RESEARCH REPORT
ANIMAL SCIENCE

Title: Identification of SNP Markers Associated with Number Born, Number Weaned and Weaning to Estrus Interval in Commercial First Parity Sows – **NPB #09-091**

Investigator: Gary Rohrer, Ph.D.

Institution: USDA-ARS Northern Plains Area, US Meat Animal Research Center

Date Submitted: June 18, 2010

Industry Summary:

The objective of this research was to identify genetic markers that would be predictive of first parity reproductive performance and rebreeding in an effort to improve the percentage of gilts retained for breeding that actually contribute to the breeding herd. Three commercial cooperators were identified and biological samples for 706 boars used in the industry, along with the performance data from their daughter's first parity were acquired. Data from 123 boars used at USMARC were also included. Performance data analyzed were number born alive, number born dead, total number born and weaning to estrus interval. DNA was extracted from each boar and genotypic data were collected using the recently developed Illumina Porcine60K BeadChip. Genotypes were called for 59,895 single nucleotide polymorphism (SNP) markers spanning the entire porcine genome. Association analyses were performed to identify significant associations between SNP markers and performance traits. In total, 14 different genomic regions were associated with a measure of reproductive performance. Two locations were associated with number born dead, nine locations were associated with number born alive and two locations were associated with weaning to estrus interval. None of the regions appeared to be associated with measures of growth rate or backfat depth, so selection for improved reproduction would not be expected to result in poorer performance of market hogs. These results have the potential to improve first parity reproductive performance in commercial herds.

Gary Rohrer
Research Geneticist
US Meat Animal Research Center
PO Box 166
Clay Center, NE 68933
Gary.rohrer@ars.usda.gov

Keywords: SNP Associations, Return to estrus, Number Born Alive, First Parity, Prolificacy

Scientific Abstract:

The objective of this research was to identify genetic markers that would be predictive of first parity reproductive performance and rebreeding in an effort to improve the percentage of gilts retained for breeding that actually contribute to the breeding herd. Three commercial cooperators were identified and biological samples for 706 boars used in the industry, along with the performance data from their daughter's first parity were acquired. Data from 123 boars used at USMARC were included. Phenotypic traits evaluated were number born alive, number born dead, total number born and weaning to estrus interval. Performance data analyzed were the average daughter deviations (daughter performance – contemporary group mean) for each boar. Boars were required to have a minimum of 4 daughter records for litter traits and 3 daughter records for weaning to estrus interval. DNA was extracted from each boar and genotypic data were collected using the recently developed Illumina Porcine 60K BeadChip. Genotypes were called for 59,895 single nucleotide polymorphism (SNP) markers spanning the entire porcine genome. Approximately 48 million genotypes were analyzed representing a 96.4% call rate. Association analyses were performed using PLINK to identify significant associations between SNP markers and performance traits. Only 3 SNP were significant using a Bonferroni adjusted threshold representing 2 different genomic regions on SSC 1 and 3, both for number born dead. Evaluation of all results significant at $P < 0.001$ in subsets of the data partitioned by breed revealed several locations that likely possess segregating quantitative trait loci (QTL). In total, 14 different genomic regions (including the two previously mentioned regions) were associated with a measure of reproductive performance. Two locations, presented as chromosome (SSC) and base position (in millions of bases, Mb), were associated with number born dead (SSC1:138.1 Mb and SSC3:57-60 Mb), nine locations were associated with number born alive (SSC1:177-182 Mb; SSC4:60-66 Mb; SSC4:93-101 Mb; SSC7:54-60 Mb; SSC7:116-120 Mb; SSC8:30.5-33.5 Mb; SSC8:109-115 Mb; SSC11:62-66 Mb and SSC15:54-62 Mb) and two locations were associated with weaning to estrus interval (SSC4:65-75 Mb and SSC14:8-12.5 Mb). None of the regions appeared to be associated with measures of growth rate or backfat depth in the USMARC composite population. The approach of using daughter deviations appears to be quite powerful and an economical approach to using the Illumina Porice 60K BeadChip. These results have the potential to improve first parity reproductive performance in commercial herds, but should be validated in other commercial pigs first.

Introduction:

Sow prolificacy is a primary indicator of her ability to remain in the herd. Reproductive performance and stayability of female pigs is critical to the competitiveness of pork in the global market. The ability to identify young females with superior reproductive longevity and productivity would have a large economic impact on commercial swine production. Rodriguez-Zas (2006) reported that a sow must produce an average of four litters for optimal economics, yet it has been estimated that over 50% of sows are culled prior to weaning four litters (D'Allaire et al., 1987; Lucia et al., 2000). More importantly, approximately one third of all females are culled prior to having two litters with over half of these removals attributed to reproduction inadequacies (D'Allaire et al., 1987; Lucia et al., 2000; Rodriguez-Zas et al., 2003). Specific reasons that sows are removed for reproductive failure were failure to return to estrus (72%), anestrus (15%), prolapse or dystocia (4%) and other (7%) in a study conducted by Engblom et al. (2007). Long weaning to estrus intervals also increases a sow's nonproductive days and reduces their profitability (Lucia et al., 2000; Engblom et al., 2007). This exceptionally high rate of reproductive culling in females with less than two parities decreases the economic potential of the herd with an overall reduction in lifetime productivity and longevity.

An ideal situation would be to utilize genetic screening to identify those animals at greater risk of failing in reproductive performance at an early age, but few studies have been reported that attempt to identify genetic

markers associated with gilt breeding performance and rebreeding performance of first parity females. Primiparous females have greater energetic demands associated with growth and energy utilization during critical periods of time such as gestation and lactation, which may have an impact on reproductive performance. In a previous study, the most significant associations between genotype and weaning to estrus interval in a Landrace-Duroc-Yorkshire composite population were for genes associated with energy metabolism (IGFBP3, PRKAG3 and PPARGC1A) along with associations with ESR1 and ESR2 markers (Rempel et al., 2010). The heritability of weaning to estrus interval was estimated to be 0.16 and the genetic correlation with age at puberty was 0.47 (Rempel et al., 2010). A scan of the entire genome would provide much greater insights into genetic mechanisms controlling number born, number weaned, weaning to estrus interval and rebreeding performance in young females.

The use of high density single nucleotide polymorphism (SNP) arrays (50,000 or more SNP) for genome-wide association analyses has been shown to be quite powerful in detecting monogenic traits in livestock (Charlier et al., 2008; Andersson 2008) as well as for quantitative traits in dairy cattle (Cole et al., 2009), beef cattle (Snelling et al., 2010) and pigs (Duijvesteijn et al., 2010). Recently, Dr. Curt Van Tassel indicated that quantitative trait locus (QTL) detection using estimated breeding values on sires from daughter records with the Illumina 50K Bovine SNP BeadChip in dairy cattle is an extremely powerful and efficient approach (National Pork Board Animal Science Genomics Advisory Meeting; Des Moines, IA; October 22, 2008).

Objectives:

The objectives of this study were:

A. Identify SNP markers associated with number born, number weaned, and return to estrus interval of first parity females in commercial production systems.

B. Determine if antagonistic pleiotropic effects exist between sow lifetime productivity SNP markers and other important production traits such as growth rate, body composition or other reproduction phenotypes.

Materials & Methods:

Phenotypic data: Three cooperators were identified to contribute biological samples and daughter phenotypic records for industry sires. Smithfield Premium Genetics Group (Rose Hill, NC) contributed 384 boars representing seven genetic lines, National Swine Registry (West Lafayette, IN) contributed 287 boars from four genetic lines and DanBred North America (Columbus, NE) contributed 35 boars from two genetic lines. In addition, 123 boars from a Landrace-Duroc-Yorkshire composite population at USMARC were included.

Phenotypic data collected from most cooperators were number born alive, number born dead, number weaned and weaning to estrus interval. Total number born was computed by addition of number born alive and number born dead. Number born dead, and subsequently total number born, was only available for part of the boars. Along with each phenotypic record on a sire's daughter, the contemporary group mean was also provided. Daughter deviations were computed by simply subtracting the contemporary group mean from each phenotype. Then an average daughter deviation was computed for each sire and each trait. These mean daughter deviations were then used as phenotypes for each boar.

Preliminary evaluation of the data indicated that a high rate of cross-fostering is implemented in the industry. The initial reason to analyze number weaned was to evaluate preweaning survival within a litter, but number weaned by a sow appeared to be a poor indicator of this phenotype. Thus, number weaned was excluded from further evaluations. After editing phenotypic data, mean daughter deviations with fewer than four daughters for number born alive and number born dead or three daughters for weaning to estrus interval were eliminated. There were a total of 812 boars analyzed for number born alive (average number of daughters was 40.6) and

605 boars analyzed for weaning to estrus interval (average number of daughters was 26.2). Fewer daughter records for weaning to estrus resulted from failure of some sows returning to estrus after their first parity and apparent incomplete reporting (or collection) in National Swine Registry animals.

Genotypic data: Biological samples used included hair follicles, blood cards, frozen semen and frozen tails. DNA from frozen semen and tails was extracted at USMARC using Promega's Wizard Prep kits for genomic DNA from tissue (Madison, WI), while all other samples were transferred to GeneSeek, Inc. (Lincoln, NE). All DNAs were assayed with Illumina's Porcine 60K BeadChip (Ramos et al., 2009) and scan results interpreted at USMARC using Illumina's Bead Studio. BeadChips for the 125 USMARC boars and 6 commercial boars were run at USMARC, while all other DNA samples were assayed by GeneSeek.

Statistical analyses: Statistical association analyses between marker genotype and phenotype were conducted using PLINK (Purcell, et al., 2007). Unique population designations were developed for each cooperator and genetic line, resulting in 13 different populations. The entire set of boars was analyzed for each trait.

In an attempt to identify additional QTL, association analyses were also conducted within subsets of boars comprised of similar germplasm for number born alive and weaning to estrus interval. Four subgroups were developed to represent purebred Duroc (n = 208), purebred Landrace (n = 182), purebred Yorkshire (n = 210) and composite populations (n = 239). The most heterogeneous subgroup was the composite population as different breeds and breed compositions were combined. These subgroups would take advantage of greater linkage disequilibrium for QTL detection. In addition, if particular QTL or SNP markers are fixed in certain populations, then the estimated importance of an association within the entire data set is diminished.

Results:

Genotyping success: A total of 706 industry sires along with 123 sires from the USMARC Landrace-Duroc-Yorkshire population were genotyped using the Illumina Porcine60K BeadChip. Data from the scanned chips were evaluated with Illumina's Bead Studio where 59,895 SNP assays were scored. From these 831 samples, approximately 48 million genotypes were called (call rate of 96.4%). The genotypic call rates from DNA extracted from semen or tail tissue were superior to other forms of biological samples (99.3 vs 95.4%).

Association Analyses: Based on the number of tests conducted for each trait some method to adjust for the multiple testing needs to be applied to prevent reporting numerous false positive associations. The most conservative method is a Bonferroni adjustment factor where 0.05 is divided by the number of tests. For this study, the Bonferroni correct nominal p-value would need to be less than $1x10^{-6}$ for a genome-wide significance. This is an extremely conservative method and based on this correction factor there were only three significant associations with number born dead and no significant associations with number born alive or weaning to estrus interval. The three significant SNP represented two different chromosomal regions. One region was located on chromosome 1 (138.1 million bases, Mb) with no other markers associated with number born dead in the vicinity. The other two SNP were located on chromosome 3 (between 57 and 60 Mb) with three other SNP significant at $P < 1x10^{-4}$.

Three SNP were associated with number born alive at $P < 1x10^{-5}$, which would be approximately equivalent to a suggestive genome-wide significance. One SNP was located on chromosome 4 (99.6 Mb), one on chromosome 8 (32.1 Mb) and the other marker's position in the genome has not yet been determined.

In an attempt to identify additional QTL without detecting numerous false positive artifacts, the significance values and estimated effects within each subgroup were evaluated for all associations in the complete data set that exceeded $P < 0.001$. For number born alive there were 158 SNP marker associations and 18 marker associations that were significant at this level for weaning to estrus interval. By sorting these associations by their location in the genome and comparing their estimated effects within each subgroup, nine locations (including the regions on chromosome 4 and 8 mentioned above) that likely possess QTL affecting number born alive and two locations for weaning to estrus interval were observed (Table 1). Scatter plots for each of these 11 regions with putative QTL are presented in Figure 1 where -log (p-value) is plotted versus the location in the genome for the complete analysis as well as for each subgroup.

As quite different performance traits were recorded by each cooperator for growth and body composition, we have only focused on an analysis using the USMARC boars. From these analyses, there were no associations between SNP markers located in the regions reported in Table 1 with 154-day weight or backfat measures. This analysis utilized data from 123 boars and each boar had approximately 65 progeny.

Discussion:

These results indicate that there are a considerable number of genomic locations that are affecting first parity female reproduction. The trait with the greatest number of associations was number born alive, likely due to the larger number of records per boar. Based on the allele frequencies in the boars sampled and the estimated effects, considerable improvement in these performance traits can be made. The QTL that could impact performance the most are those where the desirable allele is in the lowest frequency. Two locations have extremely low frequencies for the desirable allele for number born alive (chromosome 1 at 179.9 Mb and chromosome 4 at 63.1 Mb), but these results should be observed with caution as the A allele frequency is so low that the estimated effects are based on very few boars. However, other important locations with greater support for number born alive are chromosomes 4 at 99.6 Mb, 7 at 56.6 Mb, 8 at 109.6 Mb and 15 at 55.1 Mb. In addition, the location on chromosome 4 at 73.1 Mb for weaning to estrus interval also possesses great potential. As might be expected, the regions identified for number born dead, a survival trait, are nearly fixed for the desirable allele.

Only a few SNP associations reached the highly conservative Bonferroni adjusted threshold. There are a number of potential reasons for fewer significant associations than expected. First, the extent of linkage disequilibrium across multiple breeds/lines of pigs may not be as large as originally anticipated. The association analyses rely on linkage disequilibrium between SNP markers and quantitative trait nucleotides (QTN). When multiple populations are evaluated a greater marker density is required to account for less linkage disequilibrium. Originally, it was assumed that 60,000 SNP markers would be sufficient for this study, but the results would indicate that more markers may be needed for these general association studies. Frequently the most significant marker for a specific genomic region differed across subgroups/lines, despite the fact that the estimated effects for the region may have been similar.

A second potential limitation of this study was due to the fewer number of daughter records than originally anticipated. The swine industry rapidly turns over generations. This feature limits the number of daughters each sire produces and makes it difficult to identify a boar with a sufficient number of progeny records that is still available to collect a biological sample. So both number of daughters and stored biological samples can create problems. In addition, the most economical collection and storage methods don't always provide a sufficient yield of DNA for use with the Illumina Porcine 60K BeadChip.

The data set developed for this project will be a valuable resource in the future. Many of the boars chosen for this project were actively producing progeny at the time of selection. Therefore, it is our intention to update their daughter deviation information in the future to increase the reliability of these results. In addition, other statistical approaches could be evaluated. The field of genome-wide association studies is still evolving and better methods are likely to be developed.

Literature Cited:

Andersson, L. 2008. Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. Genetica doi:10.1007/s10709-008-9312-4.

Charlier, C., W. Coppieters, F. Rollin, et al. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. Nat. Genet. 40: 449-454.

Cole, J.B., P.M. Van Raden, J.R. O'Connell, C.P. Van Tassell, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor and G.R. Wiggans. 2009. Distribution and location of genetic effects for dairy traits. J. Dairy Sci. 92:2931-46.

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.

Duijvesteijn, N., E.F. Knol, J.W.M. Merks, R.P.M.A. Crooijmans, M.A.M. Groenen, H. Bovenhuis and B. Harlizius. 2010. A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC Genetics 11:42.

Engblom, L., N. Lundeheim, A-M. Dalin, and K. Andersson. 2007. Sow removal in Swedish commercial herds. Livest. Sci. 106:76-86.

Lucia, Jr., T., G.D. Dial, and W.E. Marsh. 2000. Lifetime reproductive performance in female pigs having distinct reasons for removal. Livest. Prod. Sci. 63:213-222.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. Ferreira, D. Bender, J. Maller, P. Sklar, P. de Bakker and M. Daly. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Amer. J. Hum. Genet. 81:559-75.

Ramos, A.M., R.P.M.A. Crooijmans, N.A. Affara, A.J. Amaral, A.L. Archibald, J.E. Beever, C. Bendixen, C. Churcher, R. Clark, P. Dehais, M.S. Hansen, J. Hedegaard, Z.L. Hu, H.H. Kerstens, A.S. Law, H.J. Megens, D. Milan, D.J. Nonneman, G.A. Rohrer, M.F. Rothschild, T.P.L. Smith, R.D. Schnabel, C.P. Van Tassell, J.F. Taylor, R.T. Wiedmann, L.B. Schook and M.A.M. Groenen. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. 2009. PLoS ONE 4:e6524.

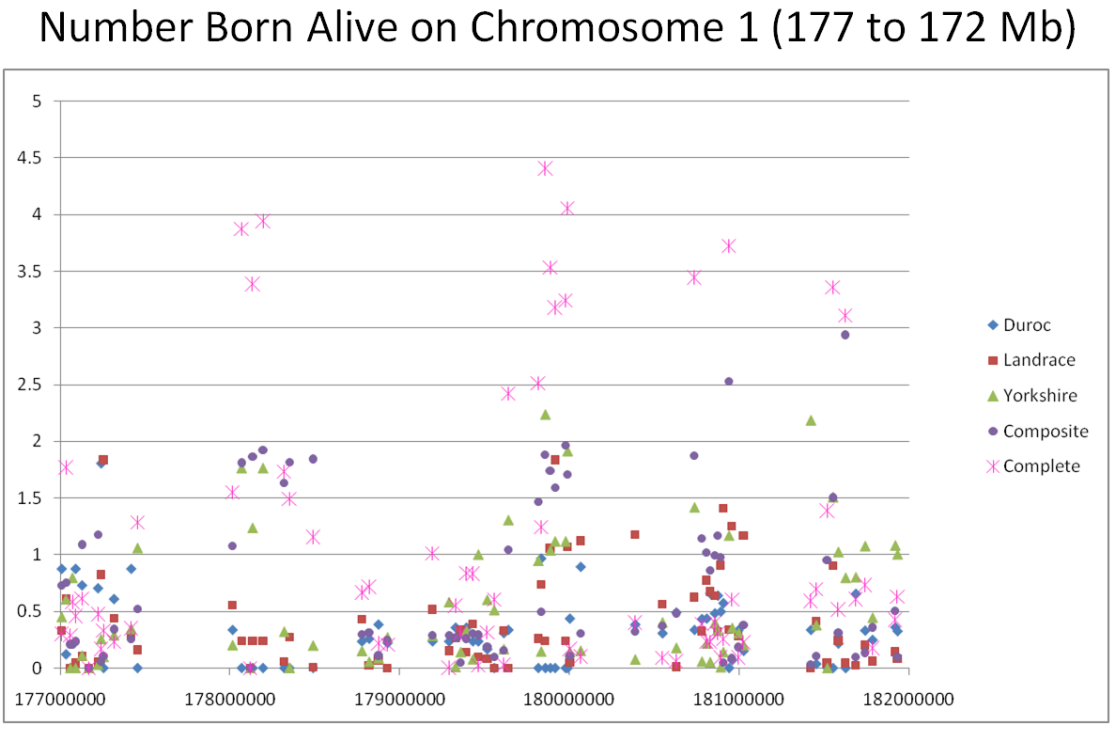
Rempel, L.A., D.J. Nonneman, T.H. Wise, T. Erkens, L.J. Peelman and G.A. Rohrer. 2010. Association analyses of candidate single nucleotide polymorphisms on reproductive traits in swine. J. Anim. Sci. 88:1-15.

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.

Rodriguez-Zas, S.L., C.B. Davis, P.N. Ellinger, G.D. Schnitkey, N.M. Romine, J.F. Connor, R. V. Knox, and B.R. Southey. 2006. Impact of biological and economic variables on optimal parity for replacement in swine breed-to-wean herds. J. Anim. Sci. 84:2555-2565.

Snelling, W.M, M.F. Allan, J.W. Keele, L.A. Kuehn, T. McDaneld, T.P. Smith, T.S. Sonstegard, R.M. Thallman and G.L. Bennett. 2010. Genome-wide association study of growth in crossbred beef cattle. J. Anim. Sci. 88:837-48.

Figure 1. Scatter plot of each test of significance for SNP within reported genomic regions. The x axis represents the position in the chromosome in megabases (Mb). The y axis is -log (p-value).



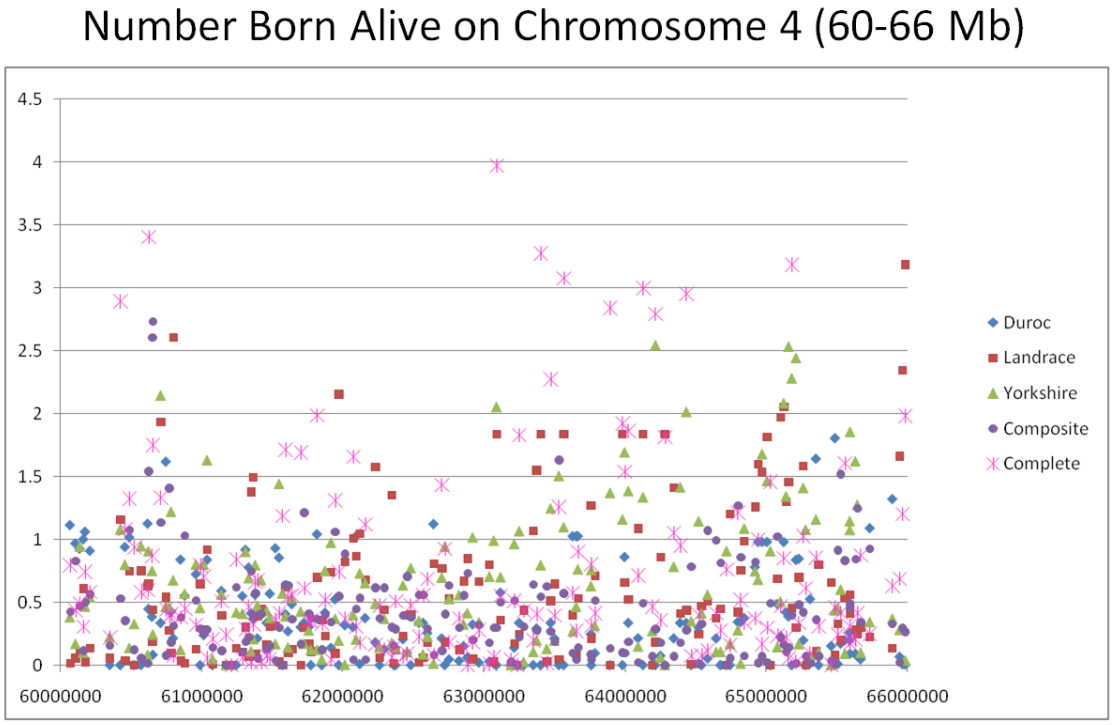
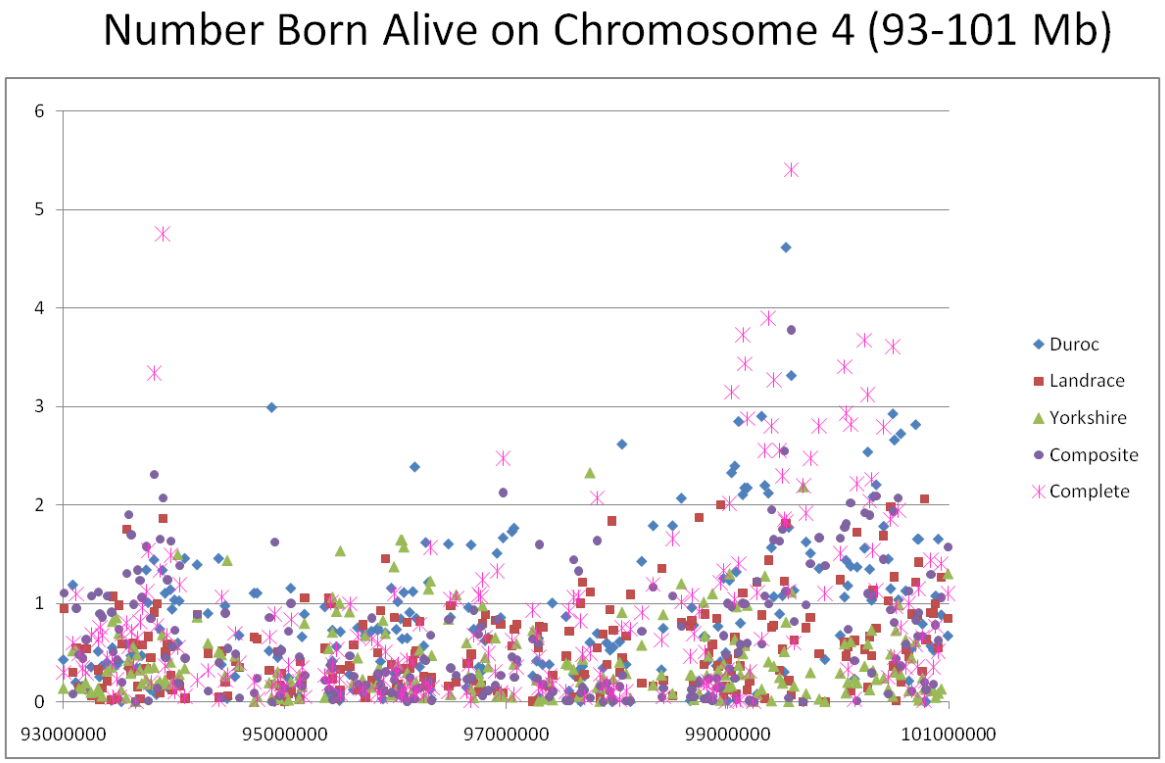


Figure 1. Continued



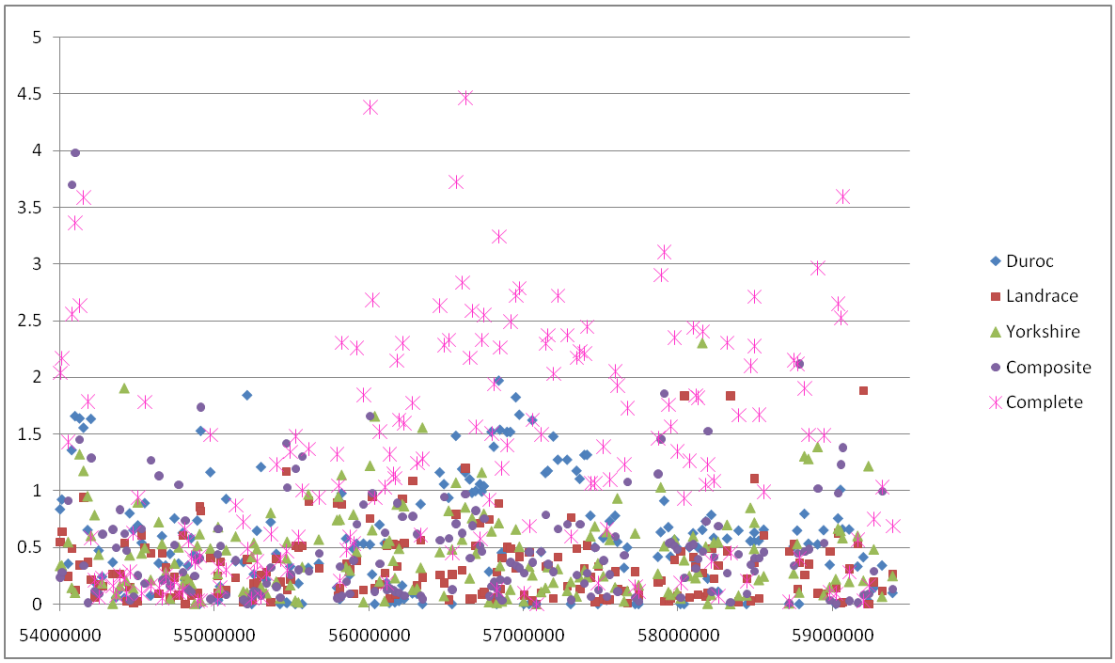
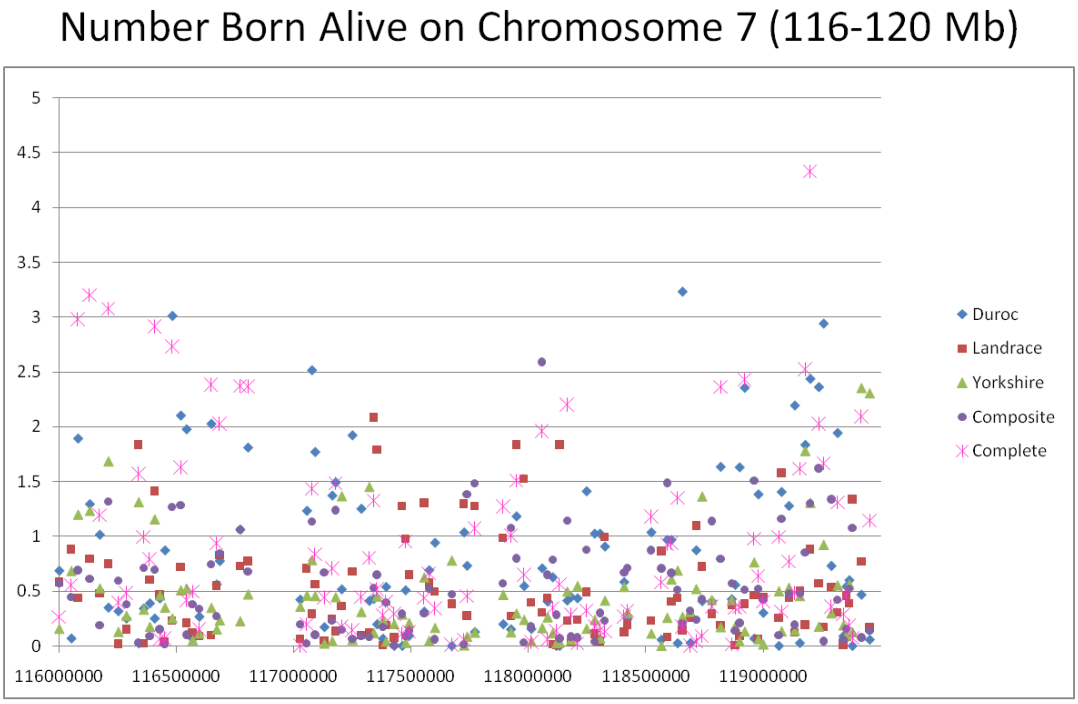


Figure 1 Continued



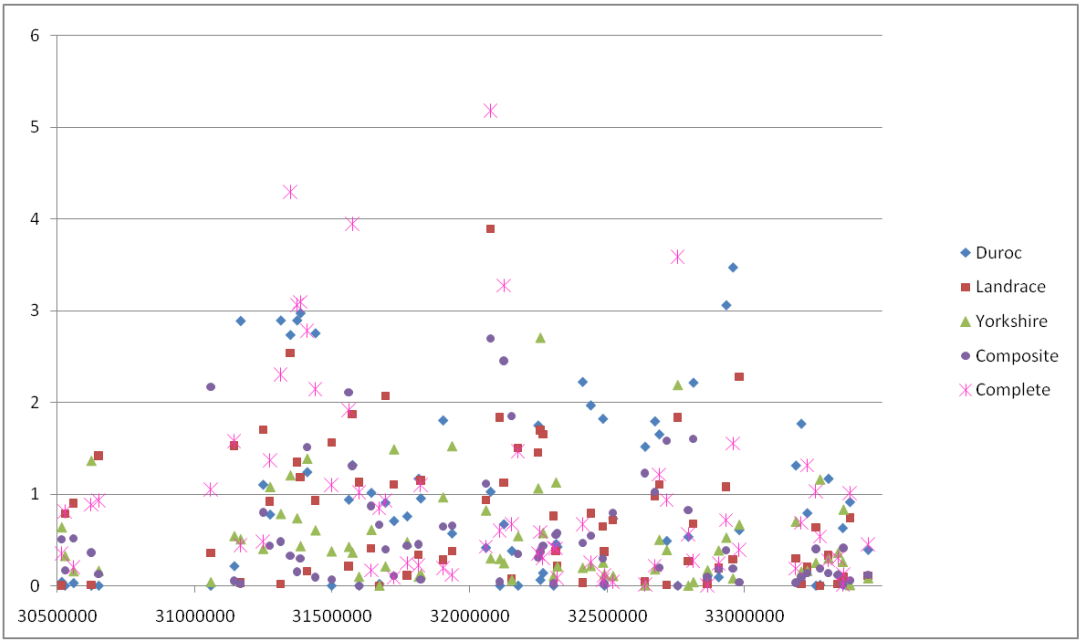
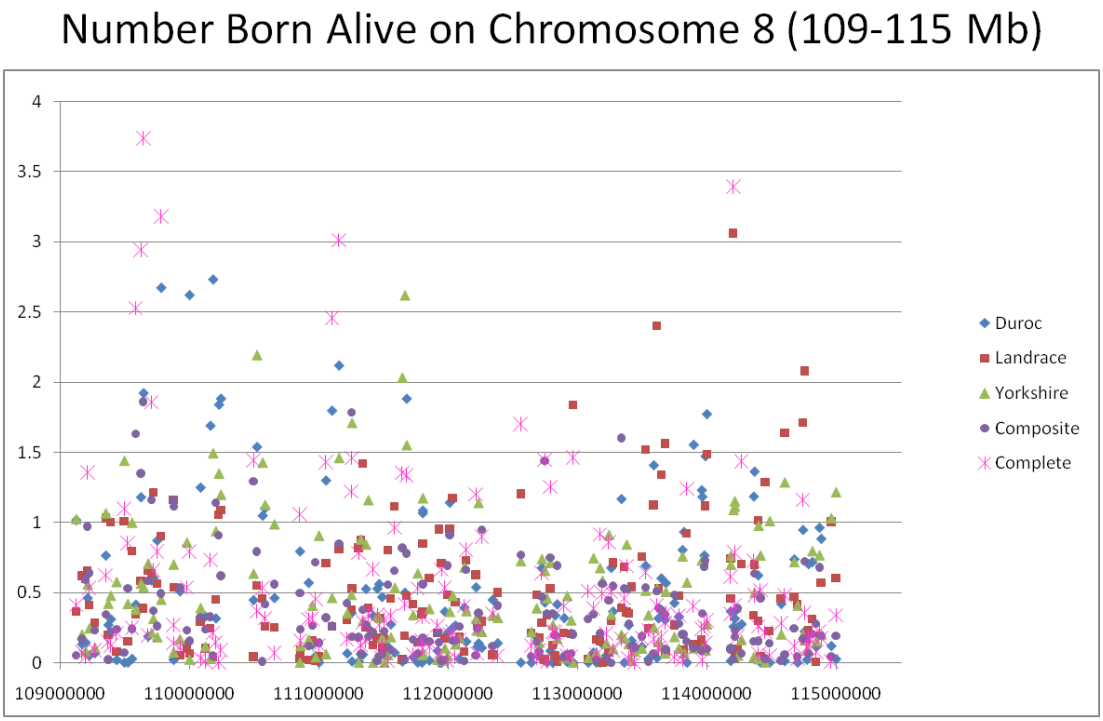


Figure 1. Continued



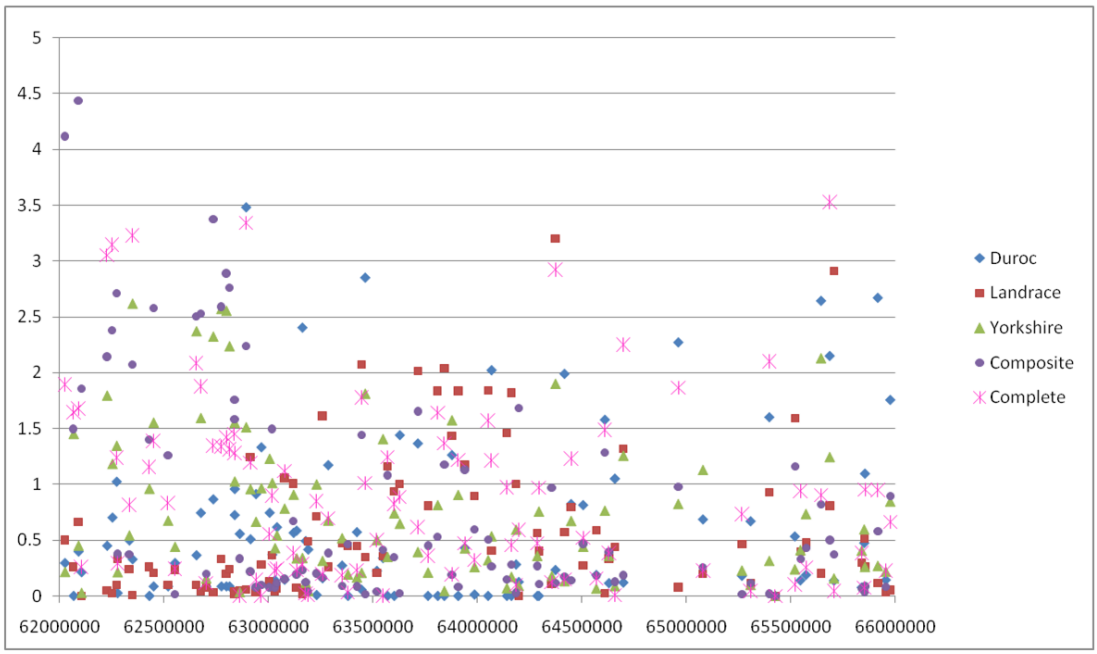
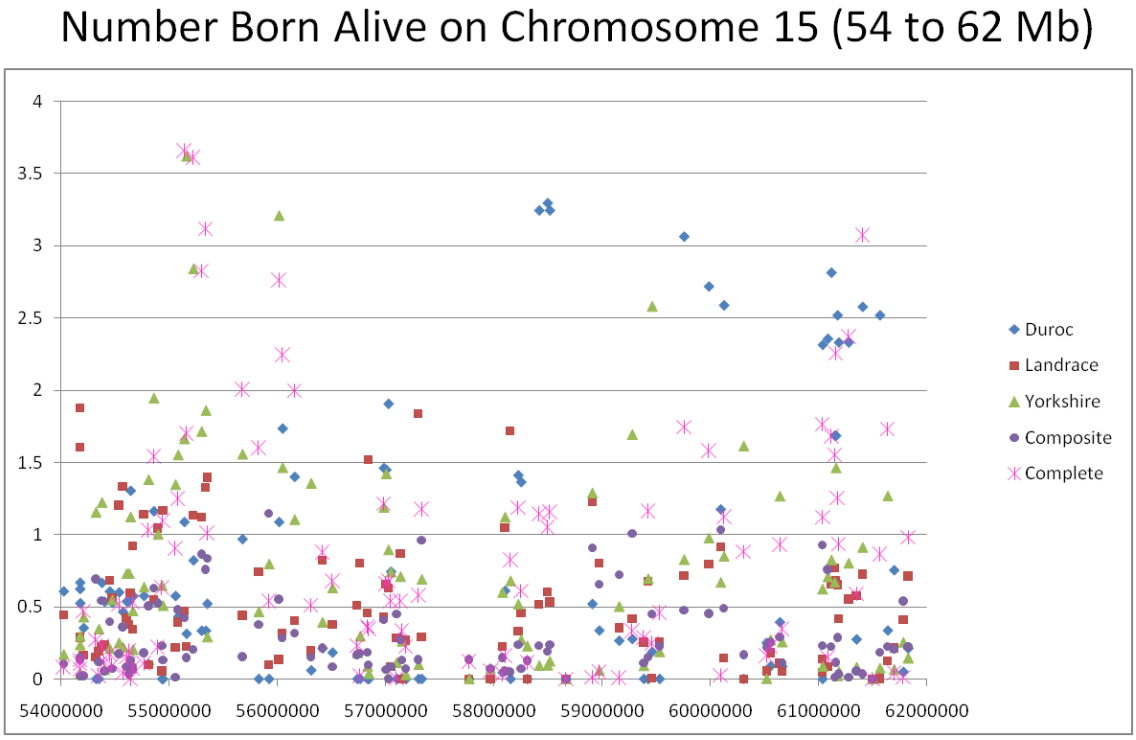


Figure 1. Continued



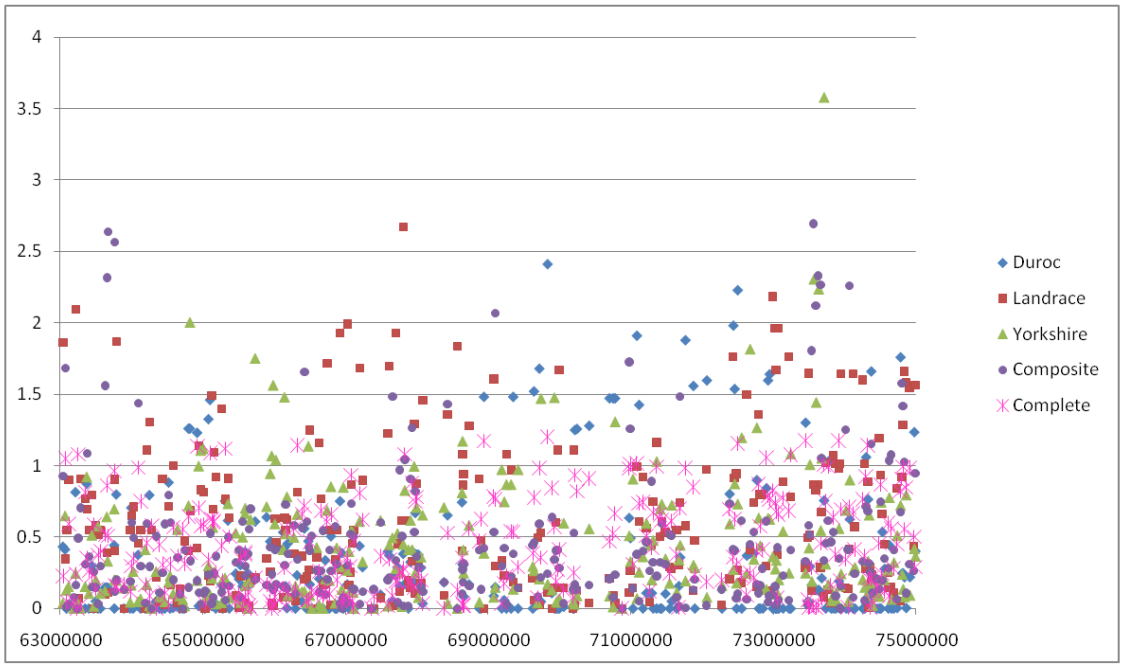


Figure 1. Continued

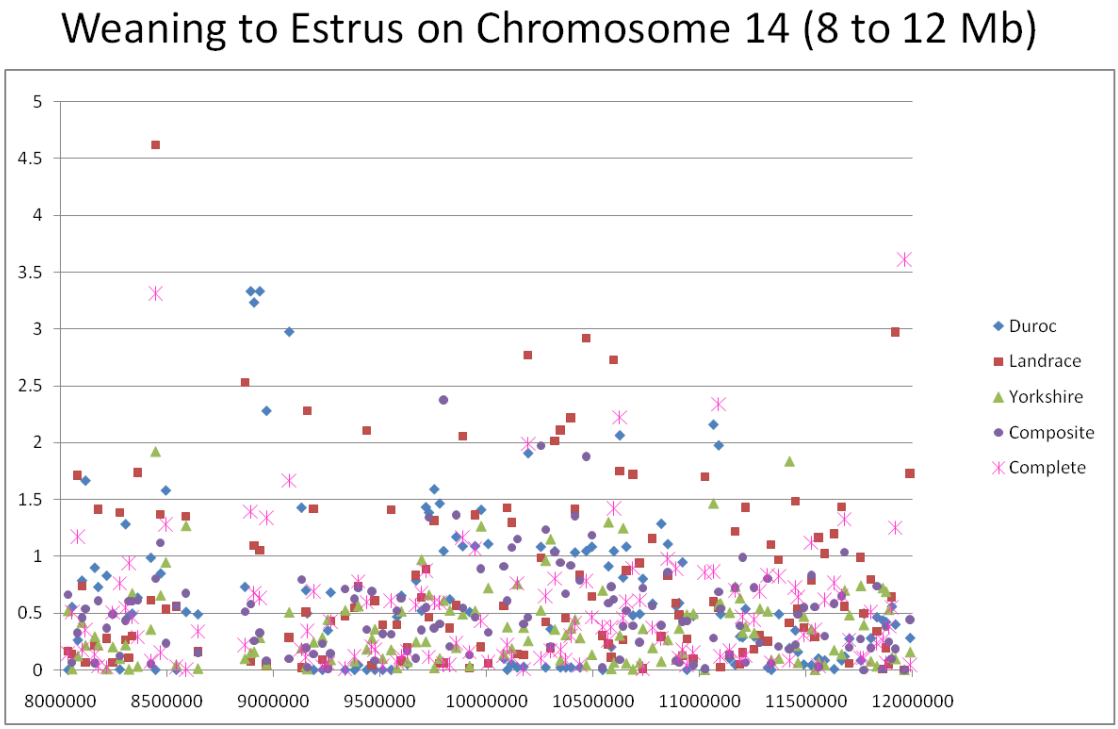


Table 1. Allele frequencies and estimated effects for the most significant SNP located in genomic regions identified.

Trait[a]	SSC	Position, Mb	Overall effect[b]	Overall f(A)	Duroc effect	Duroc f(A)	Landrace effect	Landrace f(A)	Yorkshire effect	Yorkshire f(A)	Composite effect	Composite f(A)
NBA	1	179.9	-.427 ± .103	.030	NS[c]	0	-.284 ± .502	.008	-.425 ± .152	.028	-.425 ± .170	.073
NBA	4	63.1	-.374 ± .096	.026	NS	0	NS	0	-.385 ± .146	.038	-.158 ± .234	.016
NBA	4	99.6	-.185 ± .040	.461	-.264 ± .074	.299	.017 ± .091	.505	-.143 ± .080	.531	-.345 ± .090	.484
NBA	7	56.6	-.160 ± .038	.365	-.164 ± .090	.190	-.212 ± .113	.161	-.114 ± .077	.610	-.154 ± .095	.435
NBA	7	119.2	-.165 ± .040	.645	-.218 ± .074	.441	-.149 ± .098	.808	-.167 ± .084	.681	-.194 ± .098	.685
NBA	8	32.1	-.273 ± .060	.868	-.144 ± .086	.722	-.666 ± .170	.924	.251 ± .374	.988	-.385 ± .123	.762
NBA	8	109.6	.142 ± .038	.526	.219 ± .086	.768	.072 ± .087	.557	.092 ± .087	.237	.228 ± .092	.586
NBA	11	65.7	-.175 ± .048	.776	-.228 ± .084	.751	-.162 ± .114	.811	-.170 ± .089	.765	-.112 ± .112	.782
NBA	15	55.1	.185 ± .050	.799	.290 ± .166	.949	.109 ± .113	.794	.194 ± .084	.669	.108 ± .122	.813
NBD	1	138.1	.574 ± .101	.976	NT[d]	NT	NT	NT	NT	NT	NT	NT
NBD	3	57.3	.184 ± .035	.851	NT	NT	NT	NT	NT	NT	NT	NT
WEI	4	73.1[e]	-.642 ± .169	.743	-1.004 ± .408	.659	-.883 ± .406	.768	-.199 ± .280	.778	-.438 ± .252	.762
WEI	14	11.1 [e]	-.433 ± .152	.290	-1.296 ± .501	.151	-.343 ± .317	.371	-.246 ± .219	.364	-.300 ± .274	.274

[a] Trait definitions are: NBA = Number born alive; NBD = Number born dead; and WEI = Weaning to estrus interval. [b] Estimated effects are the difference in performance when an A allele is substituted with a B allele. [c] This SNP was not segregating in this population. [d] This trait was not tested in sub groups. [e] The most significant SNP was not segregating in the subpopulations, so the next most significant SNP is reported.