[2] Residual variance obtained from the mean of posterior distribution of residual variance.

[3] Genetic variance referred to posterior mean of genetic variance for each trait.

[4] Marker heritability computed as the ratio of genetic variance to total variance.

Table 3.7 Variance components and genetic correlations from multiple-trait Bayes-A model among ADG, ADFI and FCR using only marker information (covariance estimates between traits displayed above the diagonal, genetic correlations displayed below the diagonal).

Trait	ADG	ADFI	FCR[1]
ADG, g		124.40	51.56
ADFI, g	0.82		-34.13
FCR	0.40	-0.13	

[1] FCR represents feed conversion ratio (ADG / ADFI).

Table 3.8 Realized genomic prediction accuracy for traits investigated.

Trait[1]	Heritability[2]	Prediction accuracy
ADG, g/d	0.44 (0.11)	0.24
ADFI, g/d	0.66 (0.11)	0.15
FCR	0.32 (0.09)	0.10
RFI, g/d	0.10 (0.05)	0.09
BF, cM	0.58 (0.09)	0.37
MD, cM	0.39 (0.09)	0.30
IMF, %	0.54 (0.11)	0.23
BW at birth, kg	0.44 (0.31)	0.19
BW at weaning, kg	0.13 (0.12)	0.10

[1] FCR is feed conversion ratio (ADG / ADFI); RFI is residual feed intake; BF is ultrasound back fat thickness; MD represents ultrasound muscle depth; IMF is ultrasound innermuscular fat percentage.

[2] The heritability $H^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$ is the broad-sense heritability, obtained from single-trait animal model.

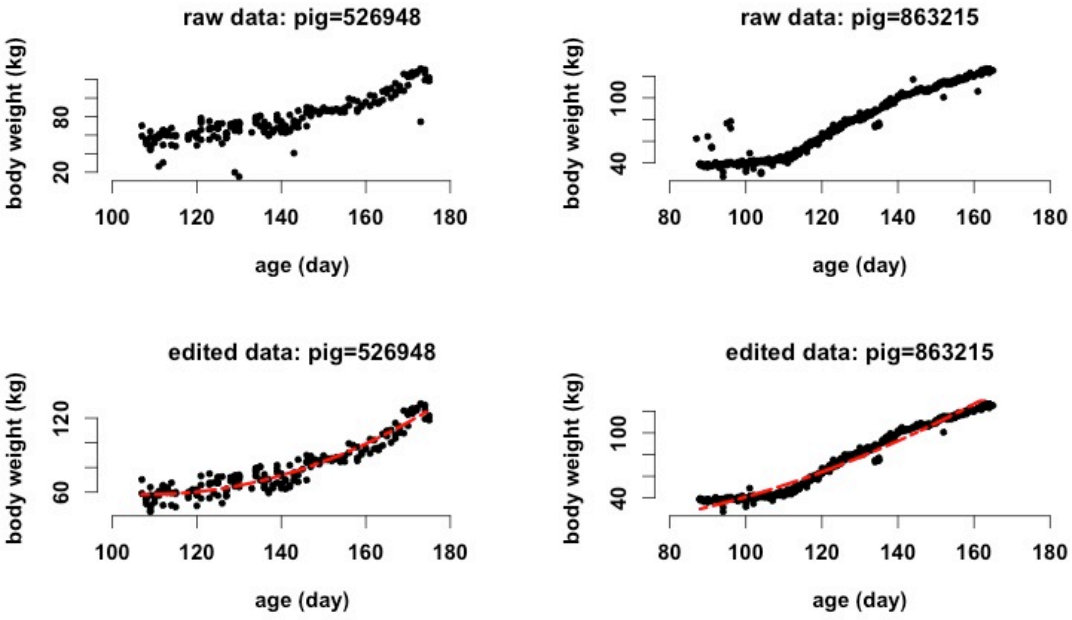

Figure 3.1 Serial pig body weights (kg) from FIRE (feed intake recording equipment) and predicted weights (kg) were plotted against age (d).

Dots in figures at the first row represent raw (unedited) pig body weights (kg/d) against age (day) for two randomly selected pigs; dots in figures at the second row represent edited pig body weights with age (day) on x axis; the dotted lines are predicted pig body weights by robust regression with age and squared age as covariates.

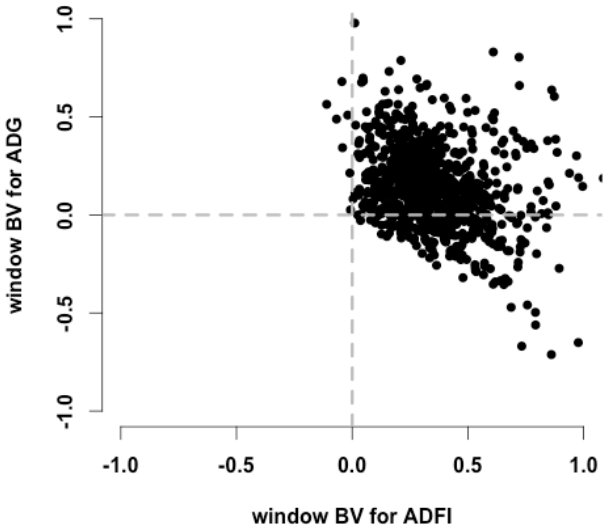

Figure 3.2 Window breeding values (BVs) for ADG and ADFI (from the result of Multivariate Bayes-A) were plotted.

The genome has been divided into 800 non-overlapping windows with 50 SNPs in each window. BV for each window across whole genome was computed and plotted. Windows were ordered by their physical position when computing the BVs.

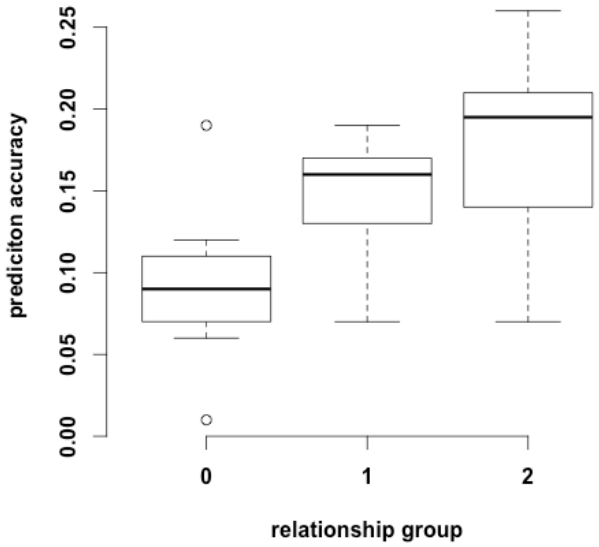

Figure 3.3 Genomic prediction accuracies for the 9 traits investigated of different relationship groups.

The nine traits includes ADFI, ADG, feed conversion ratio (FCR), residual feed intake (RFI), ultrasound backfat thickness (BF), ultrasound muscle depth (MD), ultrasound inner muscular fat content (IMF), body weight at birth (BW at birth) and at weaning (BW at weaning). Relationship groups were defined based on average relationship (using pedigree only) between individuals in validation dataset with all others in reference dataset. Relationship group: 0 indicates relationship between individual in validation and animal in reference has no relationship in average, while group 1 denoted moderate relationship (less than 0.044), group 2 represents relatively higher relationship (larger than 0.044). The value 0.044 is the mean of average relationship between validation and reference datasets.

CHAPTER 4

Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc Pigs: II. Genome-wide association.

Shihui Jiao[*], Christian Maltecca[*1], Kent A. Gray[†], Joseph P. Cassady[*]

[*] Department of Animal Science, North Carolina State University, Raleigh, NC 27695

[†] Smithfield Premium Genetics, Rose Hill, NC 28458

INTRODUCTION

Efficient use of feed resources has become a clear challenge that faces the US pork industry as feed cost continue to be the largest variable expenditure (McGlone et al., 2003; Hoque et al., 2009). Feed efficiency is typically measured as feed conversion ratio, or assessed through residual feed intake (Koch et al., 1963). Regardless, ADG and ADFI remain the two major components of nutrient utilization traits.

Substantial individual differences in feed intake exist and approximately 30% of these are explained by genetics (Lo et al., 1992; Cai et al., 2008; Hoque et al., 2008; 2009). Furthermore, it has been previously demonstrated that traits contributing to feed efficiency are genetically related to economically important traits such as backfat (Cai et al., 2008; Hoque et al., 2009). Although feed efficiency has improved to some extent by selection for growth and against backfat (Cleveland et al., 1983), further improvements require direct selection on feed intake. This is, nonetheless, complicated by the difficulty and expense of recording intake data on a large numbers of individuals. The uncovering of genetic variants that contribute to differences in feed intake and efficiency could represent a valid strategy to optimize the selection process.

Nearly two decades of advances in the QTL mapping of complex traits have led to new breeding programs incorporating molecular information. As of December 2012, 7,451 QTL were identified for economically relevant traits in pigs (Ernst et al., 2013). The availability of the Porcine60K BeadChip has greatly facilitated whole-genome association studies, contributing to increased accuracy of selection by application of marker-assisted and genomic selection (Fan et al., 2011; Onteru et al., 2011). The objective of this study was to identify genomic regions associated with variations in feed intake, average daily gain, feed efficiency and real-time ultrasound traits in a Duroc terminal sire population.

METERIALS and METHODS

Animal and phenotype data collection

Data used for this genome-wide association study came from a Duroc nucleus population owned by Smithfield Premium Genetics (SPG; Rose Hill, NC), including1047 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 weights were recorded when individuals were weaned at the mean age of 25 ± 1.93 days. Growth and feed intake were measured on each pig in a FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, KS, USA) station, starting at an average age of 85 ± 21 days and ending after approximately 45 ± 26 days. Boars born within the same week were grouped into batches and fed in the FIRE station. Individual feed intake and body weight were recorded when each pig visited the feeder. In total there were 323,639 individual visits recorded. When the boars reached approximately 118 ± 11.52 kg, ultrasound backfat thickness, muscle depth and intramuscular fat content and body weight were collected for each individual. The ultrasound images of all animals were captured over the last three ribs via an Aloka 500 (Corometrics Medical Systems, Wallingford, CT) and analyzed for intramuscular fat using the Swine Image Analysis Software (Designer Genes Technologies, Inc. Harrison, Arkansas). Descriptive statistics for each trait are listed in supplementary Table S4.1.

Feed intake data were edited based on the work of Casey et al. (2005) to account for errors and data missing in the FIRE records. The ADG phenotype was computed from linear regression of body weight records (birth, weaning and on-test BW) on age and robust regression using body weight data from FIRE records. After editing, all phenotypes were pre-adjusted for corresponding fixed effects before conducting the association analyses. The fixed effect of batch (contemporary group effect) was fitted for all traits. FCR was computed as the ratio between

ADG and ADFI. Residual feed intake was obtained as model residual from a linear mixed model including testing age, ADG and BF as covariates with batch and pig as random effects. On test body weight was included as linear covariate to the model for the pre-adjustment of ultrasound traits. Litter size and sow parity were fitted as additional fixed factors for BW at birth. The same model was used for BW at weaning, while weaning age was added as an additional covariate to the model. Additional details regarding data editing and pre-adjustment for corresponding fixed effects affecting each trait were reported in a companion paper (Jiao et al., 2014).

Genotype data editing before GWAS

A total of 1047 boars were genotyped with the Illumina PorcineSNP60K BeadChip (Illumina, Inc., San Diego, CA, USA) genotyping. SNP with call rates ≤ 0.90, MAF (minor allele frequency) ≤ 0.002, and p-value ＜ 0.0001 from a chi-square test for Hardy-Weinberg equilibrium was excluded from the genotype data set. Additionally, individual animals with call rate ≤ 0.90 were also excluded. Missing SNP genotypes were imputed for all available boars using AlphaImpute v1.1 (Hickey et al., 2011). The SNP that were not mapped to the swine genome build 10.2 and SNP on sexual chromosomes were excluded for the subsequent genome-wide association analyses. After quality control 35,140 SNP on autosomes 1 to 18 autosomes for 1,022 boars remained in the dataset.

Statistical Analysis

Four feed efficiency traits ADG, ADFI, FCR and RFI and five production traits including ultrasound BF, MD, and IMF and BW at birth and BW at weaning pre-adjusted for corresponding fixed effects and covariates, as described previously, were used to conduct the GWAS analysis. The association analyses were implemented separately for each trait with the Bayes-B model averaging approach described by (Kizilkaya et al., 2010) using GenSel v4.0

software (http://bigs.ansci.iastate.edu). The following statistical model was used for single marker regression,

$$\mathbf{y} = \mathbf{X}\beta + \sum_{i=1}^{m} Z_i u_i + \mathbf{e},$$

where **y** is the vector of pre-corrected phenotypes for each trait, **X** is an incidence matrix of fixed effects (β), which in this case consisted of only the population mean, **Z** is a (*n* individuals times *m* markers) matrix of SNP genotypes (-10, 0, 10) that were fitted as random effects (***u***) and **e** is the vector of random residual effects assumed to be normally distributed, N(0, σ_e^2). In Bayes-B, the variances of markers $(\sigma_{u_i}^2)$ are different and are estimated by the model. A hyperparameter, π governed the probability of each maker to have a non-null effect such that the conditional marker variance $(\sigma_{u_i}^2 \mid \pi)$ at each iteration is

$$\sigma_{u_i}^2 \mid \pi \sim \{ \begin{matrix} 0 & p(\pi) \\ \nu S^2 \chi^{-2}_{\nu} & p(1-\pi) \end{matrix}$$

where χ^{-2} is an inverted chi-squared distribution with ν and S^2 being the degrees of freedom and scale hyperparameters, respectively. The π parameter was fixed at 0.995 so that approximately 400 SNP with non-zero effects were fitted in the model in each iteration of the Markov chain. A total of 100,000 iterations, with a burn-in of 30,000 iterations, were run for the analyses.

Windows of 1 Mb length of non-overlapping adjacent SNP on the 18 autosomes (35,140 SNP) were constructed based on the physical map (Build 10.2). A total of 2380 windows were obtained. Average number of SNP per window was 14.8, since the SNP were not evenly

distributed across the swine genome (shown in Figure S4.1). Inferences on QTL were based on those 1-Mb length windows to account for the fact that often variance in QTL regions of moderate effect is shared between different SNP of individual marginal effect.

Over the last few years several more or less conservative measures of evidence in GWAS employing Bayesian models have been used, such as bootstrapping, posterior probability of association, or naïve Bayes factor (Fan et al., 2011; Wolc et al., 2012; Veerkamp et al., 2012); however, often little justification on the choice of method is offered. Stephens and Balding (2009) advocated posterior probability of association (PPA) over Bayes factor in a review of whole-genome association studies under Bayesian statistical models. In this study we used the posterior probability of association and variances explained by windows (in terms of percentage of variance explained per window and PPA of the window), similar to Wolc et al. (2012) to identify putative QTL regions. We conducted additional significance testing using naïve Bayes factor for 1-Mb windows (obtained as the sum of Bayes factors for single markers located within the 1-Mb window) and bootstrapping, where we built the empirical windows variance distribution under the null hypothesis. Rather than relying on a single measure of association we employed a conservative approach so that QTL regions were declared significant if concordance existed among the aforementioned three methods. It should be noted that while one of the 3 methods (bootstrapping) is non parametric and independent from the others, both Bayes Factor and posterior distribution of variances hinge on the same general concept.

The posterior distributions for 1-Mb window variances were obtained from the algorithm implemented in version 4.0 GenSel. For each 1-Mb window and each 50 iteration of the MCMC chain after burn-in, sampled values for SNP effects within the window were used to compute a sample of the posterior distribution of the direct genomic values for that window, by multiplying

the SNP effects with the individual's SNP genotypes and summing across SNP in the window for each individual. Window variances were then computed using the variance across individuals of the resulting sample breeding value ascribed to that window. The proportion of genetic variance explained by a window was obtained as the ratio of window variance against the total genetic variance (Wolc et al., 2012). Windows that captured more than 1% of genetic variance in over 50% of the samples were arbitrarily declared to explain significantly more variance than the 1-Mb window variance expected (0.042%) under a polygenic model.

The Bayes Factor was calculated for each window by using the formula of Kass and Reaftery, (1995),

$$Bayes\ factor = \frac{\Pr(y\ |H_1)}{\Pr(y\ |H_0)} = \frac{\Pr(H_1\ |y)}{\Pr(H_0|y)} \times \frac{1-\pi}{\pi},$$

where y is the observed phenotype; H_0 is the null hypothesis that there is no association between marker and QTL; and π is the prior probability of the null hypothesis, while H_1 is the hypothesis that the marker is linked to a QTL. Bayes factors greater than 1, favor the alternative over the null hypothesis. Bayes factors for 1-Mb non-overlapping windows of adjacent markers were calculated as the summation of the Bayes factors of individual markers within the window. Windows exceeding the threshold value of 500 were considered as putatively associated and further examined.

The significance level of putative candidate genomic regions was also estimated using bootstrap analysis with 1,000 replicates (Fan et al., 2011). Briefly, the proportion of genetic variance explained by each 1-Mb window was ordered by size and windows with effects exceeding 1% were selected as putative QTL regions to conduct the bootstrapping analysis. The

bootstrapping model is reported below:

$$y_j = \hat{\mu} + \sum_{i=1, i \notin QTL}^{i=35140} Z_i \hat{u}_i + \hat{\sigma}_e e_j,$$

Where y_j is a bootstrap sample for replicate j, obtained using the posterior mean of the fixed effect (general mean $\hat{\mu}$), the sum of SNP effects $\hat{u}_i$, with *i* being all the marker except the ones contained in the putative QTL windows and a vector of simulated residuals obtained by sampling a vector of independent standard normal deviations e_j, times the posterior mean of the residual SD ($\hat{\sigma}_e$). Bootstrap samples were constructed according to the null hypothesis of no QTL in the identified SNP window. Each constructed bootstrap was reanalyzed using the Bayes-B model used for the complete 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 $0.001 < P < 0.002$.

For all tests, the q value, FDR based measure of significance, was calculated for multiple hypothesis corrections using the "qvalue" package in R (Storey, 2002).

The QTL regions for each trait were declared significant when windows were significant across the 3 inference methods. If a window was significant for 1 or 2 tests and the adjacent windows were significant for the other test(s), the combination of windows were declared to be significant and further explored. In the latter case, linkage disequilibrium of the region was examined using r^2. Linkage disequilibrium was computed and plotted using Haploview (Barrett

et al., 2005). Genes within the target regions were identified using *Sus scrofa* 10.2 Build from GBrowse (www.animalgenome.org/cgi-bin/gbrowse/pig/) and gene annotations were retrieved from literatures and NCBI Gene (www.ncbi.nlm.nih.gov/gene) using the human orthologs whenever the gene function was not available for the swine specie.

Allele substitution effects and Bayes factors for SNP were obtained by fitting genome-wide single SNP into Bayes-B model (described above, data not shown). The allele substitution effects and Bayes factors for SNP within the QTL region for ADG, ADFI and BF were shown in Table S4.2. The most significant SNP (largest Bayes factor) in the QTL region for ADG, ADFI and BF was fitted as fixed effect in an animal model using ASREML 3.0 (Gilmour et al., 2009) to obtain the allele substitution effects. The animal model used was,

$$y_i = \mu + bx_i + a_i + e_i,$$

where y_i is the pre-adjusted phenotype; x_i is the number of minor alleles (0, 1, 2) as dosage; b is the corresponding regression coefficient; a_i is a random additive effect with $a_i \sim N(0, \boldsymbol{A}\sigma_a^2)$, where $\boldsymbol{A}$ is the additive relationship matrix and σ_a^2 is the additive variance; and e_i is the model residual with $e_i \sim N(0, \boldsymbol{I}\sigma_e^2)$.

For the QTL region affecting ADFI, ADG and BF, haplotypes of genomic segment flunking 5 successive SNP were constructed from the phased genotypes (phasing was performed using AlphaImpute v1.0.). Copies of haplotypes for each individual were counted from phased genotypes and were used to form design matrix for haplotype dosage effects. Haplotypes in each segment were fitted into the above model with x_i changed into design matrix for the haplotypes of the corresponding segment.

RESULTS

In this study, growth, feed intake, feed efficiency and real-time ultrasound measurements of approximately a thousand boars from a Duroc nucleus herd were collected and analyzed. In this section, results regarding windows with large variances for each trait will be presented mainly based on posterior probability of association and variances explained by windows. Further testing using Bayes factors for 1-Mb windows and bootstrapping will be introduced additionally to highlight concordant regions.

Regions identified for feed efficiency traits

Manhattan plots for percentage of genetic variance explained by each 1-Mb window for each chromosome were reported in Figure 4.1. Using a cut-off of 1% of total variance explained, putative QTL regions for all feed efficiency traits were identified, except for RFI. The top five windows for proportion of variance explained and high posterior probability of association (PPA) for feed efficiency traits were listed in Table 4.1. Four 1-Mb windows on SSC 1, 4, 11, 14 were identified for ADG, with proportion of genetic variance explained of 4.8, 5.04, 2.69 and 2.39%, respectively. The four significant windows cumulatively explained 15% of genetic variance. Two significant windows were detected for ADFI, on SSC1 and SSC10, explaining approximately 3.57% of the total genetic variance. For FCR, 1 window on SSC 4 was detected, explaining 2.65% of genetic variance for FCR. Other genomic regions exceeding the genetic variance threshold (1% of genetic variance) were identified for ADFI and RFI, but none of these windows had a probability of association large enough, possibly due to the limited sample size of the study.

To explore the characteristics of genomic regions associated with feed efficiency traits in swine, cumulative proportion of genetic variance explained by 1-Mb windows, ranking from the

most to the least variation explained were plotted in Figure 4.2. Generally, less than 2% of windows (genomic regions) can explain 50% of the genetic variation, which indicates most of the loci only have small effects on the traits. Additionally, distributions of proportion of genetic variance explained by 1-Mb windows for feed efficiency traits are shown in Figure S4.3. For all four traits many windows (approximately 90%) explained < 0.5% of the genetic variance and for RFI, more than 99.9% windows fell into this category.

Significant windows declared by Bayes factor and empirical distributions under null hypothesis using bootstrapping samples are shown in Table 4.3 and Table 4.4. Using a significant level determined by Bayes factor larger than 100, almost all the traits investigated have approximately 100 significant windows; using an higher Bayes factor as cut-off (500), a few extremely significant windows have shown strong association with each feed efficiency trait, except for RFI (Figure S4.5). The bootstrapping analysis resulted in approximately 10 windows being significant at the level of 10% FDR and approximately 5 windows at 5% for feed efficiency traits except RFI (Table 4.4).

Genomic regions collectively associated with feed efficiency traits across methods are summarized in Table 4.5. Regions associated with ADFI were mapped to SSC 1 and 10; with ADG mapped to SSC 1; FCR on SSC 4 and no QTL were found for RFI. Genes within each region or 1 Mb downstream/upstream around the target region were identified and genes involved in metabolism were selected as candidate genes and annotated.

Regions identified for production traits

Manhattan plots of percentage of genetic variance explained by 1-Mb windows for production traits are shown in Figure S4.2. Windows largely contributing to genetic variance of real-time ultrasound traits and growth traits are identified above the cut-off line (1%). For BF,

the largest window explained approximately 20% of the genetic variance. The top 5 ranking windows by proportion of genetic variance explained and PPA are listed in Table 4.2. Three windows on SSC 1, 3, 18 were significantly associated with BF, explaining collectively 26.74% of genetic variance. In the present study only one window on SSC 8 was identified for MD, with 2.64% of genetic variance explained. Six windows on SSC 3, 4, 8, 9 and 15 showed strong evidence of association with BW at weaning, explaining 20% of the genetic variance. Although no significant window was identified for IMF and BW at birth, the top five windows for each trait explained approximately 7% of the genetic variance.

In Figure 4.2, the curve illustrates the cumulative distribution of percentage of genetic variances explained by windows for production traits. The trends of the cumulative distribution of production traits are similar to feed efficiency traits with a small number of genomic regions having major effects. Again, distribution of proportion of genetic variance explained by 1-Mb windows for the 5 production traits can be found in supplementary materials (Figure S4.4).

Similar to feed efficiency traits, a large number of windows were significant for production traits if considering a window to be significant with Bayes factor larger than 100; few windows were extremely significant with a Bayes factor greater than 500 (Table 4.3). From the empirical distributions under null hypothesis constructed using bootstrap samples, 18, 13, 11, 6 and 14 windows were declared significant at a FDR level of less than 10% for BF, MD, IMF, BW at birth and at weaning, respectively (Table 4.4).

The QTL regions covering the significant windows across the 3 methods are shown in Table 4.5 for production traits.

Effects of SNP and haplotypes associated with ADG, ADFI and BF

The QTL region identified for ADFI, ADG and BF located from 166 to 170 Mb on SSC 1 (Table 4.5) with 46 markers within the region. Allele substitution effect and Bayes factors for SNP within this 4-Mb region obtained from Bayes-B model are reported in Table S4.2. Based on Bayes factors for single SNP markers, one SNP (ALGA006684) was identified to be the most significant one associated with ADFI, ADG and BF. Fitting number of copies of minor allele for ALGA006684 as dosages in an animal mode for ADFI, ADG and BF, the substitution effects (linear coefficient b) estimated for the three traits (Table 4.6).

Haplotype analyses were preformed by dividing the 4-Mb region from 166 to 170 Mb on SSC 1 into nine consecutive segments of 5 successive adjacent SNP per segment. Haplotypes' effects were obtained similarly to what proposed for single marker analysis. Haplotypes significantly associated with any of the 3 traits are reported in Table 4.7, while joint frequency of the significant haplotypes is plotted on Figure 4.4.

DISCUSSIONS

Genomic regions identified for feed efficiency traits and production traits

In this study four 1-Mb windows on SSC1 ranging from 166 Mb to 170 Mb (4 Mb in length) and showing moderate to strong linkage disequilibrium (Figure 4.3, average r2 is 0.307), were significantly associated with ADFI, ADG and BF (Table 4.5) thus suggesting that this region may harbor a pleiotropic QTL. The 4-Mb region, explaining approximately 3.35%, 6.53% and 18.62% of genetic variances by four successive 1-Mb windows for ADFI, ADG and BF in current study, was within or in near proximity of QTL regions reported by previous QTL mapping studies for ADG, feed intake and BF (Liu et al., 2007; De Koning et al., 2001; Rückert et al., 2010; Harmegnies et al., 2006; Beeckmann et al., 2003; Thomsen et al., 2004; Hernández-

Sánchez et al., 2003). Kim et al. (2000) suggested that a missense mutation in *MC4R* gene (178553488-178555752 bp from UCSC browser), located near the 4-Mb region may significantly be associated with back fat and growth rate in a number of examined pig lines as well as feed intake overall. However in the association study conducted by Fan et al. (2010) in a Yorkshire pig population the missense mutation in *MC4R* gene was associated with back fat but not with ADG and meat quality attributes.

In the current study, ADFI, ADG and BF were mapped to the same region on SSC1 (ranging from 166 to 170 Mb), where *SOCS6* (168.99 Mb) and *DOK6* (169.78) are located. Howard and coworkers (2006) reported SOCS6 being involved in development of leptin and insulin resistance in human. Furthermore it has been shown that SOCS6 may impair insulin receptor signaling (Gupta et al., 2011). Additionally, it has been concluded that the SOCS6 protein may be involved in the proteasome mediated degradation in human (Bayle et al., 2006). The DOK6 protein is an insulin receptor binding protein, highly expressed at neural and kidney development (Crowder et al., 2004; Kurotsuchi et al., 2010). Both the location and gene function support SOCS6 and DOK6 as candidates that may have pleiotropic effects and may be involved in ADG, ADFI and BF regulation. The *MC4R* gene is not far away (approximately 8 Mb downstream) from the region identified in the present study and has been suggested to be associated with growth, feed intake or back fat in different studies (Kim et al., 2000; Fan et al., 2010, Fan et al., 2011). In this Duroc population, the weak LD between the 4-Mb QTL region (from 166015433 to 169889708 bp) and its downstream region (ranging from170298129 to 180547614 bp) harboring *MC4R* gene (178553488-178555752 bp from UCSC browser) suggested the *MC4R* gene was not identified as a likely candidate (Figure S4.6).

There are 46 markers inside the 4-Mb region, significantly associated with ADFI, ADG as well as BF in current study. Sizes (allele substitution effects) of most markers in the 4-Mb region were close to 0; while sizes of a few markers bounded around 0, except SNP (ALGA0006684) for ADFI, ADG and BF with large negative effects (shown in Table S4.2). Therefore, caution need to be taken using this SNP as marker to decrease back fat, since selecting pig with the two markers may grow slow and eat less.

Allele substitution effect (the coefficient for the dosage effect of SNP) is significant for ADFI, ADG and BF when fitting the number of copies of minor allele of ALGA0006684 as fixed effect in the corresponding animal models (Table 4.6). This result highly agreed with the above analysis from Bayes-B model.

Haplotype analysis for the 4-Mb region on SSC1 revealed several haplotypes significantly associated with ADFI, ADG and BF, which may be more informative than single markers since the haplotype accounted for LD between markers (Figure 4.3). Frequency of those haplotypes and pair-wise frequency were shown in heat map (Figure 4.4), which imply the LD structure among haplotypes. As the physical distance increase between haplotypes, the LD decays as expected. Effect of haplotypes significantly associated with ADFI, ADG and BF were shown in Figure S4.8, among which two haplotypes have positive effect on ADG but negative effect on ADFI and BF. Those two haplotypes are favorable when selecting pigs with fast growth, less feed consumption and thinner backfat.

Additionally, one 1-Mb window significantly affecting ADFI was identified on SSC 10. The window is located around 73 Mb. To our knowledge no QTL has been reported in this region. The nearest QTL was mapped by Wada and coworkers (2000) in a Meishan x Gottingen cross population and located approximately 10 Mb away from the region identified in this study.

Within this target QTL region, the *PFKP* gene has been linked to obesity-related traits such as BMI, body weight and hip circumference in humans (Scuteri et al., 2007). The Gene *PFKP* encodes a phosphofructokinase acting as a viral enzyme in the glycosis, balancing the glucose utilization and thereby lipogenesis (Scuteri et al., 2007; Nakajima et al., 2002).

In this study, we reported a novel region on SSC 4 significantly associated with FCR. Windows around 4-5 Mb on SSC 4 were detected in the present study, explaining 2.65% of genetic variance for FCR. This region is approximately 3 Mb from the one reported by Duthie et al. (2008) in a three-generation full-sib population, developed by crossing Pietrain sires with a commercial dam line for FCR. An association study conducted by Fontanesi and coworkers (2010) revealed two SNP significantly associated with FCR in Italian Large White and Duroc population respectively, located approximately 102 Mb on SSC 4. Genome-wide association using 60K panel for FCR in Danish Duroc population reveled a QTL region ranging from 63.8 to 64 Mb on SSC 4 (Sahana et al., 2013).

Although based on the significance criterions used in present study, there was no region significantly affecting RFI (Table 4.5), a peak formed by several adjacent 1-Mb windows in highly linkage disequilibrium (Figure S4.7) on SSC 2, was evident from the Manhattan plot (Figure 4.1), explaining more than 2% of genetic variance. Large intervals QTL for ADG and BW were mapped to this region in several previous studies (Guo et al., 2008; Rückert et al., 2010; Thomsen et al., 2004).

Loin muscle depth QTL were mapped to 9 different chromosomes in previous studies (pigQTLdb), mainly on SSC 7 and 11. No QTL has been previously reported on SSC 8. In this study, we identified a genomic region around 115-116Mb on SSC 8, significantly associated with MD (Table 4.5). The *ARSJ* gene falls in this region. ARSJ belongs to sulfatases, a large

enzyme family, is responsible for the degradation of sulfated carbohydrates, involved in metabolism (Oshikawa et al., 2009).

Four significant windows were identified for BW at weaning on SSC 4, 8, 9 and 15, with 3.83, 2.67, 7.80 and 3.10% of genetic variance explained by each window respectively and a total of 17.4% of genetic variance. The four windows significantly associated with BW at weaning were located around 84, 119, 141 and 140 Mb (Table 4.5), respectively. QTL for BW (3 weeks, 3-10 weeks, at slaughter) have been reported previously on SSC 4 by several studies (Bidanel et al., 2001 Estellé et al., 2006; Guo et al., 2008; Marklund et al., 1999), most of which covering a large interval. The region identified affecting BW at weaning on SSC 4 in our study harbors several genes: *PLAG1*, *CHCHD7*, *RDHE2 (or SDR16C5)*, *MOS*, *RPS20*, *LYN* and *PENK.* All these genes have been found to influence both human and cattle height (Pryce et al., 2011; Nishimura et al., 2012; Gudbjartsson et al., 2008; Weedon et al., 2008; Littlejohn et al., 2012) and all have been associated with birth weight in Nellore cattle (Utsunomiya et al., 2013). In agreement with our study, BW at 17 and 22 weeks have been mapped to around 122 Mb on SSC 9 in a cross between Meishan and Large White population by Quintanilla and colleagues in 2002. Additionally, Duthie et al. (2008) reported protein accretion rate and lipid accretion rate in pigs near this region. In our analysis, this region explained up to 8% of genetic variance for BW at weaning. Genes worthly of further consideration include *PRG4* and *PGHS-2*, which may involve in cell proliferation, regulation of cell proliferation and metabolism of lipids and lipoproteins, respectively (http://www.ncbi.nlm.nih.gov/gene/10216; http://www.ncbi.nlm.nih.gov/gene/5743). Another region, identified in our study affecting BW at weaning, locates around 138-140 Mb on SSC 15, and has been previously reported to influence weaning weight in a cross between Duroc and Pietrain (Liu et al., 2008). Furthermore, it was

reported that this region may affect other growth and production traits, such as ADG, feed intake, daily feed intake and days to 105 Kg (Soma et al., 2011; Houston et al., 2005; Liu et al., 2007) in other pig populations. The *MOGAT1* gene is within this identified QTL region. This gene is closely related to *DGAT2* and *MOGAT2* genes, belonging to a gene family with at least eight members in mammals that are candidate genes for quantitative traits in cattle related to dietary fat uptake, lipid synthesis and storage (Winter et al., 2004).

Significance tests and genomic regions associated with traits of interests

In our study, significance tests were conducted via 1-Mb non-overlapping windows covering several to tens of SNP instead of using single SNP. Because of linkage disequilibrium among SNP, the effect of a QTL may be spread over a number of neighboring SNP, each capturing small variance, resulting in individual SNP effects that tend to underestimated the true QTL effects (Hayes et al., 2010; Fan et al., 2011; Wolc et al., 2012). Therefore, we used chromosome segments (1-Mb windows) to derive the distribution effects with multiple SNP that are likely to capture the real effect of QTL and significance tests were based on those chromosome segments.

The use of single-SNP tests in genome-wide association studies was reported in a large number of publications, in human, livestock or plant; however, the challenges arise when it comes to large numbers of false positive markers due to multiple testing (McCarthy et al., 2008; Goddard et al., 2009; Brachi et al., 2011). In this study, we employed three inference methods and in order to be conservative, we defined the concordant region identified across the three significance-testing methods as significant QTL regions. Comparison of results from three inference methods yield a few concordance regions while several regions mismatch among the methods, which is likely due to different power of the tests. In our case, Bayes factor analysis

was less conservative when using a value of 100 as cut-off, which is 10 times higher than the one used by Veerkamp and coworkers (2012). However, from the perspective of computation, bootstrapping is highly demanding, as previously reported by Fan et al. (2011) and Onteru et al. (2011). Further investigation of significance tests may be requested to adopt a standard tool for representing the data from a GWA scan.

To characterize the genomic regions for the nine traits related to growth, feed efficiency and other performance traits, the distributions of genetic variances explained by 1-Mb windows were investigated in the current study (Figure 4.2). The previously analyzed QTL regions identified in our study demonstrate the importance of a few loci with large effects highly influencing the traits but could not describe the whole architecture of genetic effects. Except a few regions of major effect, a large number of loci or regions having very small effects generally, which characterize the genomic regions of traits related to growth, feed efficiency and other production performance, agreed to the results from Hayes et al. (2010) for complex traits in cattle, Wolc et al. (2012) for egg weight and uniformity traits in layer chickens. The characterization of the genomic regions for traits investigated in this study were also observed clearly from the cumulative proportion of the total variance explained when windows were ranked by size (Figure S4.3, S4.4).

One justification for conducting GWA studies in livestock is to use the validated marker to select better parents through marker-assisted selection (MAS) (Smith et al., 1967). The present results provided strong evidence that one region on SSC1, 10 Mb away from *MC4R* gene, may be harboring a pleiotropic QTL simultaneously regulating ADFI, ADG and BF. Furthermore, linkage disequilibrium analysis of this region suggested that *MC4R* is not likely to be the candidate gene and that others *SOCS6* and *DOK6* might be targets for further investigations. One

2-Mb region on SSC 4 significantly associated with BW at weaning points to previously described orthologous genes affecting human and bovine height and influencing birth weight in Nellore cattle.

CONCLUSIONS

In conclusion, this study identified several genomic regions affecting growth, feed efficiency and key performance traits that will provide the basis for future swine selection and breeding targeted investigations.

LITERATURE CITED

Barrett, J. C., B. Fry, J. D. M. J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 21:263-265.

Bayle, J., S. Lopez, K. Iwaï, P. Dubreuil, and P. De Sepulveda. 2006. The E3 ubiquitin ligase HOIL-1 induces the polyubiquitination and degradation of SOCS6 associated proteins. FEBS letters. 580:2609-2614.

Beeckmann, P., J. Schröffel, G. Moser, H. Bartenschlager, G. Reiner, and H. Geldermann. 2003. Linkage and QTL mapping for Sus scrofa chromosome 1. J. Anim. Breed. Genet. 120:1-10.

Bidanel, J. P., D. Milan, N. Iannuccelli, Y. Amigues, M. Y. Boscher, F. Bourgeois, J. C. Caritez, J. Gruand, L. R. Pascale, H. Lagant, R. Quintanilla, C. Renard, J. Gellin, L. Ollivier, and C. Chevalet. 2001. Detection of quantitative trait loci for growth and fatness in pigs. Genet. Sel. Evol. 33:289-309.

Brachi, B., G. P. Morris, and J. O. Borevitz. 2011. Genome-wide association studies in plants: the missing heritability is in the field. Genome Boil. 12:232.

Cai, W., D. S. Casey, and J. C. M. Dekkers. 2008. Selection response and genetic parameters for residual feed intake in Yorkshire swine. J. Anim. Sci. 86:287-298.

Casey, D. S., H. S. Stern, and J. C. M. Dekkers. 2005. Identification of errors and factors associated with errors in data from electronic swine feeders. J. Anim. Sci. 83:969-982.

Cleveland, E. R., R. K. Johnson, R. W. Mandigo, and E. R. Peo, Jr. 1983. Index selection and feed intake restriction in swine. II. Effect on energy utilization. J. Anim. Sci. 56:570-578.

Crowder, R. J., H. Enomoto, M. Yang, E. M. Johnson, and J. Milbrandt. 2004. Dok-6, a Novel p62 Dok family member, promotes Ret-mediated neurite outgrowth. J. Bio. Chem. 279:42072-42081.

De Koning, D. J., A. P. Rattink, B. Harlizius, M. A. M. Groenen, E. W. Brascamp, and J. A. M. Van Arendonk. 2001. Detection and characterization of quantitative trait loci for growth and reproduction traits in pigs. Livest. Prod. Sci. 72:185-198.

Duthie, C., G. Simm, A. Doeschl-Wilson, E. Kalm, P. W. Knap, and R. Roehe. 2008. Quantitative trait loci for chemical body composition traits in pigs and their positional associations with body tissues, growth and feed intake. Anim. Genet. 39:130-140.

Ernst, C. W. and J. P. Steibel. 2013. Molecular advances in QTL discovery and application in pig breeding. Trends Genet. 29:215-225.

Estellé, J., M. Pérez-Enciso, A. Mercadé, L. Varona, E. Alves, A. Sánchez, and J. M. Folch. 2006. Characterization of the porcine FABP5 gene and its association with the FAT1 QTL in an Iberian by Landrace cross. Anim. Genet. 37:589-591.

Fan, B., S. Lkhagvadorj, W. Cai, J. Young, R. M. Smith, J. C. M. Dekkers, E. Huff-Lonergan, S. M. Lonergan, and M. F. Rothschild. 2010. Identification of genetic markers associated with residual feed intake and meat quality traits in the pig. Meat Sci. 84:645-650.

Fan, B., S. K. Onteru, Z. Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011.

Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. PloS ONE 6:e14726.

Fontanesi, L., C. Speroni, L. Buttazzoni, E. Scotti, L. N. Costa, R. Davoli, and V. Russo. 2010. Association between cathepsin L (< i> CTSL</i>) and cathepsin S (< i> CTSS</i>) polymorphisms and meat production and carcass traits in Italian Large White pigs. Meat Sci. 85: 331-338.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Goddard, M. E. and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programs. Nat. Rev. Genet. 10:381-391.

Gudbjartsson, D. F., G. B. Walters, G. Thorleifsson, H. Stefansson, B. V. Halldorsson, P. Zusmanovich, P. Sulem, S. Thorlacius, A. Glyfason, S. Steinberg, A. Helgadottir, A. Ingason, V. Steinthorsdottir, E. J. Olafsdottir, G. H. Olafsdottir, T. Jonsson, K. Borch-Johnsen, T. Hansen, G. Andersen, T. Jorgensen, O. Pedersen, K. K. Aben, J. Witjes, D. W. Swinkels, M. den Heijer, B. Franke, A. L. M. Verbeek, D. Becker, L. R. Yanek, L. C. Becker, L. Tryggvadottir, T. Rafnar, J. Gulcher, L. A. Kiemeney, A. Kong, U. Thorsteinsdottir, and K. Stefansson. 2008. Many sequence variants affecting diversity of adult human height. Nat. Genet. 40:609-615.

Guo, Y. M., G. J. Lee, A. L. Archibald, and C. S. Haley. 2008. Quantitative trait loci for production traits in pigs: a combined analysis of two Meishan× Large White populations. Anim. Genet. 39:486-495.

Gupta, S., K. Mishra, A. Surolia, and K. Banerjee. 2011. Suppressor of cytokine signalling-6 promotes neurite outgrowth via JAK2/STAT5-mediated signaling pathway, involving negative feedback inhibition. PloS ONE 6: e26674.

Harmegnies, N., F. Davin, S. De Smet, N. Buys, M. Georges, and W. Coppieters. 2006. Results of a whole genome quantitative trait locus scan for growth, carcass composition and meat quality in a porcine four-way cross. Anim. Genet. 37:543-553.

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet. 6:e1001139.

Hernández-Sánchez, J., P. Visscher, G. Plastow, and C. Haley. 2003. Candidate gene analysis for quantitative traits using the transmission disequilibrium test: the example of the melanocortin 4-receptor in pigs. Genetics 164:637-644.

Hickey, J. M., B. P. Kinghorn, B. Tier, J. F. Wilson, N. Dunstan, and J. H. van der Werf. 2011. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. Genet. Sel. Evol. 43:12.

Hoque, M. A., H. Kadowaki, T. Shibata, and K. Suzuki. 2008. Maternal and direct genetic parameters for production traits and maternal correlations among production and feed efficiency traits in Duroc pigs. Asian-Aust. J. Anim. Sci. 7:961-966.

Hoque, M. A., H. Kadowaki, T. Shibata, T. Oikawa, and K. Suzuki. 2009. Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livest. Sci. 121:45-49.

Houston, R. D., C. S. Haley, A. L. Archibald, and K. A. Rance. 2005. A QTL affecting daily feed intake maps to Chromosome 2 in pigs. Mamm. Genome 16:464-470.

Howard, J. K. and J. S. Flier. 2006. Attenuation of leptin and insulin signaling by SOCS proteins. Trends in Endocrino. Metab.17:365-371.

Kass, R. E. and A. E. Raftery. 1995. Bayes factors. J. AM. Stat. Asso. 90:773-795.

Kennedy, B. W., J. H. Van der Werf, and T. H. Meuwissen (1993). Genetic and statistical properties of residual feed intake. J. Anim. Sci.71:3239-3250.

Kim, K. S., N. Larsen, T. Short, G. Plastow, and M. F. Rothschild. 2000. A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. Mamm. Genome 11:131-135.

Kizilkaya, K., R. L. Fernando, and D. J. Garrick. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J. Anim. Sci. 88:544-551.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486-494.

Kurotsuchi, A., Y. Murakumo, M. Jijiwa, K. Kurokawa, Y. Itoh, Y. Kodama, T. Kato, A. Enomoto, N. Asai, H. Terasaki, and M. Takahashi. 2010. Analysis of DOK-6 function in downstream signaling of RET in human neuroblastoma cells. Cancer Sci. 101:1147-1155.

Littlejohn, M., T. Grala, K. Sanders, C. Walker, G. Waghorn, K. Macdonald, W. Coppieters, M.

Georges, R. Spelman, E. Hillerton, S. Davis, and R. Snell. 2012. Genetic variation in PLAG1 associates with early life body weight and peripubertal weight and growth in Bos taurus. Anim. Genet. 43:591-594.

Liu, G., D. G. J. Jennen, E. Tholen, H. Juengst, T. Kleinwächter, M. Hölker, D. Tesfaye, G. Ün, H.-J. Schreinemachers, E. Murani, S. Ponsuksili, J.-J. kim, K. Schellander, and K. Wimmers. 2007. A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. Anim. Genet. 38:241-252.

Liu, G., J. J. Kim, E. Jonas, K. Wimmers, S. Ponsuksili, E. Murani, C. Phatsara, E. Tholen, H. Juengst, D. Tesfaye, J. L. Chen, and K. Schellander. 2008. Combined line-cross and half-sib QTL analysis in Duroc–Pietrain population. Mamm. Genome 19:429-438.

Lo, L. L., D. G. McLaren, F. K. McKeith, R. L. Fernando, and J. Novakofski. 1992. Genetic analyses of growth, real-time ultrasound, carcass, and pork quality traits in Duroc and Landrace pigs: I. Breed effects. J. Anim. Sci. 70:2373-2386.

Marklund, L., P.-E. Nyström, S. Stern, L. Andersson-Eklund, and L. Andersson. 1999. Confirmed quantitative trait loci for fatness and growth on pig chromosome 4. Heredity 82:134-141.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. PA. Ioannidis, and J. N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. 9:356-369.

McGlone, J. J., and W. Pond. 2003. Pig Production: Biological Principles and Applications. Delmar Learning, Clifton Park, NY.

Nakajima, H., N. Raben, T. Hamaguchi, and T. Yamasaki. 2002. Phosphofructokinase Deficiency Past, Present and Future. Curr. Mol. Med. 2:197-212.

Nishimura, S., T. Watanabe, K. Mizoshita, K. Tatsuda, T. Fujita, N. Watanabe, Y. Sugimoto, and A. Takasuga. 2012. Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. BMC Genet. 13:40.

Onteru, S. K., B. Fan, M. T. Nikkilä, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Whole-genome association analyses for lifetime reproductive traits in the pig. J. Anim. Sci. 89:988-995.

Oshikawa, M., R. Usami, and S. Kato. 2009. Characterization of the arylsulfatase I (ARSI) gene preferentially expressed in the human retinal pigment epithelium cell line ARPE-19. Mol. Vis. 15:482.

Pryce, J. E., B. J. Hayes, S. Bolormaa, and M. E. Goddard. 2011. Polymorphic regions affecting human height also control stature in cattle. Genetics 187:981-984.

Quintanilla, R., D. Milan, and J. P. Bidanel. 2002. A further look at quantitative trait loci affecting growth and fatness in a cross between Meishan and Large White pig populations. Genet. Sel. Evol. 34:193-210.

Rückert, C. and J. Bennewitz. 2010. Joint QTL analysis of three connected F2-crosses in pigs. Genet. Sel. Evol. 42:40.

Sahana, G., K. Veronika, H. Henrik, N. Bjarne, and O. F. Christensen. 2013. A genome-wide

association scan in pig identifies novel regions associated with feed efficiency trait. J. Anim. Sci. 91:1041-1050.

Scuteri, A., S. Sanna, W.-M. Chen, M. Uda, G. Albai, J. Strait, S. Najjar, R. Nagaraja, M. Orrú, G. Usala, M. Dei, S. Lai, A. Maschio, F. Busonero, A. Mulas, G. B. Ehret, A. A. Fink, A. B. Weder, R. S. Cooper, P. Galan, A. Chakravarti, D. Schlessinger, A. Cao, E. Lakatta, and G. R. Abecasis. 2007. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. PLoS Genet. 3:e115.

Smith, C. H. (1967). Improvement of metric traits through specific genetic loci. Anim. Prod. 9:349-358.

Soma, Y., Y. Uemoto, S. Sato, T. Shibata, H. Kadowaki, E. Kobayashi, and K. Suzuki. 2011. Genome-wide mapping and identification of new quantitative trait loci affecting meat production, meat quality, and carcass traits within a Duroc purebred population. J. Anim. Sci. 89:601-608.

Stephens, M. and D. J. Balding. 2009. Bayesian statistical methods for genetic association studies. Nat. Rev. Genet. 10:681-690.

Storey, JD. 2002. A direct approach to false discovery rates. J. R. Stat. Soc. Ser. B 64: 479-498.

Thomsen, H., H. K. Lee, M. F. Rothschild, M. Malek, and J. C. M. Dekkers. 2004. Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine. J. Anim. Sci. 82:2213-2228.

Utsunomiya, Y. T., A. S. do Carmo, R. Carvalheiro, H. H. R. Neves, M.C. Matos, L. B. Zavarez, J. Sölkner, J. C. McEwan, J. B. Cole, C. P. Van Tassell, F. S. Schenkel, M. V. da Silva,

L. R. P. Neto, T. S. Sonstegard, and J. F. Garcia. 2013. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. BMC Genet. 14:52.

Veerkamp, R. F., M. P. Coffey, Donagh P. Berry, Y. De Haas, E. Strandberg, H. Bovenhuis, M. P. L. Calus, and E. Wall. 2012. Genome-wide associations for feed utilisation complex in primiparous Holstein–Friesian dairy cows from experimental research herds in four European countries. Animal 6:1738-1749.

Wada, Y., T. Akita, T. Awata, T. Furukawa, N. Sugai, K. Ishii, Y. Ito, E. Kobayashi, S. Mikawa, H. Yasue, N. Sugai, Y. Inage, H. Kusumoto, T. Matsumoto, M. Miyake, A. Murase, S. Shimanuki, T. Sugiyama, Y. Uchida, and S. Yanai. 2000. Quantitative trait loci (QTL) analysis in a Meishan× Göttingen cross population. Anim. Genet. 31:376-384.

Weedon, M. N., H. Lango, C. M. Lindgren, C. Wallace, D. M. Evans, M. Mangino, R. M. Freathy, J. R. B. Perry, S. Stevens, A. S. Hall, N. J. Samani, B. Shields, I. Prokopenko, M. Farrall, A. Dominiczak, DGI, WTCCC, T. Johnson, S. Bergmann, J. S. Beckmann, P. Vollenweider, D. M. Waterworth, V. Mooser, C. N. A. Palmer, A. D. Morris, W. H. Ouwehand, CGEMC, M. Caulfield, P. B. Munroe, A. T. Hattersley1, M. I. McCarthy, and T. M. Frayling. 2008. Genome-wide association analysis identifies 20 loci that influence adult height. Nat. Genet. 40:575-583.

Winter, A., M. Van Eckeveld, O. R. P. Bininda-Emonds, F. A. Habermann, and R. Fries. 2004. Genomic organization of the DGAT2/MOGAT gene family in cattle (Bos taurus) and other mammals. Cytogenet. Genome Res.102:42-47.

Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, W. G. Hill, and J. C. M. Dekkers. 2012. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. Anim. Genet. 43:87-96.

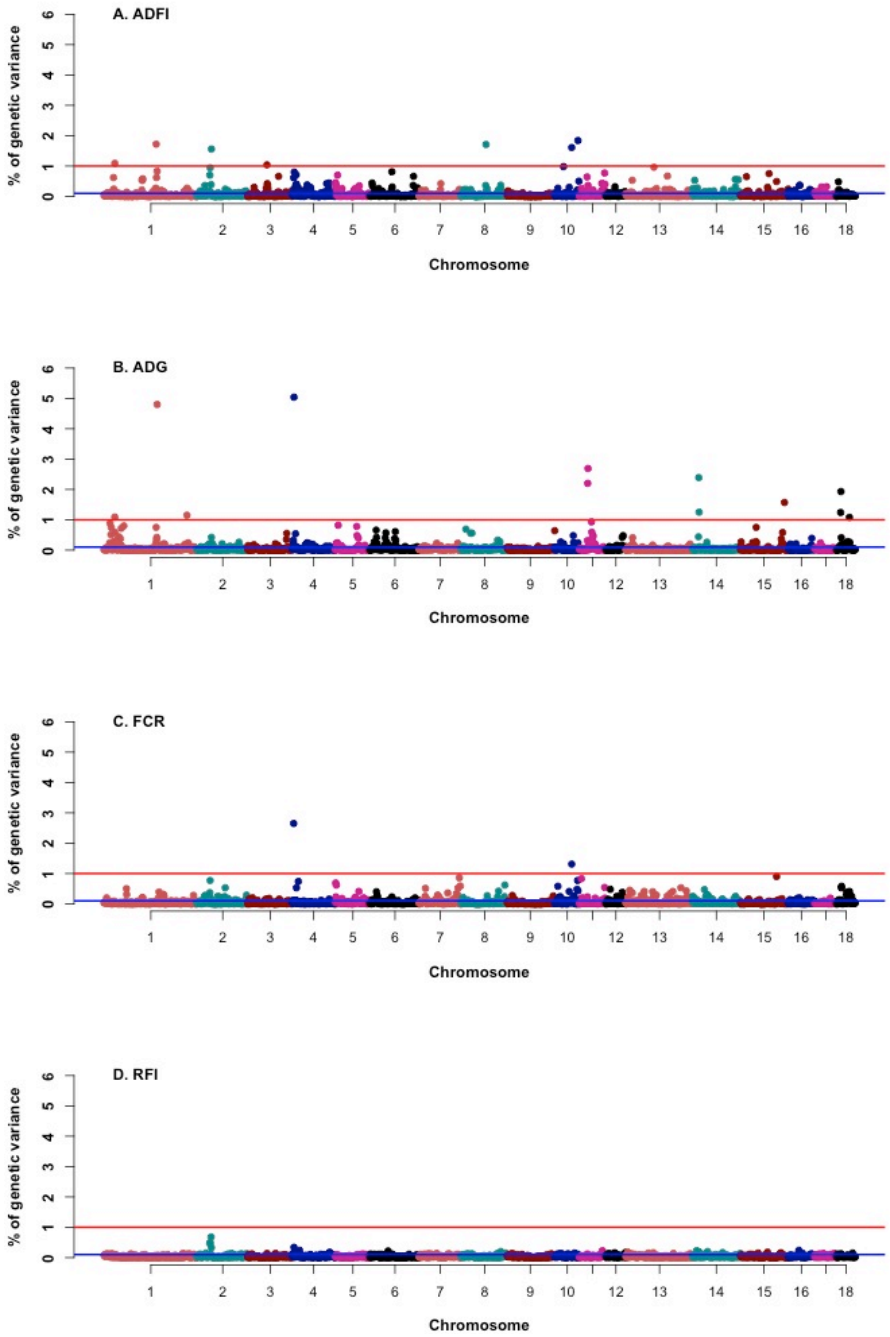

Figure 4.1 Percentage of genetic variances explained by windows (in total, 2380 1-Mb windows from autosome 1 to 18 of the swine genome) for feed efficiency traits from posterior distribution of window variances against chromosome positions.

From the top to the bottom are (A) ADFI, (B) ADG, (C) FCR (feed conversion ratio), (D) RFI (residual feed intake). In each of the Manhattan plot, the red horizontal line represents the cutoff of 1% of genetic variance explained; the blue one represented the cutoff of 0.1%.

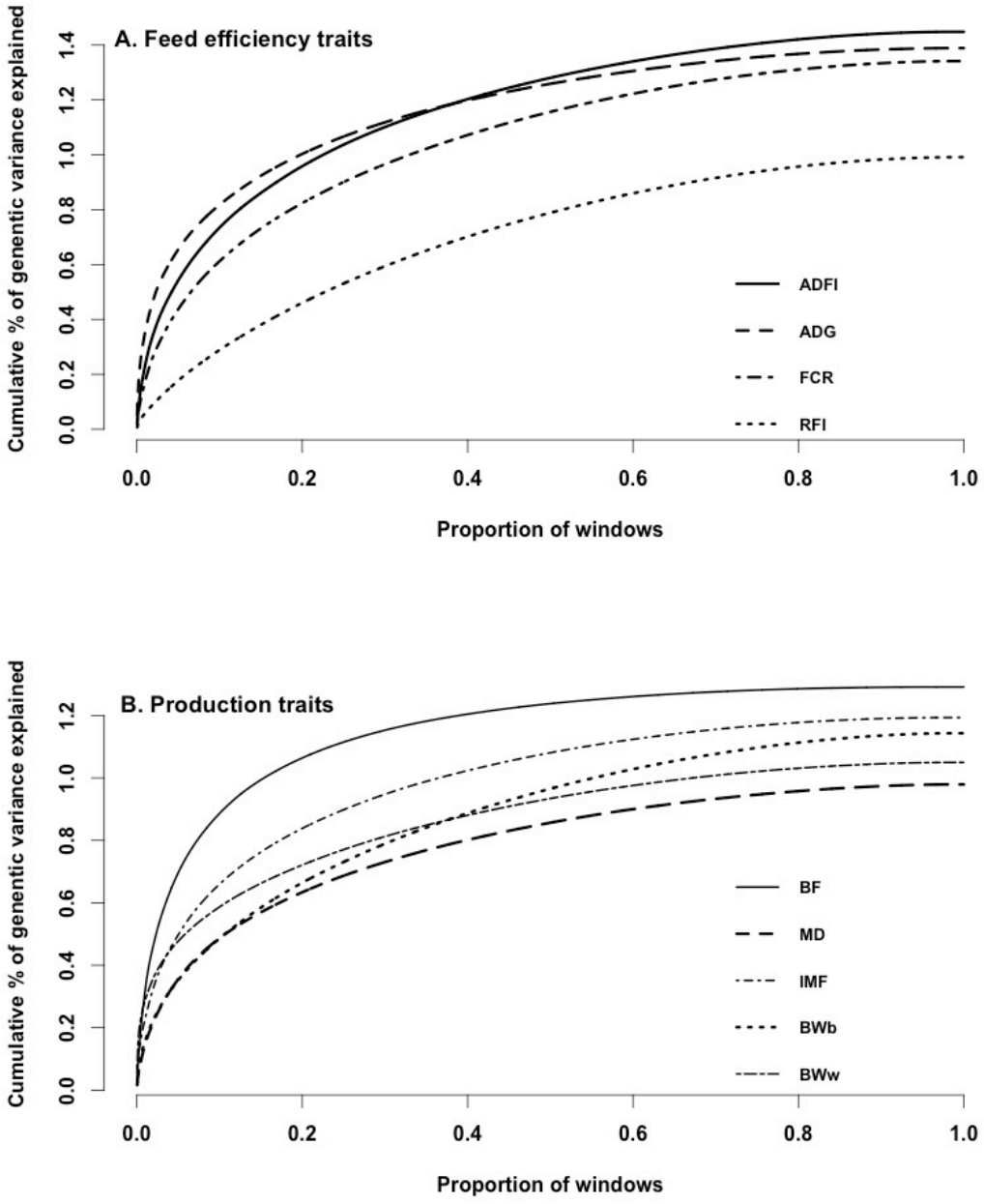

Figure 4.2 Cumulative proportion of genetic variance explained by 1-Mb windows (chromosome segments in length of 1-Mb), ranked from most to least variation explained derived from posterior distribution of window variances for each trait.

Table 4. 1 Top 5 genomic regions identified for feed efficiency traits by posterior distribution of 1-Mb window variances.

Trait[1]	Window	#SNPs	Chr	#Mb[2]	%Var[3]	Cum%Var	PPA[4]
ADFI	1497	16	10	74	1.84	1.84	0.525
	164	11	1	166	1.72	3.57	0.503
	1207	11	8	79	1.71	5.27	0.379
	1477	24	10	54	1.61	6.88	0.370
	341	20	2	46	1.56	8.45	0.314
ADG	599	22	4	6	5.04	5.04	0.823
	167	14	1	169	4.8	9.84	0.704
	1530	22	11	27	2.69	12.52	0.561
	1885	15	14	20	2.39	14.91	0.503
	1529	18	11	26	2.2	17.12	0.463
FCR	598	30	4	5	2.65	2.65	0.570
	1477	24	10	54	1.31	3.96	0.359
	2127	7	15	115	0.9	4.86	0.245
	1124	9	7	128	0.86	5.71	0.274
	1509	11	11	6	0.83	6.54	0.256
RFI	340	19	2	45	0.68	0.68	0.125
	337	28	2	42	0.5	1.18	0.140
	336	18	2	41	0.46	1.64	0.094
	599	22	4	6	0.34	1.98	0.100
	341	20	2	46	0.33	2.3	0.098

[1] FCR represents feed conversion ratio, which is Gain / ADFI; RFI is for residual feed intake.

[2] #Mb is the approximate starting position for the windows, the length of window is 1 Mb

[3] %Var stands for percentage of variance explained by the corresponding 1-Mb window

[4] PPA stands for probability of association

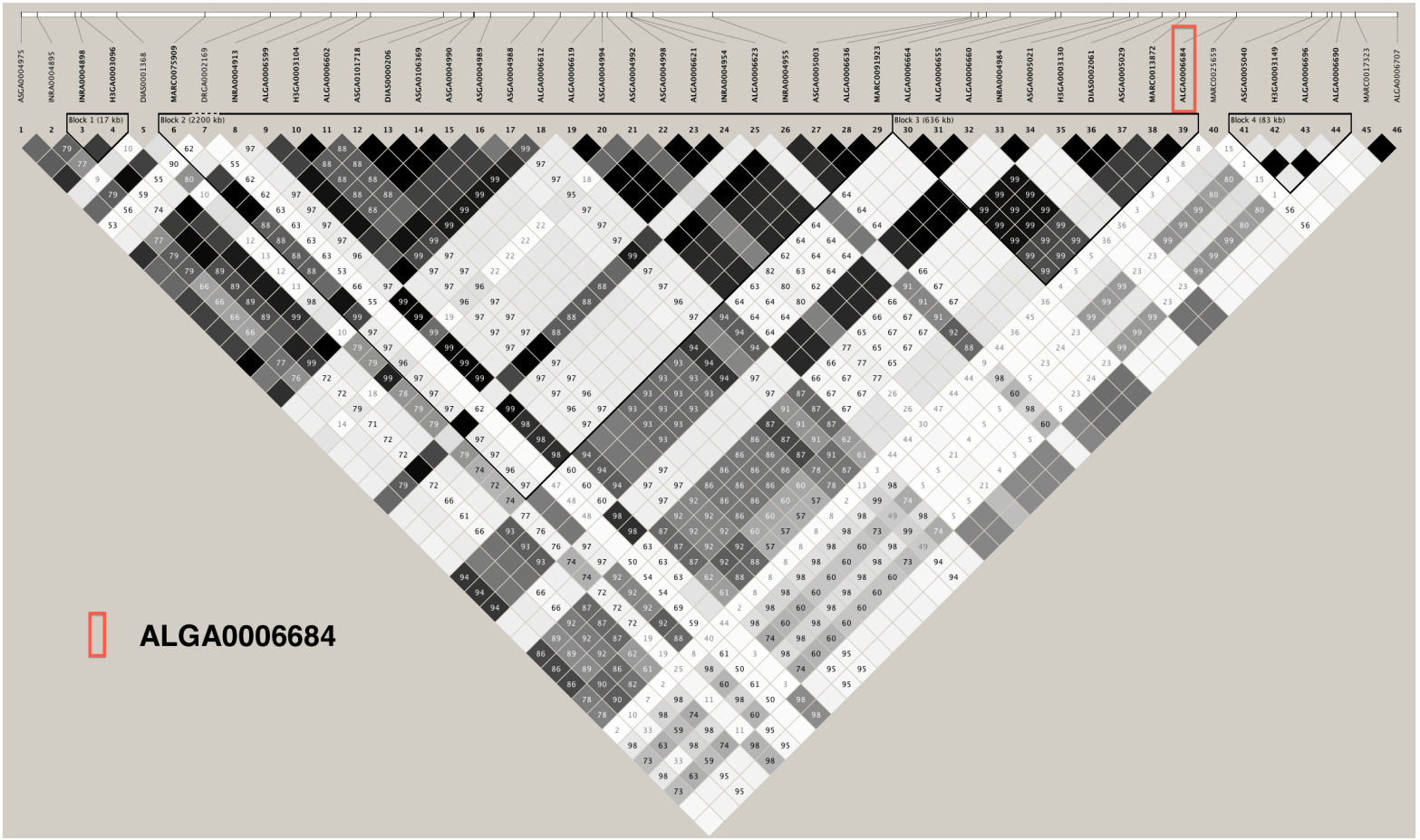

Figure 4.3 Linkage disequilibrium plot of the region approximately range from 166 to 170 Mb on SSC1, using r2 scores: white squares, r2 = 0; black ones, r2 = 1; grey one, 0 < r2 < 1. This 4-Mb region shown significant association with ADG, ADFI and BF (Ultrasound back fat thickness) and the most likely candidate genes are SOCS6 and DOK6 within this region. Note: red rectangular box were used to circle the most significant SNP (ALGA0006684) for all three traits ADFI, ADG and BF from Bayes-B models.

Table 4.2 Top 5 genomic regions identified for production traits by posterior distribution of 1-Mb window variances.

Trait[1]	Window	#SNPs	Chr	#Mb[2]	%Var[3]	Cum%Var	PPA[4]
BF	167	14	1	169	18.59	18.59	0.995
	2331	11	18	11	6.04	24.63	0.794
	457	21	3	3	2.11	26.74	0.637
	284	37	1	286	1.50	28.24	0.489
	286	19	1	288	2.30	29.53	0.393
MD	1243	21	8	115	2.64	2.64	0.508
	1244	16	8	116	1.88	4.52	0.321
	1659	26	13	6	1.75	6.27	0.345
	1099	18	7	103	1.45	7.72	0.213
	1098	11	7	102	0.86	8.58	0.166
IMF	1325	21	9	49	2.36	2.36	0.406
	1644	21	12	56	1.32	3.67	0.315
	1485	20	10	62	1.25	4.92	0.268
	1491	23	10	68	1.25	6.17	0.209
	165	15	1	167	1.21	7.37	0.255
BW at birth	2339	7	18	19	2.05	2.05	0.454
	1089	17	7	93	1.50	3.55	0.385
	1696	22	13	43	1.41	4.96	0.412
	2186	21	16	21	1.19	6.14	0.322
	509	24	3	58	1.13	7.28	0.355
BW at weaning[6]	1412	18	9	141	7.72	7.72	0.960
	677	8	4	84	3.23	10.95	0.635
	2152	17	15	140	3.00	13.95	0.731
	1247	16	8	119	2.62	16.56	0.710
	458	17	3	4	2.13	18.69	0.599
	1286	29	9	10	1.32	20.01	0.532

[1] BF is ultrasound back fat thickness; MD is muscle depth; IMF is intramuscular fat percentage.

[2] #Mb is the approximate starting position for the windows, the length of window is 1 Mb

[3] %Var stands for percentage of variance explained by the corresponding 1-Mb window

[4] PPA stands for probability of association

[4] BW at weaning had 6 windows to be significant, all the 6 windows have been listed in the table.

Table 4.3 Number of significant 1-Mb windows identified by Bayes Factor (>3.1, 100 or 500) for traits investigated.

Trait[1]	Bayes Factor > 3.1	Bayes Factor > 100	Bayes Factor > 500
ADFI	2334	228	4
ADG	2329	191	10
FCR	2344	295	2
RFI	2365	665	0
BF	2209	119	13
MD	2348	578	12
IMF	2295	103	3
BW at birth	2356	597	2
BW at weaning	2361	691	7

[1] FCR represents feed conversion ratio, which is Gain / ADFI; RFI is for residual feed intake; BF is ultrasound back fat thickness; MD is muscle depth; IMF is intramuscular fat percentage.

Table 4.4 Significant 1-Mb windows identified by bootstrapping analysis for traits investigated, using FDR < 10% and FDR < 5%.

Trait[1]	Windows < 10% FDR	Pos. of windows < 10% FDR (Chr_Mb[2])	Windows < 5% FDR
ADG	11	1_167, 1_168, 1_169, 2_85, 2_86, 9_77, 11_26, 11_27, 14_20, 14_21, 14_22	7
ADFI	9	1_166, 1_167, 1_168, 2_45, 2_46, 8_71, 10_73, 10_74, 10_75	6
FCR	11	1_27, 1_28, 2_123, 4_3, 4_4, 4_5, 9_150, 9_151, 10_46, 10_52, 10_76	4
RFI	1	11_6	1
BF	18	1_35, 1_166, 3_135, 6_69, 8_33, 8_34, 8_35, 9_64, 9_65, 14_14, 14_15, 14_16, 16_34, 16_75, 16_76	18
MD	13	1_233, 1_234, 5_98, 5_99, 6_153, 6_154, 8_116, 9_24, 9_25, 13_3, 13_4, 13_5, 13_6	0
IMF	11	1_3, 1_174, 6_36, 6_37, 7_2, 9_48, 9_49, 11_28, 14_21, 14_22, 15_127	5
BW at birth	6	3_56, 3_58, 11_21, 13_41, 18_18, 18_19	6
BW at weaning	14	4_81, 4_84, 8_15, 8_32, 8_33, 8_61, 8_118, 9_128, 9_129,9_139,9_141,12_49, 15_138, 15_139	9

[1] FCR represents feed conversion ratio, which is Gain / ADFI; RFI is for residual feed intake; BF is ultrasound back fat thickness; MD is muscle depth; IMF is intramuscular fat percentage.
[2] Chr_Mb provides the chromosome and starting physical position (in Mb) of the significant windows at FDR < 10%; the window length is 1 Mb.

Table 4.5 Summary of significant 1-Mb windows across 3 QTL inference strategies and potential candidate genes.

Trait[1]	Overlapped window[2]	Chr_(Mb)[3]	Candidate genes
ADG	1	1_(166-170)	*SOCS6, DOK6*
ADFI	2	1_(166-170), 10_(73-74)	*SOCS6, DOK6, PFKP*
FCR	1	4_(4-6)	NA
RFI	0	NA	NA
BF	1	1_(166-170)	*SOCS6, DOK6*
MD	1	8_(115-117)	*ARSJ*
IMF	0	NA	NA
BW at birth	0	NA	NA
BW at weaning	4	4_(83-85), 8_(118-120), 9_(140-142), 15_(138-141)	*PLAG1, CHCHD7, MOS, LYN, RPS20, SDR16C5, PRG4, PGHS-2, MOGAT1*

[1] FCR represents feed conversion ratio, which is Gain / ADFI; RFI is for residual feed intake; BF is ultrasound back fat thickness; MD is muscle depth; IMF is intramuscular fat percentage.

[2] Overlapped windows (QTL) for each trait were obtained from the significant regions across 3 inference methods; no overlapping windows or adjacent windows were identified for RFI, IMF or BW at birth.

[3] Physical position of potential QTL region (which includes overlapped windows (QTL) and adjacent windows identified by not all 3 significant testing methods) for all traits, given in chromosome and physical position in Mb in parenthesis.

Table 4.6 Substitution effects of minor allele for ALGA006684 (within 166-170Mb on SSC 1) for ADFI, ADG and BF (Ultrasound backfat thickness).

Marker	Position, bp	MAF[1]	ADFI			ADG			BF[2]		
			b[3]	SE	p-value	b	SE	p-value	b	SE	p-value
ALGA0006684	169437312	0.466	0.06480	0.01750	0.00012	0.01980	0.00390	0.00000	0.03000	0.00870	0.00030

[1] MAF represents minor allele frequency.

[2] BF is ultrasound back fat thickness.

[3] b is dose effect of the minor allele.

Table 4.7 Association of haplotypes of DNA segments with ADFI, ADG and BF (Ultrasound backfat thickness), SNPs were located within 166-170Mb on SSC 1.

Seg.[1]	Hap.[2]	Freqency	ADFI			ADG			BF		
			Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
1	01100	0.402	0.118	0.062	**4.800×10^{-5}**	0.018	0.014	**1.039×10^{-7}**	-0.009	0.032	**4.800×10-^{2}**
2	00100	0.394	0.008	0.051	**1.000×10^{-3}**	-0.022	0.012	**2.000×10^{-6}**	-0.020	0.028	**1.500×10^{-2}**
3	01000	0.405	-0.025	0.068	**1.000×10^{-3}**	-0.004	0.015	**2.000×10^{-6}**	-0.041	0.034	**2.400×10^{-2}**
4	00100	0.444	-0.001	0.064	**2.810×10^{-4}**	0.016	0.015	**2.000×10^{-6}**	0.004	0.033	**3.700×10^{-2}**
	10000	0.397	-0.067	0.065	1.620×10^{-1}	-0.004	0.015	4.690×10^{-1}	-0.019	0.033	**2.600×10^{-2}**
	10001	0.008	-0.036	0.113	7.510×10^{-1}	-0.064	0.024	**9.000×10^{-3}**	0.180	0.058	**2.000×10^{-3}**
5	00001	0.397	-0.067	0.065	**6.000×10-^{3}**	-0.004	0.015	**4.800×10^{-5}**	-0.019	0.033	**1.900×10^{-2}**
	00011	0.008	-0.036	0.113	7.000×10^{-1}	-0.064	0.024	**7.000×10^{-3}**	0.180	0.058	**2.000×10^{-3}**
	01000	0.444	-0.001	0.064	**3.500×10^{-2}**	0.016	0.015	**1.100×10^{-2}**	0.004	0.033	5.630×10^{-1}
6	00010	0.443	-0.102	0.104	**1.000×10^{-2}**	0.030	0.023	**3.400×10^{-5}**	-0.045	0.047	1.130×10^{-1}
	00100	0.063	-0.119	0.107	**3.200×10^{-2}**	0.035	0.024	8.400×10^{-2}	-0.057	0.049	8.300×10^{-2}
	00110	0.011	-0.029	0.114	8.420×10^{-1}	-0.041	0.025	6.000×10^{-2}	0.175	0.060	**3.000×10^{-3}**
7	00011	0.446	0.005	0.094	**3.370×10^{-4}**	0.058	0.021	**1.000×10^{-6}**	-0.038	0.044	**2.700×10^{-2}**
	01100	0.011	-0.022	0.114	9.150×10^{-1}	-0.040	0.025	9.300×10^{-2}	0.173	0.060	**3.000×10^{-3}**
8	000011	0.428	0.160	0.073	**1.000×10^{-3}**	0.051	0.016	**2.100×10^{-5}**	-0.028	0.039	**1.500×10^{-2}**
	010100	0.027	0.154	0.089	2.970×10^{-1}	0.058	0.020	**3.700×10^{-2}**	-0.072	0.048	3.880×10^{-1}
	101000	0.062	0.085	0.077	2.650×10^{-1}	0.037	0.017	**2.100×10^{-2}**	-0.059	0.041	9.600×10^{-2}
	111100	0.011	-0.013	0.117	9.030×10^{-1}	-0.043	0.026	9.900×10^{-2}	0.185	0.062	**3.000×10^{-3}**
9	000000	0.482	0.024	0.074	**2.100×10^{-2}**	-0.014	0.017	**1.868×10^{-7}**	-0.029	0.038	7.200×10^{-2}
	111100	0.006	0.388	0.155	**1.300×10^{-2}**	0.062	0.032	5.400×10^{-2}	0.007	0.063	9.130×10^{-1}

[1] Seg. represent segment, The QTL (166-170 Mb region on SSC 1) region was chunked into 9 segments, composed by 5 SNPs ordered by their physical position (Segment 9 includes 6 successive SNPs).

[2] Hap. is the combination of alleles (0 or 1) in segment, the data were phased using AlphaImpute v1.0.

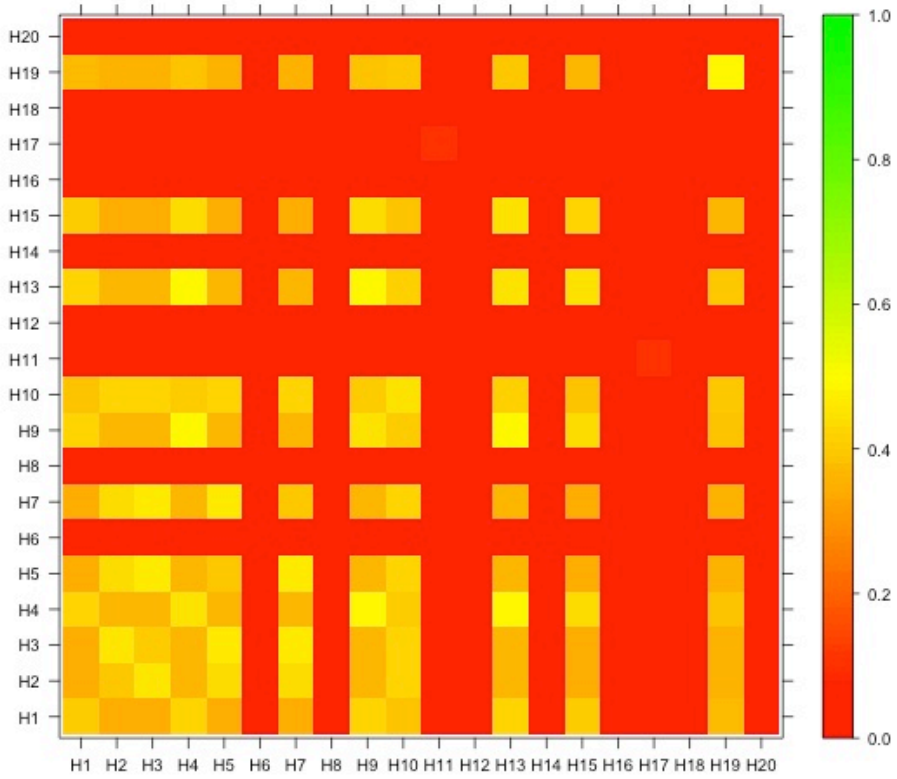

Figure 4.4 Heat map of frequency of significant haplotypes and joint frequency of pair-wise haplotypes. The 20 haplotypes (H1 to H20, ordered by their physical position on genome) were significantly associated with ADFI, ADG and BF, respectively. In the heat map, diagonal squares represent the frequency of each haplotype; the off-diagonal squares represent joint frequency (co-occurrence) of pair-wise haplotypes.

Table S4.1 Descriptive statistics for traits investigated for association analysis.

Statistics	ADFI(g)	ADG(g)	FCR[1]	RFI(g)[2]	BW at birth (kg)	BW at weaning (kg)	BF(cm)[3]	MD(cm)[4]	IMF(%)[5]
N	972	972	972	972	1047	889	1047	1047	730
min	1272.00	449.30	0.21	-843.10	1.03	2.95	0.51	2.38	2.42
1stQ	1818.00	725.30	0.36	-139.70	1.80	6.40	0.84	3.82	3.44
mean	2003.00	776.90	0.39	0.00	2.03	7.75	1.01	4.15	3.69
3rdQ	2179.00	828.60	0.42	144.20	2.26	9.07	1.14	4.43	3.96
max	3551.00	1092.90	0.65	1457.00	3.85	15.79	1.98	5.78	4.81
SD	288.71	86.70	0.06	260.96	0.34	1.93	0.22	0.47	0.40

[1] FCR represents feed conversion ratio, which is Gain / ADFI.

[2] RFI is for residual feed intake.

[3] BF is ultrasound back fat thickness.

[4] MD is muscle depth.

[5] IMF is intramuscular fat percentage.

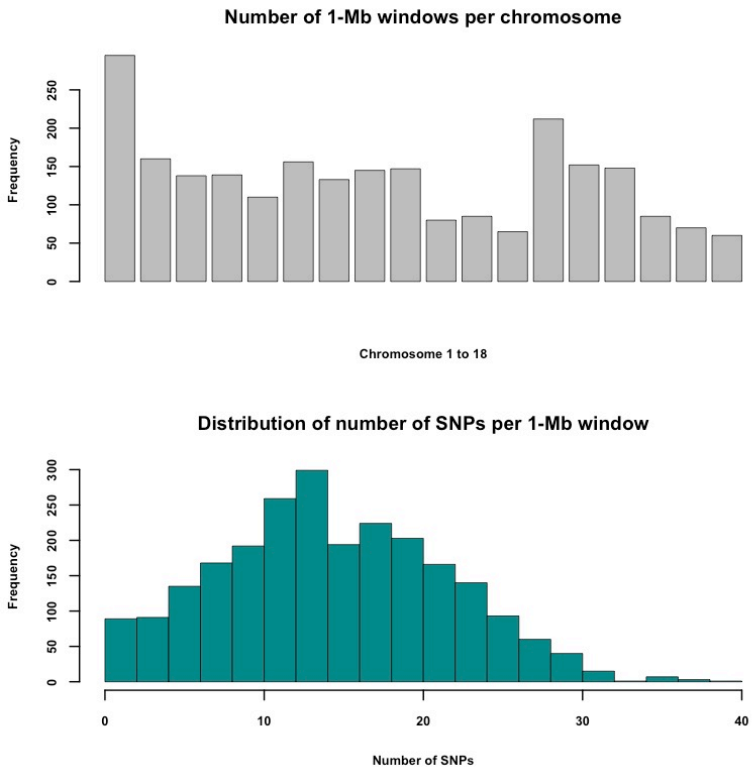

Figure S4.1 Distribution of number of 1-Mb windows per chromosome (histogram above) and distribution of number of SNPs per 1-Mb windows (histogram below).

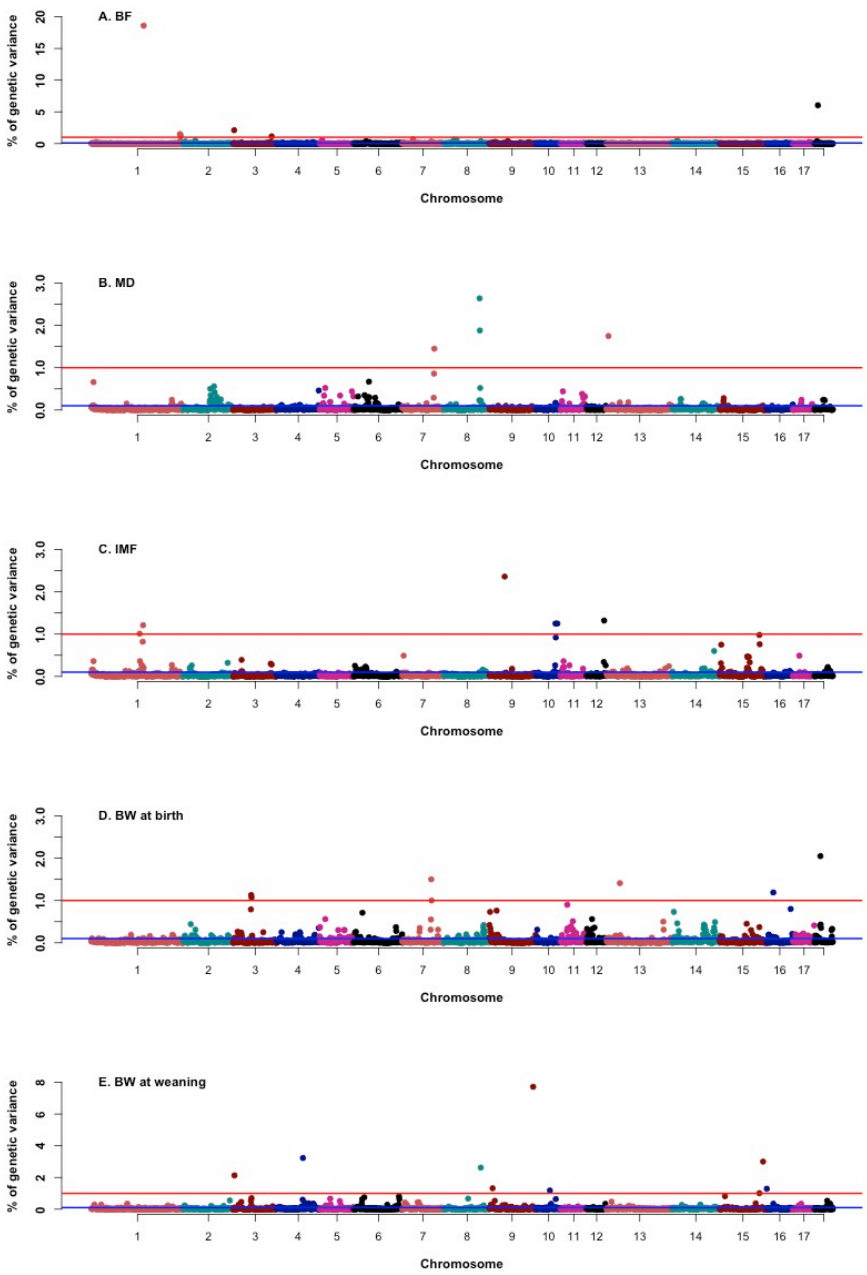

Figure S4.2 Percentage of genetic variances explained by windows (in total, 2380 1-Mb windows from autosome 1 to 18 of the swine genome) for 5 production traits from posterior distribution of window variances against chromosome positions. From the top to the bottom are (A) BF (Ultrasound back fat thickness), (B) MD (Ultrasound muscle depth), (C) IMF (Ultrasound intramuscular fat content), (D) BW at birth, (E) BW at weaning. In each of the Manhattan plot, the red horizontal line represents the cutoff of 1% of genetic variance explained; the blue one represented the cutoff of 0.1%.

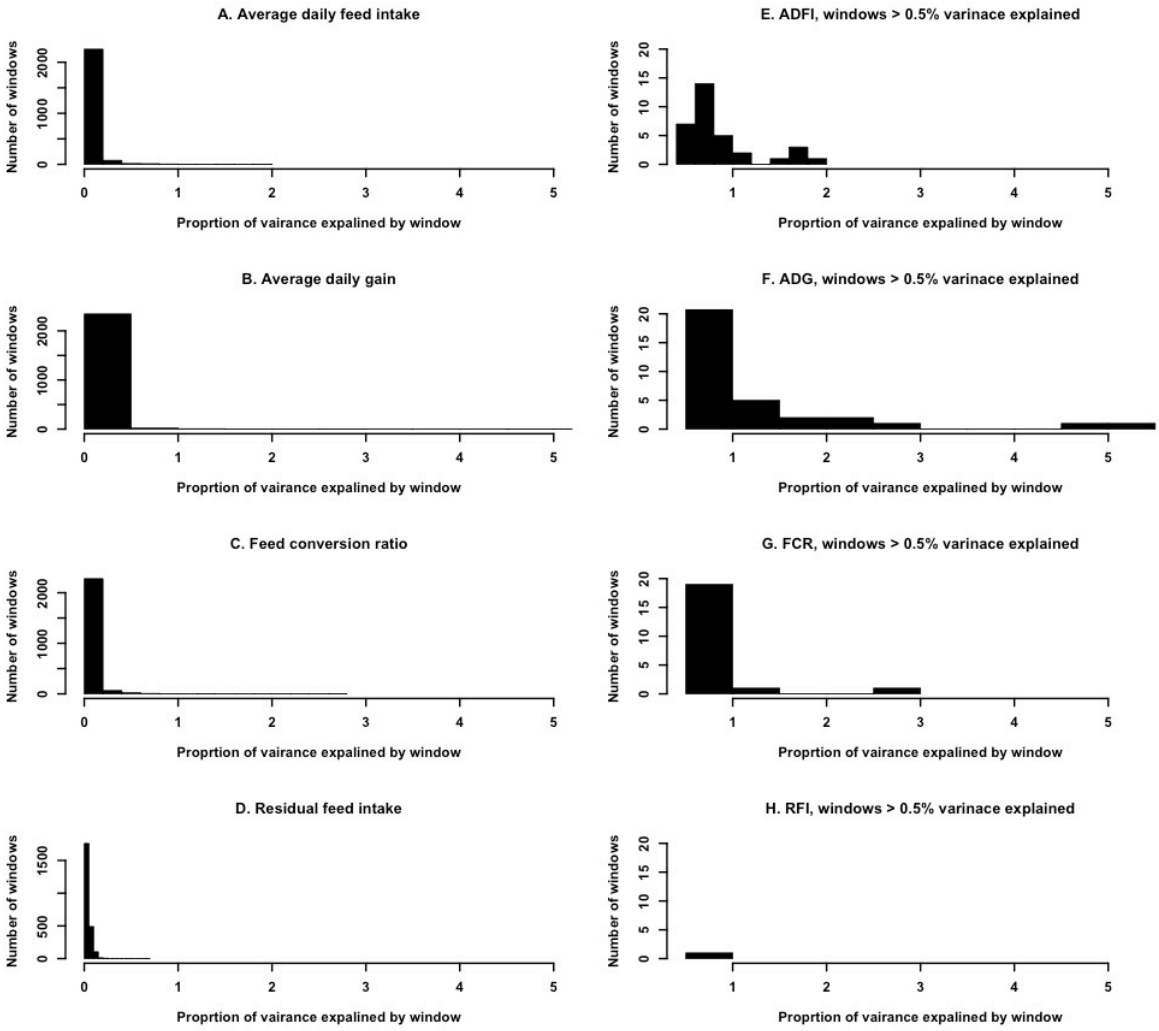

Figure S4.3 Distribution of proportion of genetic variance explained by 1-Mb windows for feed efficiency traits. From the top to the bottom are (A) Average daily feed intake, (B) Average daily gain, (C) Feed conversion ratio, (D) Residual feed intake. (E-H) are extreme right hand side of the same graphs, with the x-axis from 0.5 to 5% of variance explained.

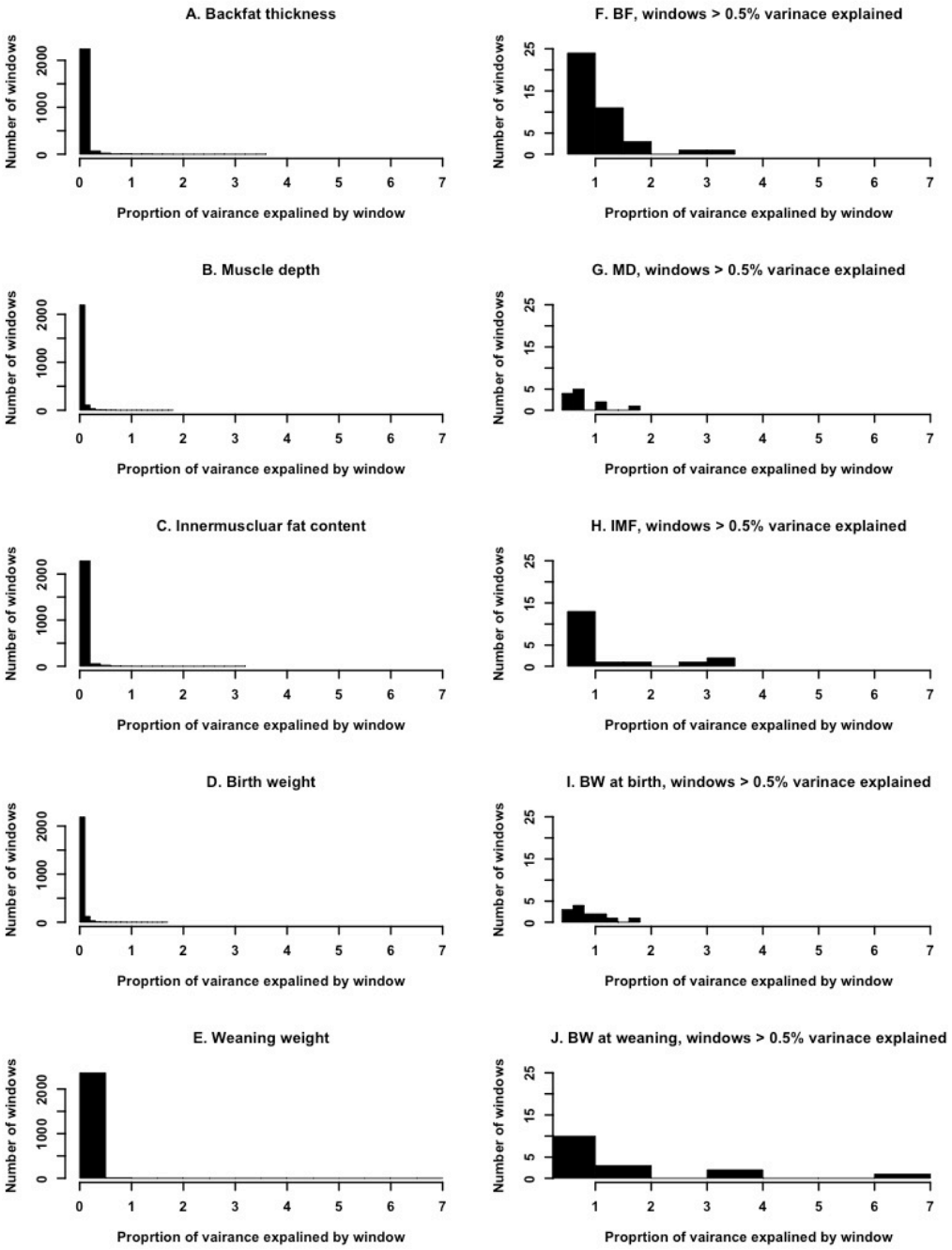

Figure S4.4 Distribution of proportion of genetic variance explained by 1-Mb windows for production traits. From the top to the bottom are (A) Back fat thickness, (B) Muscle depth, (C) Innermuscular fat content, (D) Birth weight, and (E) Weaning weight. (F-J) are extreme right hand side of the same graphs, with the x-axis from 0.5 to 7% of variance explained.

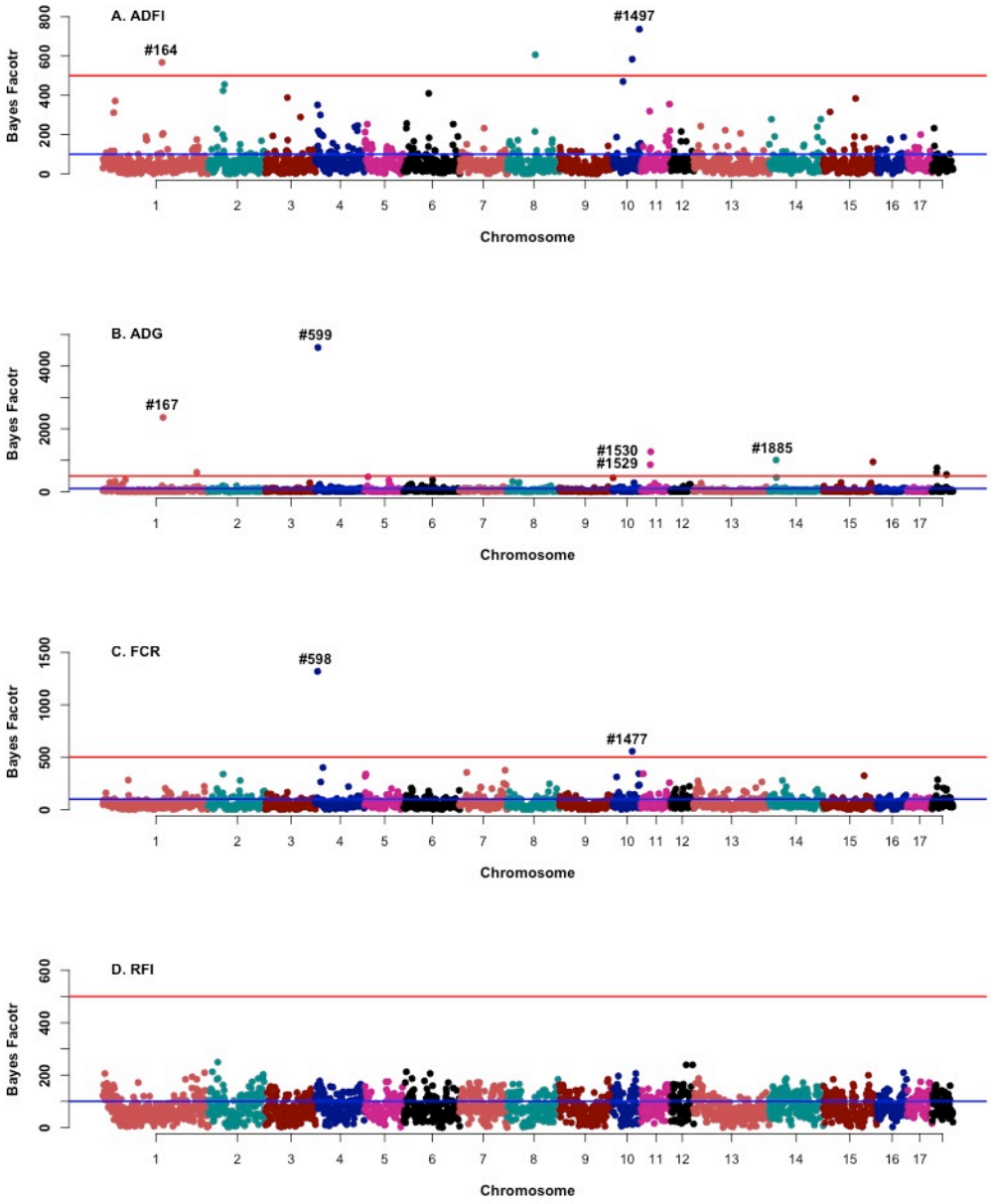

Figure S4.5 Bayes factor of windows (in total, 2380 1-Mb windows on 18 autosomes) for feed efficiency traits. In each of the Manhattan plot, cut-off line 100 and 500 were shown, and windows identified from posterior distribution of window variances were labeled on the top of the corresponding dots (except for ADG, window 1529 and 1530 were labeled on the left of the dots).

Figure S4.6 Linkage disequilibrium plot of the region between window164 and 178 (approximately range from 166 Mb to 180 Mb) on SSC1, using r2: white squares, r2 = 0; black ones, r2 = 1; grey one, 0 < r2 < 1. Regions in big blue rectangular box represents potential QTL region (window #164-169, including 46 SNPs), significantly associated with ADG, ADFI and BF (backfat). Red arrow is pointing at SNP ALGA0006684, which is most significant SNPs for ADFI, ADG and BF. Regions in small rectangular box represent region flunking by two SNPs (at 177605469 bp and SNP at 178599225 bp), where gene MC4R (at 1_178553488..178555752 bp from UCSC browser) located (blue arrow).

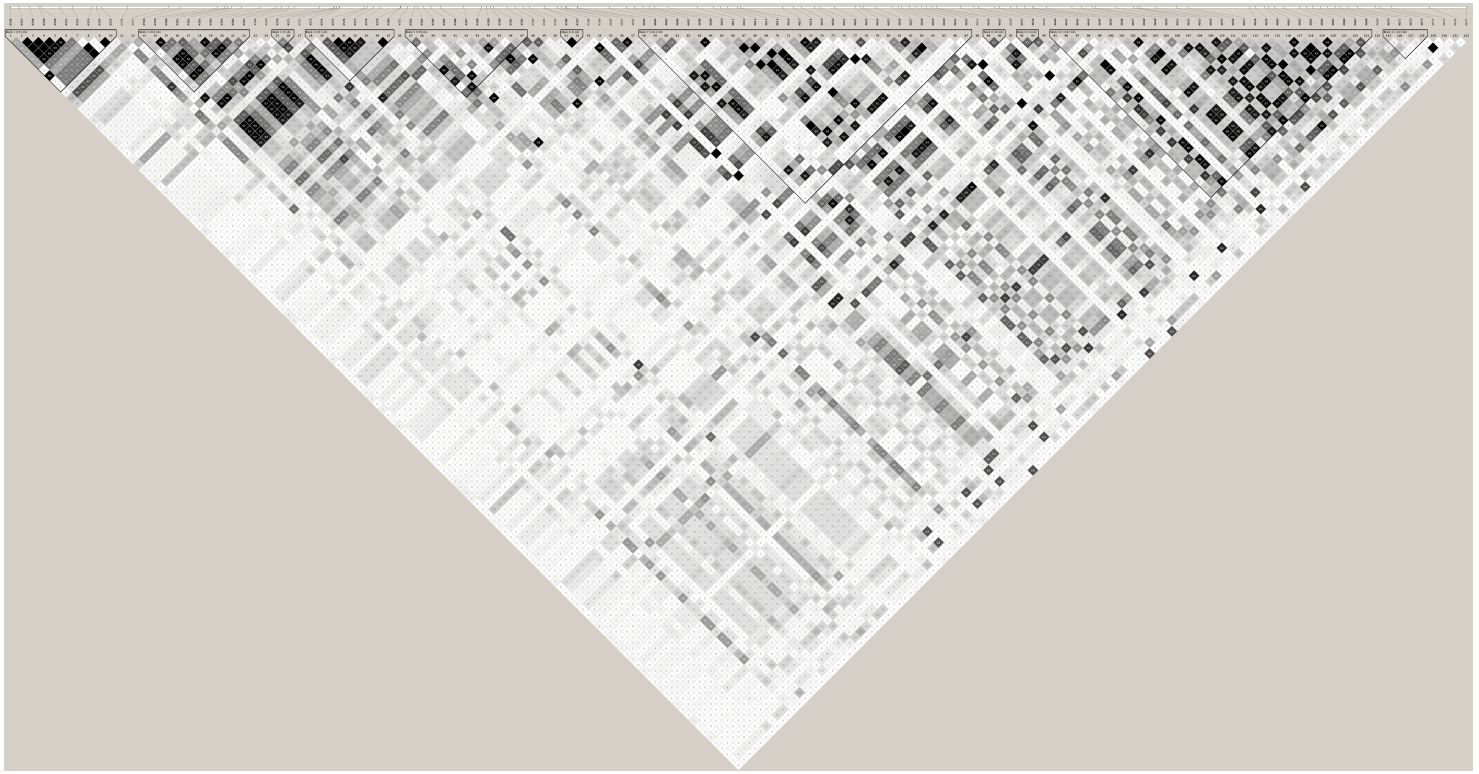

Figure S4.7 Linkage disequilibrium plot of the region between 336 to 341window (approximately 6 Mb) on SSC2, using r2: white squares, r2 = 0; black ones, r2 = 1; grey one, 0 < r2 < 1. This region associates with RFI and explaining 2.14% of genetic variance.

Table S4.2 Marker effects and Bayes factors for SNPs in 166-170Mb on SSC 1 associated with ADFI, ADG and BF (ultrasound backfat).

SNP_ID[1]	Position, bp	Window	Frequency	Eff.ADFI[2]	Eff.ADG	Eff.BF	BayF.ADFI[3]	BayF.ADG	BayF.BF
ASGA0004975	1_166015433	164	0.405	-0.000627	-0.000072	-0.000011	2.805	1.463	0.92
INRA0004895	1_166060598	164	0.465	0.004198	0.000372	0.000007	14.909	5.312	1
INRA0004898	1_166167164	164	0.479	0.0018	0.000218	0.000005	6.812	3.421	0.92
H3GA0003096	1_166184736	164	0.455	-0.000275	-0.000066	-0.000005	1.605	1.362	0.96
DIAS0001368	1_166284434	164	0.115	-0.000026	-0.000005	0.000003	0.519	0.519	0.839
MARC0075909	1_166532772	164	0.465	0.003474	0.000397	0.000012	12.32	5.459	1.02
DRGA0002169	1_166717768	164	0.014	-0.000056	-0.000053	-0.000001	0.859	1.02	1.06
INRA0004913	1_166803747	164	0.08	-0.000181	-0.000019	-0.000011	1.04	0.659	1.02
ALGA0006599	1_166888248	164	0.455	-0.000232	-0.00008	0	1.383	1.443	0.879
H3GA0003104	1_166960658	164	0.465	0.003346	0.000423	0.000013	11.962	5.943	1
ALGA0006602	1_166999933	164	0.465	0.003417	0.00046	0.00001	12.298	6.409	0.98
ASGA0101718	1_167203653	165	0.481	-0.000131	-0.000017	-0.000005	1.06	0.579	0.92
DIAS0000206	1_167254444	165	0.481	-0.000194	-0.000015	-0.000006	1.241	0.579	0.94
ASGA0106369	1_167288269	165	0.481	-0.000135	-0.000019	-0.000007	0.96	0.619	0.96
ASGA0004990	1_167292397	165	0.481	-0.000118	-0.000009	-0.000006	1.02	0.519	0.94
ASGA0004989	1_167336808	165	0.405	-0.000484	-0.000127	-0.000008	2.315	2.234	0.98
ASGA0004988	1_167458793	165	0.439	0.001336	0.000055	0.000021	5.585	1.121	1.101
ALGA0006612	1_167533528	165	0.456	-0.000255	-0.000067	-0.000001	1.443	1.362	0.879
ALGA0006619	1_167629414	165	0.081	-0.000215	-0.000023	-0.000009	1.161	0.699	1
ASGA0004994	1_167652749	165	0.081	-0.000248	-0.00002	-0.000011	1.181	0.719	1.08
ASGA0004992	1_167665746	165	0.132	-0.000061	-0.000011	-0.000003	0.599	0.519	0.879
ASGA0004998	1_167721429	165	0.081	-0.000219	-0.000013	-0.000013	1.221	0.599	1.02
ALGA0006621	1_167733085	165	0.081	-0.000221	-0.000012	-0.000017	1.101	0.639	1.02
INRA0004954	1_167736704	165	0.081	-0.000204	-0.000023	-0.000012	1.121	0.799	1.04
ALGA0006623	1_167795468	165	0.463	0.003862	0.000539	0.000017	13.516	7.41	0.98
INRA0004955	1_167961904	165	0.081	-0.000237	-0.000023	-0.000013	1.161	0.719	0.96
ASGA0005003	1_168690854	166	0.068	-0.000157	-0.00002	-0.000012	0.96	0.759	1.08

Table S4.2 Continued

ALGA0006636	1_168710857	166	0.066	-0.000174	-0.000017	-0.000005	0.96	0.619	0.96
MARC0091923	1_168733437	166	0.068	-0.000195	-0.000022	-0.00001	1.06	0.639	1.08
ALGA0006664	1_168800545	166	0.374	-0.000247	-0.00042	-0.000035	1.443	5.648	1.101
ALGA0006655	1_168927476	166	0.374	-0.000295	-0.000439	-0.00001	1.524	6.133	0.96
ALGA0006660	1_168943790	166	0.374	-0.000232	-0.000332	-0.000013	1.443	4.706	0.98
INRA0004984	1_169089277	167	0.068	-0.000192	-0.000023	-0.000009	1.06	0.759	0.98
ASGA0005021	1_169135658	167	0.068	-0.000176	-0.000019	-0.000007	1.02	0.719	0.859
H3GA0003130	1_169160364	167	0.39	-0.000272	-0.000225	-0.00001	1.645	3.359	1.04
DIAS0002061	1_169228082	167	0.39	-0.000219	-0.000242	-0.00002	1.342	3.627	1
ASGA0005029	1_169275024	167	0.39	-0.000336	-0.000336	-0.000029	1.868	4.747	1.101
MARC0013872	1_169294975	167	0.458	-0.002362	-0.002044	-0.01751	8.92	26.011	79.829
ALGA0006684	1_169437312	167	0.466	-0.006772	-0.004212	-0.04479	23.794	59.542	506.174
MARC0025659	1_169437674	167	0.375	0.000007	0.000008	0.000013	0.339	0.419	0.96
ASGA0005040	1_169649249	167	0.121	0.000061	0.000065	0.000018	0.699	1.262	1.141
H3GA0003149	1_169691789	167	0.403	0.000225	0.002707	0.000001	1.362	35.283	0.92
ALGA0006696	1_169706029	167	0.121	0.000049	0.00008	0.00002	0.659	1.403	1.161
ALGA0006690	1_169732994	167	0.403	0.000166	0.002775	0	1.08	36.587	0.879
MARC0017323	1_169771968	167	0.038	-0.000091	0.000155	0.000003	0.839	1.807	1.06
ALGA0006707	1_169889708	167	0.038	-0.000082	0.000124	0	0.799	1.645	0.98

[1] SNP_ID is SNPs name for Illumina PorcineSNP60K Bead Chip

[2] Effect.ADFI= SNP substitution effects for SNPs, obtained by Bayes-B model (Genome-wide association analysis).

[3] BayF.ADFI = Bayes factor for SNPs

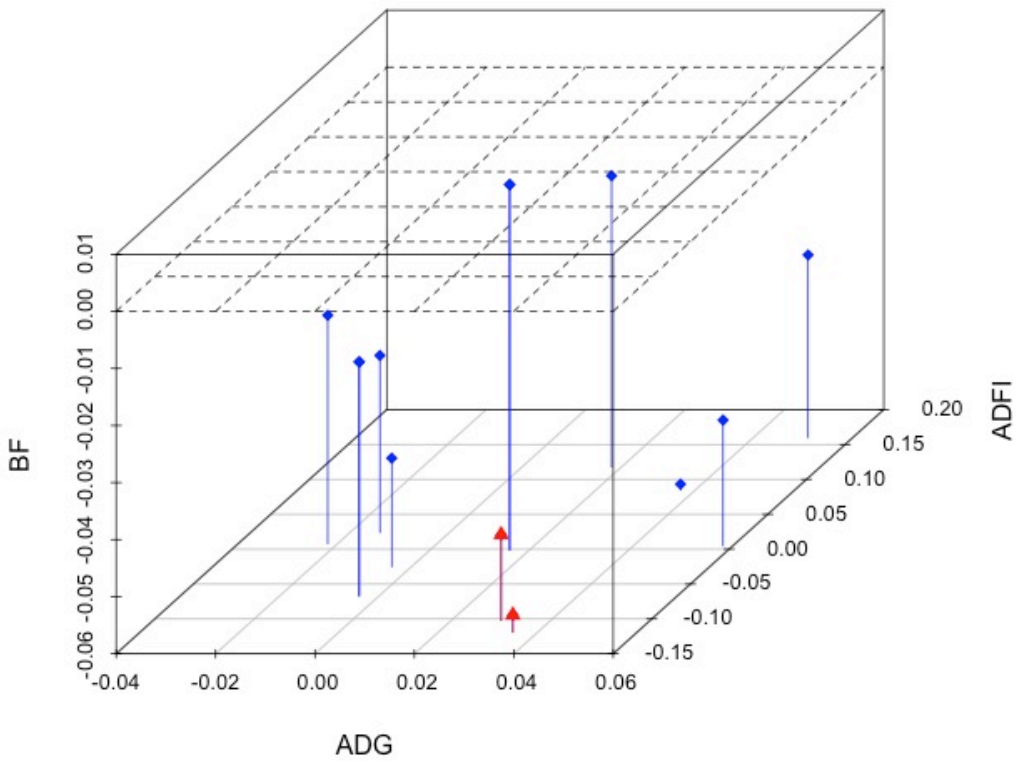

Figure S4.8 Scatter plot of significant haplotypes within 166-170 Mb on SSC 1 associated with ADG, ADFI and BF (Ultrasound backfat). The blue points and red triangular points represent the effects of haplotypes for ADFI, ADG and BF respectively. The two haplotypes marked by red color have positive effect on ADG and negative effects on ADFI and BF, which are favorable when selecting pigs with fast growth, eat less and grow less fat.

CHAPTER 5

Genetic analysis for different measures of feed efficiency and feeding behavior traits and the accuracy of genomic prediction using single-step method

Shihui Jiao[*], Francesco Tiezzi[*], Mark Knauer[*], Yijian Huang[†], Kent A. Gray[†], and Christian Maltecca[*]

[*] Department of Animal Science, North Carolina State University, Raleigh, NC 27695

[†] Smithfield Premium Genetics, Rose Hill, NC 28458

INTRODUCTION

Considerable variation has been reported for feed efficiency traits in different breeds / lines of pigs in previous studies, especially feed conversion ratio (FCR) and RFI (residual feed intake) as two measures of feed efficiency. The heritability estimates for FCR and measures of RFI were moderately high, with literature averages of 0.30 (ranging from 0.12 to 0.58) for FCR and 0.24 (ranging from 0.10 to 0.42) for RFI, respectively (Rothschild and Ruvinsky, 2010). For example, Do et al. (2013) reported estimates for FCR of approximately 0.30 for Danish Duroc, Landrace and Yorkshire while two measures of RFI were ranging from 0.34 to 0.40 across breeds. The estimate of heritability from Jiao et al. (2014) was consistent for FCR with 0.32 but much lower for RFI with 0.10. Likewise, MacNeil and Kanp (2015) reported an estimate for RFI of 0.22 in Canadian Duroc boars. However, careful consderation should be taken when comparing those estimates, as feed efficiency does not refer to one trait, but rather encompasses all traits associated with the efficiency of feed utilization. Furthermore, differences in the formulae used to compute a particular feed efficiency trait may lead to differences in estimates or even selection outcomes (Arthur and Herd, 2005). Little is known about differences of estimates of heritability for different measures of feed efficiency, the genetic correlation of those measures with one other and other traits especially when incorporating pedigree and marker information in large populations.

This lack of information might reflect the fact that collecting individual feed intake is expensive. The use of electronic feeding stations (Maselyne et al., 2015) has greatly facilitated the process of collection of individual feed intake for group-housed pigs in large

populations, such as FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, Kansas USA) and IVOG (individual feed intake recording in group housing, Insentec B.V., Marknesse, the Netherlands). Extra benefits have been seen when using the two feeding systems as the electronic feeding system allows to record not only the consumption of feed but also related feeding behavior measures. In spite of this, limited research has been done on the relationships between behavior measures and different measures of feed efficiency.

As genotyping has become affordable, the swine industry started to genotype a large part of their nucleus populations in order to take advantage of genomic selection. The single-step method (ssBLUP) has been widely adopted due to its advantages of simplicity and applicability to both genotyped and non-genotyped animals (Legarra et al., 2014). The present work had three objectives. The first was to estimate genetic parameters for different measures of feed efficiency and other traits (including growth, off-test measurements and feeding behavior traits) with pedigree and marker information. The second objective was to quantify the effect of social interaction among pen mates for traits collected using FIRE system. The third objective was comparing accuracy of prediction for all those traits of interests using traditional BLUP or single-step genomic BLUP approach.

MATERIALS and METHODS

Field data

Data provided by Smithfield Premium Genetics (SPG; Rose Hill, NC) was collected from June 2004 through May 2013 for Duroc purebred boars on a nucleus farm. It included 14901 boars from 4801 litters obtained by matings of 3094 dams and 325 sires (Table S5.1).

At birth, the date of birth and birth weight were recorded. Piglets were weaned at an average age of 25 d and the individual weaning weights were recorded.

Pigs weaned within the same week and with the same gender were grouped into contemporary groups (CG, concatenation of year and week, n = 223) and then moved to finishing barn after the nursery phase (approximately 7 weeks). During the test period (growing phase in the finishing period), individual feed intake and related measures were recorded using single-space FIRE feeders placed in each pen of 14 rooms (n = 240 physical Pen + Room, eg. Pen_Room) and described in details by Chen et al. (2010) and Jiao et al. (2014a). FIRE feeders were equipped with a weigh scale (ACCU-ARM Weigh Race) to measure body weight of each pig accessing the feeder. In addition, feed consumption and time were recorded each time a pig visited the feeder. When on test, boars of a group of ~12 (ranging from 6 to 19) were distributed into a pen with a single-space FIRE feeder. Boars started the test at approximately 93 d of age (on-test age) with test duration of 67 days. Boars were removed from test at an average age of 170 d and 4958077 single individual visits to feeder for the 14901 testing boars were recorded. Data collected by FIRE feeders contain errors (Casey et al., 2005, Zumbach et al., 2010) and the error visits were adjusted using multiple imputation technique described by Jiao et al. (2015, submitted) and robust regression detailed by Zumbach et al. (2010).

Feed intake and feeding behavior traits summed for each individual by day or averaged by visit across the whole testing period were employed in subsequent analyses. Seven feeding behavior traits were analyzed: average daily feed intake (ADFI), average daily occupation time (AOTD), average daily feeding rate (AFRD), daily average number of visits

(ANVD), average feed intake per visit (AFIV), average occupation time per visit (AOTV) and average feeding rate per visit (AFRV).

During the off-test procedure, all boars were weighed to get off-test body weight (off-test BW) and had longitudinal ultrasound images captured over the last three ribs. Images were captured via an Aloka 500 (Corometrics Medical Systems, Wallingford, CT) and analyzed for intramuscular fat using the Swine Image Analysis Software (Designer Genes Technologies, Inc. Harrison, Arkansas). The four off-test traits considered in the current study were off-test BW, backfat thickness (BF), muscle depth (Muscle) and intramuscular fat percentage (IMF).

Growth rates for each testing boars were computed in three different ways: life time average daily gain (ADG_Life) using regression coefficient of off-test BW and birth weight; post-weaning average daily gain (ADG_PostWean) as regression coefficient of off-test BW and weaning weight; on-test average daily gain (ADG_FIRE) using body weight recorded using FIRE systems (Jiao et al., 2014a). Additionally, the metabolic mid-test body weight (MMW) was computed using the estimates of intercept from robust regression as described by Nguye et al. (2005).

Eleven feed efficiency measures were employed in the present analysis. Three measures of feed conversion FCR1, FCR2 and FCR3 were computed as ratios of growth to ADFI (in g), with ADG_FIRE (g), ADG_Life (g) and ADG_PostWean (g) as numerators, respectively. Six measures of RFI (including RFI1, RFI2, RFI3, RFI4, RFI5 and RFI6) were modeled starting from residuals from different regression models with sets of different production traits as covariates (Table 5.1). Residual growth (RG) was determined similarly to

RFI, and computed as the difference of growth (ADG_FIRE) and predicted growth (Crowley et al., 2010; Willems et al, 2013). Residual feed intake and body weight gain (RIG) was the difference of RG and RFI6, as described in Crowley et al. (2010) and Willems et al. (2013). The RG was defined as the residual of the following model,

$$y_i = \beta_0 + CG_{ij} + \beta_1 MMW_i + \beta_2 ADFI_i + \beta_3 BF_i + \beta_4 Muscle_i + e_i,$$

where y_i was ADG_FIRE of the ith boar; CG_{ij} was the ith boar in the jth CG; MMW_i, $ADFI_i$, BF_i, and $Muscle_i$ were MMW, ADFI, BF and Muscle for the ith boar as covariates in the model; β_0 represented the intercept of the regression model and β_1 to β_4 were regression coefficients for the corresponding covariates; and e_i was model residual.

Descriptive statistics of the data are shown in Table 5.2, including the number of observations, minimums, maximums, means and standard deviations (SD) for all the phenotypes. Distribution of sires in CG and pen are summarized in Table S5.2 and Table S5.3.

Marker data

Animals were genotyped using the high-density marker panel Illumina PorcineSNP60 Beadchip (Illumina Inc., San Diego, CA USA; n = 3699) and the low-density panel GGP-Porcine containing approximately 10,000 SNP (GeneSeek Inc., Lincoln, NE USA; n = 4621). Imputation of genotypes from low-density panels to high-density marker panels was achieved using Beagle software (Browning and Browning, 2009). The multiple quality control edits conducted for animal and SNP included removal of animals with call rate < 0.9, SNP with call rate < 0.9, SNP with minor allele frequency (MAF) < 0.01 and SNP with p-value < 0.0001 of a chi-square test for Hardy-Weinberg equilibrium. Details of imputation

procedures can be found in Howard et al. (2015). Additionally, animals with parent and progeny conflicts and possible duplicated genotype samples were also excluded from the analysis using PreGSF90 program (Misztal et al., 2012). After editing, the genotype dataset included 33967 SNP on 18 autosomes for 8701 animals.

Dataset splits

In order to compare the accuracy of genomic prediction using single-step approach and traditional BLUP, forward validation was used. The youngest generation of boars born in 2011 and 2012 with both genotypes and phenotype records were used as testing set, including 506 individuals. The training data contained 14395 boars with phenotypes and 8195 animals genotyped.

Statistical Analysis

Variance components were estimated using AIREMLF90 in the BLUPF90 family of programs (Misztal et al., 2002). All analyses were completed with either single-trait or two-trait animal models.

To estimate heritability for each trait in Table 5.1, three single-trait animal models were used. The first statistical model which is the simplest model used to describe the data was (Model 1),

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e},$$

where **y** is the vector of observations (behavior, growth, feed efficiency and off-test traits); b is a vector of fixed effects including contemporary groups (CG, n = 223), parity of dam (1, 2, 3+), and fitted covariates for corresponding trait; **a** is a vector of additive genetic effects of the boar; **e** is a vector of random residuals; and X, Z are corresponding

incidence matrices. On-test age and test duration were fitted as covariates for behavior traits ADG_FIRE and feed efficiency traits, except FCR2 and FCR3; off-test age was fitted as covariates for off-test BW and off-test BW was fitted as covariate for BF, Muscle and IMF.

Assumptions are slightly different in pedigree analysis and marker analysis for additive effects. In pedigree analysis, additive effect is assumed $\mathbf{a} \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2)$ and **A** is the numeric relationship matrix. Alternatively with marker information, additive effect $\mathbf{a} \sim N(\mathbf{0}, \mathbf{H}\sigma_a^2)$ and **H** is the blended relationship matrix described by Legarra et al. (2009) and Christensen and Lund (2010)

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} 0 & 0 \\ 0 & \lambda\,(\mathbf{G^{-1}} - \mathbf{A_{22}^{-1}}) \end{bmatrix},$$

where $\mathbf{A^{-1}}$ is the inverse of the numerator relationship matrix (**A**) including all animals; **G** is the genomic relationship matrix; $\mathbf{A_{22}^{-1}}$ is the inverse of the **A** matrix for only genotyped animals; λ parameter was set to 0.95 corresponding to weighting 0.95 for $\mathbf{G^{-1}}$ and 0.05 for $\mathbf{A_{22}^{-1}}$. Observed allele frequencies were used to center and scale the observed genotype matrix. Standard assumptions for residual were employed so that $\mathbf{e} \sim N(\mathbf{0}, I\sigma_e^2)$.

To account for the possible spatial effect of the pen a second single-trait animal model was fitted for each trait. The model had the following form (Model 2),

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{Z_1}\mathbf{a} + \boldsymbol{Z_2}\mathbf{p} + \mathbf{e},$$

where the model is identical to Model 1, except for the extra term added: **p** is a vector of pen-room (concatenation of pen and room, n = 240) effect, treated as random effects and

$\boldsymbol{Z_2}$ as incidence matrix for **p**. The assumptions for random effects of this model in pedigree analysis were

$$\begin{bmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{A}\sigma_a^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I_p}\sigma_p^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_e}\sigma_e^2\end{bmatrix}\right)$$

where σ_a^2, σ_p^2 and σ_e^2 are variance components for additive, spatial pen-room and residual effects and **A** is the numerator relationship matrix derived from pedigree information. Likewise, when incorporating marker information, the assumptions of the model changed with the numerator relationship matrix **A** replaced by the blended relationship matrix **H**.

A third single-trait animal model was fitted to take into account the interactions among pen mates and a social common pen effects instead of the spatial pen-room effect for each trait was fitted. The statistical model was (Model 3),

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{Z_1}\mathbf{a} + \boldsymbol{Z_3}\mathbf{s} + \mathbf{e},$$

where **s** is a vector of social common pen effect (n = 921) and $\boldsymbol{Z_3}$ is the incidence matrix for **s**. All other model terms remained the same as in model 1. The assumptions of this third model were similar to model 2, except replacing $\boldsymbol{I_p}\sigma_p^2$ in the (co)variance structure in model 2 with $\boldsymbol{I_p}\sigma_s^2$ in model 3.

Preliminary results from the use of seven different models were evaluated and the above three described model were chosen due to the better fits (Table S5.4)

The two-trait animal model used for behavior traits and other traits of interests were identical to the single-trait model (Model 1), except adding covariance among

additive effects of the two traits in the model. Only model 1 was used to obtain genetic variance among behavior traits and other traits for simplicity and easier convergence.

The estimates of variance components from single-trait and two-trait models were obtained by using ASREMLF90 program. Estimates of heritability and genetic correlations and their standard errors were computed using the approach suggested by Tsuruta (2015).

To predict the estimated breeding values (EBVs) or genomic estimated breeding values (GEBVs) for animals in testing set, phenotype of those animals were masked as missing. The EBVs for boars in the testing set were calculated using a traditional single-trait BLUP prediction model with pedigree information, identical to Model 3 described above. Analyses were carried out with BLUPF90 (Misztal et al., 2002) with the variance components fixed as estimates from Model 3 by AIREMLF90.

Accuracy for EBVs or GEBVs from traditional BLUP or single-step GBLUP were computed using formula as follows (Forni et al., 2011)

$$r = \sqrt{1 - (SEP^2 / \sigma^2_a)},$$

where r is the accuracy of prediction; SEP (standard error of prediction) as output from BLUPF90 program by taking the diagonals of the inverse left hand side (LHS) of the mixed model equations; and σ^2_a is the additive genetic variance of the trait. Thus, accuracy of prediction was defined by taking the average of the accuracy of EBVs or GEBVs for boars in the testing dataset.

RESULTS

Table 5.1 presents the different regression models used to compute the six measures of RFI. The models were different as they contained different sets of production traits as covariates in the LHS of the regression. The covariates fitted in the models were ADG, MMW, BF and MD, respectively. Nearly all estimated regression coefficient were significant, except for Muscle in the model for RFI5 and RFI6. For all measures after fitting the model, the residuals were taken as corresponding measures of residual feed intake.

Figure 5.1 shows the average relationship among pen mates in each pen. The average relationship of pen mates ranged from 0 to 0.08, which is consistent with the fact that few litter mates living in the same pen, provided the average number of litter mates in the dataset is approximately 3 (See Table S5.1 for distribution of pigs per litter). Consequently, it seems there is no need to consider litter effects in the analysis. In fact, adding litter effects did not improve model fitting (Table S5.4).

Table 5.3 displays the variance components and heritability estimates for seven feeding behavior traits, three growth traits, eleven feed efficiency traits and four off-test traits derived from three different single-trait animal models with pedigree information. With the simplest models (Model 1), the estimates of heritability for all traits of interests were moderate to high: feeding behavior traits ranged from 0.18 ± 0.03 for ADFI to 0.68 ± 0.02 for ANVD; estimate of heritability for growth trait during testing period were 0.35 ± 0.02 (ADG_FIRE) and slightly higher for growth rates of life time (ADG_Life, 0.48 ± 0.02) and of post-weaning (ADG_PostWean, 0.47 ± 0.03); measures of feed efficiency traits ranged from 0.11 ± 0.02 for FCR2 to 0.53 ± 0.02 for RIG; heritability estimate for IMF was 0.11 ±

0.02, the least heritable of the off-test traits while BF was 0.56 ± 0.04 the highest heritable off-test traits. Compared to results derived from Model 1, estimates of additive genetic variance for all the traits in Model 2 were slightly reduced except ADFI, ADFR and BW_Off; the magnitudes of estimates of the additional variance components capturing spatial Pen_Room efffects were smaller when compared with additive genetic variance, leading to slightly decreased estimates of heritability for most of traits and the ratio of Pen_Room variance against total phenotypic variance ranges from 0.01 to 0.18. When the additional social common pen effects was fitted to Model 3 for traits investigated, it was likely that the variances associated with the social common pen effect were pulled from residual variances in the reduced model (Model 1), as the estimates of additive genetic variance were slightly reduced or increased in Model 3 compared to the ones in Model 1(the reduced model). The estimates of social common pen variances in Model 3 were obviously larger when compared to the ones of spatial variances in Model 2, which can be reflected by the variance ratios. The estimates of heritability were reduced substantially for all feeding behavior traits, growth rate on test (ADG_FIRE) and feed efficiency traits in Model 3, however, estimate of heritability remained moderate to high for feeding behavior traits, growth traits, two feed efficiency traits (RG and RIG) and off-test traits except IMF.

Table 5.4 displays the estimates of additive genetic correlation between various feed efficiency measures and other production traits, including growth and off-test traits using the simplest two-trait animal models (Model 1 described above). Three measures of FCR were positively correlated with growth traits with moderate genetic correlations, ranging from

0.17 ± 0.05 for FCR1 with ADG_PostWean, to 0.35 ± 0.07 for FCR1 and ADG_Life. Unsurprisingly, measures of RFI were not genetic independent with measures of growth rate, and weak to moderate genetic correlations were found between measures of RFI and the three growth rate traits, except RFI5 and ADG_FIRE (-0.19 ± 0.14). Also week negative genetic correlations were found between measures of RFI and BF, which imply that selection for RFI may lead to an undesirable increase in BF. Favorable genetic correlations were revealed between RG and growth measures, from 0.30 ± 0.05 to 0.45 ± 0.08. However, the weak positive correlation genetically between RG and BF (0.21 ± 0.11) was unfavorable. Similar trends were found in RIG.

Table 5.5 shows the estimates of additive genetic correlation between feeding behavior traits and all other traits analyzed. Strong to moderate positive or negative genetic correlation were found among feeding behavior traits: strong positive genetic correlations were found between ADFI and other behavior traits (ranging from 0.42 ± 0.04 for ADFI and AOTV to 0.83 ± 0.12 for ADFI and FIV), between AOTD with other behavior traits, ranging from 0.31 to 0.75, except with AFIV (0.08 ± 0.02), between ADFI and AFIV (0.63 ± 0.30) and between AFIV and AOTV (0.89 ± 0.01) as well as between AFIV and FRV (0.63 ± 0.02); high negative genetic correlation were found between ANVD and AFIV (-0.78 ± 0.06) and between ANVD and AOTV (-0.72 ± 0.03). Except for AOTD, all behavior traits showed high to moderate genetic correlations with growth traits, ranging from -0.28 ± 0.03 for ANVD and ADG_FIRE to approximately 0.80 for ADFI and growth traits. The additive genetic correlations between behavior traits and RFI6 were found to be positive from 0.03 ± 0.01 (AOTV and RFI6) to 0.88 ± 0.10 (ADFI and RFI6). The genetic correlation between

behavior traits and other RFI measures were not provided in Table 5.4 since high similarity in genetic correlation between behavior traits and RFI measures. Similarly, FCR3 were omitted in Table 5.4 and removed from later analysis as the same trend with FCR2 (FCR2 and FCR3 were highly correlated phenotypically and genetically and may be seen as same trait). In contrast, the additive genetic correlation between feeding behavior traits and other feed efficiency measures were found moderately to highly negative. There is no clear trend found in genetic correlation between feeding behavior and off-test traits, except for the trivial genetic correlation between Muscle and the behavior measures.

Table 5.6 displays the heritability and variance ratio estimates for feeding behavior, growth, feed efficiency and off-test traits obtained by incorporating marker information in Model 3. The model included the social common pen effect as additional random effect. Compared with the estimates in Table 3, fractional changes were found in additive or social common pen variance components or in the estimates of heritability or variance ratio.

Table 5.7 represents the accuracies of prediction for the 506 boars in the validation dataset using traditional BLUP or Single-step genomic BLUP based on 16077 animals in training dataset and 8195 animals with marker information. The accuracies of EBVs or GEBVs were defined as the mean of accuracy for boars in testing dataset and SD were also provided to represent the variation in accuracy for estimated EBVs or GEBVs. Increased accuracies in GEBVs over EBVs were found for all traits investigated, but with different magnitude: the largest improvement was found in single-step genomic prediction (0.66) over the traditional BLUP (0.35) for BF with the increased accuracy 0.31 while the least increase

(approximately 0.06) in prediction accuracy using single-step over BLUP were observed for AFIV, AFRV and Muscle.

DISCUSSIONS

Measures of feed efficiency

In the present study, eleven feed efficiency measures were evaluated, bringing new understanding into the genetics underlying different measures of efficiency of feed utilization (Table 5.3 and Table 5.6) and the interrelationship of those various measures with other production traits, such as growth and off-test traits (Table 5.4). Different measures of feed efficiency have been used in the present analysis for several reasons. First, feed efficiency is not a directly measurable single trait, but must be calculated from its component traits and it includes all traits associated with the efficiency of feed utilization, such as FCR or its reciprocal (feed : gain ratio), RFI (Koch et al., 1963), RG and RIG (Crowley et al., 2010). The two measures FCR and RFI have been commonly reported in literatures, but RG and RIG were seldom used as efficiency measures in swine research (MacNeil and Kanp, 2015) rather in poultry (Willems et al., 2013) or in beef cattle (Crowley et al., 2010). Second, RFI is commonly used as a measure of feed efficiency at a given level of production, therefore the predicted feed intake can be defined by different sets of production traits, as suggested by several authors (Johnson et al., 1999; Hoque et al., 2009; Do et al., 2013). However, only two or three different models were used to define residual feed intake in those analysis (Johnson et al., 1999; Hoque et al., 2009; Do et al., 2013) instead of all possible models with all combination of production traits in the present study (Table 5.1). Third and finally, the

interrelationships between feed efficiency traits with other traits may not be consistent across different stages of maturity (Archer et al., 2002) or different populations or breeds (Do et al., 2013). Assessing various measures of feed efficiency in one large Duroc purebred population may shed lights on how to choose a particular useful feed efficiency measure in the future selection program.

Heritabilities of the various feed efficiency measures have been investigated in the present study using three different animal models with different random effects modeled. The estimates for the two commonly used efficiency measures FCR and RFI (except two of the FCR meatuses, FCR2 and FCR3) were higher than the literature averages (Rothschild and Ruvinsky, 2010; Hoque and Suzuki, 2009) but only slightly higher than the right bound when using the simplest animal model with animal effect as the only random effect. Fitting common litter effect was suggested by several authors (Hoque et al., 2008; Hoque et al., 2009; Johnson et al., 1999) to further partition the additive genetic variance for FCR and RFI, but only explained less than 5% variation in those measures in preliminary analysis thus were removed as suggested by MacNeil and Kanp (2015). When adding Pen_Room effect as random in the simplest animal model, the additive variance and heritability estimates were reduced slightly and approximately 10 to 17% of the variation was found to be due to spatial Pen_Room effects for FCR and RFI. A smaller proportion of variation (roughly 3%) was reported by Do et al. (2013) for measures of FCR and RFI, resulting from ‘pen’ effects but with a vague definition of ‘pen’. In their paper is not clear whether the ‘pen’ is referring to the common space shared by all pen mates (social common pen defined in our analysis) or the physical space (pen) shared by different groups of pigs longitudinally (Pen_Room

defined in the present study). In our work, when the social common pen effect was fitted as random effect, a dramatic decrease in the additive genetic variances was observed for all measures of FCR and RFI, compared to the estimates derived from Model 1. Consequently, the heritability estimates dropped rapidly for all the measures of FCR and RFI ranging from 0.3 to 0.6, while the social common pen accounted for as much as 70% of the total variance for each of the FCR and RFI measure. Similar results have not been reported in previous studies for feed efficiency traits in pigs due to the fact that the social common pen effects were not fitted into the linear mixed models when evaluating those traits. However, large proportions of variation due to social group effects have been reported for feed intake and growth rate in pigs with estimates of 27.5 % and 42.2 % of total variance expressed relative to the phenotypic variance (Bergsma et al., 2008). Similar results were found for feeding behavior traits (Chen et al., 2010).

Heritability estimates for RG (0.46 ± 0.04) was similar to FCR when Model 1 was fitted, however, decreased slightly when spatial Pen_Room (0.45 ± 0.03) or social common pen effect (0.38 ± 0.04) was added. MacNeil and Kanp (2015) reported a lower estimate for RG (0.21) with an animal model similar to our Model 1 in a Canadian Duroc population (n = 3291). However, it is difficult to compare the results fairly since standard error were not reported in their study and a different set of production traits were used as covariates to obtain RG in their study. Although seldom reported in pigs, the estimates of RG have been reported in other species, such as in Crowley et al. (2010) for beef cattle (0.28 ± 0.04) and Willems et al. (2013) for turkey (0.19 ± 0.04). Similar to RG, RIG was less affected by the model choice and the estimates of heritability were 0.18 ± 0.01 when accounting social

interaction among pen mates in the present analysis. Few studies have been performed on RIG in pigs, but estimates were found in beef cattle (0.30 ± 0.06) and turkey (0.23 ± 0.03) as reported by Berry and Crowley (2012) and Willems et al. (2013).

Additive genetic correlations between various measures of feed efficiency with other trait, especially growth and off-test traits in the present study may provide new insight into the interrelationships between those measures and other economically important traits (See Table 5.4). The weak but positive genetic correlation between measures of FCR and growth and BF indicates that selection for FCR may lead to small favorable indirect response for growth as well as the undesirable increase in BF. Those results differ from previously reported estimates (Jiao et al., 2014) with data collected in a similar population. This may due to the smaller number of animals included in the previous analysis. However, our estimates agree well with the findings in other studies. Do et al. (2013) reported similar genetic correlation between FCR and growth as well as FCR and BF in Danish Duroc and Yorkshire population but not in Landrace. Similar results were also observed from Hoque and Suzuki (2008) in Duroc and Landrace population. Measures of RFI were not genetically independent with growth (weak positive correlation) or backfat (weak negative genetic correlation) in the present population, consistent with results found in Johnson et al. (1999), Do et al. (2013) as well as Mrode and Kennedy (1993). However, Hoque et al. (2009) reported lower estimates of genetic correlation for BF and one measure of RFI (accounted for both growth and backfat) but significant higher estimates for BF and another measure of RFI (only accounted for growth). Given the genetic correlation in this Duroc population, selection against RFI may lead to slower growing pigs with thicker backfat. The moderate genetic

correlation found between RG and growth as well as weak a correlation between RG and BF indicates that RG might be a good feed efficiency measure. Selection for RG would result in rapidly growing pigs with only a negligible increase in backfat. The selection for RIG may achieve similar goal given the similar genetic correlations to RG.

The genetics of different measures of feed efficiency and interrelationships with other important economic traits were explored in the present study, providing new information and new understanding of the various measures. However, it is not straightforward to answer the question to which is the best measure to choose to improve feed efficiency. The commonly used measure of feed efficiency, such as FCR and RFI have inherent flaws. For example, the problem with FCR (or Feed : gain ratio) is that 1) its close correlation with both feed intake and rate of gain (Carstens et al., 2003); and 2) it may lead to animals with heavier mature weights and greater maintenance requirements and animals with similar FCR may differ greatly in their rate of gain and feed intake (Smith et al., 2010); 3) selection based on ratio traits presents problems related to the prediction of change in component traits in future generations (Gunsett, 1984). In contrast, RFI is independent of level of production, such as size, growth or backfat thickness, but it is still genetically correlated with the production traits (Kennedy et al., 1993); Moreover, it may lack acceptance by producers because slow growing animals eating relatively less of feed may actually have good RFI. Crowley et al. (2010) and Berry and Crowley (2012) argued that RG and RIG may be better measures of feed efficiency because improved RG is associated with faster growth rate on average, given same level of feed intake and RIG is combining RFI and RG with both advantages. It is indicated that RG and RIG might be two good indicator traits of efficiency of feed utilization

in pigs (based on its heritability and genetic correlation with other production traits) but the underlying genomic basis to variation in the trait and the physiological process involved in those traits remain to be determined.

Social interaction by modeling pen sharing by pen mates

Social interaction appear to be a significant source of variation for traits that were measured by electronic feeding systems, such as feeding behavior, growth (ADG_FIRE) and feed efficiency traits (Table 5.3). Only a handful of results have been reported on this topic but similar outcomes were reported in pigs by Bergsma et al. (2008), Chen et al. (2010) and Arango et al. (2005). Our study suggests that social interaction among pen mates exists and arise form the fact that only a single-space feeder is placed in the pen shared by a groups of pigs. It might be possible that pen mates have to compete to get the chance to access feed. Thus, social interaction should not be removed from those models. Bijma et al. (2007) showed that a small non-genetic covariance among pen mates may substantially bias the genetic parameter estimation. Conversely, the inclusion of social interaction, especially the variation due to social interaction, could improve predictions of breeding values for direct genetic effects, as suggested by Hsu et al. (2010).

In the present study, instead of modeling the heritable social effect for each animal, a social common pen effect was modeled as nonheritable social effect as an alternative equivalent but simpler way to fit correlated residuals within pens (Bergsma et al., 2008; Bijma et al., 2007). However, it would be possible to fit social genetic effects (indirect genetic effects) due to (n – 1) group members into the traditional animal model with additive genetic effect for a certain individual. It is suggested that one advantage of modeling

heritable social effects is that heritable social effects can be utilized to increase response in their selection (Ellen et al., 2010; Muir, 2005). However, modeling heritable social effects in this dataset was not possible as models presented convergence problems. It remains though an interesting problem that should be investigated further.

Feeding behavior traits

Limited research has been done to investigate feeding behavior traits in largely phenotyped population. In the current analysis, feeding behavior traits were moderate to highly heritable (Table 5.3 and Table 5.5) in agreement with Do et al. (2013), Rohrer et al. (2013) and Labraue et al. (1997). Similar to previous reported interrelationships with other traits, our results showed that some of the feeding behavior measures were strongly genetically correlated with growth and off-test traits, especially those highly correlated with feed efficiency traits, such as AOTD and ANVD. Based on results of this study, it would be possible to use feeding behavior measures as selection criterion to improve feed efficiency traits since some feeding behavior traits can be recorded without measuring individual feed intake (Maselyne et al., 2015; Maselyne et al., 2014) with less equipment and will be less costly and easier to maintain (Brown-Brandl et al., 2013). Furthermore, the availability of marker information might provide more latitude to investigate the underlying genetic architectures of feeding behavior traits and feed efficiency or possible pleiotropic genomic regions explaining the strong genetic correlation between some of the feeding behavior and feed efficiency measures.

Accuracy of genomic prediction

Single-step GBLUP (or ssGBLUP) has been used widely in different species for routinely genetic evaluations (Misztal et al., 2013) with its advantages being its easy implementation and the use of standard BLUP routines with genotyped or non genotyped animals. In the current analysis, validation of prediction accuracy were conduced by splitting data into training and testing and the result showed mean accuracies for animals in the testing set were higher for GEBVs from ssGBLUP than EBVs from BLUP for all traits investigated. The increased accuracies of predictions by using ssGBLUP incorporated with genotypes over pedigree-based BLUP were also reported by Christensen et al. (2012) and Forni et al. (2011) in analyses using swine data. Due to the increased accuracy in prediction, Denmark has routine genetic evaluation has been made by using ssGBLUP since 2011 (Legarra et al., 2014). However, the increase of prediction accuracies due to additional genomic information used in ssGBLUP although is encouraging, the use of this genomic selection approach need to be accompanied with validations for further implementation as suggested by Misztal et al. (2013).

CONCLUSIONS

Nonheritable social interaction has been observed for traits associated with measures recorded by electronic feeding system (FIRE) and it is suggested that there is a need to include the social effects to reduce bias for genetic parameter estimation when the variance explained by social interaction has been found significant. After accounting for social interaction, RG and RIG have been found as two good measures of feed efficiency due to moderate heritability and genetic correlation with other economically important traits, such

as growth and off-test traits. Feeding behavior traits were found moderately heritable and some of the behavior measures were highly correlated with feed efficiency traits, which is worth further investigation. Increased accuracies have been shown when apply ssGBLUP over pedigree-based BLUP for feeding behavior, feed efficiency, growth and off-test traits in a validation setting. Further research is needed to address the questions that which feed efficiency measure is useful and should be used as selection criterion to improve feed efficiency in pigs and whether it is appropriate to select feeding behavior traits for correlated response in feed efficiency given the underlying genetic architectures of those traits.

LITERATURE CITED

Arango J., I. Misztal, S. Tsuruta, M. Culbertson, and W. Herring. 2005. Estimation of variance components including competitive effects of Large White growing gilts. J. Anim. Sci. 83:1241-1246.

Arthur, P. F., and R. M. Herd. 2005. Efficiency of feed utilisation by livestock-Implications and benefits of genetic improvement. Can. J. Anim. Sci. 85(3):281-290.

Archer, J. A., A. Reverter, R. M. Herd, D. J. Johnston, and P. F. Arthur. 2002. Genetic variation in feed intake and efficiency of mature beef cows and relationships with postweaning measurements. Proc. 7th Wld. Congr. Genet. Appl. Livest. Prod. 31:221–224.

Bergsma, R., E. Kanis, E. F. Knol, P. Bijma. 2008. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genet.178:1559-1570.

Bijma P, W. M. Muir, E. D. Ellen, J. B. Wolf, J. A. M. Van Arendonk. 2007. Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. Genet. 175:289-299.

Berry, D. P., J. J. Crowley JJ. 2009. Residual intake and body weight gain: a new measure of efficiency in growing cattle. J Anim. Sci. 90:109-115.

Brown-Brandl, T.M., G. A. Rohrer, and R. A. Eigenberg. 2013. Analysis of feeding behavior of group housed growing-finishing pigs. Comput. Electron. Agric. 96:246-252.

Browning, B. L., and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84(2):210-223.

Casey, D. S., H. S. Stern, and J. C. M. Dekkers. 2005. Identification of errors and factors associated with errors in data from electronic swine feeders. J. Anim. Sci. 83:969-982.

Chen, C. Y., I. Misztal, S. Tsuruta, W. O. Herring, J. Holl, and M. Culbertson. 2010. Influence of heritable social status on daily gain and feeding pattern in pigs. J. Anim. Breed. Genet. 127(2):107-112.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Gen. Sel. Evol. 42:2.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal. 6(10):1565-1571.

Crowley, J. J., M. McGee, D. A. Kenny, D. H. Crews, R. D. Evans, and D. P. Berry. 2010. Phenotypic and genetic parameters for different measures of feed efficiency in different breeds of Irish performance-tested beef bulls. J. Anim. Sci. 88(3):885-894.

Do, D. N., A. B. Strathe, J. Jensen, T. Mark, and H. N. Kadarmideen. 2013. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. J. Anim. Sci. 91(9):4069-4079.

Ellen, E. D., V. Ducrocq, B. J. Ducro, R. F. Veerkamp, and P. Bijma. 2010. Genetic parameters for social effects on survival in cannibalistic layers: combining survival analysis and a linear animal model. Genet. Sel. Evol. 42:27.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol. 43:1.

Gunsett, F.C. 1984. Linear index selection to improve traits defined as ratios. J. Anim. Sci. 59:1185-1193.

Hoque, M. A., H. Kadowaki, T. Shibata, T. Oikawa, and K. Suzuki. Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livest. Sci. 121(1): 45-49.

Hoque, M. A. and K. Suzuki. 2008. Genetic parameters for production traits and measures of residual feed intake in Duroc and Landrace pigs. Anim. Sci. J. 79:543-549.

Hoque, M. A., and K. Suzuki. 2009. Genetics of residual feed intake in cattle and pigs: a review. Asian-Australas. J. Anim. Sci. 22:747-755.

Hsu, W. L., K. J. Rodger, and L. D. Van Vleck. 2010. Effect of pen mates on growth, backfat depth, and longissimus muscle area of swine. J. Anim. Sci. 88(3):895-902.

Jiao, S., C. Maltecca, K. A. Gray, and J. P. Cassady. 2014. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. J. Anim. Sci. 92 (6):2377-2386.

Jiao, S., F. Tiezzi, Y. Huang, K. A. Gray, and C. Maltecca. 2015. The Use of Multiple Imputation for the accurate measurements of individual feed intake by electronic feeders. J. Anim. Sci. (submitted).

Johnson, Z. B., J. J. Chewning and R. A. Nugent. 1999. Genetic parameter for production traits and measures of residual feed intake in Large White swine. J. Anim. Sci. 77:1679-1685.

Kennedy, B.W., J. H. S. van der Werf, T. H. E. Meuwissen.1993. Genetic and statistical properties of residual feed intake. J. Anim. Sci. 71:3239–3250.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486–494.

Labroue, F., R. Gueblez, and P. Sellier. 1997. Genetic parameters of feeding behaviour and performance traits in group-housed Large White and French Landrace growing pigs. Genet. Sel. Evol. 29:451-468.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656-4663.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single step, a general approach for genomic selection. Liv. Sci. 166:54-65.

MacNeil, M., and R. A. Kemp. 2015. Genetic parameter estimation and evaluation of Duroc boars for feed efficiency and component traits. Can. J. Anim. Sci. 95(2):155-159.

Maselyne, J., W. Saeys, and A. Van Nuffel. 2015. Review: Quantifying animal feeding behaviour with a focus on pigs. Physiol. Behav. 138:37-51.

Maselyne, J., W. Saeys, B. De Ketelaere, K. Mertens, J. Vangeyte, E. F. Hessel, B. Sonck, and W. Saeys. 2014. Validation of a High Frequency Radio Frequency Identification (HF RFID) system for registering feeding patterns of growing-finishing pigs. Comput. Electron. Agric.108:209-220.

Misztal, I., A. E. Aggrey, and W. M. Muir. 2013. Experiences with a single-step genome evaluation1. Poul. Sci. 92(9):2530-2534.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). Communication No. 28-07 in Proc. 7th World Cong. Genet. Appl. Livest. Prod., Montpellier, France.

Mrode, R. A., and B. W. Kennedy. 1993. Genetic variation in measures of food efficiency in pigs and their genetic relationships with growth rate and backfat. Anim. Prod. 56:225-232.

Muir, W.M. 2005. Incorporation of competitive effects in forest tree or animal breeding programs. Genet.170:1247-1259.

Nguyen, N. H., C. P. McPhee, and C. M. Wade. 2005. Responses in residual feed intake in lines of Large White pigs selected for growth rate on restricted feeding (measured on ad libitum individual feeding). J. Anim. Breed. Genet. 122(4):264-270.

Rohrer, G. A., T. Brown-Brandl, L. A. Rempel, J. F. Schneider, and J. Holl. 2013. Genetic analysis of behavior traits in swine production. Livest. Sci. 157(1):28-37.

Rothschild, M. F., and A. Ruvinsky. 2010. The genetics of the pig. 2nd ed. CABI, Cambridge, MA.

Smith, S. N., M. E. Davis, and S. C. Loerch. 2010. Residual feed intake of Angus beef cattle divergently selected for feed conversion ratio. Livest. Sci. 132(1):41-47.

Tsuruta, S. 2015. Approximate variance for heritability estimates. http://nce.ads.uga.edu/html/projects/AI_SE_revised.pdf. Accessed April 30, 2015.

Willems, O. W., S. P. Miller, and B. J. Wood. 2013. Assessment of residual body weight gain and residual intake and body weight gain as feed efficiency traits in the turkey (Meleagris gallopavo). Genet. Sel. Evol. 45(1):26.

Zumbach, B., I. Misztal, C. Y. Chen, S. Tsuruta, M. Łukaszewicz, W. O. Herring, and M. Culbertson. 2010. Use of serial pig body weights for genetic evaluation of daily gain. J. Anim. Breed. Genet. 127:93–99.

Table 5.1 Different measures (1-6) of residual feed intake (RFI), in g.

Trait	Model[1]	b_1 (s.e.)	b_2 (s.e.)	b_3 (s.e.)	b_4 (s.e.)
RFI1	$CG_j + b_1 ADG_i + e_i$	2.74 (0.09)**[2]			
RFI2	$CG_j + b_1 ADG_i + b_2 MMW_i + e_i$	2.21 (0.09)**	4.87 (0.30) **		
RFI3	$CGj + b_1 ADG_i + b_3 BF + e_i$	1.97 (0.11)**		35.75 (2.77)**	
RFI4	$CGj + b_1 ADG_i + b_2 MMW_i + b_3 BF_i + e_i$	1.54 (0.11) **	4.57 (0.30)**	32.24 (2.73)**	
RFI5	$CGj + b_1 ADG_i + b_3 BFi + b_4 Muscle_i + ei$	1.93 (0.11) **		36.13 (2.79)**	1.35 (1.23)
RFI6	$CG_j + b_1 ADG_i + b_2 MMW_i + b_3 BF_i + b_4 Muscle_i + e_i$	1.49 (0.11) **	4.57 (0.30)**	32.68 (2.75)**	1.51 (1.21)

[1] The differences of the models were the left hand sides with different production traits such as ADG (average daily gain during testing period, ADG_FIRE), MMW (metabolic mid body weight), BF (ultrasound backfat thickness) and Muscle (muscle depth) as covariates for right hand side ADFI (average feed intake). CG_j representing the $_j$th contemporary group. The regression coefficients b_1, b_2, b_3 and b_4 were estimated for ADG (g), MMW (kg), BF (mm) and Muscle (mm), respectively.

[2] ** representing p-value of t-test for regression coefficient < 0.01.

Table 5.2 Descriptive statistics for feeding behavior measures, growth, feed efficiency and off-test traits.

Category	Trait[1]	No. of Observation	Min	Mean	Max	SD
Behavior	ADFI, g	11798	78.68	2154.60	8250.63	609.84
	AOTD, s	11798	155.00	3715.00	11494.00	840.24
	ANVD	11798	1.19	5.77	20.86	1.78
	ADFR, g/min	11798	1.58	36.72	288.57	12.64
	AFIV, g	11798	21.98	471.62	1259.73	181.66
	AOTV, s	11798	48.92	778.39	2025.41	233.22
	AFRV, g/min	11798	1.58	36.72	288.86	12.65
Growth	ADG_Life, g/d	15221	307.00	656.80	976.00	76.32
	ADG_FIRE, g/d	6500	351.00	886.10	1450.00	222.87
	ADG_PostWean, g/d	14254	344.40	733.80	1115.20	86.99
Efficiency	FCR1, 100%	6485	7.60	41.86	506.80	16.04
	FCR2, 100%	10943	8.90	32.02	678.90	20.17
	FCR3, 100%	10502	9.80	35.92	762.80	23.23
	RFI1, g	6464	0.00	-2207.74	5644.27	459.35
	RFI2, g	6464	0.00	-2225.27	5644.75	456.52
	RFI3, g	6464	0.00	-2172.20	5778.53	453.42
	RFI4, g	6464	0.00	-2189.72	5777.60	450.68
	RFI5, g	6464	0.00	-2169.33	5776.80	453.38
	RFI6, g	6464	0.00	-2187.05	5776.00	450.65
	RG, g	6464	0.00	-191.80	134.10	40.56
	RIG, g	6464	0.00	-6.23	12.97	1.49
Off-test	BW_Off, kg	15209	68.04	114.58	168.74	12.94
	BF, mm	15218	4.32	11.03	25.23	5.68
	Muscle, mm	15216	23.37	42.19	70.70	2.81
	IMF, 100%	11351	1.41	3.64	7.01	0.49

[1] ADFI, average daily feed intake, in g. AOTD, average occupation time, in s. ANVD, average number of visits. ADFR, average daily feeding rate, in g/min. AFIV, average feed intake per visit, in g. AOTV, average occupation time per visit across testing period, in s. AFRV, average feeding rate per visit across testing period, in g/min. ADG_FIRE, average daily gain (ADG) using FIRE (Feed Intake recording equipment), in g/d. ADG_Life, life time ADG, in g/d. ADG_PostWean, post-weaning ADG, in g/d. FCR, feed conversion ratio, daily gain/daily feed intake, unit in 100%. RFI, residual feed intake, in g. RG, residual growth, in g. RIG, difference of residual growth and residual feed intake, unit in g. BW_Off, body weight at off-test, in kg. BF, ultrasound backfat at off-test, in mm. Muscle, ultrasound muscle depth, in mm. IMF, intramuscular fat percentage, in 100%.

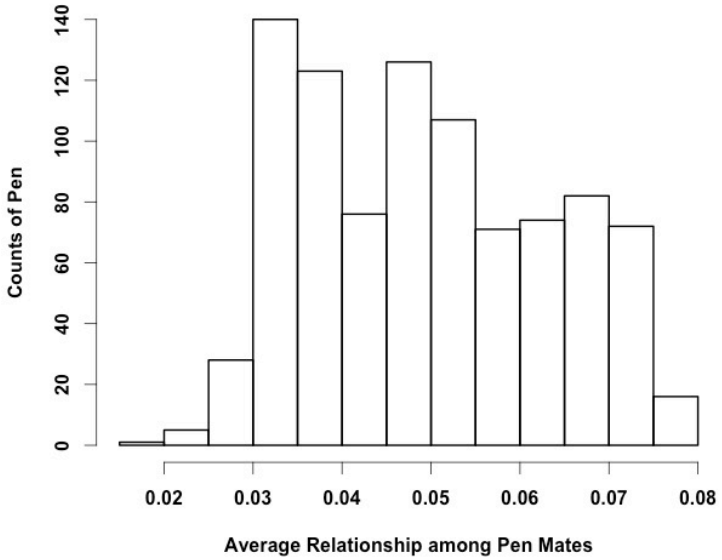

Figure 5.1 The average relationship among pigs sharing the same pen as pen mates. The pedigree for pigs with phenotypes was traced back to 3 generations to construct the numerator relationship matrix (A). The average relationship for each pen was computed by taking the mean of the off-diagonal elements of the A matrix for pigs living in the same pen.

Table 5.3 Estimates of variance components[1] and heritability (h^2) for behavior, growth, efficiency and off-test traits with pedigree information. Variance components include additive genetic variance ($\boldsymbol{\sigma_a^2}$), spatial Pen_Room effect ($\boldsymbol{\sigma_p^2}$), social common pen effect ($\boldsymbol{\sigma_s^2}$) and phenotype variance ($\boldsymbol{\sigma_{phe}^2}$).

Category	Trait[2]	Model 1[3]		Model 2[4]				Model 3[5]			
		σ_a^2	h^2 (SE)	σ_a^2	σ_p^2	h^2 (SE)	σ_p^2/σ_{ph}^2 [6]	σ_a^2	σ_s^2	h^2 (SE)	σ_s^2/σ_{ph}^2 [7]
Behavior	ADFI	4.30E+04	0.18 (0.03)	4.75E+04	2.72E+04	0.19 (0.03)	0.11	4.53E+04	1.91E+05	0.14 (0.01)	0.59
	AOTD	3.53E+05	0.71 (0.03)	2.27E+05	5.74E+03	0.50 (0.04)	0.01	3.22E+05	9.27E+04	0.52 (0.04)	0.15
	ANVD	1.46E+00	0.68 (0.02)	9.84E-01	7.69E-02	0.42 (0.03)	0.03	8.83E-01	4.52E-01	0.36 (0.03)	0.18
	ADFR	8.99E+01	0.35 (0.03)	9.09E+01	3.23E+01	0.35 (0.01)	0.13	6.58E+01	1.69E+02	0.19 (0.03)	0.48
	AFIV	1.10E+04	0.49 (0.01)	1.06E+04	1.88E+03	0.45 (0.02)	0.08	8.65E+03	6.86E+03	0.37 (0.03)	0.29
	AOTV	2.75E+04	0.58 (0.04)	2.66E+04	1.56E+03	0.57 (0.04)	0.03	2.52E+04	7.59E+03	0.51 (0.04)	0.15
	AFRV	2.25E+04	0.48 (0.03)	2.06E+04	5.49E+03	0.43 (0.05)	0.11	9.98E+03	3.06E+04	0.18 (0.03)	0.56
Growth	ADG_FIRE	1.35E+04	0.35 (0.03)	1.35E+04	7.82E+03	0.31 (0.01)	0.18	1.22E+04	2.91E+04	0.23 (0.01)	0.55
	ADG_Life	2.38E+03	0.48 (0.02)	2.36E+03	8.00E+01	0.47 (0.02)	0.02	2.23E+03	3.84E+02	0.44 (0.01)	0.08
	ADG_PostWean	2.94E+03	0.47 (0.03)	2.91E+03	1.36E+02	0.46 (0.04)	0.02	2.73E+03	5.31E+02	0.42 (0.03)	0.08
Efficiency	FCR1	8.55E+01	0.44 (0.04)	8.12E+01	3.02E+01	0.39 (0.03)	0.14	1.63E+01	1.79E+02	0.06 (0.04)	0.69
	FCR2	3.56E+01	0.11 (0.02)	3.74E+01	3.46E+01	0.11 (0.01)	0.10	1.21E+01	2.52E+02	0.03 (0.04)	0.62
	FCR3	4.82E+01	0.11 (0.03)	5.11E+01	4.81E+01	0.11 (0.02)	0.11	1.53E+01	3.27E+02	0.03 (0.02)	0.62
	RFI1	1.25E+05	0.54 (0.04)	1.09E+05	4.03E+04	0.45 (0.03)	0.16	2.35E+04	1.94E+05	0.08 (0.03)	0.69
	RFI2	1.28E+05	0.55 (0.03)	1.15E+05	3.86E+04	0.47 (0.04)	0.16	1.67E+04	2.10E+05	0.06 (0.04)	0.71
	RFI3	1.21E+05	0.53 (0.04)	1.06E+05	3.94E+04	0.44 (0.02)	0.17	1.57E+04	1.99E+05	0.06 (0.03)	0.71
	RFI4	1.23E+05	0.54 (0.04)	1.12E+05	3.85E+04	0.47 (0.03)	0.16	1.12E+04	2.17E+05	0.04 (0.04)	0.73
	RFI5	1.20E+05	0.53 (0.03)	1.06E+05	3.93E+04	0.44 (0.04)	0.16	1.60E+04	1.99E+05	0.06 (0.04)	0.71
	RFI6	1.23E+05	0.54 (0.03)	1.11E+05	3.84E+04	0.46 (0.05)	0.16	1.15E+04	2.16E+05	0.04 (0.03)	0.73
	RG	1.15E+03	0.46 (0.04)	1.13E+03	3.79E+01	0.45 (0.03)	0.02	9.37E+02	2.60E+02	0.38 (0.04)	0.11
	RIG	1.49E+00	0.57 (0.03)	1.42E+00	2.62E-01	0.53 (0.02)	0.10	4.97E-01	1.35E+00	0.18 (0.01)	0.48

Table 5.3 Continued

Off-test	BW_Off	7.99E+00	0.37 (0.02)	8.00E+00	2.43E-01	0.37 (0.01)	0.01	6.87E+00	2.49E+00	0.31 (0.01)	0.11
	BF	2.64E+00	0.56 (0.04)	2.64E+00	7.82E-02	0.55 (0.03)	0.02	2.56E+00	2.41E-01	0.53 (0.03)	0.05
	Muscle	6.88E-02	0.41 (0.03)	6.86E-02	1.46E-03	0.41 (0.03)	0.01	6.54E-02	1.01E-02	0.38 (0.04)	0.06
	IMF	3.56E-03	0.11 (0.02)	3.78E-03	3.46E-04	0.11 (0.01)	0.01	4.58E-04	2.80E-02	0.01 (0.03)	0.69

[1] Variance components include additive genetic variance (σ_a^2), spatial Pen_Room effect (σ_p^2), social common pen effect (σ_s^2) and phenotype variance (σ_{phe}^2).

[2] Traits analyzed: ADFI, average daily feed intake, in g. AOTD, average occupation time, in s. ANVD, average number of visits to feeder. ADFR, average daily feeding rate, in g/min. FIV, average feed intake per visit across testing period, in g. OTV, average occupation time per visit across testing period, in s. FRV, average feeding rate per visit across testing period, in g/min. ADG_Life, average daily gain from birth to off-test period, in g/d. ADG_FIRE, average daily gain across testing period using FIRE (Feed Intake recording equipment), in g/d. ADG_PostWean, average daily gain from post-weaning to off-test, in g/d. FCR1, FCR2 and FCR3 are three measures of feed conversion ratio, daily gain/daily feed intake, unit in 100%. RFI1 to RFI6 are six measures of residual feed intake, in g. RG, residual growth, in g. RIG, difference of residual growth and residual feed intake, unit in g. BW_Off, body weight at off-test, in kg. BF, ultrasound backfat thickness at off-test, in mm. Muscle, ultrasound muscle depth at off-test, in mm. IMF, intramuscular fat percentage, in 100%.

[3] Model 1 $\mathbf{y} = X\mathbf{b} + Z\mathbf{a} + \mathbf{e}$, where additive effect $\mathbf{a} \sim N(\mathbf{0}, A\sigma_a^2)$, A is the numeric relationship matrix and residual is $\mathbf{e} \sim N(\mathbf{0}, I\sigma_e^2)$. fixed effects b include contempary groups (257 levels of concatenation of year and week), sow parity (1, 2, 3+) and age start testing and length of testing period as covariates.

[4] Model 2 $\mathbf{y} = X\mathbf{b} + Z_1\mathbf{a} + Z_2\mathbf{p} + \mathbf{e}$, where additive genetic effect a follows $\mathbf{a} \sim N(\mathbf{0}, A\sigma_a^2)$, A is the numeric relationship matrix, p is the room_pen (n = 251 levels) effect as $\mathbf{p} \sim N(\mathbf{0}, I\sigma_{Pen}^2)$ and $\mathbf{e} \sim N(\mathbf{0}, I\sigma_e^2)$. The room_pen represented the physical position of pens equipped with FIRE and may capture the special effects of the pen.

[5] Model 3 $\mathbf{y} = X\mathbf{b} + Z_1\mathbf{a} + Z_2\mathbf{s} + \mathbf{e}$, where additive genetic effect a follows $\mathbf{a} \sim N(\mathbf{0}, A\sigma_a^2)$, A is the numeric relationship matrix, s is social common pen (n = 921 levels) effect as $\mathbf{s} \sim N(\mathbf{0}, I\sigma_{Pen}^2)$ and $\mathbf{e} \sim N(\mathbf{0}, I\sigma_e^2)$. The pen group only included boars living together as pen mates.

[6] σ_p^2/σ_{ph}^2. Variance ratio for spatial Pen_Room effect against phenotype variance.

[7] σ_s^2/σ_{ph}^2. Variance ratio for social common pen effect against phenotype variance.

Table 5.4 Estimates of additive genetic correlation (SE)[1] from two-trait animal model between feed efficiency and other production traits with pedigree information.

Trait[2]	Feed Efficiency[3]										
	FCR1	FCR2	FCR3	RFI1	RFI2	RFI3	RFI4	RFI5	RFI6	RG	RIG
	0.29	0.19	0.21	0.13	0.21	0.19	0.20	0.11	0.15	0.30	0.46
ADG_FIRE	(0.13)	(0.04)	(0.04)	(0.07)	(0.15)	(0.05)	(0.13)	(0.04)	(0.08)	(0.05)	(0.02)
	0.17	0.34	0.35	0.22	0.24	0.19	0.24	0.17	0.12	0.45	0.38
ADG_Life	(0.05)	(0.07)	(0.11)	(0.05)	(0.07)	(0.05)	(0.02)	(0.03)	(0.04)	(0.08)	(0.03)
	0.17	0.32	0.30	0.19	0.24	0.10	0.12	0.19	0.16	0.44	0.38
ADG_PostWean	(0.03)	(0.06)	(0.07)	(0.03)	(0.06)	(0.04)	(0.05)	(0.06)	(0.04)	(0.03)	(0.04)
	0.18	0.22	0.24	0.17	0.15	0.18	-0.20	0.14	0.10	0.27	0.25
BW	(0.03)	(0.06)	(0.08)	(0.03)	(0.04)	(0.02)	(0.16)	(0.07)	(0.11)	(0.03)	(0.05)
	-0.17	-0.24	-0.24	-0.21	-0.27	-0.14	-0.11	-0.10	-0.14	0.21	0.19
BF	(0.09)	(0.11)	(0.07)	(0.05)	(0.13)	(0.02)	(0.03)	(0.02)	(0.03)	(0.11)	(0.07)
	0.22	0.13	0.13	0.22	0.13	0.11	0.20	-0.14	-0.10	0.26	0.31
Muscle	(0.13)	(0.03)	(0.04)	(0.06)	(0.04)	(0.04)	(0.08)	(0.09)	(0.08)	(0.04)	(0.14)
	0.12	0.09	0.11	0.09	0.11	0.08	0.14	0.13	0.21	0.13	-0.11
IMF	(0.07)	(0.09)	(0.07)	(0.03)	(0.04)	(0.05)	(0.05)	(0.10)	(0.09)	(0.09)	(0.03)

[1] Estimates of additive genetic correlation and their SE were provided (in parentheses).

[2] ADG_FIRE, average daily gain across testing period using FIRE (Feed Intake recording equipment), in g/d. ADG_Life, average daily gain from birth to off-test period, in g/d. ADG_PostWean, average daily gain from post-weaning to off-test, in g/d. BW_Off, body weight at off-test, in kg. BF, ultrasound backfat thickness at off-test, in mm. Muscle, ultrasound muscle depth at off-test, in mm. IMF, intramuscular fat percentage, in 100%.

[3] FCR1, FCR2 and FCR3, the measures of feed conversion ratio, daily gain/daily feed intake, unit in 100%. RFI1 to RFI6 are measures of residual feed intake, in g. RG, residual growth, in g. RIG, difference of residual growth and residual feed intake, unit in g.

Table 5.5 Estimates of additive genetic correlation[1] from two-trait animal model between behavior and other traits with pedigree information.

Category	Trait[3]	Feeding behavior[2]						
		ADFI	AOTD	ANVD	ADFR	AFIV	AOTV	AFRV
Behavior	AOTD	0.75 (0.04)						
	ANVD	0.47 (0.03)	0.77 (0.05)					
	ADFR	0.67 (0.02)	0.42 (0.03)	-0.28 (0.07)				
	AFIV	0.83 (0.12)	0.08 (0.02)	-0.78 (0.06)	0.63 (0.30)			
	AOTV	0.42 (0.04)	0.68 (0.02)	-0.72 (0.03)	-0.43 (0.06)	0.89 (0.01)		
	AFRV	0.71 (0.04)	0.31(0.03)	-0.33 (0.05)	NaN[4]	0.63 (0.02)	-0.43 (0.02)	
Growth	ADG_FIRE	0.82 (0.04)	0.11 (0.04)	-0.28 (0.03)	0.57 (0.09)	0.54 (0.06)	0.22 (0.05)	0.23 (0.04)
	ADG_Life	0.88 (0.05)	0.06 (0.02)	-0.56 (0.03)	0.60 (0.06)	0.78 (0.02)	0.42 (0.02)	0.60 (0.16)
	ADG_PostWean	0.87 (0.06)	0.06 (0.02)	-0.56 (0.05)	0.59 (0.09)	0.78 (0.02)	0.42 (0.02)	0.59 (0.19)
Efficiency	FCR1	-0.42 (0.03)	-0.65 (0.04)	-0.37 (0.03)	-0.48 (0.03)	-0.34 (0.03)	-0.10 (0.03)	-0.48 (0.03)
	FCR2	-0.89 (0.02)	-0.69 (0.06)	-0.37 (0.03)	-0.46 (0.04)	-0.34 (0.03)	-0.33 (0.04)	-0.46 (0.04)
	RFI6	0.88 (0.10)	0.72 (0.06)	0.23 (0.05)	0.78 (0.04)	0.57 (0.12)	0.03 (0.01)	0.78 (0.13)
	RG	-0.39 (0.03)	-0.29 (0.03)	-0.50 (0.03)	-0.41 (0.06)	-0.56 (0.03)	-0.35 (0.03)	-0.41 (0.16)
	RIG	-0.39 (0.06)	-0.61 (0.02)	-0.50 (0.05)	-0.52 (0.03)	-0.14 (0.03)	-0.20 (0.03)	-0.52 (0.03)
Off-test	BW	0.53 (0.03)	0.06 (0.02)	-0.50 (0.08)	0.58 (0.03)	0.10 (0.02)	0.42 (0.02)	0.60 (0.13)
	BF	0.61 (0.04)	0.17 (0.02)	-0.25 (0.08)	0.10 (0.02)	0.15 (0.03)	-0.02 (0.02)	0.12 (0.02)
	Muscle	-0.02 (0.05)	-0.08 (0.03)	-0.14 (0.10)	0.02 (0.02)	0.11 (0.03)	0.05 (0.03)	0.04 (0.03)
	IMF	0.34 (0.05)	0.10 (0.03)	0.01 (0.03)	-0.08 (0.02)	0.18 (0.03)	0.08 (0.03)	0.04 (0.03)

[1] Estimates of additive genetic correlation and their SE (in parentheses) were provided.

[2] Feeding behavior traits: ADFI, (average) daily feed intake, in g. AOTD, occupation time, in s. ANVD, number of visits. ADFR, daily feeding rate, in g/min. AFIV, feed intake per visit, in g. AOTV, occupation time per visit, in s. AFRV, feeding rate per visit, in g/min.

[3] ADG_FIRE, average daily gain (ADG) using FIRE, in g/d. ADG_Life, life time ADG, in g/d. ADG_PostWean, post-weaning ADG, in g/d. FCR, feed conversion ratio, daily gain/daily feed intake, unit in 100%. RFI, residual feed intake, in g. RG, residual growth, in g. RIG, difference of residual growth and residual feed intake, unit in g. BW_Off, body weight at off-test, in kg. BF, ultrasound backfat at off-test, in mm. Muscle, ultrasound muscle depth, in mm. IMF, intramuscular fat percentage, in 100%.

[4] NaN. Not available, the model was not converaged.

Table 5.6 Estimates of variance components and heritability (h^2) for behavior, growth, efficiency and off-test traits using single-trait animal model incorporated with pedigree and marker information.

Category	Trait[1]	$\sigma_a^{2\,2}$	$\sigma_S^{2\,3}$	h^2 (SE)	$\sigma_S^2/\sigma_{Ph}^{2\,4}$
Behavior	ADFI	49231.12	163781.30	0.14 (0.02)	0.51
	AOTD	344215.22	89743.65	0.52 (0.03)	0.16
	ANVD	0.91	0.54	0.36 (0.03)	0.20
	ADFR	69.43	171.00	0.19 (0.04)	0.44
	AFIV	9765.13	7215.57	0.37 (0.03)	0.28
	AOTV	27549.01	7416.46	0.51 (0.03)	0.18
	AFRV	10001.28	31080.88	0.18 (0.03)	0.54
Growth	ADG_FIRE	9876.21	30012.02	0.23 (0.03)	0.48
	ADG_Life	2407.50	394.67	0.44 (0.03)	0.08
	ADG_PostWean	2732.80	587.52	0.42 (0.03)	0.98
Efficiency	FCR1	15.92	191.23	0.06 (0.03)	0.53
	RFI6	9986.93	174510.11	0.04 (0.02)	0.72
	RG	1031.20	248.41	0.38 (0.03)	0.16
	RIG	0.49	1.47	0.18 (0.01)	0.44
Off-Test	BW	6.99	2.62	0.31 (0.02)	0.21
	BF	2.65	0.20	0.53 (0.03)	0.04
	Muscle	0.07	0.02	0.38 (0.01)	0.08
	IMF	0.00	0.00	0.01 (0.03)	0.36

[1] ADFI, average daily feed intake, in g. AOTD, average occupation time, in s. ANVD, average number of visits to feeder. ADFR, average daily feeding rate, in g/min. AFIV, average feed intake per visit across testing period, in g. AOTV, average occupation time per visit across testing period, in s. AFRV, average feeding rate per visit across testing period, in g/min. ADG_FIRE, average daily gain across testing period using FIRE (Feed Intake recording equipment), in g/d. ADG_Life, average daily gain from birth to off-test period, in g/d. ADG_PostWean, average daily gain from post-weaning to off-test, in g/d. FCR1 is the measure of feed conversion ratio, daily gain/daily feed intake, unit in 100%. RFI6 is a measure of residual feed intake, in g. RG, residual growth, in g. RIG, difference of residual growth and residual feed intake, unit in g. BW_Off, body weight at off-test, in kg. BF, ultrasound backfat thickness at off-test, in mm. Muscle, ultrasound muscle depth at off-test, in mm. IMF, intramuscular fat percentage, in 100%.

[2] σ_a^2 is the estimate of additive genetic variance component.

[3] σ_S^2 is the estimate of variance component for social common pen effect of the pen sharing by pen mates.

[4] σ_S^2/σ_{Ph}^2 is the ratio of social common pen effect and the total phenotypic effects.

Table 5.7 Accuracies[1] of predictions for testing dataset (n = 506, pigs born in 2011 and 2012, genotyped) using single-step (GEBV) and pedigree (EBV) with the training dataset (n = 16077, pigs born from 2002 to 2010 with phenotype data and additional n = 8195 animals genotyped in the training dataset).

Category	Trait[2]	GEBV[3] (SD)	EBV[4] (SD)
Behavior	ADFI	0.36 (0.02)	0.21 (0.04)
	AOTD	0.43 (0.06)	0.33 (0.05)
	ANVD	0.52 (0.02)	0.39 (0.04)
	ADFR	0.37 (0.04)	0.29 (0.06)
	AFIV	0.37 (0.03)	0.31 (0.11)
	AOTV	0.57 (0.06)	0.32 (0.03)
	AFRV	0.33 (0.03)	0.27 (0.02)
Growth	ADG_FIRE	0.54 (0.08)	0.42 (0.12)
	ADG_Life	0.65 (0.02)	0.51 (0.04)
	ADG_PostWean	0.63 (0.05)	0.46 (0.04)
Efficiency	FCR	0.32 (0.05)	0.19 (0.07)
	RFI	0.41 (0.06)	0.25 (0.05)
	RG	0.47 (0.07)	0.32 (0.09)
	RIG	0.46 (0.05)	0.31 (0.06)
Off-test	BW	0.68 (0.09)	0.47 (0.10)
	BF	0.66 (0.13)	0.35 (0.07)
	Muscle	0.36 (0.05)	0.30 (0.07)
	IMF	0.37 (0.11)	0.12 (0.05)

[1] Average accuracy calculated from the SEP from the inverse of the left hand side (LHS) (Forni et al., 2011).

[2] ADFI, average daily feed intake, in g. AOTD, average occupation time, in s. ANVD, average number of visits. ADFR, average daily feeding rate, in g/min. AFIV, average feed intake per visit, in g. AOTV, average occupation time per visit across testing period, in s. AFRV, average feeding rate per visit across testing period, in g/min. ADG_FIRE, average daily gain (ADG) using FIRE (Feed Intake recording equipment), in g/d. ADG_Life, life time ADG, in g/d. ADG_PostWean, post-weaning ADG, in g/d. FCR, feed conversion ratio, daily gain/daily feed intake, unit in 100%. RFI, residual feed intake, in g. RG, residual growth, in g. RIG, difference of residual growth and residual feed intake, unit in g. BW_Off, body weight at off-test, in kg. BF, ultrasound backfat at off-test, in mm. Muscle, ultrasound muscle depth, in mm. IMF, intramuscular fat percentage, in 100%.

[3] Using single-step animal model incorperating marker information (both animals in testing and training).

[4] Using traditional animal model with pedigree information.

Table S5.1 Distribution of progenies from different sires, dams, or litters in the dataset.

No. of progenies[1]	No. of sires (%)[2]	No. of dams (%)[3]	No. of litters (%)[4]
1	7 (2)	565 (18)	1435 (30)
2	11 (3)	659 (21)	1382 (29)
3	9 (3)	525 (17)	966 (20)
4	15 (5)	401 (13)	592 (12)
5	13 (4)	316 (10)	271 (6)
6	10 (3)	209 (7)	119 (2)
> 6	264 (80)	409 (14)	36 (1)
> 10	232 (70%)	100 (3)	-[5]
> 20	171 (52)	2 (0)	-
> 50	72 (22)	-	-
> 100	22 (7)	-	-
Total	325	3094	4801

[1] Values represent the number of observations (progenies) in the dataset.

[2] Values represent the number of sires (percentage of sires) with the particular progeny counts in the first column in the dataset.

[3] Values represent the number of dams (percentage of dams) with the particular progeny counts in the first column in the dataset.

[4] Values represent the number of litters (percentage of litters) with the particular littermate counts in the first column in the dataset.

[5] The sign '-' means not exist.

Table S5.2 Distribution of the male progenies of sires (n = 325) in the contemporary groups[1](CGs, n= 223) in the dataset.

No. of sires	No. of CGs include Male Progenies of the Sire
78	1
45	2
51	3
22	4
32	5
18	6
9	7
9	8
2	9
7	10
8	11
5	12
4	13
3	14
4	16
1	17
3	18
3	19
1	20
3	21
3	22
1	23
1	24
3	27
1	28
1	29
4	31
1	32
1	33
1	35

[1] Contemporary group. The concatenation of birth year and birth week of the pig.

Table S5.3 Distribution of the male progenies of sires (n = 325) in the pen[1] (n=921).

No. of sires	No. of pen housing the male progenies of the sire
14	1
26	2
26	3
20	4
18	5
20	6
18	7
15	8
9	9
9	10
8	11
4	12
14	13
6	14
7	15
7	16
6	17
9	18
5	19
5	20
10	21
2	22
3	23
3	24
3	25
1	26
3	27
2	28
2	29
5	30
6	31
2	32
1	33
4	34
1	35
2	36
3	37

Table S5.3 Continued

1	38
25	> 40

[1] Pen, the smallest unit sharing by a group of animals, each of which equipped with a FIRE (feed intake recording equipment) feeder during test period.

Table S5.4. Model comparison and selection for average daily feed intake (ADFI).

Models[1]	Model Fitting[2]		Estimates of Variance Components[3]							Variance ratios[4]			
	AIC	-2logL	σ_a^2	σ_e^2	σ_{cg}^2	σ_p^2	σ_s^2	σ_l^2	h^2	$\sigma_{cg}^2/\sigma_{Ph}^2$	σ_p^2/σ_{ph}^2	σ_s^2/σ_{Ph}^2	σ_l^2/σ_{ph}^2
M1	184278	184274	43015	195170					0.18				
M2	183798	183792	47546	179320		27246			0.19		0.11		
M3	178885	178879	45294	85814			191280		0.14			0.59	
M4	187483	187477	47754	192220	93406				0.14	0.28			
M5	182695	182689	72706	78289			190480		0.21			0.56	
M6	184202	184208	19989	191770				22085	0.09				0.09
M7	178821	178813	23671	89444			191490	12831	0.07			0.6	0.04

[1] Models compared for ADFI were single-trait animal models, different in fitted terms in left hand side (LHS) and model assumption. M1 with contemporary groups (CG), on-test age, duration of test and dam parity as fixed effects and animal effect (**a**) and residual (**e**) as random effects with **a** ~ N (**0**, $A\sigma_a^2$), A is the numeric relationship matrix and **e** ~ N (**0**, $I\sigma_e^2$). M2 with CG, on-test age, duration of test and dam parity as fixed effects and physical Pen_Room effect (**p**), animal effect (**a**) as well as residual (**e**) as random effects with **p** ~ N (**0**, $I\sigma_p^2$). M3 with CG, on-test age, duration of test and dam parity as fixed effects and social common pen effect (**s**), animal effect (**a**) as well as residual (**e**) as random effects with assumption **s** ~ N (**0**, $I\sigma_s^2$). M4 with on-test age, duration of test and dam parity as fixed effects and CG (**cg**), animal effect (**a**) as well as residual (**e**) as random effects with **cg** ~ N (**0**, $I\sigma_{cg}^2$). M5 with on-test age, duration of test and dam parity as fixed effects and social common pen effect (**s**), animal effect (**a**) as well as residual (**e**) as random effects with **s** ~ N (**0**, $I\sigma_s^2$). M6 with CG, on-test age, duration of test and dam parity as fixed effects and common litter effect (**l**), animal effect (**a**) as well as residual (**e**) as random effects with **l** ~ N (**0**, $I\sigma_l^2$). M7 with CG, on-test age, duration of test and dam parity as fixed effects and social common pen effect (**s**), common litter effect (**l**), animal effect (**a**) as well as residual (**e**) as random effects with **s** ~ N (**0**, $I\sigma_s^2$), **l** ~ N (**0**, $I\sigma_l^2$). [2] The fit of the models were assessed by Akaike information criterion (AIC) and minus twice of loglikelihood (-2logL).[3] The variance components include σ_a^2 (additive genetic variance), σ_e^2 (residual variance), σ_{cg}^2 (variance for random CG effect), σ_p^2 (variance for spatial Pen_Room effect), σ_s^2 (variance for social common pen effect), σ_l^2 (variance for random common litter effect) as well as σ_{ph}^2 (total phenotypic variance).[4] Variance ratios include h^2 (heritability), $\sigma_{cg}^2/\sigma_{Ph}^2$ (ratio of CG over total), σ_p^2/σ_{ph}^2(ratio of Pen_Room variance over total), σ_s^2/σ_{Ph}^2 (ratio of common pen variance over total), σ_l^2/σ_{ph}^2 (ratio of common litter variance over total phenotypic variance).

CHAPTER 6

Genome-wide association for different measures of feed efficiency and feeding behavior

Shihui Jiao[*], Francesco Tiezzi[*], Yijian Huang[†], Kent A. Gray[†], and Christian Maltecca[*]

[*] Department of Animal Science, North Carolina State University, Raleigh, NC 27695

[†] Smithfield Premium Genetics, Rose Hill, NC 28458

ABSTRACT: Feed efficiency is of great economically importance for the swine industry. To better understand the genetic mechanisms for different measures of feed efficiency and feeding behavior, a genome-wide association study (GWAS) has been preformed to investigate the underlying genetic architecture of each trait of interest. Genomic regions have been identified for ANVD (average daily number of visit) on SSC 7, AOTV (average occupation time per day) on SSC 9 and FCR (feed conversion ratio) on SSC 4, all exceeding a conservative Bonforroni genome-wide significant threshold 1.4×10^{-6}. Additionally, regions with obvious peaks appraching the significant threshold for AOTV on SSC 3, residual growth (RG) on SSC 1 and a combined measure of residual feed intake and residual growth (RIG) on SSC 15 were identified. In total, 27 candidate genes within those regions were identified involved mostly in metabolisms. Gene networks constructed including all candidate genes suggested a close relationship among these in common pathways or in co-expression, involving metabolism and regulation of muscle fiber.

Key words: GWAS, feed efficiency, feeding behavior

INTRODUCTION

Feed efficiency is of primary economic importance for the swine industry. The need to improve feed efficiency has long been recognized in pigs, providing that feed cost is the single largest input cost in swine production (Henman, 2003; Hoque et al., 2009) and individual feed intake can be more easier quantified in pigs and poultry production than ruminants. Although the efficiency of feed utilization in pigs has been improved to some degree in the past through indirect selection for growth and backfat thickness (Cleveland et

al., 1983; McPhee et al., 1988), improvement for more efficient pig is further required in the future, given the increasing demand for high-value protein with the ever increasing global population and in the face of a much more intense competition from other uses for land and water recourses (Hayes et al., 2013).

The genetic improvement achieved in swine feed efficiency has been partially hampered in the past by the fact that individual feed intake is often difficult and expensive to measure (Hayes et al., 2013). Electronic feeding systems have been developed and are currently commercially available, however, they are still expensive. Thus only a limited number of males in nucleus population are normally recorded during the growing-finishing phase of production (Casey, 2003). Therefore, indicator traits strongly genetically correlating with measures of feed efficiency and easier to measure will help to accelerate the genetic improvement of efficiency. It has been speculated that some measures of feeding behavior can be used as indicator traits, given their strong genetic correlation with measures of feed efficiency reported previously and by Do et al., (2013a). A new device to measure such behaviors has reported by Maselyne et al. (2015) and appears to be appealing, since it will be less expensive, easier to maintain than electronic feeding systems used in group housed pigs, nonetheless its effectiveness as a selection tool remains to be demonstrate. Additionally, feed efficiency is not referring to a single measure, but encompasses all measures associated with the efficiency of feed utilization and reflects the complexity of the trait (Arthur and Herd, 2005). Although different measures of feed efficiency have been found moderately heritable (Jiao et. al unpublished), such as feed conversion ratio (FCR), residual feed intake (RFI) and residual growth (RG), differences between those measures of efficiency exists, especially in

their relationships with other traits thus suggesting that they may be related to different complementary genetic mechanisms.

The availability of marker information from mid-density SNP panels has allowed the implementation genomic selection in livestock populations. This has led to promising results for the selection of low heritability and difficult traits to measure, including feed efficiency in chickens, dairy cattle and pigs (Gonzalez-Recio et al., 2009; Pryce et al., 2012; Christensen et al., 2012; Jiao et al., 2014a). As suggested by Hayes et al. (2010) the accuracy of prediction for future phenotypes or genetic merit using high-density SNP makers in genomic selection will also depend on the genetic architecture of the trait of interest. A limited amount of information is currently available in swine regarding the underlying genetic architecture of different measures of feed efficiency and feeding behavior. Therefore, the objective of this study was to perform genome-wide association for different measures of feed efficiency and feeding behavior in a purebred Duroc population.

MATERIALS AND METHODS

Field data

Individual feed intake data were collected from June 2004 through May 2013 for 14901 growing-finishing Duroc purebred boars using of FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, Kansas, USA) on a nucleus farm owned by Smithfield Premium Genetics (SPG; Rose Hill, NC). Those boars were born in 4801 litters from the mating of 3094 dams and 325 sires. After being weaned at the average age of 25d, the pigs were placed into contemporary groups (CG, concatenation of year and week, n =

223) and then moved to finishing barn after approximately 7 weeks of the nursery phase. During the test period (growing-finishing period), individual feed intake and related behavior measures were recorded using single-space FIRE feeders placed in each pen and the details for management and testing procedures can be found in Chen et al. (2010) and Jiao et al. (2014a). Briefly, boars of a group of ~12 (ranging from 6 to 19) were distributed into a pen with a single-space FIRE feeder. Boars started test at approximately 93 d of age (on-test age) with test duration of 67 days. Boars were removed from test at an average age of 170 d and 4958077 single individual visits to feeder for the 14901 testing boars were recorded.

Since the data collected by FIRE feeders were found to contain errors and extreme values, the feed intake and behavior measures were edited by multiple imputation (Jiao et al, 2015, unpublished) before use. The body weights collected by FIRE system were edited using robust regression (Zumbach et al., 2010) to obtain average daily gain during the testing period (ADG). After editing, the measures from each feeding visit were summarized into behavior measures, including average daily occupation time (AOTD), average daily feeding rate (AFRD), ANVD, average feed intake per visit (AFIV), and AOTV.

Average daily feed intake (ADFI) was also obtained after data editing, then different measures of feed efficiency were obtained using regression models. Details of those models can be found in Jiao et. al (2015, unpublished). Four different measures of feed efficiency were investigated in the present study, including feed conversion ratio (FCR), residual feed intake (RFI), residual growth (RG) and combined measure of residual feed intake and residual growth (RIG). The descriptive statistics for all phenotypes investigated were shown in Table 6.1.

Marker data

Animals were genotyped using the high-density marker panel Illumina PorcineSNP60 Beadchip (Illumina Inc., San Diego, CA USA; n = 3699) and the low-density panel GGP-Porcine containing approximately 10,000 SNP (GeneSeek Inc., Lincoln, NE USA; n = 4621). Imputation for genotypes from low-density panels to high-density marker panels was achieved using Beagle software (Browning and Browning, 2009). The multiple quality control edits conducted for animal and SNP included removal of animals with call rate < 0.90, SNP with call rate < 0.90, SNP with minor allele frequency (MAF) < 0.02 and SNP with p-value < 0.0001 of a chi-square test for Hardy-Weinberg equilibrium. Details procedures of genotype imputation can be found in Howard et al. (2015). After editing, the genotype dataset included 33039 SNP on 18 autosomes for 8701 animals. However, 1250 boars had both phenotype and genotype records and were retained for the subsequent association analysis.

Statistical analysis

Simple single marker linear model were used to conduct the genome-wide association analysis for each trait of interest (Zhou and Stephens, 2012) with pre-adjusted phenotypes as response. The pre-adjustment included correcting the systematic fixed effects and social interactions. The model was shown below,

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{Z_1}\mathbf{a} + \boldsymbol{Z_3}\mathbf{s} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of observations (behavior and feed efficiency measures); $\mathbf{b}$ is a vector of fixed effects including contemporary groups (CG, n = 223), parity of dam (1,

2, 3+), and fitted covariates including on-test age and test duration; **a** is a vector of additive genetic effects of the boars; **e** is a vector of random residuals; and **X** and $\boldsymbol{Z_1}$ are corresponding incidence matrices; **s** is a vector of social common pen effects (n = 921) effect, treated as random effects and $\boldsymbol{Z_2}$ as incidence matrix for **s**. The assumption for random effects of this model in pedigree analysis was

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{s} \\ \mathbf{e} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{A}\sigma_a^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I_s}\sigma_s^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I_e}\sigma_e^2 \end{bmatrix} \right)$$

where σ_a^2, σ_s^2 and σ_e^2 were variance components for additive, social common pen and residual effects and A was the numerator relationship matrix derived from pedigree information. The pre-adjusted phenotype was then computed as the summation of animal effect and model error.

The single markter regression model was used to perform the GWAS as described by Zhou and Stephens (2012),

$$\boldsymbol{y}^* = \mathbf{x\beta} + \mathbf{Zg} + \mathbf{e},$$

where $\boldsymbol{y}^*$ is the pre-adjusted phenotype; x is the fixed marker genotype denoted as 0, 1, 2 to represent the copy number of minor allele for each SNP marker; **β** is the estimated fixed marker effect; **g** is a vector of polygenic effects to capturing the other marker effects and account for potential population structure; **Z** is the corresponding incidence matrix and **e** is a vector of model residuals. The assumption of the model for random effects are **g** ~ N (**0**, $\mathrm{G}\sigma_g^2$) and **e**~ N ($\mathbf{0}$, $\mathrm{I}\sigma_e^2$). The loglikelihood ratio test was conducted at each time a SNP

marker was fitted into the mixed model using GEMMA (Zhou and Stephens, 2012) and the p-value was used for each marker for later significant test.

In this study, the Bonforroni adjustment for type I error were used for multiple testing in order to set the genome-wide significance threshold. The threshold was 1.4×10^{-6}, computed as 0.05 / 33039. Candidate genes in the significant or close significant regions were annotated and gene functions were summarized. Consequently, gene network were constructed for those candidate genes using GeneMANIA (Warde-Farley et al., 2010).

RESULTS

A genome-wide association analysis was performed for each single trait of interest using 33039 quality controlled autosomal SNP markers.

Figure 6.1 displays the Manhattan plots for the two behavior measures ANVD and AOTV and one feed efficiency measure FCR with significant genomic regions identified. A peak formed by nearby SNP markers exceeding the genome-wide significant threshold line was observed for ANVD on SSC 7. Additionally, a weaker signal nearly reaching the significance threshold was observed for ANVD in the middle region of SSC 3. A significant region on SSC 9was identified for AOTV. For FCR, a peak exceeding the significance threshold was identified on SSC 4.

Figure 6.2 illustrates the Manhattan plots for the other 3 feed efficiency measures. Although no significant SNP markers were identified for the 3 traits, clear peaks were observed for RG on SSC 1 and RIG on SSC 15. There was however no obvious signal identified for residual feed intake. Figure 6.3 shows the Manhattan plots for 3 other behavior

traits where no significant markers were identified. Little similarity was also found among the 4 different measures of feed efficiency..

Significant SNP markers were listed in Table 6.2 along with markers approaching significance (defined as p-value < 2.5×10^{-5}). The marker name, reference SNP ID in SNP database at NCBI (http://www.ncbi.nlm.nih.gov/), physical position in base pair, allele frequency, marker effect, p-value and candidate genes in upstream /downstream 1 Magabase (Mb) of each SNP were also provided. Allele frequency of significant markers SNP marker ranged from 0.17 to 0.49 while the estimated marker effects were also provided at least providing information of the direction of marker effect by examining the sign of the estimated effects.

Table 6.3 provides the physical position and gene function for each potential candidate gene identified for feeding behavior and measures of feed efficency. The candidate genes were those located approximately 1 Mb upstream or downstream of the marker identified. Most of the genes have functions involved in metabolisms: metabolism of lipid such as *CLPS* and *GBA2*; or metabolism of protein like *RPL10A*, *ALG9*, *NANS*, ALG2, and *SEC61B* gene. Some candidate genes were identified due to their involvement in the hormone signaling, such as HK2 protein involving in HIF-1 signaling, HSD17B7 in steroid biosynthesis, or CREB3 protein playing roles in insulin secretion and AMPK signaling pathway. The top 9 candidate genes were identified for feeding behavior ANVD and AOTV, whereas the rest genes in Table 6.3 and Table 6.4 were found associated with feed efficiency, including FCR, RG and RIG.

Candidate genes identified in this study were used to construct gene networks to examine their relatedness. Figure 6.4 shows the gene network constructed for the feeding behavior traits ANVD and AOTV. Most of the candidate genes, except *LAYN* and *CLPS*, were connected together due to co-expression and co-localization, suggesting their co-functioning in the common pathways. Similarly, the gene network displayed in Figure 6.5 included the candidate genes identified for feed efficiency measures. All genes except *CD72* were shown closely bonded to each other directly or through other intermediate genes because of co-expression and shared protein domains. The gene network constructed from all genes regardless of trait is shown in Figure 6.6. As expected a more complex and less defined network was identified.

DISCUSSIONS

Whole genome association for measures of feed efficiency

Different measures of feed efficiency were investigated in the current GWAS study. Only few regions of large effect involving feed efficiency were identified, confirming the polygenic nature of the trait. These results are consistent with the foundlings of Do et al. (2014a, 2014b) and Onteru et al. (2013) for measures of RFI and our previous results (Jiao et al. 2014) on a similar population for FCR and RFI. Addtionally, little similarity was found across the 4 measures of efficiency. This was not surprising since albeit all measures point to the same overall metabolic efficiency differences exists in the formulae used to compute each of them and this might partially explain the differences in estimates. Furthemore, different heritability estimates and genetic correlations with other traits have been reported in previous studies (Rothschild and Ruvinsky, 2010; Cai et al., 2008; MacNeil and Kanp, 2015; Jiao et

al, 2014a). Finally, different heritability for the same measures of efficiency and genetic correlations estimates with other traits were also reported in different breeds, even sometimes of opposite directions (Do et al., 2013b) suggesting a high heterogeneity of the trait.

A handful of genome-wide scans have been conducted in the past few years in Danish Duroc and Yorkshire (Do et al., 2014a; Do et al., 2014b; Sahana et al., 2013), US Yorkshire selected for lower RFI of generations (Onteru et al., 2013) and in a related Duroc population (Jiao et al., 2014b) for feed efficiency traits in pigs, focusing mainly on FCR and RFI. The current analysis identified a genomic region at approximately 94 Mb of SSC 4 for FCR, A region at 4-5 Mb of SSC4 was previously reported by Jiao et al. (2014b) which did not reach significance level in the current analysis. The use of different animals, accounting for social interaction in the current study, different adjustment methods used to correct feed intake phenotypes and differences in genotype panels, as also found in Howard et al. (2015) could partially explain these differences. Sahana et al. (2013) also reported significant markers on SSC 4 for FCR in a Danish Duroc population using 3071 animals with de-regressed EBV as response, and the identification they reported were at approximately 63-64 Mb and 100-114 Mb, where the region we identified in the current study for FCR in US Duroc just located nearby. The present study located *RGS4*, *RGS5* and *HSD17B7* as candidate genes for FCR. Both RGS4 and RGS5 protein are involved in the inactivation of MAPK (mitogen-activated protein kinase) pathway, which growth factors are known to signal positively through (Chang and Karin, 2001). The protein coded by *HSD17B7* gene is involved in steroid biosynthesis, cholesterol bio synthesis, metabolic pathways, metabolism of lipids and lipoproteins and was highly expressed in obese men and women (Wang et al., 2013).

No signal was identified for RFI in this study, which agreed with result from our previous association analysis using individuals from the same population (Jiao et al., 2014b). However, significant markers were identified by Onteru et al. (2013) and Do et al. (2014a; 2014b) for residual feed intake. Regions on SSC 7 at 16-17 Mb and 39-40 Mb and on SSC 14 at 59-60 Mb and 90-91Mb were reported significantly associated with RFI in Yorkshire (Onteru et al., 2013). But the regions identified for RFI in Danish Duroc were identified at SSC 1, 8, 9, 13, and 18 and in Yorkshire located at SSC 1, 3, 7, 8, 9, 10, 15, 17 (Do et al., 2014a; 2014b). Both of those markers or regions explained only a small amount of phenotypic variation, indicating measures of residual feed intake may be highly polygenic.

Regions approaching significance for RG on SSC 1 and RIG on SSC 15 were identified for the first time in pigs in the present study. Candidate genes identified in this region contain *CREB3* and *NOR-1* involving in insulin secretion or signaling (Table 6.2). The candidate genes potentially associated with feed efficiency measures were bonded closely (Figure 6.5), indicating they may function in common pathways for the regulation of metabolism.

Genome-wide association for feeding behavior

To the best of our knowledge, only one GWAS study has been performed so far on feeding behavior traits in swine by Do et al. (2013b). In our study genomic regions on SSC 3, SSC 7 and SSC 9 were associated with ANVD and AOTV. Candidate genes in these regions were *CLPS* and *PPARD* with *CLPS* functioning as a regulator of lipid metabolism and *PPARD* involved in activating in digestion of dietary lipid and absorption. In addition, mutations in *PPARD* were found in obesity in human (Shin et al., 2004). Do et al. (2013b)

reported significant additional significant markers on SSC 14 strongly associated with ANVD along with another region on SSC 12. None of these regions were identified in the current study. Other regions were found significantly associated with other behavior traits in their analysis, although none overlapping with our results. The use of different populations, the correction for social interactions in the current analysis and a more lenient threshold in their study might partially explain the differences between the two experiments.

CONCLUSIONS

In conclusion, a list of SNP associated with different measures of feed efficiency and feeding behavior were provided in this study, offering valuable information for better understanding the genetic architecture of those traits. Furthermore, the mapping of genes for different measures of feed efficiency suggesting the different measures were likely affected by different set of genes. This may also bring new insight to those traits and further facilitate future selection decisions

LITERATURE CITED

Arthur, P. F., and R. M. Herd. 2005. Efficiency of feed utilisation by livestock-Implications and benefits of genetic improvement. Can. J. Anim. Sci. 85(3):281-290.

Cai, W., D. S. Casey, and J. C. M. Dekkers. 2008. Selection response and genetic parameters for residual feed intake in Yorkshire swine. J. Anim. Sci. 86:287-298.

Casey, D. S. 2003. The use of electronic feeders in genetic improvement programs for swine. PhD Diss. Iowa State Univer., Ames.

Chang, L., and M. Karin. 2001. Mammalian MAP kinase signaling cascades. Nature. 410(6824): 37-40.

Chen, C. Y., I. Misztal, S. Tsuruta, B. Zumbach, W. O. Herring, J. Holl, and M. Culbertson. 2010. Estimation of genetic parameters of feed intake and daily gain in Durocs using data from electronic swine feeders. J. Anim. Breed. Genet. 127:230-234.

Christensen, O. F. P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal. 6:10:1565-1571.

Cleveland, E. R., R. K. Johnson, R. W. Mandigo, and E. R. Peo, Jr. 1983. Index selection and feed intake restriction in swine. II. Effect on energy utilization. J. Anim. Sci. 56:570-578.

Do, D. N., A. B. Strathe, J. Jensen, T. Mark, and H. N. Kadarmideen. 2013a. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. J. Anim. Sci. 91(9):4069-4079.

Do, D. N., A. B. Strathe, T. Ostersen, J. Jensen, T. Mark, and H. N. Kadarmideen. 2013b.

Genome-wide association study reveals genetic architecture of eating behavior in pigs and its implications for humans obesity by comparative mapping. Plos One, 8(8):14102-14109.

Do, D. N., A. B. Strathe, T. Ostersen, S. D. Pant, and H. N. Kadarmideen. 2014a. Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. Front. Genet. 5:307.

Do, D. N., T. Ostersen, A. B. Strathe, T. Mark, J. Jensen, and H. N. Kadarmideen. 2014b. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. BMC Genet. 15(1): 27.

Falconer D. S., and T. F. C. Mackay.1996. Introduction to Quantitative Genetics. Longman.

González-Recio, O., D. Gianola, G. J. M. Rosa, K. A. Weigel, and A. Kranis. 2009. Genome-assisted prediction of a quantitative trait measured in parents and progeny: application to food conversion rate in chickens. Genet. Sel. Evol. 41(3):1.

Hayes, B. J., H. A. Lewin, and M. E. Goddard. 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. Trends Genet. 29 (4):206-214.

Henman, D. 2003. Nutritional management in integrated pig units.p11-132. in J. Wiseman, M. A. Varely, and B. Kemp, eds. Perspectives in pig science. Nottingham University Press, Nottingham, UK.

Hoque, M. A., H. Kadowaki, T. Shibata, T. Oikawa, and K. Suzuki. 2009. Genetic parameters for measures of residual feed intake and growth traits in seven generations of Duroc pigs. Livest. Sci. 121:45-49.

Jiao, S., C. Maltecca, K. A. Gray, and J. P. Cassady. 2014a. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. J. Anim. Sci. 92 (6):2377-2386.

Jiao, S., C. Maltecca, K. A. Gray, and J. P. Cassady. 2014b. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genome-wide associaiton. J. Anim. Sci. 92 (7):2846-2860.

Howard, J. T., S. Jiao, F. Tiezzi, Y. Huang, K. A. Gray, and C. Maltecca. 2015. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. BMC Genet. 16(1):59.

MacNeil, M., and R. A. Kemp. 2015. Genetic parameter estimation and evaluation of Duroc boars for feed efficiency and component traits. Can. J. Anim. Sci. 95(2):155-159.

Maselyne, J., W. Saeys, and A. Van Nuffel. 2015. Review: Quantifying animal feeding behaviour with a focus on pigs. Physiol. Behav. 138:37-51.

Onteru, S. K., D. M. Gorbach, J. M. Young, D. J. Garrick, J. C. M. Dekkers, and M. F. Rothschild. 2013. Whole genome association studies of residual feed intake and related traits in the pig. PloS one. e61756.

Pryce, J. E., J. Arias, P. J. Bowman, S. R. Davis, K. A. Macdonald, G. C. Waghorn, W. J. Wales, Y. J. Williams, R. J. Spelman, and B. J. Hayes. 2012. Accuracy of genomic

predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. J. Dairy Sci. 95(4): 2108-2119.

Rothschild, M. F., and A. Ruvinsky. 2010. The genetics of the pig. 2nd ed. CABI, Cambridge, MA.

Sahana, G., K. Veronika, H. Henrik, N. Bjarne, and O. F. Christensen. 2013. A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. J. Anim. Sci. 91:1041-1050.

Shin, H.D., B. L. Park, L. H. Kim, H. S. Jung, Y. M. Cho, M. K. Moon, Y. J. Park, H. K. Lee, and K. S. Park. 2004 Genetic polymorphisms in peroxisome proliferator-activated receptor delta associated with obesity. Diabetes 53:847-851.

Wang, F., V. Vihma, J. Soronen, U. Turpeinen, E. Hämäläinen, H. Savolainen-Peltonen, T. S. Mikkola, J. Naukkarinen, K. H. Pietiläinen, M. Jauhiainen, H. Yki-Järvinen, and M. J. Tikkanen. 2013. 17 β -Estradiol and estradiol fatty acyl esters and estrogen-converting enzyme expression in adipose tissue in obese men and women. J. Clin. Endocrinol. Metab. 98(12):4923-4931.

Warde-Farley, D., S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G. D. Bader, and Q. Morris. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucl. Acids Res. 38 (suppl 2): 214-220.

Zhou, X., and M. Stephens. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44(7):821-824.

Zumbach, B., I. Misztal, C. Y. Chen, S. Tsuruta, M. Łukaszewicz, W. O. Herring, and M. Culbertson. 2010. Use of serial pig body weights for genetic evaluation of daily gain. J. Anim. Breed. Genet. 127: 93-99.

Table 6.1 Descriptive statistics for feeding behavior and feed efficiency measures.

Trait[1]	N	Mean	SD	Min	Max
AOTD, s	11798	3714.61	840.25	154.98	11494.27
ANVD	11798	5.77	1.78	1.19	20.86
ADFR, g/min	11798	36.72	12.65	1.58	288.57
FIV, g	11798	471.62	181.66	21.98	1259.73
OTV, s	11798	778.39	233.22	48.92	2025.41
FCR, g/g	6485	41.86	16.04	7.60	506.80
RFI6, g	6464	-0.26	450.65	-2187.05	5776.00
RG, g	6463	-0.01	40.56	-191.80	134.10
RIG, g	6462	0.00	1.49	-6.23	12.97

[1] AOTD, average occupation time, in s. ANVD, average number of visits to feeder. ADFR, average daily feeding rate, in g/min. AFIV, average feed intake per visit across testing period, in g. AOTV, average occupation time per visit across testing period, in s. AFRV, average feeding rate per visit across testing period, in g/min. FCR is the measure of feed conversion ratio, daily gain/daily feed intake, unit in 100%. RFI is a measure of residual feed intake, in g. RG, residual growth, in g. RIG, difference of residual growth and residual feed intake, unit in g.

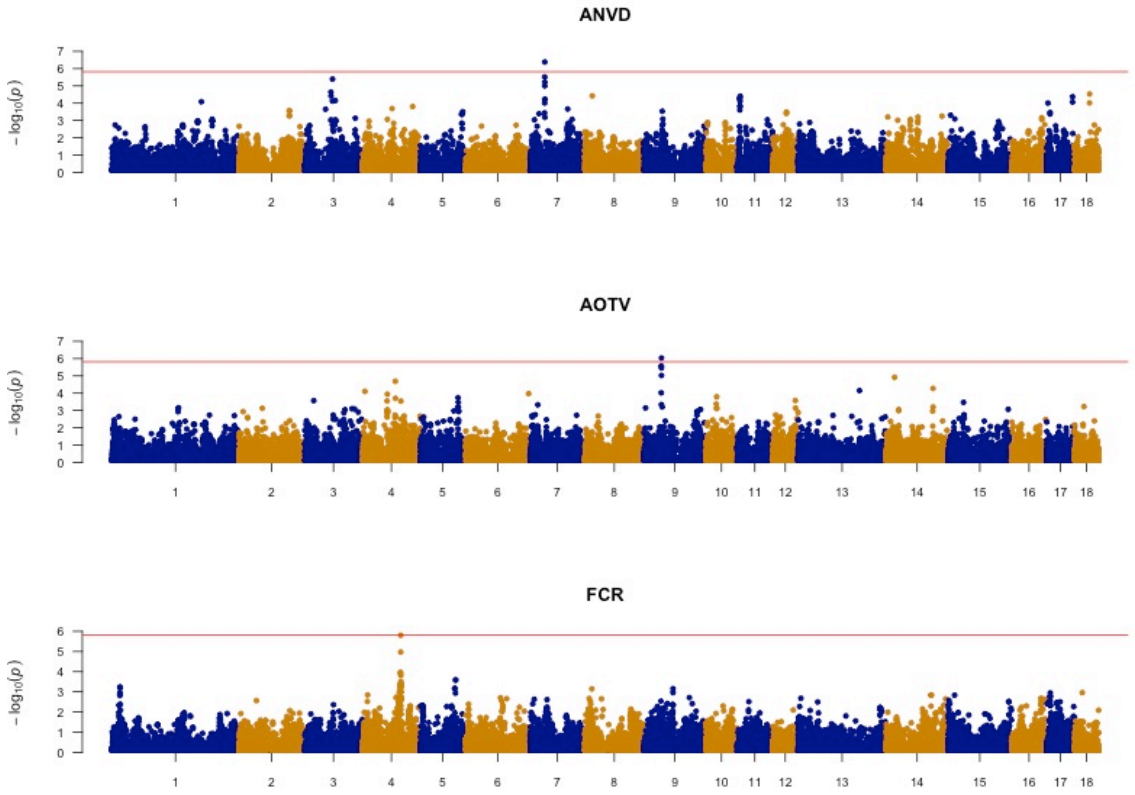

Figure 6.1 The Manhattan plots of average daily number of visits to feeders (ANVD), average occupation time per visit (AOTV) and feed conversion ratio (FCR). The x-axis is the physical position (bp) of the SNP (n = 33039) on 1 to 18 autosomes and y is the negative logarithm p-value with base 10. The red line is the genome-wide significance threshold that is set with 5.820 (by compute Bonferroni adjusted p-value for significance).

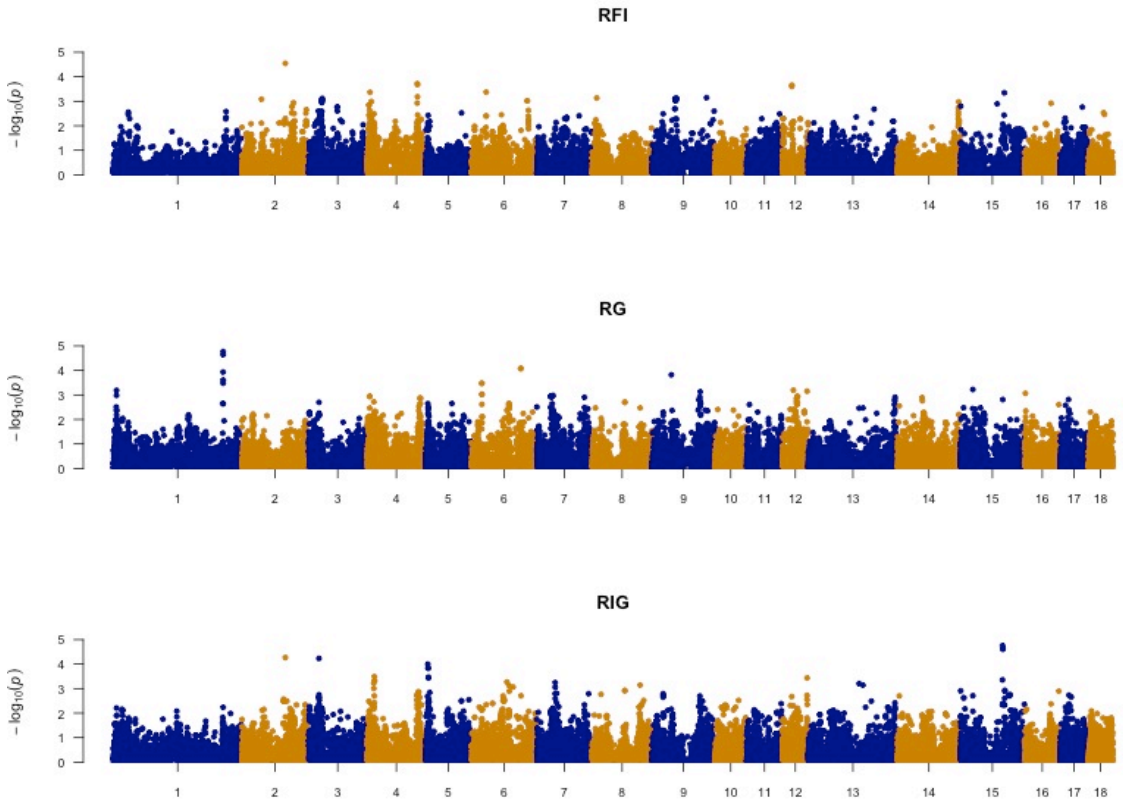

Figure 6.2 The Manhattan plots of measures of feed efficiency: residual feed intake (RFI), residual growth (RG) and residual feed intake and growth (RIG). The x-axis is the physical position (bp) of the SNP (n = 33039) on 1 to 18 autosomes and y is the negative logarithm p-value with base 10. None of the SNPs exceeded the significant test (the genome-wide significance threshold is set to be 5.820.

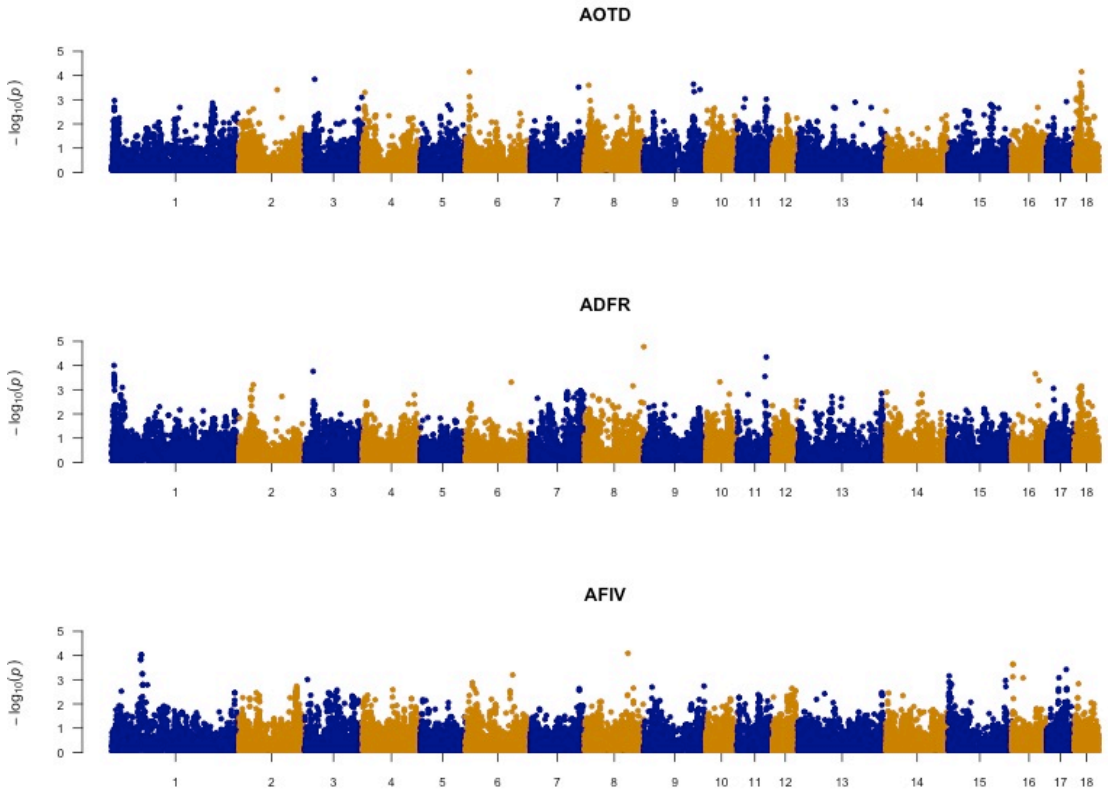

Figure 6.3 The Manhattan plots feeding behavior traits: average daily occupation time (AOTD), average daily feeding rate (ADFR) and average feed intake per visit (FIV). The x-axis is the physical position (bp) of the SNP (n = 33039) on 1 to 18 autosomes and y is the negative logarithm p-value with base 10. None of the SNPs exceeded the significant test (the genome-wide significance threshold is set to be 5.820.

Table 6.2 The significant (or close to significant) markers and potential candidate genes.

Trait[1]	Marker Name[2]	Ref. SNP ID[3]	Chr.[4]	Position (bp)	Frequency[5]	Beta[6]	P-Value	Candidate Gene
ANVD	ASGA0032595	rs80859588	7	36497507	0.49	-188.48	4.16E-07	*CLPS*
								SRPK1
								RPL10A
	ALGA0019577	rs81371887	3	69935645	0.17	151.07	4.04E-06	*HK2*
AOTV	ALGA0052446	rs81409200	9	43488001	0.47	31.29	9.27E-07	*RDX*
								FDX1
								POU2AF1
								LAYN
								ALG9
FCR	ASGA0020651	rs80953146	4	94916276	0.38	-2.27	1.58E-06	*RGS5*
								RGS4
								HSD17B7
RG	INRA0006750	rs322884221	1	269092739	0.26	5.99	1.90E-05	*VCP*
								CD72
								TPM2
								CREB3
								GBA2
								NPR2
								HINT2
								NANS
								TGFBR1
								ALG2
								SEC61B
								NOR-1
								KLF4
RIG	ALGA0086411	rs81454029	15	103608693	0.35	-0.71	1.74E-05	*COL3A1*

Table 6.2 Continued

MARC0032153	NaN	15	104101286	0.26	-0.54	2.43E-05	*COL5A2*

[1] ANVD, average number of visits to feeder. AOTV, average occupation time per visit across testing period, in s. FCR is the measure of feed conversion ratio, daily gain/daily feed intake, unit in 100%. RG, residual growth, in g. RIG, difference of residual growth and residual feed intake, unit in g.

[2] The markers are named in Illumina PorcineSNP60 Beadchip (Illumina Inc., San Diego, CA).

[3] The Ref. SNP ID represents the reference SNP ID in NCBI SNP database.

[4] The Chr. represents chromosome number.

[5] Frequency showed the minor allele frequency of significant SNP markers.

[6] Beta represents the regression coefficient of markers.

Table 6.3 The potential candidate gene(s) and the gene function(s).

Candidate Gene[1]	Gene Position[2]	Gene function (s)[3]
CLPS	chr7: 36546797-36571810	Digestion of dietary lipid; fat digestion and absorption; metabolism of lipids and lipoproteins
SRPK1	chr7:36575699-36638641	Innate immune response; viral process; sperm chromatin condensation
RPL10A	chr7:36106599-36123980	Metabolism of proteins; Nonsense Mediated Decay (NMD)
HK2	chr7: 71331772-71407448	Carbohydrate digestion and absorption; glucose metabolism; HIF-1 signaling pathway
RDX	chr9:42953206-43009694	Axon guidance; Recycling pathway of L1 (L1 functions in many aspects of neuronal development)
FDX1	chr9:43138316-43262180	Gormone biosynthetic process; cholesterol metabolic process; involved in steroid, vitamin D, and bile acid metabolism
POU2AF1	chr9:43842916-43866146	Humoral immune response
LAYN	chr9:44120371-44140925	Involved in the enhancement of inflammation
ALG9	chr9:44324268-44410329	Metabolism of proteins, N-Glycan biosynthesis
RGS5	chr4:94999826-95053683	Inactivation of MAPK pathway
RGS4	chr4:95096206-95103332	Inactivation of MAPK pathway
HSD17B7	chr4:95571663..95593948	Steroid Biosynthesis; Cholesterol biosynthesis; metabolic pathways, metabolism of lipids and lipoproteins
VCP	chr1:263379925-263394992	Cellular response to heat stress and stress: diseases of signal transduction
CD72	chr1:264070947-264078460	Axon guidance: B cell receptor signaling pathway: developmental biology
TPM2	Chr1:264127284-264134652	Striated muscle contraction; cardiac muscle contraction; adrenergic signaling in cardiomyocytes
CREB3	chr1:264176301-264181290	Insulin secretion; AMPK signaling pathway(inhibition of energy-consuming biosynthetic pathways, such as protein, fatty acid and glycogen synthesis, and activation of ATP-producing catabolic pathways, such as fatty acid oxidation and glycolysis.)

Table 6.3 Continued

GBA2	chr1:264181426-264193076	Metabolism of lipids and lipoproteins; metabolic pathways;
NPR2	chr1:264234817-264254186	bone development; cGMP biosynthetic process; regulation of blood pressure Purine metabolism
HINT2	chr1:264257330-264259905	Steroid biosynthetic process
NANS	chr1:268051524-268082265	Metabolism of proteins; sialic acid metabolism; amino sugar and nucleotide sugar metabolism
TGFBR1	chr1:269090665-269145072	Hippo signaling pathway
ALG2	chr1:269210468-269215674	Metabolic pathways, metabolism of proteins, N-Glycan biosynthesis
SEC61B	chr1:269216031-269222679	Metabolism of proteins; immune System
NOR-1	chr1:269817145-269852707	Modulate hepatic glucose production; insulin signaling in adipocytes and oxidative metabolism in skeletal muscle
KLF4	chr1:279056822-279061560	SRF and miRs in Smooth Muscle Differentiation and Proliferation
COL3A1	chr15:104067018-104105776	Protein digestion and absorption; collagen biosynthesis and modifying enzymes; vesicle-mediated transport
COL5A2	chr15:104122882-104272688	Protein digestion and absorption; collagen biosynthesis and modifying enzymes; vesicle-mediated transport

[1] Gene name in NCBI.

[2] Gene position in the format of chromosome number : 5'-end base pair position – 3'-end base pair position.

[2] Gene function(s) is (are) summarized from NCBI gene and KEGG gene.

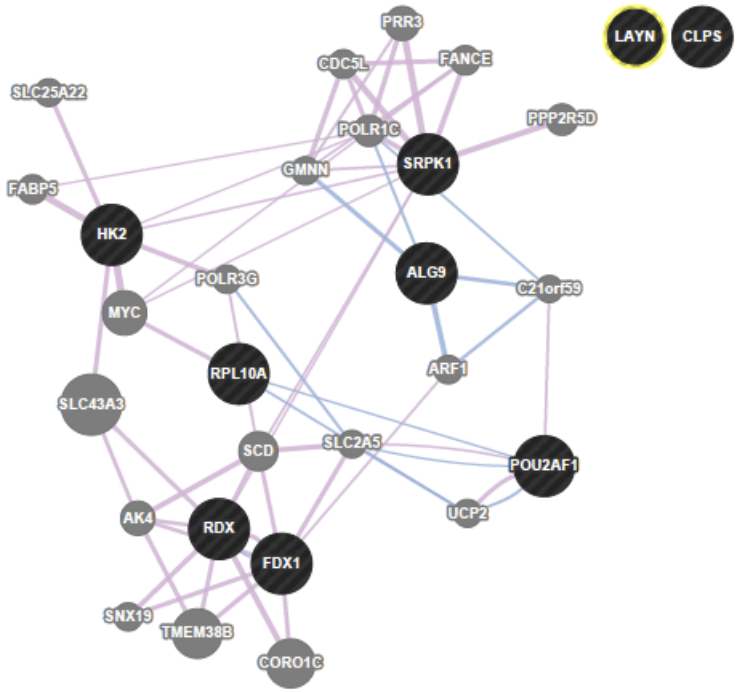

Figure 6.4 The gene network with candidate genes identified for feeding behavior traits.

Black dots represent candidate genes identified for ADNV (average daily number of visits) and AOTV (average occupation time per visit). The grey dots illustrat the intermediate genes needed to link the candidate genes in quiry and the colored lines representing the network construction critera, where blue lines represent co-localization of those genes and purple lines regresent the co-expression of those genes.

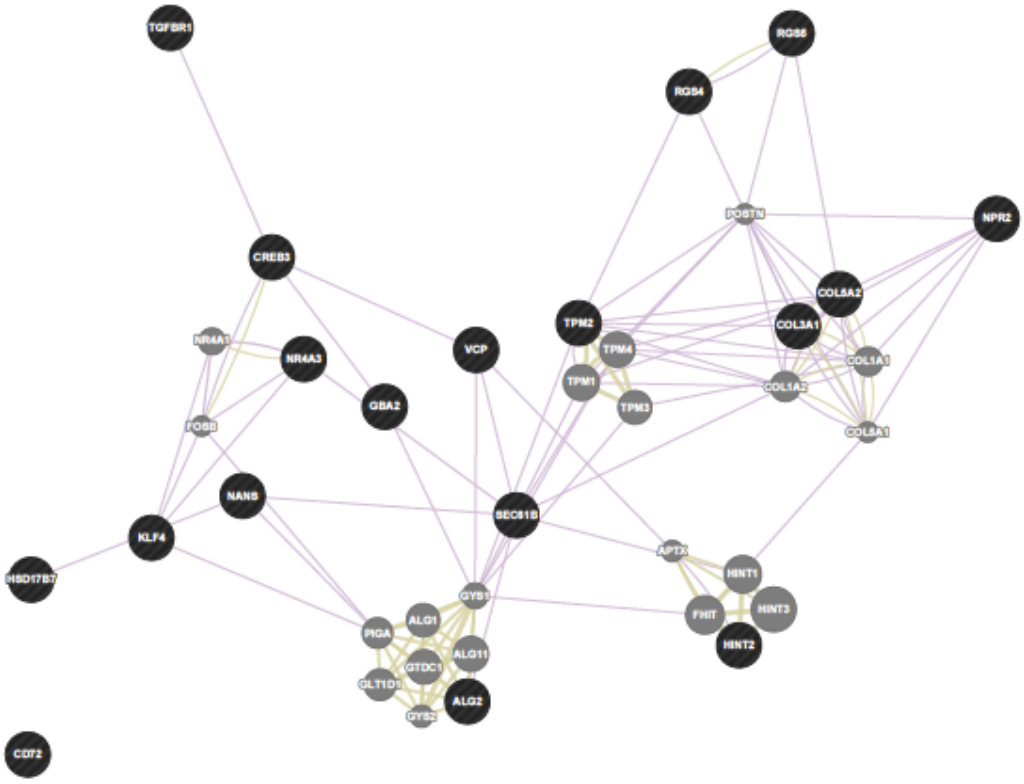

Figure 6.5 The gene network with candidate genes identified for measures of feed efficiency.

Black dots represent candidate genes identified for FCR (feed conversion ratio), RG (residual growth), and RIG (combind measure of residual feed intake and residual growth). The grey dots illustrat the intermediate genes needed to link the candidate genes in quiry and the colored lines representing the network construction critera, where the yellow lines represent shared protein domains of those genes and purple lines regresent the co-expression of those genes.

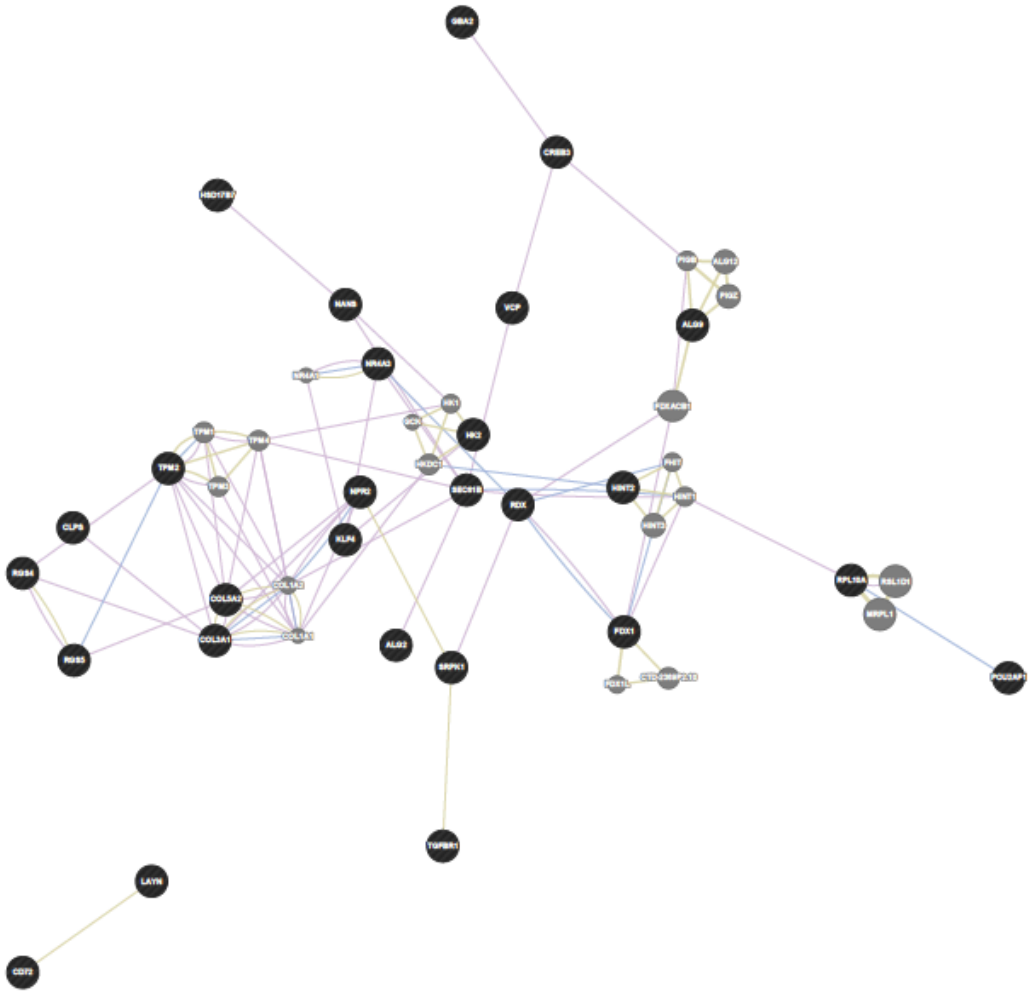

Figure 6.6 The gene network with candidate genes identified for feeding behavior and measures of feed efficiency.

Black dots represent candidate genes identified for ADNV (average daily number of visits), AOTV (average occupation time per visit), FCR (feed conversion ratio), RG (residual growth), and RIG (combind measure of residual feed intake and residual growth). The grey dots illustrat the intermediate genes needed to link the candidate genes in query and the colored lines representing the network construction criteria, where the yellow lines represent shared protein domains of those genes, purple lines represent the co-expression of those genes and blue lines represent the co-localization of those genes.