8-2015

Evaluation of reduced subsets of single nucleotide polymorphisms for the prediction of age at puberty in sows

Katherine L. Lucot
University of Nebraska-Lincoln, Katherine.lucot@gmail.com

Matthew L. Spangler
University of Nebraska-Lincoln, mspangler2@unl.edu

Melanie D. Trenhaile
University of Nebraska-Lincoln, mtrenhaile@huskers.unl.edu

Stephen D. Kachman
University of Nebraska-Lincoln, steve.kachman@unl.edu

Daniel C. Ciobanu
University of Nebraska-Lincoln, dciobanu2@unl.edu

Published in *Animal Genetics* 46:4 (August 2015), pp. 403–409; doi: 10.1111/age.12310

Accepted April 15, 2015; published online June 9, 2015.

Evaluation of reduced subsets of single nucleotide polymorphisms for the prediction of age at puberty in sows

K. L. Lucot,[1] M. L. Spangler,[1] M. D. Trenhaile,[1] S. D. Kachman,[2] and D. C. Ciobanu[1]

1 Animal Science Department, University of Nebraska–Lincoln, Lincoln, NE 68583
2 Department of Statistics, University of Nebraska–Lincoln, Lincoln, NE 68583

Corresponding author — D. C. Ciobanu, email dciobanu2@unl.edu

Abstract
Genomic information could be used efficiently to improve traits that are expensive to measure, sex limited or expressed late in life. This study analyzed the phenotypic variation explained by major SNPs and windows for age at puberty in gilts, an indicator of reproductive longevity. A genome-wide association study using 56,424 SNPs explained 25.2% of the phenotypic variation in age at puberty in a training set (n = 820). All SNPs from the top 10% of 1-Mb windows explained 33.5% of the phenotypic variance compared to 47.1% explained by the most informative markers (n = 261). In an evaluation population, consisting of subsequent batches (n = 412), the predictive ability of all SNPs from the major 1-Mb windows was higher compared to the variance captured by the most informative SNP from each of these windows. The phenotypic variance explained in the evaluation population varied from 12.3% to 36.8% when all SNPs from major windows were used compared to 6.5–23.7% explained by most informative SNPs. The correlation between phenotype and genomic prediction values based on SNP effects estimated in the training population was marginal compared to their effects retrained in the evaluation population for all (0.46–0.81) or most informative SNPs (0.30–0.65) from major windows. An increase in genetic gain of 20.5% could be obtained if genomic selection included both sexes compared to females alone. The pleiotropic role of major genes such as *AVPR1A* could be exploited in selection of both age at puberty and reproductive longevity.

Keywords: fertility, genomic prediction, swine

Introduction

Improving reproductive longevity in maternal lines using a traditional phenotypic selection approach is challenging due to the fact that it is lowly heritable, sex limited and expressed late in life. There is a need to predict the genetic merit of gilts for longevity in breeding herds for economic and animal welfare reasons. One trait that has been shown to predict reproductive longevity is the age at which a gilt reaches puberty (Serenius & Stalder 2007; Tart *et al.* 2013). The earlier a gilt expresses puberty, the more likely she will generate more parities and, as a result, more piglets born during her lifetime. Profiling expression of first estrus is labor intensive and rarely used at the commercial level. The objective of this research was to evaluate genomic predictors of age at puberty in training and evaluation sets of similar genetics based on the University of Nebraska (UNL) reproductive longevity resource population. A genome-wide association study (GWAS) was conducted using high-density genotypes and phenotypes for age at puberty using a Bayes B model to uncover regions of the genome and SNPs that influence phenotypic variation and estimate the efficacy of genomic predictors.

It is generally agreed that an increase in the number of markers used in a GWAS will result in larger proportions of the phenotypic variance explained because of the improved genome coverage. However, this concept has limitations in recent evolutionarily developed populations that have had less time to accumulate recombinations and are characterized by slow linkage disequilibrium decay and extended relationships between DNA markers and QTL. In recent work, Visscher (2014) showed that an increase in the number of significant SNPs from a discovery population (n = 250,000, ~1 mil. SNPs) when validated in five independent sets led to an increase in the average proportion of variance explained from 16% ($n_{SNPs} \approx 700$, $P < 5 \times 10^{-8}$) to 21% ($n_{SNPs} \approx 1900$, $P < 5 \times 10^{-5}$) to 29% ($n_{SNPs} \approx 9500$, $P < 5 \times 10^{-3}$). When no SNP selection was performed, the whole SNP panel of approximately 1 mil-

lion SNPs explained 50% of the variation, leading to the conclusion that most of the variation is explained by 9500 SNPs. Interestingly, an increase in the number of significant SNPs had limited to no influence on the correlation between genomic predictors and height. In addition, two recent studies showed that there is marginal advantage in using large-density SNP panels (770K or 500K SNPs) compared to the traditional 50K SNP panel in genomic evaluations of dairy cattle (VanRaden *et al.* 2011; Saatchi & Garrick 2014).

Because known examples of genetic variants that explain substantial amounts of variation for different quantitative traits are well characterized (e.g. Milan *et al.* 2000; Van Laere *et al.* 2003), we hypothesized that a small subset of major SNPs with relatively large effects explain a substantial part or most of the proportion of phenotypic variance captured by the entire Porcine SNP60 BeadChip. We expected that SNPs associated with major effects are more likely to be in high linkage disequilibrium with important QTLs and, as a result, more robust predictors of phenotypic variance in independent data sets. To test this hypothesis, the full set of SNPs and various subsets of SNPs from 1-Mb windows that explained the greatest proportions of variation were used to estimate genomic prediction values (GPVs) in the evaluation populations to assess the potential of genomic information to explain phenotypic variation for age at puberty outside the training set.

Materials and methods

Population

Phenotypic and genotypic data were obtained from a population of crossbred gilts (n = 1232) produced in 11 batches (B1 to B11) as described by Tart *et al.* (2013). The experimental pigs were derived from Nebraska Index Line (NIL) and commercial Large White × Landrace dams inseminated with semen from two unrelated commercial Landrace genetic lines (batch B1–B4: Landrace 1; batch B5–B11: Landrace 2). As a result, there is a partial degree of genetic relationship between the animals across batches. NIL was derived from commercial Large White 9 Landrace crossbreds and was selected for reproductive traits for over 32 years (Hsu 2011). Experimental gilts were inseminated with pooled commercial Duroc semen.

Nutrition, culling and phenotypes

Gilts received the same diet and management from day 0 through day 123. Starting at day 123, gilts were fed either an *ad libitum* corn–soybean meal diet or a caloric-restricted diet consisting of a daily allotment of feed that was ~80% of that consumed by gilts on the *ad libitum* regimen, as described by Tart *et al.* (2013). Beginning at approximately day 135, age at puberty was determined as the first day the gilts displayed signs of estrus in the presence of a boar. Estrus detection continued through approximately day 240 or until all gilts in a pen expressed estrus twice. At day 240, gilts were moved to the breeding barn and stayed in production through four parities, except B6, which due to limited space was maintained for three parities. Lifetime number of parities and lifetime total number born were measured as the total number of parities and piglets produced until culling respectively. Culling was performed after one unsuccessful service, after four parities, for soundness and health issues or for other reproductive failures.

Genotyping

DNA was isolated from an ear notch or tail clip using DNeasy and Puregene kits (Qiagen). DNA quality and quantity were tested by gel electrophoresis and Nanodrop (Thermo Scientific) spectrophotometry. Genotyping was performed using the Porcine SNP60 BeadChip (Illumina). All genotypes with a quality score below 0.4 were removed and replaced with mean allelic frequencies from the experimental population (B1–B11). SNPs and individual samples with a call rate below 0.8 were excluded, leaving 56,424 high-quality SNPs and 1232 experimental samples used in the analysis.

Genome-wide association analyses and genomic predictions

Phenotypic-based heritabilities were estimated using a linear mixed model; fixed effects included batch, diet and dam line, and random effects included sire and litter. Sire line was not included in the model due to partial confounding with batch. Genome-wide association was performed via GENSEL software (Fernando & Garrick 2008) using a Bayes B model, with genetic line, diet and batch fitted as fixed effects. The *a priori* value of π was set equal to 0.99 in most of the analyses (when the number of SNPs > 130), and the MCMC (Markov chain Monte Carlo) included 41,000 iterations with the first 1000 discarded as burn-in.

The influence of major 1-Mb windows and SNPs relative to predicting phenotypic variation were assessed using B1– B7 (n = 820) as a training data set and the subsequent B8–B11 (n = 412) batches as an evaluation data set. The 1-Mb windows and SNPs were ranked based on the proportion of variation explained in the training data set, with the windows and SNPs respectively that explained the most variation considered as major. The proportion of phenotypic variation explained in the evaluation data set was based on all SNPs or the single most informative or major SNP within the top 1%, 5%, 10%, 20%, and 50% of the 1-Mb windows explaining the most variation uncovered in the training data set. GPVs were based on subsets

of major SNP and 1-Mb windows obtained in the training data set using the mean posterior SNP effects estimated in the training data set or SNP effects re-estimated in the evaluation data set.

Response to direct selection for age at puberty per year was calculated by $\Delta G = h_g i\sigma_g/t$, where h_g is the square root of the proportion of the phenotypic variation explained by SNPs (genomic heritability), i is the selection intensity, σ_g is the genetic standard deviation (Table 1) and t is the generation interval (Table 1). The proportion of females and males selected was assumed to be 30% and 10% respectively, corresponding to a standardized average selection intensity of 1.457. If males were not selected, as would be the norm without a genetic selection tool, the selection intensity dropped to 1.159. The generation interval was assumed to be 1.5 years. The predictive accuracy of the genetic variance by GPV was estimated by dividing the sample correlation between the GPV and the observed phenotypic values by the square root of heritability (h^2).

The effect of age at puberty and *AVPR1A* genotype (*G31E*) on the probability that females would produce first two parities was evaluated by fitting age at puberty as a covariate in a generalized linear mixed model using a logit link function that included batch, diet and *AVPR1A* SNP genotypes as fixed effects and litter as a random effect.

Candidate gene identification

The porcine genome was divided into 1-Mb non-overlapping windows and extended by 1 Mb on each side in the major QTL regions to search for candidate genes using the *Sus scrofa* Build 10.2 assembly and tools available in Ensemble's BioMart – http://uswest.ensembl.org/biomart/martview – and DAVID – http://david.abcc.ncifcrf.gov/

Results and Discussion

Genome-wide association analyses of major regions and SNPs associated with age at puberty

Age at puberty is a complex trait with one of the largest heritability estimates for a reproductive trait (Rothschild *et al.* 1998). The heritability of age at puberty in the resource population was 0.42 (SE = 0.09; B1–B11, n = 1232), an estimate similar to previous reports (Bidanel 2011). A GWAS for age at puberty based on the entire data set (B1–B11, n = 1232) and all high-quality SNPs from the Porcine SNP60 BeadChip (n = 56,424) explained 28.3% of the phenotypic variance (Table 1). The three major 1-Mb windows located on SSC4 (7 Mb), SSC12 (2 Mb) and SSC3 (71 Mb) explained 3.3% of the genetic variation. These three regions include potential candidate genes, such as *NDRG1* (*n-myc downstream regulated 1*; SSC4, 7.8–7.9 Mb), known to play a role in DNA damage response and peripheral nervous system myelin maintenance; *BAIAP2* (*brain-specific angiogenesis inhibitor 1-associated protein 2*; SSC12, 1.5–1.6 Mb), shown to affect signal transduction, dendrite development and neuron projection; and *M1AP* (*meiosis 1 associated protein*; SSC3, 71.5–71.6 Mb), known to affect female gamete generation (http://www.geneontology.org).

We hypothesized that the selection of subsets of major SNPs, such as those identified in the regions described above, would explain the majority of the phenotypic variance captured by the entire SNP panel. To demonstrate this, major 1-Mb windows and SNPs were determined based on their contribution to genetic variation in the training data set (B1–B7). The proportion of the phenotypic variance explained in the training set was estimated based on subsets of the top 1%, 5%, 10%, 20%, and 50% 1-Mb windows including all SNPs in the respective windows or including the most informative SNP or the SNP associated with the largest effect in each window (Tables S1 and S2).

When all SNPs located in the major subsets of windows were used in the analysis, the proportion of variance peaked when the top 10% of the windows were used, explaining 33.5% of the phenotypic variance (Table 2, Fig. 1). The proportion of the phenotypic variance explained when the entire panel of high-quality SNPs (n = 56,424) was used was 25.2%. The proportion of the variance explained was larger when the most informative SNP from each of the windows was used; the most informative SNPs (n = 261) from the top 10% windows explained 47.1% of the phenotypic variance, the largest proportion of variation explained by markers. Standard deviation of the proportion of phenotypic variation explained across subsets of SNPs was limited, varying from 2.8 to 4.1%. The large proportion of the variance explained by SNPs, especially by the most informative SNP from each major window across different subsets, was expected because the same data set was used in ranking and also in the evaluation of subsets of the SNPs and windows. However, the fact that the subsets of SNPs contained in major windows explained more variation than the entire SNP panel can be attributed to the signal-to-noise ratio in the

Table 1. Phenotypic means, standard deviation (SD), posterior means of variance components and proportion and the posterior standard deviation (PSD) of the phenotypic variance of age at puberty (AP, B1–B11) and lifetime number of parities (LT-NP, B1–B8) explained by 56,424 SNP effects.

Trait	n	Mean (SD)	Genetic variance	Residual variance	Total variance	Phenotypic variance explained by SNPs (% PSD)
AP	1232	168.9 (19.9)	90.33	229.30	319.63	28.3 (3.0)
LT-NP	899	2.05 (1.60)	0.19	2.27	2.45	7.5 (1.6)

high-density SNP-based GWAS. The fact is that not all SNPs represented in the Porcine SNP60 panel are predictive for a given trait and, consequently, represent a source of noise. We expect that association analyses using subsets of significant SNPs that explained the predominant genetic variance are less affected by noise carried by the majority of the SNPs from the whole panel that are not associated with the trait of interest.

Evaluation of genomic predictors for age at puberty

Practical applications of genomic information in animal breeding could be more valuable if SNP panels could capture functional effects across populations and data sets. Major SNPs identified in B1–B7 (n = 820) were evaluated in subsequent batches (B8–B11, n = 412) of similar genetics by using SNPs selected based on their effects estimated in training and by re-estimating their effects in the evaluation set. When SNP effects were re-estimated in B8–B11, the proportion of phenotypic variance explained by all SNPs located in subsets of major windows varied from 12.3% to 36.8%, reaching the peak when all SNPs within the top 20% of the windows were used followed by a decline when more SNPs were added (Table 2). In contrast, the most informative SNPs located in the major windows captured less phenotypic variance, with proportions varying from 6.5% (top 1%) to 23.7% (top 50%). This was expected given that the SNPs identified as being the most informative in the training set are likely not the direct sources of the association and the linkage disequilibrium status between them and the QTLs were redefined in the evaluation population.

Traits that could potentially benefit the most from genomic selection include those that are expensive to measure, expressed late in life or limited by sex. The ability of the GPVs to predict the phenotype in the evaluation data set was substantially reduced if the SNP effects used to calculate GPV were estimated in the training population. For example, the largest correlation between the phenotype and GPVs ($r_{Y,GPV}$) was 0.17 and was obtained when all SNPs from the top 50% of the major windows were used in the analysis (Table 3). The correlation was negligible

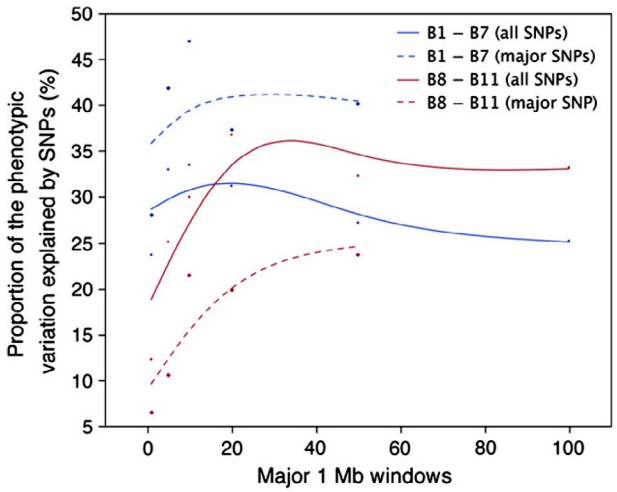

Figure 1. Proportion of the phenotypic variation of age at puberty in gilts explained by subsets of major SNPs and 1-Mb windows in the training (B1–B7) and evaluation (B1–B11) data sets.

when the most informative SNPs from the subsets of major windows were used. Predictive accuracies of the genetic variance by GPV had increased with larger subsets of major 1-Mb windows up to 50% (0.26) followed by a plateau (Fig. 2). These results were expected because previous reports in different species showed that SNP effects, or GPVs, are not well transferable across populations (e.g. Kachman *et al.* 2013). When GPVs were calculated based on the same subsets of major SNPs but with their effect estimated in the evaluation data set, there was a general increase in the ability of the GPV to predict the phenotype. The $r_{Y,GPV}$ varied from 0.46, when all SNPs from the major 1% of the windows were included, to 0.81, when the top 50% of the windows were included (Table 3). In general, using more than 50% of the windows showed a limited increase in the predictive ability. The use of the most informative SNP in the subsets of major windows did not improve the phenotypic variance explained; the $r_{Y,GPV}$ varied from 0.30 (26 SNPs/1% windows) to 0.65 (1 299 SNPs/50% windows). Results illustrate that, although the SNP effects are not robust across populations, informative subsets of SNP can be identified in training data and used in other populations if the effects are re-estimated.

Table 2. Proportion and posterior standard deviation (PSD) of the phenotypic variation of age at puberty explained by subsets of major SNPs and 1-Mb windows in the training (B1–B7) and evaluation data sets (B8–B11).

Percent major 1-Mb windows (%)	Number of SNPs within windows	Number of major SNPs	Proportion of phenotypic variation explained by SNPs (%, PSD, training set: B1–B7, n = 820)		Proportion of phenotypic variation explained by SNPs (%, PSD, evaluation set: B8–B11, n = 412)	
			All SNPs	Major SNPs	All SNPs	Major SNPs
1	652	26	23.7 (3.4)	28.0 (2.8)	12.3 (4.1)	6.5 (2.3)
5	3422	131	33.0 (3.8)	41.8 (3.2)	25.1 (4.8)	10.5 (3.7)
10	7246	261	33.5 (4.0)	47.1 (3.1)	30.0 (4.8)	21.4 (4.3)
20	14,429	521	31.2 (4.0)	37.3 (4.0)	36.8 (5.0)	19.9 (4.8)
50	33,297	1299	27.2 (3.6)	40.1 (4.1)	32.3 (4.4)	23.7 (4.6)
100	56,424	–	25.2 (3.2)	–	33.2 (4.3)	–

Table 3. Correlation (r) between the phenotype (y) and genomic prediction value (GPV) for age at puberty in the evaluation data set (B8–B11). GPVs were calculated based on the SNP effects estimated either in the training (B1–B7) or in the evaluation (B8–B11) data sets.

Percent major 1-Mb windows (%)	Number of SNPs within windows	Number of major SNPs	$r_{Y,GPV}$ based on SNP effects calculated in the training data set (B1–B7)		$r_{Y,GPV}$ based on SNP effects calculated in the evaluation data set (B8–B11)	
			All SNPs	Major SNPs	All SNPs	Major SNPs
1	652	26	0.02	0.00	0.46	0.30
5	3422	131	0.15	–0.01	0.67	0.47
10	7246	261	0.10	0.07	0.75	0.60
20	14,429	521	0.13	0.04	0.70	0.54
50	33,297	1299	0.17	0.10	0.81	0.65
100	56,424	–	0.17	–	0.82	–

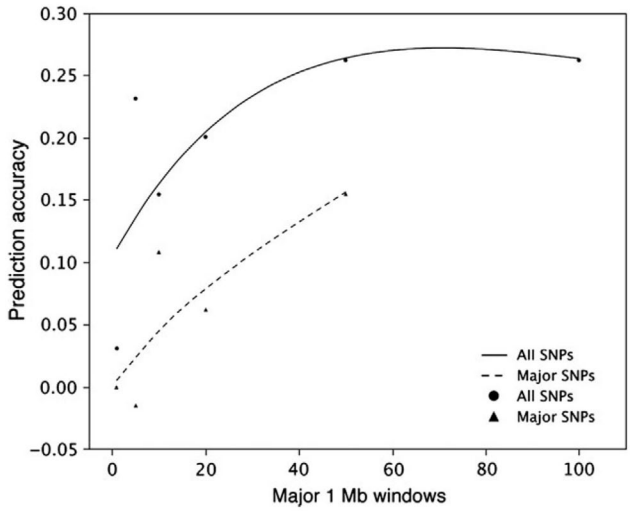

Figure 2. Prediction accuracy of the genetic variance by genomic prediction values for age at puberty in gilts by subsets of major SNPs and 1-Mb windows in the evaluation data sets (B8–B11).

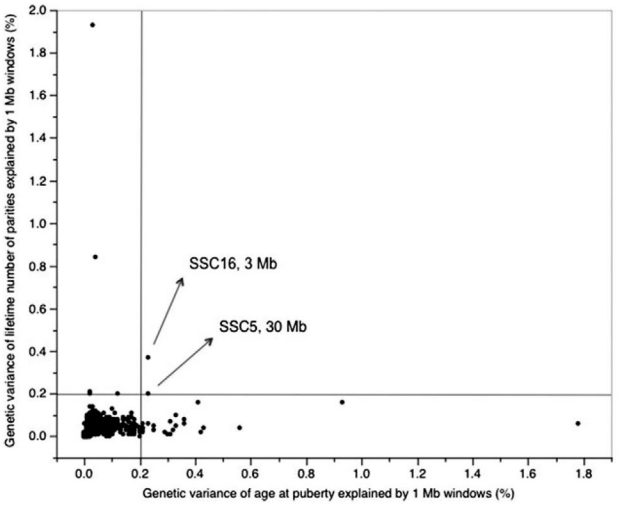

Figure 3. Proportion of genetic variance for age at puberty (X: B1–B11) and lifetime number of parities (Y: B1–B8) explained by 1-Mb windows. The reference lines are the 99% quantiles for the proportion of genetic variance explained by 1-Mb windows.

However, this option could be restricted if the availability of the phenotypic information in the evaluation population is limited.

Opportunities for the improvement of age at puberty through genomic selection

One of the advantages of using genomic information in animal breeding programs is the opportunity to perform selection in both sexes for traits limited in expression to only one of the sexes. For example, if genomic information were applied in selection for age at puberty, there would be an increase in genetic gain of 20.5% if selection includes both sexes compared to the scenario of using females alone, as in the case of mass selection on phenotypes. The genetic gain reached the peak when all SNPs or the most informative SNPs from the major 10% of the windows were used in selection.

Direct selection for early expression of age at puberty will have positive impacts by simply increasing the number of gilts that cycled multiple times and are reproductively mature prior to breeding. In addition, we recently provided evidence that age at puberty represents an early indicator of reproductive longevity and that potential pleiotropic DNA markers can be used to improve both traits (Tart *et al.* 2013). Evidence of common loci and pathways that influence both traits was demonstrated by negative relationships between the GPVs of age at puberty and lifetime number of parities (–0.41; $P < 0.0001$) and other lifetime productivity phenotypes such as lifetime number born alive (–0.45; $P < 0.0001$) and lifetime total number born (–0.45; $P < 0.0001$). Due to the limited effect of the full SNP panel in explaining phenotypic variation for lifetime reproductive longevity, our study analyzed the ability of the major loci associated with age at puberty to predict lifetime number of parities rather than using genome-wide prediction models. Potential pleiotropic loci were determined by analyzing the proportion of genetic variance for age at puberty and lifetime number of parities explained by 1-Mb windows. In Fig. 3, windows that affect both traits are represented on the diagonal of the *x*- and *y*-axes compared to trait-specific win-

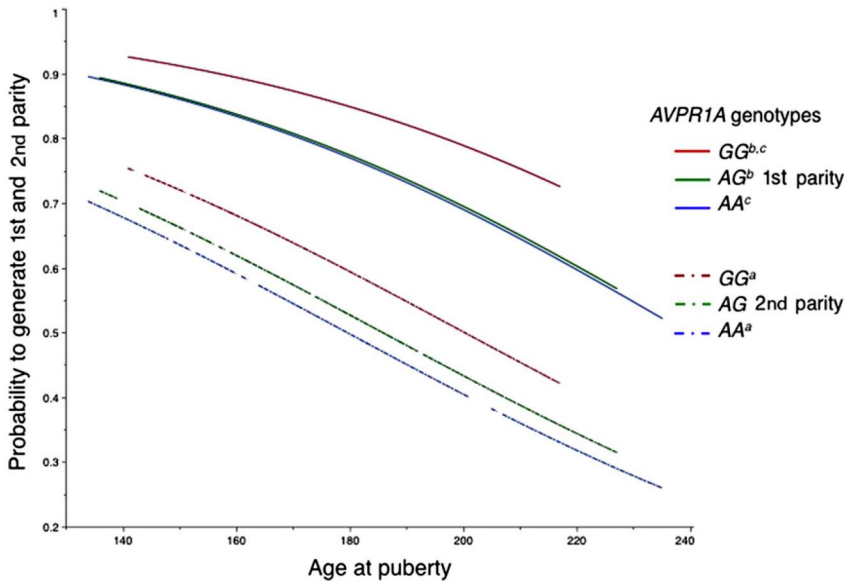

Figure 4. The genotypes of the *G31E* polymorphisms in *AVPR1A* influence the success rate of sows generating the first two litters (genotypes with the same subscript differ:
a, $P < 0.10$;
b,c, $P < 0.05$).

dows displayed parallel to the respective axes. Two of the windows that were associated with both traits (proportion of genetic variance >99% quantile) are located on SSC5 (30 Mb) and SSC16 (3 Mb). The window mapped on SSC5 (30 Mb) includes *AVPR1A* (*arginine vasopressin receptor 1A*; SSC5, 30 Mb), a gene involved in social and sexual behavior (Walum *et al.* 2008; Gobrogge *et al.* 2009).

The effect of age at puberty and the genotype of a nonsynonymous SNP (*G31E*), located in *AVPR1A,* on the probability that females would produce up to two parities were evaluated in a generalized linear mixed model. Across all genotypes, a delay in age at puberty was associated with a reduction in the probability of a female to generate a first or second litter ($P < 0.0001$). The *GG* genotype was associated with higher probability of the gilts to produce the first parity compared to *AG* and *AA* genotypes ($P < 0.05$) and also a higher probability of the sows to produce the second parity compared to the *AA* genotype ($P < 0.10$). Interestingly, the difference in probabilities between the favorable *GG* genotype and the other genotypes in generating the first parity increases with the delay in the expression of age at puberty (Fig. 4). For example, the difference in probability to generate parity 1 between gilts with the favorable *GG* genotype and gilts with the alternate *AA* homozygote genotype was 5% higher in the group of gilts that expressed early estrus (SD < –1) compared to 12.9% higher in the group that expressed estrus late (SD > 1). The difference in probabilities of generating the second parity between genotypes remained similar across the age at puberty spectrum.

Conclusions

Although age at puberty has been shown to be one of the early predictors of reproductive longevity, estrus detection is labor intensive and not used in commercial settings. In this case, the use of genomic information could be applied early in life to predict whether a gilt will be successful reproductively. In our resource population, we have shown that a limited number of informative major markers could reach or exceed the proportion of the variation explained by a full SNP data set, in our case the Porcine SNP60 BeadArray. Evaluation of the predictive potential of major SNPs in a subset of subsequent batches of gilts demonstrated that all SNPs from major 1-Mb windows were able to explain more phenotypic variation compared to single most informative SNPs, most likely due to the changes in linkage disequilibrium status between major SNPs and the QTL. In addition, retraining the subset of major SNP in the evaluation population improved the potential of the GPV to explain phenotypic variation as compared to using GPV derived from SNP effects in the training data. In general, although some of the major SNPs and windows had consistently ranked high and some displayed pleiotropic properties influencing both age at puberty and reproductive longevity, their individual ability to explain substantial phenotypic variation was limited.

Acknowledgments — This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2013-68004- 20370 from the USDA National Institute of Food and Agriculture and also by grants from the National Pork Board, Nebraska Pork Producers Association and University of Nebraska Agricultural Research Division and Hatch funds. We thank Justin Bundy and the UNL Swine Farm team, Jeffrey Perkins, Tom McGargill, Kurt Wenninghoff, Cory Willison and Jessica White for their technical assistance.

References

Bidanel J.P. (2011) Biology and genetics of reproduction, in *Genetics of the Pig* (ed. Rothschild M.F. & Ruvinsky A.), pp 218–42. CAB International. Wallingford, UK.

Fernando R.L. & Garrick D.J. (2008) *GENSEL – User Manual for a Portfolio of Genomic Selection Related Analyses*. Iowa State University, Ames.

Gobrogge K.L., Liu Y., Young L.J. & Wang Z. (2009) Anterior hypothalamic vasopressin regulates pair-bonding and drug-induced aggression in a monogamous rodent. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19144–9.

Hsu W.-L. (2011) *Analysis of Long-term Selection (28 generations) for Reproduction, Growth, and Carcass Traits in Swine*. PhD thesis, University of Nebraska.

Kachman S.D., Spangler M.L., Bennet G.L. *et al.* (2013) Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genetics Selection Evolution* 45, 30.

Milan D., Jeon J.T., Looft C. *et al.* (2000) A mutation in *PRKAG3* associated with excess glycogen content in pig skeletal muscle. *Science* 288, 1248–51.

Rothschild M.F. & Bidanel J.P. (1998) Biology and genetics of reproduction. In: *Genetics of the Pig* (Ed by Rothschild M.F. & Ruvinsky A.), pp 313–343. CAB International, Wallingford, UK.

Saatchi M. & Garrick D.J. (2014) Improving accuracies of genomic predictions by enriching 50K genotypes with markers from 770K genotypes at QTL regions, *Proceedings of the Midwest ADSA/ASAS Meeting*, March 18, 2014, Des Moines.

Serenius T. & Stalder K.J. (2007) Length of productive life of crossbred sows is affected by farm management, leg conformation, sow's own prolificacy, sow's origin parity and genetics. *Animal* 1, 745–50.

Tart J.K., Johnson R.K., Bundy J.W. *et al.* (2013) Genome-wide prediction of age at puberty and reproductive longevity in sows. *Animal Genetics* 44, 387–97.

Van Laere A.S., Nguyen M., Braunschweig M. *et al.* (2003) A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* 425, 832–6.

VanRaden P.M., O'Connell J.R., Wiggans G. & Weigel K.A (2011) Genomic evaluations with many more genotypes. *Genetics Selection Evolution* Mar 2 43, 10.

Visscher P. (2014) Quantitative genetics of complex traits in human populations, *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*, August 17–22, 2014, Vancouver, BC.

Walum H., Westberg L., Henningsson S. *et al.* (2008) Genetic variation in the *vasopressin receptor 1a* gene (*AVPR1A*) associates with pair-bonding behavior in humans. *Proceedings of the National Academy of Sciences of the United States of America* 105, 14153–6.

Supporting information

Additional supporting information is attached to the repository cover page for this article.

Table S1. Genetic variance of age at puberty explained by 1-Mb genome windows.

Table S2. Genetic variance of age at puberty explained by 56,424 SNPs.