5-2014

Genomic Predictions for Age at Puberty and Reproductive Longevity in Sows Using Bayesian Methods

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

GENOMIC PREDICTIONS FOR AGE AT PUBERTY AND REPRODUCTIVE LONGEVITY IN SOWS USING BAYESIAN METHODS

by

Katherine Lindsay Lucot

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Animal Science

Under the Supervision of Professor Daniel Ciobanu

Lincoln, Nebraska

May, 2014

GENOMIC PREDICTIONS FOR AGE AT PUBERTY AND REPRODUCTIVE LONGEVITY IN SOWS USING BAYESIAN METHODS

Katherine L. Lucot, M.S.

University of Nebraska, 2014

Advisor: Daniel Ciobanu

Including marker-assisted selection in breeding programs is potentially more efficient than traditional selection for improving traits that are expensive or difficult to measure. One of the challenges of genomics is the lack of robustness of marker effects across populations and over time (generations) and the cost to commercial producers of high-density arrays. The objective of this study was to analyze differences in the proportion of phenotypic variation explained by different fractions of major 1 Mb windows and SNPs. Using a population of Nebraska Index Line and commercial Large White x Landrace females (n = 1,234) generated in 11 batches, we conducted a genome-wide association analysis for age at puberty (AP) using a Bayes B algorithm with a π value of 0.99 and the concatenation of diet and batch fitted as a fixed effect. A total of 56,424 SNPs explained 0.28 of the phenotypic variation for AP. Analysis of the genetic variance explained by 1 Mb windows across the genome and major SNPs, uncovered major regions associated with AP. The proportion of the phenotypic variation explained by all SNPs within the top 1%, 5%, 10% and 20% windows varied from 0.22 (1% windows; 645 SNPs) to 0.39 (10% windows; 19,362 SNPs). In contrast, the proportion of

the phenotypic variation explained by the most informative SNP from these windows varied from 0.18 (1% windows; 24 SNPs) to 0.48 (20% windows; 259 SNPs). Different π values (0, 0.25, 0.50, 0.75 and 0.99) had a limited effect on the proportion of phenotypic variation explained by the top 1% (0.20 to 0.23) and 10% (0.36 to 0.37) windows. The first seven batches were used as training data (R1 - R7, n = 822) to evaluate the ability of major SNPs and windows to predict AP in subsequent batches. The pooled simple correlation between genomic prediction values (GPV) and adjusted AP phenotypes was 0.18 in R8 - R11 (n = 412) when 56,424 SNPs were used. When GPV were derived using the most informative SNP from each of the top 10% windows or all SNPs from the top 10% windows identified in training, $r_{GPV,AP}$ was 0.18 and 0.12, respectively. Weaker correlations were obtained when the most informative SNP or all of the SNPs from the top 1% windows were used for prediction (0.01 and 0.06, respectively). These results showed that a limited number of SNPs were able to explain proportions of phenotypic variation similar to that obtained from high-density SNP panels.

Keywords: Genomic prediction, puberty, Swine.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Daniel Ciobanu, for his guidance and humor throughout my time as a graduate student. During my graduate career I have grown as a researcher as well as a person. If it were not for his daily jokes and tenacious drive, I would not be presenting this thesis. I would also like to thank the rest of my committee members, Dr. Rodger Johnson, Dr. Stephen Kachman, and Dr. Merlyn Nielsen, for their assistance and boundless knowledge.

This project would never have been possible if it were not for the technical support and help of Justin Bundy and the staff at the Swine Research Unit in Mead, Nebraska. I am also thankful for the aid of laboratory undergraduate staff. I would also like to thank my office mates and fellow graduate students Melanie Trenhaile and especially Taylor Benjamin Engle, who has become like a brother to me.

In addition to my colleagues, I would like to thank my parents Jeffrey and Jennifer Lucot and my little sister Emily, for their infinite support, encouragement, and love throughout my academic career. I would also like to thank Dr. Phillip Miller for his fatherly advice and Sherri and Mike Pitchie, who have basically become my adopted parents during my time at the University of Nebraska-Lincoln. Lastly I would like to thank Adrienne Bhatnagar, Austin Benes, and Dana and Mark Didde, for their unwavering friendship.

Katherine Lindsay Lucot

TABLE OF CONTENTS

CHAPTER 1

1.1 LITERATURE REVIEW

1.1.1 Economic Importance of Reproductive Traits

At the commercial level swine producers have been able to increase productivity by eliminating as much environmental variation as possible as well as developing genetic composites with the highest lean growth. Unfortunately through continually selecting for lean growth and the acceleration of the reproductive cycle, sow culling rates have increased to 50% or more (Hoge and Bates, 2011). One factor that still remains a challenge is the variation in female reproductive longevity. By being able to increase reproductive productivity in sows, swine producers will not have to cull as many animals each year (Faust et al., 1993). When sows are kept around longer the cost to replace them decreases and there is increased sow output (Kroes and Van Male, 1979) with superior genetic potential.

Faust et al. (1993) analyzed the effect of selection for reproductive traits at the nucleus, multiplier, and commercial levels. Over the one, five, and 10-year intervals used for culling, producers were able to reduce sow costs on pig production by $0.75 to $0.91 per pig annually or $7.50 to $9.10 per pig after 10 years of selection in all three of the commercial levels (Faust et al., 1993). Faust et al. (1993) suggested that the first thing commercial producers should do from an economic standpoint, would be to minimize the sow replacement rate. Secondly, they should purchase genetically superior replacement stock to allow for genetic response to continue.

Rodriguez-Zas et al. (2003) looked at the reproductive longevity of 32 herds from Central Illinois over a 7-year time period. They found that in the swine industry 40% to 50% of sows are removed before their third or fourth parity, which is in agreement with previous reports (D'Allaire et al., 1987). As a result more sows are culled right at the age when most replacements just begin to recover their initial cost (Stalder et al., 2003). When comparing lines with superior and inferior reproductive longevity, in their herd, Rodriguez-Zas et al. (2003) found a difference of $52.39 in net income per sow, and a difference of $13.49 in net income per sow between the top two longevity lines. This shows the potential economic advantage that can be had when properly selecting for reproductive longevity that will allow for sows to stay in the breeding herd longer. Stalder et al. (2003) further confirmed the economic importance of sow longevity by analyzing sow farrowing records over a 5-year period. They found that on average the cost of replacing a sow is roughly $200. This cost becomes far higher, when multiplied across the size of a commercial herd.

1.1.2 Genetic Variation of Reproductive Traits

Due to low heritability, selection for reproductive traits represents a difficult task. However, reducing negative environmental effects as well as improved using methods of genetic selection were successful in improving traits such as age at puberty, ovulation rate, litter size, and longevity (Bidanel, 2011). Heritability estimates for reproductive traits range from 0-0.73 for age at puberty, 0-0.76 for total number born, to 0-0.66 for number born alive (Rothschild and Bidanel, 1998; Johnson et al. 1999; Nikkilä et al.,

2013; Table 1.1a). Heritability estimates of sow reproductive longevity range from 0.05 to 0.27, suggesting that genetic improvement can be made (Serenius and Stalder, 2004; Yazdi et al., 2000; Knauer et al., 2010; Nikkilä et al., 2013).

When analyzing factors that affect sow reproductive longevity, age at puberty was found to be significant (Knauer et al., 2010a). Similar results have been reported by Yazdi et al. (2000), Tart et al. (2013b), and İrfan (2013), showing earlier puberty is associated with longer productive life of females. A moderate heritability of age at puberty suggests that selection can be made to decrease age at puberty while improving reproductive longevity (Knauer et al. 2010b; Bidanel et al. 1995; Serenius and Stalder 2006a).

Table 1.1a: Heritability (h^2) estimates for female reproductive traits in swine. Adapted from Rothschild and Bidanel (1998).

Trait	*Number of Estimates*	*Average h^2*	*Range*
Age at Puberty	16	0.37	0.0 – 0.73
Ovulation Rate	18	0.32	0.10 – 0.59
Total Number Born	103	0.11	0.0 – 0.76
Number Born Alive	118	0.10	0.0 – 0.66
Number Weaned	54	0.08	0.0 – 1.0

1.1.3 Selection Response for Reproductive Traits

The most efficient and effective way to improve a trait in a population is through genetic selection. Even though heritability estimates for reproductive traits in general are smaller, selection for improvement has been shown to be effective. Genetic selection for

reproductive traits has proven to be successful in many studies, for traits such as age at puberty, ovulation rate, uterine capacity, and number of piglets born alive (Ruíz-Flores and Johnson, 2001; Zimmerman and Cunningham, 1975). Lamberson et al. (1991) found that by selecting for decreased age at puberty over nine generations, they were able to reduce the mean age at puberty from 174.1 days to 158.5 days. Over five generations of selection for increased ovulation rate, the selected gilts had an average increase of corpa lutea of 0.40 ± 0.09 per generation compared to -0.16 ± 0.05 decrease in corpus lutea per generation in the control line (Zimmerman and Cunningham, 1975). The selection differentials increased to 1.637 for fully formed pigs per generation in a selection line for ovulation rate and litter size, in contrast the control line increased to 0.152 for fully formed pigs per generation (Hsu, 2011).

Serenius et al. (2006b) looked at sow longevity in six commercial lines from six different suppliers: American Diamond Swine Genetics (Prairie City, IA), Danbred North America (Seward, NE), Dekalb-Monsanto DK44 (St. Louis, MO), Dekalb-Monsanto GPK347 (St. Louis, MO), Newsham Hybrids (West Des Moines, IA), and National Swine Registry (West Lafayette, IN) (Serenius et al., 2006b). All of these lines were a composite of Landrace and Large White-Yorkshire crosses, with potentially an introduction of an outside breed during the development of the line (Serenius et al., 2006b). These sows were selected for sow longevity, or the length of their productive life from the time they were entered into the breeding herd until culling or censoring date (800 d). The GPK347 sows had the lowest risk of being culled compared to the other lines in the study. The Nebraska Index Line comprised 50% of the makeup of the GPK347 line; the other 50% was an unrelated Monsanto maternal line. The National

Swine Registry line had a 1.37 time greater risk of being culled compared to the GPK347 line. They also found that later the age at first farrowing increased the risk of being culled in the Danbred line. From this study they concluded that genetic differences between lines contribute to the range of sow longevity, and therefore selection for this trait is possible.

1.1.4 Mapping Loci that Influence the Variation of Reproductive Traits

Reproductive traits such as age at puberty, lifetime number of parities, and number born alive are polygenic and have low heritabilities. Since reproductive performance in the swine industry has economic and welfare importance, there is a strong interest in uncovering quantitative trait loci (QTL) and genes that regulate their variation (Rothschild et al., 1996).

1.1.4.1 QTL Mapping

The animal genome QTL database currently has 9,862 QTL reported for the *Sus scrofa* genome (Hu et al., 2005; PigQTLdb accessible at http://wwwanimalgenome.org/cgi-bin/QTLdb/SS/index), representing about 219 different traits. Out of these there are 284 QTL identified for reproductive traits. This database gathers QTL results from all public studies, differing densities of linkage maps, and sizes of experimental resource populations.

Rohrer et al. (1999) were able to identify QTL affecting age at puberty, ovulation rate, weight of ovaries, uterine capacity, and litter size using a genome-wide scan, as well as a resource population based on a Meishan × White composite population. QTLs were identified on *Sus scrofa* (SSC) chromosomes 4, 8, 13, and 15 for ovulation rate, SSC1 and SSC6 for litter size, and SSC8 for length of uterine horn (Rohrer et al., 1999).

Cassady et al. (2001) used an F_2 resource population created at the University of Nebraska from a cross of a line selected 10 generations for ovulation rate and embryonic survival and a randomly selected control line. They wanted to identify chromosomal regions that harbored QTL affecting reproductive traits in swine. The informativeness of a marker was determined to be the proportion of F_1 meioses for which the allelic line of origin could be determined. Using this method they found evidence for a QTL affecting ovulation rate on SSC9 ($P < 0.05$), a QTL for number of fully formed piglets ($P < 0.05$) and live pigs at birth ($P < 0.10$) were identified on SSC11, and QTLs for number of stillborn pigs were identified on SSC5 and SSC13.

Campbell et al. (2003) were able to identify additional genetic markers for the data set used by Rohrer et al. (1999), using a comparative mapping approach. They selected genes which were mapped to *Homo sapiens* (HAS) chromosome 4, which allowed for a high-resolution comparative map for SSC8, which also provided enough markers to determine whether the QTL for litter size and ovulation rate were segregating in commercial swine populations (Campbell et al., 2003). Onteru et al. (2011) performed a genome-wide association study on 683 female Large White × Landrace composites. Using a Bayes C model they were able to find QTL regions that were highly significantly ($P < 0.01$) associated with total number born and number born alive for the first two

parities (Onteru et al., 2011). Some of the QTL overlap with genes such as: *MEF2C* on *Sus scrofa* chromosome 2 (SSC2), *PLSCR4* and *PLSCR5* on SSC13, which are in the same QTL regions as previously identified to play a role in fat deposition, meat quality, and ovulation rate (Onteru et al., 2011). Schneider et al. (2012b) used similar methods as Onteru et al. (2011) and were able to find 11 QTL for total number born (three on SSC1, three on SSC4, one on SSC13, one on SSC14, two on SSC15, and one on SSC17) and 14 QTL were found for number born alive (four on SSC1, one on SSC4, one on SSC6, one on SSC10, one on SSC13, three on SSC15, and three on SSC17).

Looking at their results from a GWAS performed on 852 gilts, from either NIL or commercial Large White × Landrace crossbreds, Tart et al. (2013) found that the combined SNP effects explained 26% of the phenotypic variation amongst the gilts in age at puberty (Table 1.1b). 77.6% of the SNPs used in the analyses were characterized by minor allelic frequency of at least 0.10. When the genetic variance, explained by 1 Mb windows of the swine genome, was analyzed based on the posterior distributions of the SNP effects, major regions associated with reproductive traits were uncovered. Some of these regions include: SSC1 (31, 94.2-94.9, 287, 94.1-94.9 Mb), SSC3 (16-16.9, 71.1-72 Mb), SSC6 (115.1-116, 144 Mb), SSC8 (36, 37 Mb), SSC9 (21.1-22, 139 Mb), SSC12 (1.2-2, 2.1-3, 11-11.9 Mb), SSC13 (117-117.9, 142.1-142.9 Mb), and SSC14 (19, 28, 66, 68 Mb). Most of these regions harbored large clusters of SNPs associated with age at puberty onset. With the use of marker-assisted selection there is potential for genetic response to selection to increase.

Table 1.1b: Proportion of phenotypic variation explained by SNPs using a Genome-wide Association Study with a Bayesian approach.

Trait	*Proportion of Phenotypic Variation Explained by SNPs*	*Authors*
Age at Puberty	0.26	Tart et al., 2013
Number Born Alive	0.06	Schneider et al., 2012a
Total Number Born	0.04	Schneider et al., 2012a
Lifetime Number Born Alive	0.02 0.15	Tart et al., 2013 Onteru et al., 2010
Lifetime Total Number Born	0.01 0.15	Tart et al., 2013 Onteru et al., 2010
Lifetime Number of Parities	0.19	Tart et al., 2013

1.1.4.2 Candidate Gene Approach

Since there are numerous genes known to affect reproduction in swine, locating them along the *Sus scrofa* genome allows for researchers to find genetic markers and functional mutations associated with phenotypic variation. Previous studies uncovered DNA markers located in 15 candidate genes that affect reproductive traits (Table 1.1c); most of the markers affected litter size traits like total number born or number born alive. For example, *ESR1*, estrogen receptor 1, and *FSHB*, follicle stimulating hormone beta, both have been identified to affect total number born and number born alive (Rothschild et al., 1996; Linville et al., 2001).

Rothschild et al. (1996), found that pigs homozygous for the favorable allele for the Estrogen Receptor gene produced 2.3 more pigs per parity compared to pigs homozygous for the unfavorable allele. The additive substitution effect was 1.2 pigs per

parity per favorable allele and dominance was not detected. Linville et al. (2001) used pigs from the Nebraska Index Line (Johnson et al., 1999) in which selection in one group for eight generations for index of ovulation and embryo survival followed by two-stage selection for ovulation rate and litter size was imposed and compared it with randomly selected controls (Linville et al., 2001). In the selected group, allele A had a frequency of 0.94 and allele B had a frequency of 0.06, with the overall allele frequency for *ESR1* of 0.98. Rothschild et al. (1996) reported that the *ESR1* B allele was significantly associated with increased number of piglets born.

Linville et al. (2001) used a candidate gene approach to determine whether specific loci explained responses in ovulation rate, number of fully formed pigs, number of pigs born alive, stillborn, and mummified pigs at birth observed in two lines selected for ovulation rate and litter size in comparison to a randomly selected control line. Allelic frequencies were determined by the total count of an allele in a line divided by two times the observations in that line. The A allele was determined to be the favorable allele for *PTGS2* (prostaglandin-endoperoxide synthase 2) and *RBP4* (retinol-binding protein 4), whereas the B allele was determined to be the favorable allele for *ESR1* (estrogen receptor), *FSHβ* (follicle stimulating hormone beta), *EGF* (epidermal growth factor), and *PRLR* (prolactin receptor). The selection lines exceeded the control line by 20% to 50% in ovulation rate and litter size at birth; however, none of the markers explained a significant proportion of the responses.

Table 1.1c: DNA markers in candidate genes associated with female reproductive traits in pigs. Adapted from Bidanel (2011).

SSC	Gene	Associated Reproductive Trait(s)*	References
1	Estrogen receptor 1 (*ESR1*)	TNB, NBA, TN	Rothschild et al., 1996; Short et al., 1997; Van Rens et al., 2002; Goliasova and Wolf, 2004; Horogh et al., 2005; Munoz et al., 2007
1	Paired box 5 (*PAX5*)	AP	Kuehn et al., 2009
2	Follicle stimulating hormone beta (*FSHB*)	TNB, NBA, NW, LWW, GL	Li et al., 2008
2	Erythropoietin receptor (*EPOR)*	Uterine Capacity	Vallet et al., 2005; Nonneman et al., 2006
6	Leptin receptor (*LEPR*)	Liter size	Chen et al, 2004a
6	Fucosyl transferase 1 (*FUT1*)	TNB, NBA	Horak et al., 2005; Buske et al., 2006
6	Ring finger protein 4 (*RNF4*)	TNB, NBA	Niu et al., 2009
7	Properdin (*BF*)	NB, NBA	Buske et al., 2005
8	Gonadotropin releasing hormone receptor (*GNRHR*)	OR	Jiang et al., 2001
8	Osteopontin (*OPN*)	TNB, NBA	Korwin-Kossakkowska et al., 2002
8	Leukaemia inhibitory factor (*LIF*)	NBA	Spotter et al., 2009
10	Aldo keto reductase 1C2 (*AKR1C2*)	AP, OR, TN	Nonneman et al., 2006
14	Retinol binding protein (*RBP4*)	TNB, NBA	Rothschild et al., 200; Spotter et al., 2009
16	Prolactin receptor (*PRLR*)	TNB, NBA, AP, OR	Vincent et al., 1998; Drogemuller et al., 2001; Van Rens and Van der Lende, 2002; Van Rens et al., 2003
18	Leptin (*LEP*)	TNB, NBA	Korwin-Kossakowska et al., 2002; Chen et al., 2004b

*AP – age at puberty; OR – Ovulation rate; GL – Gestation length; TN – Number of teats; LP – Lifetime total number born; LNBA – Lifetime number born alive; TNB – Total number born; NBA – Number born alive; NW – Number weaned; LWW – Litter weaning weight.

1.1.5 Bayesian Analyses in Genome-wide Association Studies

The current approach used in selection of economically important quantitative traits is based on phenotypic records of the individual and its relatives. A common method for estimating breeding values for these animals is by best linear unbiased prediction (BLUP; Henderson, 1984). With the development of molecular genetics, one justification for its use in research on livestock is the expectation that information found at the DNA level can lead to faster genetic gain compared to just the use of phenotypic data (Meuwissen et al., 2001).

Quantitative traits are affected by many genes; as a result the benefit of marker-assisted selection is limited by the proportion of phenotypic variation explained by the QTLs. Ideally all QTLs affecting the trait would be used in marker-assisted selection, but a high density marker map defines a large number of chromosomal regions, resulting in more effects to be estimated than phenotypic data points available (Meuwissen et al., 2001). This same problem arises if we assume that comparative mapping efforts will identify approximately 50,000 markers across all genes in the swine genome (Aparicio, 2000). But if we attempt to estimate the allelic effects of these genes we once again do not have enough degrees of freedom to fit all of the effects simultaneously (Lande and Thompson, 1990). BLUP allelic effects can be calculated if there are more effects than data points as long as we assume that all genes explain *a priori* an equal amount of variance (Meuwissen et al., 2001). Having each gene have the same variance is an unrealistic assumption because the dense panels of markers allow for exploitation of linkage disequilibrium between QTL and genome-wide markers (de los Campos et al.,

2013). In Bayesian models, variances are assumed to come from a prior distribution (Meuwissen et al., 2001).

Bayesian models are broken down into the Bayesian alphabet: Bayes A, Bayes B, Bayes C, Bayes Cπ, and Bayes Dπ. Each Bayesian approach has its own parameters to define the way the analysis runs. Bayes A: has data modeled at two levels, the model of the data and the model of the variances of the loci. The data model is equal to that of BLUP except that the chromosomal segments are all different and are estimated from the variances (Meuwissen et al., 2001). The model at the variance level, estimates based off of the combined information from the prior distribution of the variances as well as the data. All markers are also used in the model at once, meaning they all have an effect, which can potentially lead to over parameterization of the model. Bayesian analyses also use a π value, which dictates what percentage of the markers will not have an affect. In Bayes A the π value is set to 0.00 meaning that all markers have an effect in the model.

Bayes B assumes, that in the distribution of genetic variances across the loci that there are many loci that indeed do not actually have an effect at all, while only some loci have an effect (Meuwissen et al., 2001); meaning that each SNP has its own individual variance, some with an effect and most without. These variances can be weighted by changing the degrees of freedom. A π value is included in the analysis as well, but it can range anywhere from 0 to 1. The π value must be greater than 0.00, otherwise it is not a Bayes B analysis; instead it is one of the other letters, like Bayes A.

Bayes C assumes that all of the SNPs have the same genetic variance as well as the π value is known. In Bayes Cπ the priors of all SNP effects have a common variance (Habier et al., 2011), but the π value is unknown, meaning that the analysis will

continually fit different π values until it finds the one that fits your data best. Bayes Dπ denotes that each SNP has its own variance, but similar to Bayes Cπ, the π value is unknown (Habier et al., 2011). The type of data you are working with or what you are trying to achieve from your data dictates which Bayesian model should be used. All of the methods have some things in common; such as you must set the number of iterations you would like performed (chain length), the total number of iterations that will be ignored (burn in), variation due to genetics, and lastly variation due to everything else (residual; Meuwissen et al., 2001).

1.1.6 Genomic Predictions

In the commercial swine industry, the majority of selection for genetic response happens at the nucleus level. It has been shown though that selecting at the nucleus level has limitations in predicting (correlation of 0.8 or less) how the crossbreds at the multiplier level will perform (Lutaaya et al., 2001).

One potential solution to this problem is the use of genomic selection (Schneider et al., 2012b). Genomic selection uses the genomic information from each animal and allows for the development of genomic breeding values (GBV) without actually having to phenotype the individuals (Shumbusho et al., 2013). This method allows for selection of traits, which are potentially difficult or expensive to measure, the potential to select animals early in life, as well as the possibility of selecting individuals for traits limited by sex. Genomic selection has the potential to speed up annual genetic gain. It also allows

for selection at a younger age, which will reduce generation interval. This benefit is greater in cattle than pigs but may be important in pigs for reproduction and longevity.

At the commercial/multiplier level the swine genome was scanned to look for markers that affect traits such as reproductive longevity or carcass traits (Dekkers, 2007). The objective of the this study was to simulate the use of high-density marker genotypes on the purebred population in the nucleus herds and samples of their commercial crossbred descendants in the field, to then use performance data of genotyped animals to estimate breed effects of the marker haplotypes on the commercial crossbreds' performance. Dekkers (2007) then used markers associated with the reproduction and carcass traits in a crossbred commercial population to try and predict the phenotype in the purebred population. Purebred selection, based off of phenotypes from the nucleus, resulted in a genetic response of 0.38 σ_P in the generational performance of the crossbreds. When the commercial crossbreds', which were paternal half siblings to the purebreds, phenotypic data were included in the selection, response increased by 22% to 0.46 σ_P. One deficiency in this method is the rate of inbreeding per generation increased from 2.1% to 3.0%. Dekkers (2007) also showed that selection-based estimated breeding values derived from the simulated purebred phenotypic data did not result in substantial increases in the genetic improvement of the simulated commercial crossbred performances. When the same selection was performed on marker-based estimated breeding values from commercial crossbreds data, the genetic response exceeded > 0.52. This method did a better job of predicting the phenotypes of the purebreds, instead of having the purebreds predict the crossbreds.

When trying to predict cattle phenotypes in progeny Weber et al. (2012) found that the accuracy of their molecular breeding values was dependent upon the size of the training set, the prediction population, marker density, heritability of the trait, as well as the statistical method. They used multibreed populations to train on as well as to evaluate on, which led towards a bias of the most highly represented breeds, compared to the underrepresented breeds. They concluded that they would expect molecular breeding values trained and evaluated within one breed would generate greater accuracies. This was in agreement with Habier et al. (2007, 2010) who found the accuracies of molecular breeding values to be directly related to genetic relationship with the training population, therefore single-breed training populations may maintain a greater average relationship compared to multibreed training populations. Weber et al. (2012) concluded that to increase the accuracy for genomic predictions genotyping density should be increased as well as the training population size, to allow for greater representation of smaller breeds in the training population.

Meuwissen et al. (2001) used simulation data to see how accurately they could predict the phenotypes of animals in their data set based on genomic information. The accuracy of genomic selection is affected by the accuracy of the marker effect estimates and the correlations between genotyped markers and the underlying QTL (Goddard, 2009). They found that using an analysis method that assumes prior distribution for the variances is more accurate when using a marker map that spans the whole genome and that selecting based on genomic breeding values has potential to substantially increase the rate of genetic gain (Meuwissen et al., 2001).

Using different marker panels and methods for predicting genomic breeding values across and within breeds, Erbe et al. (2012), found that their ability to predict the phenotypes of Jersey dairy cattle from a training population of Jersey and Holstein cattle, increased when uninformative SNPs from the model were removed or their effect was set to zero. They believed that this increased their ability to predict because when using all of the SNPs they greatly increased the number of SNP effects to be estimated without increasing the number of animal records. This increased the size of estimation errors, which in turn eroded the accuracy of the genomic estimated breeding value of the animals (accuracy of 0.52 to 0.43; Erbe et al., 2012).

Cleveland and Hickey (2013) wanted to find a practical way to implement genomic selection in swine using molecular breeding values. They used data from 4,763 moderately to highly related pigs from a single nucleus pig line, and all animals were genotyped. Using the Porcine SNP60K BeadChip (Illumina) they evaluated total number born for all of the females (Cleveland and Hickey, 2013). Genomic breeding values were calculated for all animals using a single-step genomic evaluation (Aguilar et al., 2010), which uses all available SNPs to construct a genomic relationship matrix, which is then combined with a standard numerator relationship matrix that includes all ungenotyped individuals. From their results Cleveland and Hickey concluded that the uneven distribution of imputation accuracy across the swine genome suggested that the low-density SNP panels could be improved. Some of their strategies for this included developing a SNP panel which used SNPs based off of *a priori* evidence of the usefulness of that SNP or based off of knowledge of the characteristics of particular genomic regions, such as higher recombination spots (Cleveland and Hickey, 2013).

From this conclusion they proposed the idea that each animal could be genotyped using a SNP panel composed of only informative SNPs, which would in turn increase the accuracy of the genomic breeding value, even if the total number of SNPs on the panel was smaller than the amount on the Porcine SNP60K BeadChip.

1.1.7 Conclusion

With reproductive failure being of major concern to the swine industry, research continues to be done to try and reduce the occurrence; for economic reasons as well as welfare concerns. Age at puberty is the earliest indicator of longevity and has a moderate heritability; therefore development of methods to reduce age at puberty or detect it sooner would be a large advantage for producers so they would know which gilt is more likely to be successful longevity wise.

1.2 LITERATURE CITED

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

Aparicio, S. A. R. J. 2000. How to count human genes. Nat Genet. 25:129-130.

Bidanel, J. P., J. Gruand, and C. Legault. 1995. Genetic variability of age and weight at puberty, ovulation rate and embryo survival in gilts and relations with production traits. Genet Sel Evol. 28:103-115.

Bidanel, J. P. Biology and genetics of reproduction. In M. F. Rothschild and A. Ruvinsky (eds.) 2011. The genetics of the pig. p 228-232. C.A.B. International, Oxfordshire, UK.

Buske, B., I. Sternstein, M. Reissmann, P. Reinecke, and G. Brockmann. 2006. Analysis of association of *GPX5, FUT1* and *ESR2* genotypes with litter size in a commercial pig cross population. Archiv für Tierucht. 49:259-268.

Campbell, E. M. G., D. Nonneman, and G. A. Rohrer. 2003. Fine mapping a quantitative trait locus affecting ovulation rate in swine on chromosome 8. J. Anim. Sci. 81:1706-1714.

Cassady, J. P., R. K. Johnson, D. Pomp, G. A. Rohrer, L. D. Van Vleck, E. K. Spiegel, and K. M. Gilson. 2001. Identification of quantitative trait loci affecting reproduction in pigs. J. Anim. Sci. 79:623-633.

Chen, C., T. Chang, and H. Su. 2004a. Genetic polymorphisms in porcine leptin gene and their associatin with reproduction and production traits. Austrailian Journal of Agricultural Research. 55:699-704.

Chen, C., C. T. Chang, and H. Y. Su, 2004b. Characterization of porcine leptin receptor polymorphisms and their association with reproduction and production traits. Animal Biotechnology. 15:89-102.

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

D'Allaire, S., T. E. Stein, and A. D. Leman. 1987. Culling patterns in selected Minnesota swine breeding herds. Can. J. Vet. Res. 51:506-512.

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. 2007. Marker-assisted selection for commercial crossbred performance. J. Anim. Sci. 85:2104-2114.

Drogemuller, C., H. Hamann, and O. Distl. 2001. Candidate gene markers for litter size in different German pig lines. J. Anim. Sci. 79:2565-2570.

Dwane R. Zimmerman and P. J. Cunningham. 1975. Selection for Ovulation Rate in Swine: Population, Procedures, and Ovulation Response. J. Anim. Sci. 40:61-69.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goaswami, 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. 95:4114-4129.

Faust, M. A., O. W. Robison, and M. W. Tess. 1993. Genetic and economic analyses of sow replacement rates in the commercial tier of a hierarchical swine breeding structure. J. Anim. Sci. 71:1400-1406.

Goddard M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica (The Hague) 136:245–257.

Goliasova, E. and J. Wolf. 2004. Impact of the ESR gene on litter size and production traits in Czech Large White pigs. Animal Genetics. 35:293-297.

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

Habier, D., J. Tetens, F. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:5.

Habier, D., R. L. Fernando, K. Kizilkaya, D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics. 12:186:1471-2105.

Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, Ontario, Canada.

Hoge, M. D., and R. O. Bates. 2011. Developmental factors that influence sow longevity. J. Anim. Sci. 89:1238-1245.

Horak, P., T. Urban, and J. Dvorak. 2005. The *FUT1* and *ESR* genes – their variability and associations with reproduction in Prestic Black-Pied sows. Jouranl of Animal Breeding and Genetics. 122:210-213.

Horogh, G., A. Zsolnai, I. Komlosi, A. Nyiri, I. Anton, and L. Fesus. 2005. Oestrogen receptor genotypes and litter size in Hungarian Large White pigs. Journal of Animal Breeding and Genetics. 122:56-61.

Hsu, Wan-Ling. (2011). Analysis of Long-term Selection (28 generations) for Reproduction, Growth, and Carcass Traits in Swine. Theses and Dissertations in Animal Science, Paper 34. Retrieved February, 21, 2014, from the Animal Science Department at DigitalCommons@University of Nebraska – Lincoln.

Hu, Z. L., S. Dracheva, W. Jang, D. Maglott, J. Bastiaansen, M. F. Rothschild, and J. M. Reecy. 2005. A QTL resource and comparison tool for pigs: PigQTLDB. Mammalian Genome. 16(10):792-800.

İrfan T. U. R. 2013. General reproductive properties in pigs. Turk J Vet Anim Sci. 37:1-5.

Jiang, Z., J. P. Gibson, A. L. Archibald, and C. S. Haley. 2001. The porcine gonadotropin-releasing hormone receptor gene (*GNRHR*): genomic organization, polymorphisms, and association with the number of corpora lutea. Genome. 44:7-12.

Johnson, R. K., M. K. Nielsen, and D. S. Casey. 1999. Responses in ovulation rate, embryonal survival, and litter traits in swine to 14 generations of selection to increase litter size. J. Anim. Sci. 77:541-557.

Korwin-Kossakowska, A., M. Kamyczek, D. Cieslak, M. Pierzchaa, and J. Kury. 2002. The effect of leptin (*LEP*), leptin receptor (*LEPR*), and osteopontin (*OPN*) gene polymorphism on selected reproduction traits of synthetic Line 990 sows. Animal Science Papers and Reports. 20:159-168.

Knauer, M., K. J. Stalder, T. Serenius, T. J. Baas, P. J. Berger, L. Karriker, R. N. Goodwin, R. K. Johnson, J. W. Mabry, R. K. Miller, O. W. Robison, and M. D. Tokach. 2010a. Factors associated with sow stayability in 6 genotypes. J. Anim. Sci. 88:3486-3492.

Knauer, M. T., J. P. Cassady, D. W. Newcom, and M. T. See. 2010b. Estimates of variance components for genetic correlations among swine estrus traits. J. Anim. Sci. 88:2913-2919.

Kroes, Y., and J. P. Van Male. 1979. Reproductive lifetime of sows in relation to economy of production. Livest. Prod. Sci. 6:179-183.

Kuehn, L. A., D. J. Nonneman, J. M. Klindt, and T. H. Wise. 2009. Genetic relationsihps of body composition, serum leptin, and age at puberty in gilts. J. Anim. Sci. 87:477-483.

Lamberson, W. R., R. K. Johnson, D. R. Zimmerman, and T. E. Long. 1991. Direct responses to selection for increased litter size, decreased age at puberty, or random selection following selection for ovulation rate in swine. J. Anim. Sci. 69:3129-3143.

Lande, R., and R. Thompson. 1990. Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics. 124:743-756.

Li, F. E., S. Q. Mei, C. Y. Deng, S. W. Jiang, B. Zuo, R. Zheng et al. 2008. Association of microsatellite flanking *FSHB* gene with reproductive traits and reproductive tract components in pigs. Czech Journal of Animal Science. 53:139-144.

Linville, R. C., D. Pomp, R. K. Johnson, and M. F. Rothschild. 2001. Candidate gene analysis for loci affecting litter size and ovulation rate in swine. J. Anim. Sci. 79:60-67.

Lutaaya, E., I. Misztal, J. W. Mabry, T. Short, H. H. Timm, and R. Holzbauer. 2001. Genetic parameter estimates form joint evaluation of purebreds and crossbreds in swine using the crossbred model. J. Anim. Sci. 79:3002-3007.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics. 157(4):1819-1829.

Munoz, G., C. Ovilo, J. Estelle, L. Silio, A. Fernandez, and C. Rodriguez. 2007. Association with litter size of new polymorphisms on *ESR1* and *ESR2* genes in a Chinese-European pig line. Genetics Selection Evolution. 39:195-206.

Nikkilä, M. T., K. J. Stalder, B. E. Mote, M. F. Rothschild, F. C. Gunsett, A. K. Johnson, L. A. Karriker, M. V. Boggess and T. V. Serenius. 2013. Genetic associations for gilt growth, compositional, and structural soundness traits with sow longevity and lifetime reproductive performance. J. Anim. Sci. 91:1570-1579.

Nonneman, D. J., T. H. Wise, J. J. Ford, L. A. Kuehn, and G. A> Rohrer. 2006. Characterization of the aldo-keto reductase 1C gene cluster on pig chromosome 10: possible associations with reproductive traits. BMC Veterinary Research 2:28.

Onteru, S. K., B. Fan, Z-Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. A whole-genome association study for pig reproductive traits. Animal Genetics. 43:18-26.

Onteru, S. K., B. Fan, M. T. Nikkilä, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2010. Whole-genome association analyses for lifetime reproductive traits in pigs. J. Anim. Sci. 89:988-995.

Rodriguez-Zas, S. L., B. R. Southey, R. V. Knox, J. F. Connor, J. F. Lowe, and B. J. Roskamp. 2003. Bioeconomic evaluation of sow longevity and profitability. J. Anim. Sci. 81:2915-2922.

Rohrer, G. A., J. J. Ford, T. H. Wise, J. L. Vallet, and R. K. Christenson. 1999. Identification of quantitative trait loci affecting female reproductive traits in a multigeneration Meishan-White composite swine population. J. Anim. Sci. 77:1385-1391.

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

Rothschild, M. F., C. Jacobson, D. Vaske, C. Tuggle, L. Wang, T. Short, G. Eckardt, S. Sasaki, A. Vincent, D. McLaren, O. Southwood, H. Van Der Steen, A. Mileham, and G. Plastow. 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proc. Natl. Acad. Sci. USA. 93:201-205.

Rothschild, M. F., L. Messer, A. Day, R. Wales, T. Short, O. Southwood, and G. Plastow. 2000. Investigation of the retinol-binding protein 4 (*RBP4*) gene as a candidate gene for increased litter size in pigs. Mammalian Genome. 11:75-77.

Ruíz-Flores, A. and R. K. Johnson. 2001. Direct and correlated responses to two-stage selection for ovulation rate and number of fully formed pigs at birth in swine. J. Anim. Sci. 79: 2286-2297.

Schneider, J. F., L. A. Rempel, and G. A. Rohrer. 2012a. Genome-wide association study of swine farrowing traits. Part I: Genetic and genomic parameter estimates. J. Anim. Sci. 90:3353-3359.

Schneider, J. F., L. A. Rempel, W. M. Snelling, R. T. Wiedmann, D. J. Nonneman, and G. A. Rohrer. 2012b. Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data'. J. Anim. Sci. 90:3360-3367.

Serenius, T., and K. J. Stalder. 2004. Genetics of length of productive life and lifetime prolificacy in the Finnish Landrace and Large White pig populations. J. Anim. Sci. 82:3111-3117.

Serenius, T. and K. J. Stalder. 2006a. Selection for sow longevity. J. Anim. Sci. 84:E166-E171.

Serenius, T., K. J. Stalder, T. J. Baas, J. W. Mabry, R. N. Goodwin, R. K. Johnson, O. W. Robison, M. Tokach, and R. K. Miller. 2006b. National Pork Producers Council Maternal Line National Genetic Evaluation Program: A comparison of sow longevity and trait associations with sow longevity. J. Anim. Sci. 84:2590-2595.

Short, T. H., M. F. Rothschild, O. I. Southwood, D. G. McLaren, A. De Vries, H. Van der Steen et al. 1997. Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. J. Anim. Sci 75: 3138-3142.

Shumbusho, F., J. Raoul, J. M. Astruc, I. Palhiere, and J. M. Elsen. 2013. Potential benefits of genomic selection on genetic gain of small ruminant breeding programs. J. Anim. Sci. 91:3644-3657.

Spotter, A., S. Muller, H. Hamann, and O. Distl. 2009. Effect of polymorphisms in the genes for *LIF* and *RBP4* on litter size in two German pig lines. Reproduction in Domestic Animals. 44:100-105.

Stalder, K. J., R. C. Lacy, T. L. Cross, and G. E. Conatser. 2003. Financial impact of average parity of culled females in a breed-to-wean swine operation using replacement gilt net present value analysis. J. Swine. Health Prod. 11:69-74.

Tart, J. K., R. K. Johnson, J. W. Bundy, N. N. Ferdinand, A. M. McKnite, J. R. Wood, P. S. Miller, M. F. Rothschild, M. L. Spangler, D. J. Garrick, S. D. Kachman, and D. C. Ciobanu. 2013. Genome-wide prediction of age at puberty and reproductive longevity in sows. Animal Genetics. 44:4:387-397.

Vallet, J. L., B. A. Freking, K. A. Leymaster, and R. K. Christenson. 2005. Allelic variation in the erythropoietin receptor gene is associated with uterine capacity and litter size in swine. Animal Genetics. 36:97-103.

Van Rens, B. T. T. M. and T. Van der Lende. 2002. Litter size and piglet traits of gilts with different prolactin receptor genotypes. Theriogenology. 57:883-893.

Van Rens, B. T. T. M., P. N. De Groot, and T. Van der Lende. 2002. The effect of estrogen receptor genotype on litter size and placental traits at term in F2 crossbred gilts. Theriogenology. 57:1635-1649.

Van Rens, B. T. T. M., G. J. Evans, and T. Van der Lende. 2003. Components of litter size in gilts with different prolactin receptor genotypes. Theriogenology. 59:915-926.

Vincent, A. L., G. Evans, T. H. Short, O. I. Southwood, G. S. Plastow, C. K. Tuggle, and M. F. Rothschild. 1998. The prolactin receptor gene is associated with increased litter size in pigs. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia. 27:15-18.

Weber, K. L., R. M. Thallman, J. W. Keele, W. M. Snelling, G. L. Bennett, T. P. L. Smith, T. G. McDaneld, M. F. Allan, A. L. Van Eenennaam, and L. A. Kuehn. 2012. Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. J. Anim. Sci. 90:4177-4190.

Yazdi, M. H., L. Rydhmer, E. Ringmar-Cederberg, N. Lundeheim, and K. Johansson. 2000. Genetic study of longevity in Swedish Landrace sows. Livest. Prod. Sci. 63:255-264.

CHAPTER 2: Genomic Predictions for Age at Puberty and Reproductive Longevity in Sows Using Bayesian Methods

2.1 INTRODUCTION

Improving reproductive longevity in maternal lines, using traditional quantitative genetic approaches, is challenging due to the low heritability and the contribution of many polymorphic genes, each with relatively small effects that play a role in the phenotypic variation amongst the animals. Measuring reproductive traits is not only time consuming and laborious but it is also costly. A slight improvement in these reproductive traits could have an important economic impact due to the decrease of input costs for replacement of breeding gilts. There is a need for the ability to predict future phenotypes of gilts in the breeding herds, not only to save money on input costs but also for animal welfare.

Currently there are limited approaches to accurately predict if a gilt will have a successful reproductive career. One trait that has been shown to help predict reproductive longevity is the age at which a gilt reaches puberty (Serenius and Stalder, 2007). Previous studies have shown that the younger a gilt expresses puberty the more likely she will be successful in her reproductive lifetime, with more piglets born as well as more successful parities (Tart et al., 2013b). Using the UNL Longevity gilts (R1 – R6), Johnson et al. (2011) showed that the probability of a sow producing a fourth parity if they produced a first parity increased the sooner their age at puberty was (Figure 2.1a; Figure 2.1b; Figure 2.1c; Figure 2.1d)

The objective of this project was to evaluate if the prediction of phenotypes of future generations of gilts could be achieved based on genomic information obtained from a training population. Using high-density genotypes, a genome-wide association study was employed based on a Bayes B model, uncovering regions and SNPs that influence phenotypic variation of age at puberty, lifetime total number of piglets born, and lifetime total number of parities. Posterior SNP effects from the training population were used to estimate genomic prediction values (GPV) and predict the phenotypes in subsequent generations of validating data sets.

2.2 Materials and Methods

2.2.1 Population

Phenotypic and genotypic data were collected on a population of Nebraska Index Line and commercial Large White × Landrace females (n = 1,234), generated in 11 batches. Large White × Landrace crossbreds originated from two commercial breeding programs in 1981. These sows were split into two lines: Line 1 and Line 2. Line 1 was split into Line 1 and Line 5 while Line 2 was split into Line 2 and Line 4. Lines 4 and 5 were then combined to create Line 45, which then became known as the Nebraska Index Line. The Nebraska Index Line (NIL) was selected for reproductive traits, over 32 years (Hsu, 2011). Two lines of gilts, selection line and control line, underwent eight generations of two-stage selection for ovulation rate and litter size (Ruíz-Flores and Johnson, 2001). Starting at generation eight through 16, gilts were selected for ovulation rate and boars were selected from litters where their dams were selected for ovulation rate. Generations 17 through 19 were selected for increased number of pigs born alive and increased birth weight. In generation 20 the males and females were reciprocally crossed and litter traits were measured. In generations 21 through 28, pigs were selected for increased litter size, increased growth rate, decreased backfat, and increased longissimus muscle (Hsu, 2011). During the entire selection period, the selection differentials increased in NIL. During generations 20 through 26 in the selected lines, the selection differentials increased to 1.637 for fully formed pigs per generation, where as the control line increased to 0.152 for fully formed pigs per generation (Hsu, 2011). The

NIL females were crossed with NIL males and parity one offspring were then selected for NIL replacements. Parity two offspring were then crossed with commercial Landrace. R1 – R4 females were NIL × Newsham Choice Genetics Landrace and R5 – R11 were NIL × Danbred North America Landrace. The NIL × commercial Landrace dams (Longevity dams) were then crossed with pooled commercial Duroc semen from Danbred North America. The piglets produced by the Longevity dams, through parity four, were recorded for litter traits and then sold to market.

2.2.2 Nutrition, Culling, and Phenotypes

At day 0 piglets were weighed and at day 21 piglets were weighed and weaned. At day 45 gilts were randomly selected (≈ 140 females) and weighed. Starting at day 123 of age, gilts were fed one of three different diet regimens (for R5 – R11). The diets consisted of either a full caloric diet of corn-soybean meal, a corn-soybean meal diet with a 20% caloric reduction, or lastly a corn-soybean meal diet with a 20% caloric reduction as well as a Lysine reduction (Tart et al., 2013a). R1 – R4 were fed one of two diets: an ad libitum corn-soybean meal diet or a 20% caloric restricted diet. Beginning at day 140, age at puberty was determined as the first day a gilt displayed signs of estrus, such as standing, redness of the vulva, or discharge, in the presence of a boar. Estrus detection continued through day 240, until all gilts in the pen expressed estrus twice (Tart et al., 2013a; Figure 2.2a). At day 240, gilts were moved to the breeding barn. Gilts stayed in production through four parities, except batch 6, which was only three parities (Tart et al.,

2013a). Lifetime traits, number of parities and total number born, were measured as the total number of parities a gilt had before she was culled, the maximum allowed being four, and the summation of all piglets born to her (Figure 2.2b; Figure 2.2c).

In our herd gilts were culled after one unsuccessful service, after four parities, for soundness and health issues, or other reproductive failures such as abortion. Litter sizes, after each parity, were summed to obtain lifetime productivity.

2.2.3 Genotyping

Tissue samples were collected from every gilt, via an ear notch or tail clip. DNA was isolated using the Qiagen DNeasy and Puregene kits. DNA quality and quantity were tested by gel electrophoresis and Nanodrop (Thermo Scientific) spectrophotometry (Tart et al., 2013a). Genotyping was performed using the Porcine SNP60K BeadChip (Illumina). All genotypes with a quality score below 0.4 were removed and replaced with allelic frequencies. SNPs and individual samples with a call rate below 0.8 were excluded as well.

2.2.4 Statistics and Genome-wide Association Studies (GWAS)

A genome-wide association study (GWAS) for age at puberty, lifetime number of parities, and lifetime total number born was performed via GenSel (http://bigs.ansci.iastate.edu/bigsgui/login.html?redir=/bigsgui/) software using Bayes B, with sire line, diet, and batch included as fixed effects. Using a π value of 0.99, 41,000

iterations were run with the first 1,000 as burn in. Potential pleiotropic sources of genetic variation for age at puberty and reproductive longevity were analyzed using a pair-wise correlation of GPV, comparing the genetic variance explained by 1 Mb windows across the swine genome following Bayesian analysis (Tart et al., 2013a).

2.2.4.1 Heritability

Heritabilities were estimated using a linear mixed model. Batch and diet were considered as fixed effects and sire and litter were random effects. Heritability estimates were obtained by multiplying sire variance by four, to get the breeding value variance, which we then divided by the phenotypic variance (summation of sire, litter, and residual variances) acquired from the model. The same model was used for age at puberty, lifetime total number born, and lifetime total number of parities. Standard errors were calculated as 4 × standard error of the interclass correlation. The interclass correlation for half-siblings is the sire variance over the phenotypic variance, which is equal to ¼ heritability (h^2; Mackay, 1996). This made our overall equation to be:

$$\text{s.e. } h^2 = 4\sqrt{\left[\frac{2\left[1+(n-1)t\right]^2(1-t)^2}{n(n-1)(N-1)}\right]}$$

Where N is equal to the number of sires (AP N= 91; LT-NP N = 67; LT-TNB N = 67) and n is equal to approximately how many offspring per sire (AP n ≈ 15; LT-NP ≈ 14; LT-TNB ≈ 14).

2.2.4.2 Bayes Cπ

To make sure that our input π value was correct, we ran a Bayes Cπ analysis. We input a π value of 0.00 to start with. Bayes Cπ assumes that the π value is unknown and the model selects different π values to find which one best fits the data, even though you already have prior information on your variances. You are able to determine the appropriate π value for the data when it converges following the number of iterations you specified.

2.2.4.3 Bayes B

All of the analyses were performed for age at puberty used a Bayes B model with a genotypic variance of 90.37 and a residual variance of 253.12 (Table2.2a). Bayes B assumes that each SNP has its own variance, and we can weight these variances by changing the degrees of freedom (Habier et al., 2011). We also input a π value of 0.99 for our original Bayesian (B) analyses combining our experimental batches. The π value of 0.99, assumes that 99% of the SNPs used in the analysis have no affect. The following genetic and residual variances and π values were used in the Bayes B analyses.

Table 2.2a: Prior variances and π values for Bayesian Analyses

Trait	*Genetic Variance*	*Residual Variance*	*π Value*
Age at Puberty	90.37	253.12	0.99
Lifetime Total Number Born	3.41	301.83	0.99
Lifetime Number of Parities	0.23	1.00	0.99

We also evaluated the influence of 1 Mb windows and SNPs on the phenotypic variance of age at puberty and the ability to predict phenotypes. In these analyses the 1 Mb windows were ranked based on their genetic variance, with the window with the most variance, considered as major and being the 'top' window. We also ranked the markers in each window, based on their estimated genetic variance. The marker with the greatest estimated variance in a window was classified as the 'top' marker in that window. In these analyses the 1 Mb windows and SNPs were ranked based on genetic variance; we then re-ran the Bayesian (B) analyses with either a) all SNPs in the top percentages of windows, or b) the top SNP in the top percentages of windows. Different percentages of top windows were used: 1%, 5%, 10%, 20%, 30%, 40%, 50%, and 100%.

2.2.4.4 Bayes B with Varying π Values

Multiple Bayes B analyses with different π values were performed to see if the π value affected the proportion of variation explained by the markers, as well as the correlation ($r_{\hat{G}P}$) between the GPV and the phenotypes. We used various π values from 0.00 to 0.99 and different combinations of batches as training populations to then evaluate different data sets.

2.2.4.5 Prediction of Phenotype Based on Individual Genomic Prediction Values (GPV)

In a prediction analysis, each animal's genomic prediction value is calculated based on the posterior mean of the SNP effects. Animals with extreme individual and litter average GPV were selected and used in the prediction of phenotypes. Predictions were based on the phenotypic expression of age at puberty. Genomic prediction values were sorted from smallest (early age at puberty) to largest (late age at puberty) and we selected the three animals with the smallest GPV and the three with the largest GPV. After we selected the six animals we then ranked all of the evaluation animals based off their age at puberty, with the earliest puberty being first, and the latest age at puberty being the last animal. We estimated average GPV for each litter and selected litters with the most extreme GPV and within litter we selected three animals from each end of the GPV distribution with the earliest and three with latest individual GPV. These animals with extreme individual and litter GPVs were used in the prediction of phenotypes.

2.2.4.6 Candidate Gene Identification

Using the Bayesian analyses' output, major 1 Mb windows that included large clusters of SNPs that explained the largest amount of phenotypic variation (Figure 2.2d; Figure 2.2e; Figure 2.2f), were extended by 1 Mb on each side to search for candidate genes using the *Sus scrofa* Build 10.2 assembly. Ensembl's

(http://www.ensembl.org/index.html) BioMart and DAVID were then used in analysis of gene ontology and pathways of major pleiotropic QTL regions (Tart et al., 2013a).

2.3 Results and Discussion

2.3.1 Heritabilities

Our calculated heritability was 0.42 (s.e. = 0.09) for age at puberty (n = 1,234), 0.01 (s.e. = 0.06) for lifetime number of parities (n = 903), and 0.05 (s.e. = 0.06) for lifetime total number born (n = 903). These estimates are similar with previous reports (Rothschild and Bidanel, 1998).

2.3.2 Bayesian Analyses

In our Bayes Cπ analysis we used 150,000 iterations, however since our data set is small, our π value never converged, and as a result we relied on previous π data provided by previous reports (Onteru et al., 2011; Tart et al., 2013; Figure 2.3a).

Following our Bayes Cπ analysis we performed multiple Bayes B analyses with varying proportions of π values ranging from 0.00 to 0.99. Using the top 1% of the windows, we trained on R1 – R4 to evaluate on R5 – R10, and the correlation decreased from 0.04 when using a π value of 0.00, to 0.01, when using a π value of 0.99. The same trend was observed when using all of the other batch combinations (train on R5 – R7 evaluate R8 – R10, train on R1 – R7 evaluate R8 – R10, and lastly train on R1 – R10 evaluate R1 – R10; Table 2.3a). When using the top 10% of the windows, the markers were trained on R1 – R4 and evaluated on R5 – R10. We found that increasing the

percentage of windows used, increased the $r_{\hat{G}P}$, but once again the π value did not affect the $r_{\hat{G}P}$ substantially; when using a π value of 0.00 the $r_{\hat{G}P}$ was 0.05 and when we increased the π value to 0.99, the $r_{\hat{G}P}$ decreased to 0.03. The same trend was found once again in all of the other batch combinations (Table 2.3b).

2.3.3 Genome-wide Association Study (GWAS)

A GWAS analysis was performed using 11 batches of pigs (n = 1,234) while including only the high quality SNPs from the Porcine SNP60K BeadChip (n = 56,424). Batches one through four (R1-R4) had a different sire line compared to batches five through 11 (R5-R11). The batches were grouped in different combinations to analyze differences in the proportion of the variance explained, location of the major widows and the influence of the sire line on the analysis. A Bayes B analysis was ran for R1 – R4 (n = 547), R5 – R7 (n = 275), R1 – R7 (n = 822), R1 – R10 (n = 1,117), and R1 – R11 (n = 1,234).

When we used the entire data set R1 – R11 (n = 1,234) the markers explained 28% of the phenotypic variation. The genetic and residual variances (Table 2.3c) and proportion of phenotypic variation explained by the SNPs were achieved after running a Bayes B analysis for age at puberty, lifetime total number born, and lifetime total number of parities.

Table 2.3c: Posterior means of variance components of sow reproductive traits based on 56,424 SNP effects estimated by GWAS.

Trait*	*n*	Genetic Variance	Residual Variance	Total Variance	Proportion of Phenotypic Variance Explained by SNPs
AP	1,234	90.02	229.74	319.76	0.28
LT-TNB	903	3.17	473.72	476.89	0.007
LT-NP	903	0.20	1.00	1.20	0.16

A Bayesian analysis for age at puberty on R1 – R11 uncovered major 1 Mb windows and individual SNPs associated with the largest genetic variance for age at puberty (Appendix). The top three windows, which were found on SSC4 (7 Mb), SSC12 (2 Mb), and SSC3 (71 Mb), explained 3.28% of the phenotypic variation amongst all gilts (n = 1,234). These three windows harbor potential candidate genes including *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, response to bacterium, dendrite development, and neuron projection and *M1AP* (Meiosis 1 Associated Protein, SSC3 71.5-71.6 Mb) known to affect female gamete generation.

The 1 Mb windows were ranked based on their estimated genetic variance and then we selected the top 1%, 5%, 10%, and 20% of all of the windows across the *Sus scrofa* genome, which is approximately 2,600 windows. By using all of the SNPs in these top windows we aimed to demonstrate that it is not required to use all of the SNPs in the entire Porcine SNP60K BeadChip to explain most of the phenotypic variation for age at puberty. When using all SNPs (n = 56,424) and 100% of the windows, 28% of the

phenotypic variation was explained for R1 – R11, while using only 10% of the windows and all SNPs, 39% of the phenotypic variation was explained for R1 – R11 (Table 2.3d).

Using the top SNPs (n = 24) in the top 1% of the windows for R1 – R11, we were able to explain 18% of the phenotypic variation. When using the top SNPs (n = 519) in the top 20% of the windows, we were able to explain 48% of the phenotypic variation (Table 2.3e). Using batches, R1 – R7, which included a smaller number of gilts, we were still able to explain approximately as much phenotypic variation as R1 – R11 (25% compared to 28%; Table 2.3f).

It was believed that with an increase in the number of markers used we would be able to explain more of the phenotypic variation expressed because they would provide a better coverage of the *Sus scrofa* genome. Since there are known examples of genetic variants that explain a substantial amount of variation of different traits, we wanted to test this by using our entire population set of R1 – R11 and different percentages of top windows, comparing all SNPs versus the top SNP in each top window being used. A Bayesian (B) analysis for R1 – R11 was ran including all markers (n = 56,424), resulting with 28% of the phenotypic variation being explained. Using different combinations of top percentages of windows and markers we found that when using all markers from the top 1% of the top windows we explained 22% of the phenotypic variation while we explained 28% of the variation when we used all of the 1 Mb windows. If we used all of the markers from the top 10% windows we explained 39% of the variation, leading us to believe that by adding more markers, without increasing the number of animals in our data set, we were over parameterizing our model (Table 2.3g). To try and reduce the over parameterization, we then used the single top marker in the different percentages of

top windows, for R1 – R11. By doing this we then showed that when using only the top 1% of windows and the top marker in each window we could explain 19% of the phenotypic variation. When we increased the percentages of top windows to 20% we explained 48% of all phenotypic variation. At 50% of the top windows and the top marker being used we explained 51% of the phenotypic variation showing that the proportion of variation explained begins to plateau and eventually decrease when all windows and all markers are used, leaving us with 28% of the phenotypic variation being explained (Table 2.3h).

A deficiency presented by using this method of selecting 'top' windows and 'top' markers is, we are using previously found information from the original Bayesian (B) analysis when all of the markers (n = 56,424) and all of the windows are used. The original analysis is what allowed us to discover which windows had the most genetic variation as well as which markers had the most genetic variation, and the potential to be the best predictors for our GPV.

2.3.4 Genomic Predictions

When running a Bayesian (B) analysis on a training population, each SNP contributes its effect to the GPV of each individual. These posterior SNP effects can be used to predict the phenotypes of future generations. We used different batches of training populations to try and predict the phenotypes of future generations of gilts. In the first analysis, the SNPs were trained on R1 – R7 and evaluated on R8 – R11. When

all SNPs were trained on R1 – R7 and used to predict R8 – R11 the correlation ($r_{\hat{G}P}$) between the GPV and the phenotype was 0.18.

When all SNPs (n = 630) for the top 1% of windows for R1 – R7 were used to predict R8 – R11 the $r_{\hat{G}P}$ was 0.06, while when the top SNP (n = 26) for the top 1% of windows were used the $r_{\hat{G}P}$ decreased to 0.01. When evaluating R8 – R11 using all SNPs from the top 1% windows trained on R1 – R7 (n = 630 SNPs) we were still able to explain 18% of the phenotypic variation. By increasing the percentage of top windows and SNPs to 10%, we were able to increase our $r_{\hat{G}P}$ for all SNPs used to 0.12 in R8 – R11 and when using just the top SNPs, $r_{\hat{G}P}$ increased to 0.29. The proportion of variation explained in R8 – R11 when still using the top 10% of the windows and all SNPs was 29% (Table 2.3i).

Since R1 – R4 were generated using a different sire line, we estimated how much of an effect this played in our ability to predict future generations of gilts. If we excluded these batches from the training population and used batches R5 – R7 to try to predict the phenotypes in R8 – R11 we were able to explain 36% of the phenotypic variation using the top SNPs from the top 1% of windows. When we increased the percentage of windows to 10% and used the top SNP, we were able to explain 45% of the phenotypic variation. These same SNPs were used to then predict R8 – R11, with the $r_{\hat{G}P}$ being 0.18 when using the top 1% of windows and all SNPs, and when using just the top SNP, the $r_{\hat{G}P}$ was 0.14. The $r_{\hat{G}P}$ increased to 0.21 when using all the SNPs in 10% of the windows, and 0.26 when using just the top SNP (Table 2.3j).

2.3.5 Prediction of Phenotype Based on Individual Genomic Prediction Values (GPV)

Different approaches were used to predict individual phenotypes of our future generations of gilts using estimated SNP effects generated from prior generations of sows. These approaches were tested for the selection of individuals for a gene expression study prior to phenotypic expression. We used different combinations of batches as training data sets to predict individuals that expressed age at puberty early or late during boar exposure. We originally ran a Bayesian (B) analysis for R1 – R10, using the entire SNP data set. Using the output from R1 – R10 we tried to evaluate R11. Success of the prediction was determined if the phenotype of the selected individuals were in the extreme quartiles of the data. When using individual GPVs we were able to predict 1/3 gilts with early pubertal phenotypic expression (EP) and 3/3 with late pubertal phenotypic expression (LP). Our ability to predict both extreme phenotypes decreased when we selected animals based on both litter average GPVs and individual GPV, predicting 0/3 EP and 3/3 LP (Figure 2.3b). After using the entire SNP set, we evaluated the accuracy in prediction when the top SNP in the top 10% of windows were used (Figure 2.3c). Using the individual GPVs we successfully predicted the phenotype in 0/3 of the EP and 3/3 of the LP. When using both GPVs for litter averages and individual GPV, only LPs were predicted accurately (3/3), with none of the early phenotypes being predicted correctly (0/3).

When using R5 – R10 to evaluate R11 using all of the markers and all of the windows and selection based on individual GPVs, the ranking of the 1/3 of the EPs and

3/3 of the LPs were predicted. When using GPVs for litter averages we were not able to predict the EPs (0/3), but could predict the phenotype ranking of all LPs (3/3; Figure 2.3d). By decreasing our number of markers, using the top 10% of windows and the top SNP, our ability to predict which gilts would express an extreme phenotype increased; we predicted 2/3 EPs using both individual GPVs and litter average GPVs (Figure 2.3e). All of the LPs (3/3) were accurately predicted when using litter average GPVs, but decreased when individual GPVs were used (Figure 2.3e).

Lastly R1 – R9, as well as R5 – R9 were used to predict the phenotype ranking of R10. These combinations were first analyzed using all of the windows and all of the markers, followed by using the top 10% of windows and the top SNP, and lastly using GPV for litter averages for both, all SNPs and all windows and using the top 10% of windows and the top SNP (Figure 2.3f; Figure 2.3g; Figure 2.3h; Figure 2.3i).

2.3.6 Lifetime Reproductive Traits

Our research was focused on age at puberty because it has been found that there are negative phenotypic correlations between age at puberty and lifetime reproductive traits such as, lifetime number of parities, and lifetime total number born alive (Tart et al., 2013b). In addition we found significant negative correlations between the GPV for age at puberty and lifetime number of parities (-0.40; $P < .0001$), lifetime total number born alive (-0.45; $P < .0001$), and lifetime total number born (-0.45; $P < .0001$) (Table 2.3k; Figure 2.3j) indicating that common sources of genetic variation may influence the variation across these traits.

2.3.6.1 Lifetime Number of Parities

In commercial settings a large percentage of sows produce two parities, even with such lax culling practices compared to our research farm. In our data set (R1 – R8), more than half of the gilts did not generate more than 2 parities (55%; Table 2.3l; Figure 2.2c/2.3k).

Due to lifetime number of parities' low heritability, we investigated how much of the phenotypic variation could be explained by markers. Using different combinations of batches, R1 – R8 (n = 903), R1 – R4 (n = 548), R5 – R7 (n = 273) and R1 – R7 (n = 821), we employed a Bayesian (B) analysis using the entire SNP data set. Using the entire set of gilts R1 – R8 the SNPs explained 16% of the phenotypic variation. In the R1 – R4 and R1 – R7 the SNPs explained 17% of the phenotypic variation, while when using R5 – R7 we explained 14% of phenotypic variation (Table 2.3m). Using these same batch combinations we tested our ability to predict the phenotype of lifetime number of parities for future generations. When batches R1 – R7 were used to evaluate R8, the $r_{\hat{G}P}$ was 0.072 (Table 2.3m). When R1 – R4 was used to evaluate R5 – R8 there was a $r_{\hat{G}P}$ of 0.11, while when R5 – R7 was used to evaluate R8, the $r_{\hat{G}P}$ was 0.10.

A GWAS based on Bayesian approaches, mapped several major regions of the swine genome that explained phenotypic variation of lifetime number of parities (R1 – R8, n = 903). For example, the top three windows were found on SSC11 (4 Mb) and SSC16 (3 Mb and 20 Mb; Appendix). These top three windows explained a combined total of 5.17% of all of the genotypic variation amongst all gilts (n = 903). Some

interesting potential candidate genes found near these windows are *GSX1* (Genomic Screened Homeo Box 1), *C1QTNF3* (C1q And Tumor Necrosis Factor Related Protein 3), and *U6* (Uncharacterized Protein 6). *GSX1* can be found on SSC11 (4.81 – 4.82 Mb) and plays a role in spinal cord association neuron differentiation, hypothalamus development, as well as adenohypophysis development. *C1QTNF3* is located on SSC16 (20.7 – 20.8 Mb) and is involved in different metabolic functions such as glucose homeostasis, fat cell differentiation, and negative regulation of gluconeogenesis and inflammatory response. Lastly *U6* can be found on SSC16 (3 Mb) and is an uncharacterized protein associated with protein binding.

2.3.6.2 Lifetime Total Number Born

Another reproductive trait with important economic value is lifetime total number born. Using R1 – R8 we were able to see that a large number of gilts get culled, for reproductive reasons, before or at parity two and have a lifetime total number born piglets of approximately 25 piglets or less (Figure 2.2b/2.3l). To see how much of this phenotypic variation could be explained by markers, we used the same combinations of Bayesian (B) analyses as in lifetime number of parities. Using all of the windows and all of the markers R1 – R8 explained approximately 1% of the phenotypic variation (Table 2.3n), with the same found for R1 – R4, R5 – R7, and R1 – R7. Batches R1 – R4 evaluated R5 – R8 with a $r_{\hat{G}P}$ of 0.09. Batches R5 – R7 and R1 – R7 evaluated R8 with a correlation of 0.035 and 0.040 respectively (Table 2.3n).

Using our full dataset for lifetime total number born we looked at which windows explained the most variance as well as harbored potential candidate genes. Using the Bayesian (B) analysis output, we looked at 1 Mb windows across the *Sus scrofa* genome for R1 – R8. The top three windows in this analysis explained a combined total of 1% of the phenotypic variation. These windows were located on SSC11 (4 Mb), SSC13 (160 Mb), and SSC16 (20 Mb). Two of these windows (SSC11, 4 Mb, and SSC13, 160 Mb) are in common with the top windows in lifetime total number of parities. A potential candidate found on SSC13 (160.4 Mb) is *CD47* (Cluster of Differentiation 47 Molecule). *CD47* has a role in positive regulation of inflammatory response, phagocytosis, and T cell activation, as well as opsonization. By looking at windows with the most variation we have the potential to find candidate genes that are responsible for the phenotypic variation amongst our breading heard.

2.4 CONCLUSION

One of the reasons sow reproductive performance is of high significance to producers is because of the increased costs of detecting age at puberty and the high culling rates due to reproductive failure. Even if the environment and management practices are controlled, we can still see a large amount of phenotypic variation within nucleus and commercial populations. In an ideal setting producers could genotype their breeding herds and know ahead of time whether or not a gilt will be successful reproductively. We have shown that not all 60,000 SNPs are required to explain a substantial proportion of the phenotypic variation. For example, using only a limited number of informative markers that explain the largest percentage of variation, we are able to explain more phenotypic variation than we are when we use all of the markers. Most likely, by using SNPs associated with major effects, we are able to decrease the background noise by not over parameterizing the model and more accurately explain phenotypic variation and predict the phenotypes of the future generations using genomic prediction values. With the cost of genotyping decreasing there is potential for a smaller SNP chip to be developed. If only 260 informative DNA markers are used (the number used in the top 10% windows using just the top SNP), the cost of genotyping would be decreased significantly. This would allow producers to predict age at puberty for their gilts when they are still piglets, allowing for significant savings. Since age at puberty is also negatively correlated with lifetime reproductive traits, by predicting which gilts will reach puberty sooner using pleiotropic DNA markers, the producers will also simultaneously be predicting which gilts will have a successful reproductive lifespan, of

course not accounting for management practices and environmental effects. Since there has been success in using predictions from crossbred animals to predict purebreds, these same SNP chips developed on a commercial population could be used to implement superior selection techniques at the nucleus level, allowing for more genetic response to occur.

2.5 LITERATURE CITED

Habier, David, Rohan L. Fernando, Kadir Kizilkaya, and Dorian J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics. 12:186:1471-2105.

Hsu, Wan-Ling. (2011). Analysis of Long-term Selection (28 generations) for Reproduction, Growth, and Carcass Traits in Swine. Theses and Dissertations in Animal Science, Paper 34. Retrieved February, 21, 2014, from the Animal Science Department at DigitalCommons@University of Nebraska – Lincoln.

Johnson, R., P. Miller, and D. Ciobanu. 2011. Genetics and development of replacement gilts. Presentation at the Billy Day Symposium at the ADSA-ASAS Midwest Meeting.

Johnson, Z. B., J. J. Chewning, and R. A. Nugent, 3rd. 2002. Maternal effects on traits measured during postweaning performance test of swine from four breeds. J. Anim. Sci. 80:1470-1477.

Mackay, Trudy F. C. "Chapter 10: Heritability." *Introduction to Quantitative Genetics*. By D. S. Falconer. 4th ed. Essex, England: Longman, 1996. 180. Print.

Onteru, S. K., B. Fan, Z-Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. A whole-genome association study for pig reproductive traits. Animal Genetics. 43:18-26.

Ruíz-Flores, A. and R. K. Johnson. 2001. Direct and correlated responses to two-stage selection for ovulation rate and number of fully formed pigs at birth in swine. J. Anim. Sci. 79: 2286-2297.

Serenius, T. and K. J. Stalder. 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. 1:745-750.

Tart, J. K., K. L. Lucot, J. W. Bundy, P. S. Miller, M. F. Rothschild, M. L. Spangler, D. J. Garrick, R. K. Johnson, S. D. Kachman, and D. C. Ciobanu. 2013a. Genome-wide prediction of pleiotropic regions that influence age at puberty and reproductive longevity in sows. Poster session presented at: Plant & Animal Genome XXI. 21st Annual International Plant & Animal Genome Conference; 2013 Jan 12-16; San Diego, CA.

Tart, J. K., R. K. Johnson, J. W. Bundy, N. N. Ferdinand, A. M. McKnite, J. R. Wood, P. S. Miller, M. F. Rothschild, M. L. Spangler, D. J. Garrick, S. D. Kachman, and D.

C. Ciobanu. 2013b. Genome-wide prediction of age at puberty and reproductive longevity in sows. Animal Genetics. 44:4:387-397.

TABLES

Table 2.3a: The correlation between the genomic predictor and phenotypes (age at puberty) of the animals. Markers were trained on designated training populations and then evaluated on evaluation populations. Different π values were used to see if they affected correlation values.

Training	**Evaluation**	**$r_{\hat{G}P}$ (1% of top windows* using different π values)**				
		0.00	**0.25**	**0.50**	**0.75**	**0.99**
R1 – R4	R5 – R10	0.04	0.04	0.04	0.03	0.01
R5 – R7	R8 – R10	0.20	0.20	0.20	0.20	0.19
R1 – R7	R8 – R10	0.08	0.08	0.08	0.07	0.03
R1 – R10	R1 – R10	0.54	0.54	0.54	0.54	0.49

*Includes only mapped markers

Table 2.3b: The correlation between the genomic predictor and phenotypes (age at puberty) of the animals. Markers were trained on designated training populations and then evaluated on evaluation populations. Different π values were used to see if they affected correlation values.

Training	**Evaluation**	**$r_{\hat{G}P}$ (10% of top windows* using different π values)**				
		0.00	**0.25**	**0.50**	**0.75**	**0.99**
R1 – R4	R5 – R10	0.052	0.052	0.052	0.052	0.033
R5 – R7	R8 – R10	0.26	0.26	0.26	0.26	0.24
R1 – R7	R8 – R10	0.15	0.15	0.14	0.14	0.0088
R1 – R10	R1 – R10	0.78	0.79	0.79	0.79	0.76

*Includes only mapped markers

Table 2.3d: Proportion of phenotypic variation for age at puberty explained by markers when using different percentages of top windows trained individually on different batch combinations.

Batches	*# of Animals*	*Top Windows Used (%)* & All Markers*				
		1%	*5%*	*10%*	*20%*	*100%*
R5 – R7	n = 275	0.25	0.33	0.33	0.31	0.22
R1 – R7	n = 822	0.27	0.37	0.36	0.32	0.25
R1 – R11	n = 1234	0.22	0.38	0.39	0.35	0.28

* Includes only mapped SNPs

Table 2.3e: Varying percentages of top windows used to identify the proportion of phenotypic variation explained by the markers for age at puberty.

Training Population	*SNPs*	*% of Top Windows Used*				
		1%	*5%*	*10%*	*20%*	*100%*
R1 – R11	*Proportion of Variance Explained by Markers*	0.22	0.38	0.39	0.35	0.28
	All SNPs Used in Windows	645	4505	8751	19362	56,424
R1 – R11	*Proportion of Variance Explained by Markers*	0.18	0.37	0.44	0.48	
	Top SNPs Used in Windows	24	129	259	519	

* Includes only mapped SNPs

Table 2.3f: Varying percentages of top windows used to identify the proportion of phenotypic variation explained by the markers for age at puberty.

Training Population	*SNPs*	*% of Top Windows Used*				
		1%	*5%*	*10%*	*20%*	*100%*
R1 – R7	*Proportion of Variance Explained by Markers*	0.27	0.37	0.36	0.32	0.25
	All SNPs Used in Windows	630	3463	7125	16582	56,424
R1 – R7	*Proportion of Variance Explained by Markers*	0.26	0.43	0.48	0.51	
	Top SNPs Used in Windows	26	131	266	616	

* Includes only mapped SNPs

Table 2.3g: Proportion of phenotypic variance explained by SNPs when using different percentages of top windows and all SNPs for R1 – R11 for age at puberty.

R1 – R11: Top Windows & All SNPs		
Percent of Top Windows	*Proportion of Variance Explained by Markers*	*Standard Deviation*
1%	0.22	0.024
5%	0.38	0.028
10%	0.39	0.028
20%	0.35	0.028
30%	0.34	0.028
50%	0.31	0.028
100%	0.28	0.030

* Includes only mapped SNPs

Table 2.3h: Proportion of phenotypic variance explained by markers when using the top SNP from different percentages of top windows from R1 – R11 for age at puberty.

R1 – R11: Top Windows & Top SNP		
Percent of Top Windows	*Proportion of Variance Explained by Markers*	*Standard Deviation*
1%	0.19	0.022
5%	0.37	0.027
10%	0.44	0.029
20%	0.48	0.027
30%	0.49	0.028
50%	0.51	0.027
100%	0.28	0.030

* Includes only mapped SNPs

Table 2.3i: Training on R1 – R7 and evaluation on R8 – R11 for age at puberty.

Training	*% Windows*	*Markers Used*	*Proportion of Variation Explained by SNPs*	*Predict*	$r_{\hat{G}P}$
R1 – R7	100%	All SNPs	0.25	R8 – R11	0.18
R1 – R7	Top 1%	All SNPs	0.27	R8 – R11	0.06
R1 – R7	Top 1%	Top SNPs	0.26	R8 – R11	0.01
R8 – R11	Top 1%	All SNPs from R1 – R7	0.18		
R8 – R11	Top 1%	Top SNPs from R1 – R7	N/A		
R1 – R7	Top 10%	All SNPs	0.36	R8 – R11	0.12
R1 – R7	Top 10%	Top SNPs	0.48	R8 – R11	0.29
R8 – R11	Top 10%	All SNPs from R1 – R7	0.29		
R8 – R11	Top 10%	Top SNPs from R1 – R7	0.32		

* Includes only mapped SNPs

Table 2.3j: Training on R5 – R7 and evaluating on R8 – R11 for age at puberty.

Training	*% Windows*	*Markers Used*	*Proportion of Variation Explained by SNPs*	*Predict*	$r_{\hat{G}P}$
R5 – R7	Top 1%	All SNPs	0.33	R8 – R11	0.18
R5 – R7	Top 1%	Top SNPs	0.36	R8 – R11	0.14
R8 – R11	Top 1%	All SNPs from R5 – R7	0.19		
R8 – R11	Top 1%	Top SNPs from R5 – R7	N/A		
R5 – R7	Top 10%	All SNPs	0.32	R8 – R11	0.21
R5 – R7	Top 10%	Top SNPs	0.45	R8 – R11	0.26
R8 – R11	Top 10%	All SNPs from R5 – R7	0.32		
R8 – R11	Top 10%	Top SNPs from R5 – R7	0.20		

* Includes only mapped SNPs

Table 2.3k: Pairwise correlations between age at puberty and lifetime reproductive traits.

R1 – R8						
Variable	By Variable	Correlation	Count	Lower 95%	Upper 95%	Significant Probability
AP	LTNP	-0.40	900	-0.46	-0.35	<.0001
AP	LTNBA	-0.45	900	-0.50	-0.39	<.0001
AP	LTTNB	-0.45	900	-0.50	-0.40	<.0001

LTNP – lifetime number of parities; LTNBA – lifetime total number born alive; LTTNB – lifetime totally number born.

* Includes only mapped SNPs

Table 2.3l: Distribution for lifetime number of parities.

Category (Parities)	% Of Animals (n)			
	R1 – R4	R5 – R7	R1 – R7	R1 –R8
0 Parities	28%	18%	24%	24%
1 or Fewer	55%	32%	47%	47%
2 or Fewer	63%	41%	56%	55%
3 or Fewer	70%	72%	71%	69%
4 Parities	30%	28%	29%	31%
Total Animals (n)	548	273	821	903

Table 2.3m: Various batches used to train and evaluate for lifetime number of parities.

Lifetime Number of Parities				
Bayes (Training)	Evaluation	# Of Animals (n)	Proportion of Variation Explained by SNPs	$r_{\hat{G}P}$
R1 – R4	R5 – R8	548	0.17	0.11
R5 – R7	R8	273	0.14	0.10
R1 – R7	R8	821	0.17	0.072
R1 – R8	R1 – R8	903	0.16	0.68

* Includes only mapped SNPs

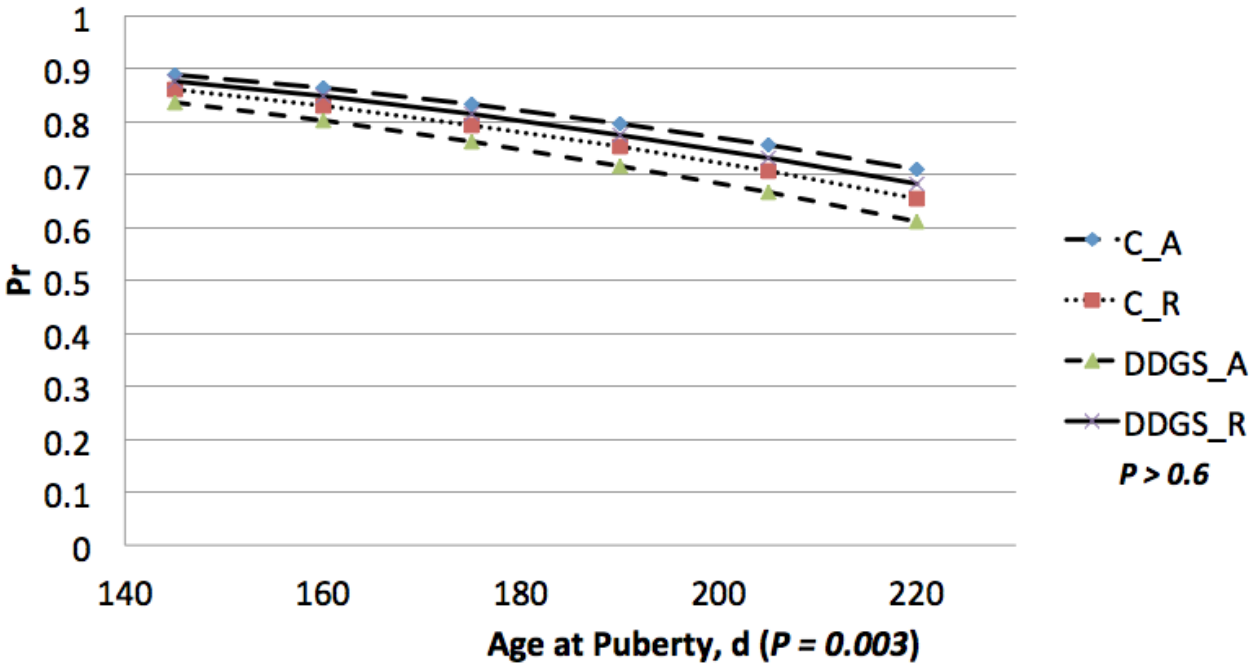
Table 2.3n: Various batches used to train and evaluate for lifetime total number born.

Lifetime Total Number Born				
Bayes (Training)	Evaluation	# Of Animals (n)	Proportion of Variation Explained by SNPs	$r_{\hat{G}P}$
R1 – R4	R5 – R8	548	0.0070	0.086
R5 – R7	R8	273	0.0061	0.035
R1 – R7	R8	821	0.0069	0.040
R1 – R8	R1 – R8	903	0.0066	0.56

* Includes only mapped SNPs

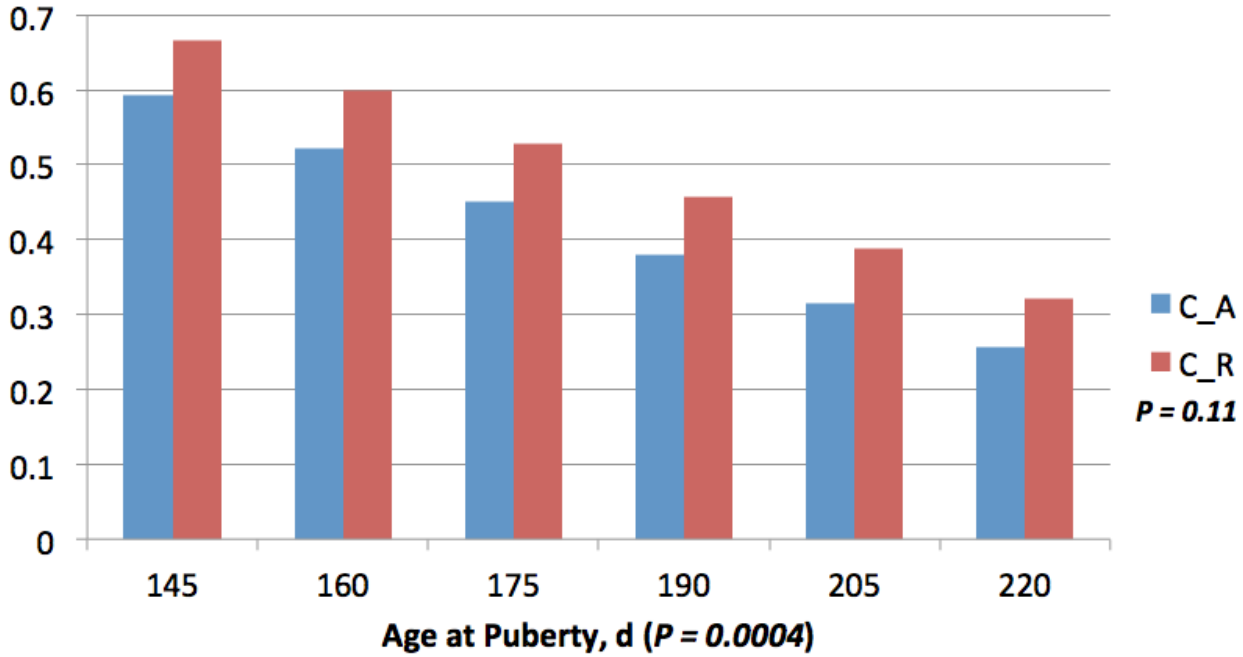
FIGURES

Figure 2.1a: The probability of breeding gilts producing a parity 1 litter.



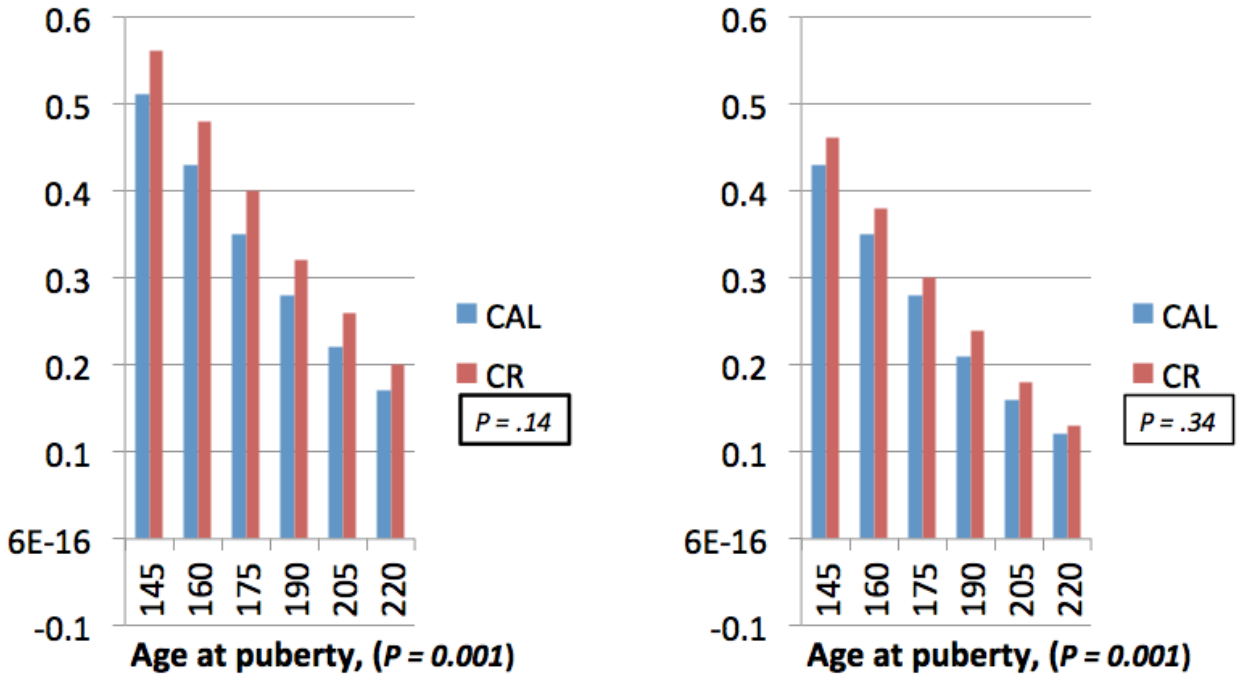
C_A –ad libitum diet; C_R –calorically restricted diet; DDGS_A – ad libitum diet; DDGS_R – calorically restricted diet

Figure 2.1b: The probability of breeding gilts producing a parity 2 litter.



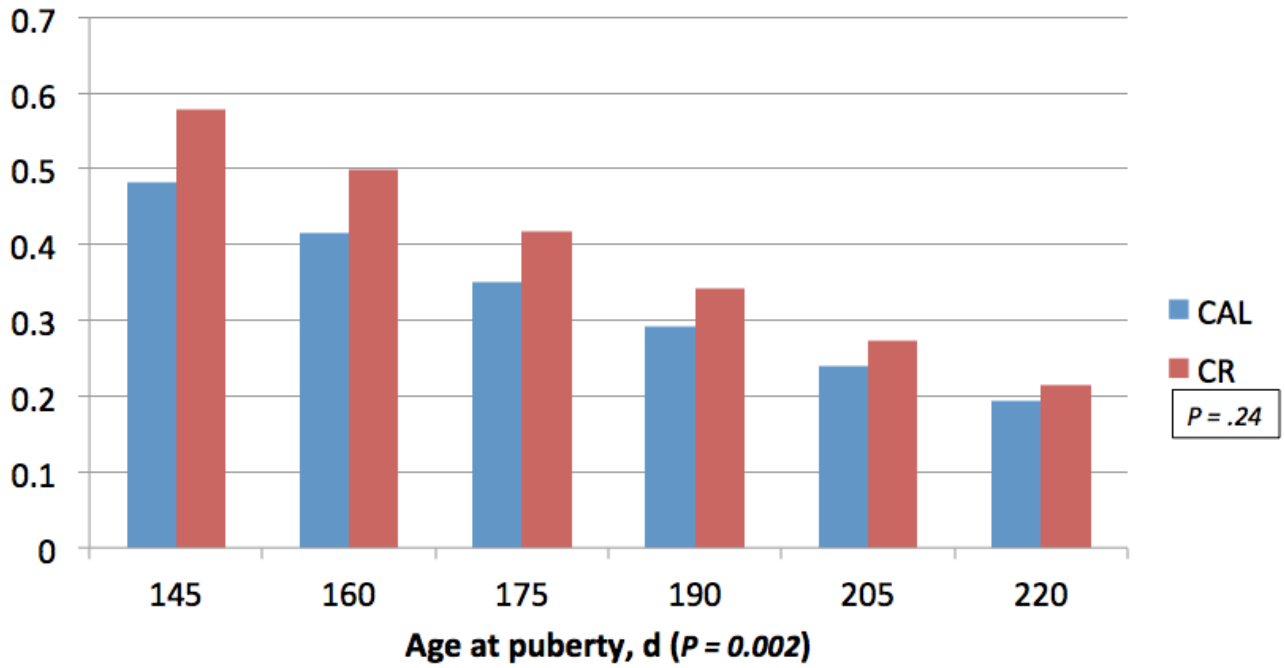
C_A –ad libitum diet; C_R –calorically restricted

Figure 2.1c: The probability of breeding gilts producing a parity 3 (graph on left) and parity 4 litter (graph on right).



C_A –ad libitum diet; C_R –calorically restricted

Figure 2.1d: The probability of breeding gilts producing a parity 4 litter if they produced a parity 1 litter.



C_A –ad libitum diet; C_R –calorically restricted

Figure 2.2a: Distribution of age at puberty (n = 1,234)

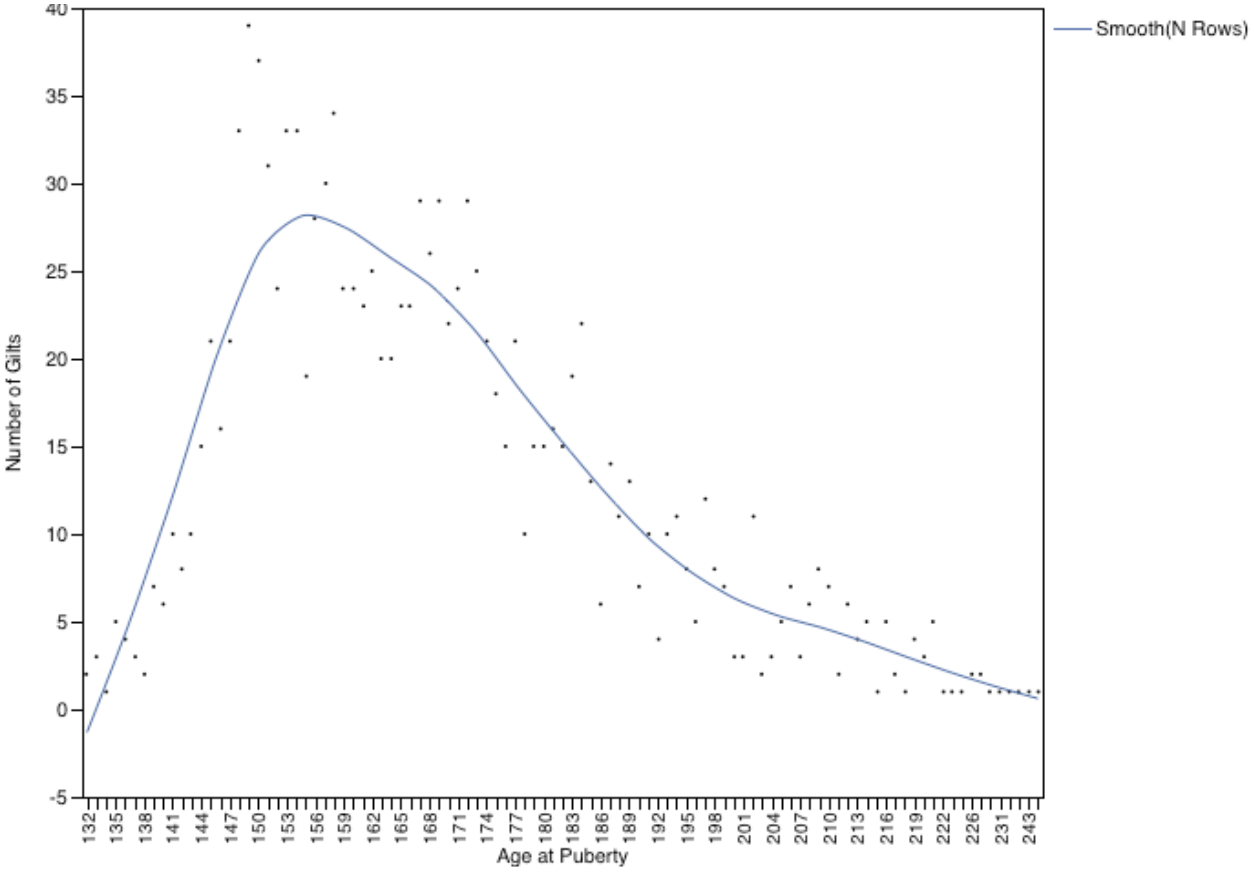


Figure 2.2b/2.3l: Distribution of lifetime total number of piglets born per sow R1 – R8 (n = 903).

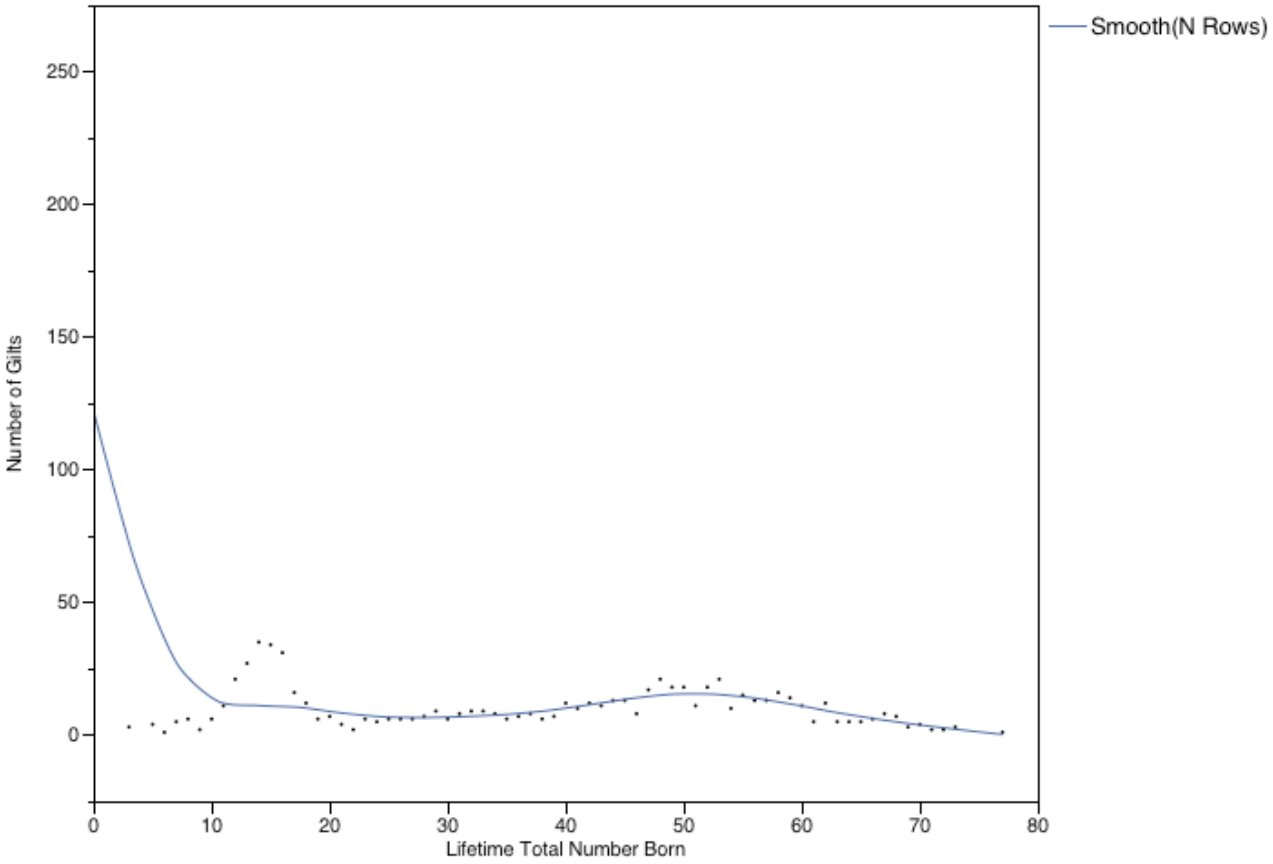


Figure 2.2c/2.3k: Distribution of lifetime number of parities using R1 – R8 (n = 903).