Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Genome-Wide Association Study For Growth And Feed Intake in Duroc boars Utilizing Random Regression Models

***J. T. Howard*[*], F. Tiezzi[*], S. Jiao[*], K. A. Gray† and C. Maltecca[*]**
[*]North Carolina State University, †Smithfield Premium Genetics

ABSTRACT: Average daily body weight (DBW_{Avg}; n=676) and adjusted daily feed intake measurements (DFI_{Adj} ; n=846) from FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, KS, USA) stations were used to identify genomic regions that impact these traits and their relative change across the growth trajectory in Duroc boars. Single-step genomic BLUP with Legendre polynomials (order=2) were utilized to model the trajectory and estimate variance components. The EBVs for Legendre regression coefficients were used to estimate SNP variance for the intercept, slope and curvature. Estimated heritability across age ranged from 0.03 to 0.40 and 0.06 to 0.24 for DBW_{AVG} and DFI_{ADJ}. Multiple genomic regions were identified that impact the intercept and shape of the growth or feed intake trajectory. The use of age dependent genomic information allows for selection to be placed across the trajectory to alter growth and feed intake curves.
Keywords: swine; random regression; genome-wide association study

Introduction

Growth and feed intake are economically important traits for many livestock species and the estimation of individual growth or feed intake curves allows for selection to be placed across varying points of the trajectory. The use of longitudinal data analysis techniques to analyze growth or feed intake allows for the (co)variance structure to change across time. Previous research has proved the advantages from random regression models for growth and feed intake (Haraldsen et al. (2009); Schnyder et al. (2002); Wetten et al. (2012)) in order to select for different trajectories.

The use of genomic information to infer the estimated breeding value of an individual, referred to as genomic-EBV (DGV), has become a routine practice in multiple livestock species due to the rapid expansion and cost-effectiveness of genotyping technology. The majority of all traits utilized when estimating DGV are measures occurring at a single time point or averaged across time points. The use of genomic information in longitudinal type traits has been performed in chickens (Wolc et al. (2013)) and dairy cattle (Tetens et al. (2012)), although a limited amount of research has been done in swine. Also genome-wide association mapping across the growth trajectory has the potential to identify regions that have an effect only at a certain age along with regions that impact the trait consistently.

Previous research has used deregressed EBV across various ages regressed on genotypes to identify regions affecting milk yield (Tetens et al. (2012)) and replace the pedigree derived relationship matrix with a genomic relationship matrix to estimate DGV (Wolc et al. (2013)). An alternative approach would be to use single-step genomic BLUP (Aguilar et al. (2010)) and estimate SNP variance for the intercept, slope and curvature from the estimated Legendre regression coefficients in one step, as outlined in Wang et al. (2012). The objective of this study is to use single-step genomic BLUP with Legendre polynomials to model growth and feed intake curves and to conduct a genome-wide association study (GWAS) on the Legendre regression coefficients.

Materials and Methods

Data. Electronic FIRE (Feed Intake Recording Equipment, Osborne Industries, Inc., Osborne, KS, USA) station feed intake and weight measurements on 1,047 Duroc boars from July 22, 2007 to March 16, 2011 were initially utilized as described by Jiao et al. (Submitted). It has been previously reported electronic feeders are prone to errors for weight (Zumbach et al. (2009)) and feed intake (Casey et al. (2005)). Therefore feed intake editing techniques developed by Casey (2003) and Casey et al. (2005) were used to identify and remove errors. The remaining feed intake observations within each day were summed and adjusted (DFI_{adj}) to account for feed consumed during the visits that were removed based on the regression equation outlined by Casey (2003, Chapter 2). Lastly, animals with less than 25 DFI_{adj} observations and DFI_{adj} records greater than 4.5 kg were removed. Utilizing robust regression weight was fit to a quadratic regression of on-test day and linear regression of on-test age and weights that were zero were removed from the analysis as outlined by Zumback et al. (2009). Lastly, on-test ADG was computed by regressing age on weight and values less than .4 kg or greater than 2.0 kg were removed and the remaining weights were averaged by day (DBW_{Avg}). The final number of animals was 846 (n= 52,719 observations) and 676 (n= 40,988 observations) for DFI_{Adj} and DBW_{Avg}, respectively. The average (±SD) number of observations was 62.3 (±16.1) and 60.3 (±15.6) for DFI_{Adj} and DBW_{Avg}, respectively. Genotyping of the DNA samples was performed using Illumina PorcineSNP60K BeadChip (Illumina Inc., San Diego, 150 CA, USA) as described by Jiao et al. (Submitted). The SNP unmapped to the swine genome build 10.2 and SNP on sexual chromosomes were excluded

from the analysis, resulting in 35,140 SNP utilized in the analysis.

Statistical Analysis. Legendre polynomials (order = 2) were used to model trajectory of growth and feed intake across age. Analysis was carried out utilizing gibbs3f90 (Misztal (2008)) for 800,000 iterations with the first 200,000 discarded as burn-in and post-burn in samples were extracted every 60 iterations. Age was blocked into 7 classes (~14 days per class) and a heterogeneous variance structure was fit across time. The model for DFI_{adj} and DBW_{Avg} was,

$$Y_{ijmno} = \mu + CG_i + Pen_j + Parity_m + \sum_{k=0}^{2} \phi_{nok}\beta_k + CG_i + Pen_j + Parity_m + \sum_{k=0}^{2} \phi_{nok}\beta_k + \sum_{k=0}^{2} \phi_{nok}u_{nk} + \sum_{k=0}^{2} \phi_{nmk}pe_{nk} + e_{ijmno},$$

where Y_{ijmn} was DFI_{adj} or DBW_{Avg}, μ was average, CG_i was the fixed effect of concatenation of birthyear and season, Pen_j was the fixed effect of Pen, $Parity_m$ was the fixed effect of parity(1,2,3+), β_k was the fixed regression coefficient, u_{nk} was the effect of $animal_n$, pe_{nk} was the permanent environmental effect of $animal_n$ and e was the residual. The effect ϕ_{nok} is the kth Legendre polynomial for $animal_n$ at age_o. It was assumed $u \sim N(0, H \otimes G)$, where G was a 3x3 (co)variance matrix for the animal Legendre polynomials and $pe \sim N(0, I \otimes P)$, where P was a 3x3 (co)variance matrix for animal permanent environmental Legendre polynomials. Construction of the H matrix consisted of blending a 3-generation pedigree-derived numerator relationship matrix and a genomic relationship combined with weighting factors of 0.995 and 0.005, respectively (Aguilar et al. (2010)). The genomic relationship matrix was created using the method outlined by Van Raden (2008) were marker contribution is weighted by its expected variance.

Genome-Wide Association Mapping. Estimated breeding values for Legendre regression coefficient were used to estimate SNP effects as outlined by Wang et al. (2012), $\hat{u} = DZ'[ZDZ']^{-1}\widehat{a_n}$, where D is a diagonal matrix of weights for variance of SNP, Z is a matrix relating genotypes of each locus to $animal_n$, and $\widehat{a_n}$ is the EBV for Legendre regression coefficient for $animal_n$. Individual variance for SNP_i was estimated as: $\sigma^2_{\hat{u}_i} = \hat{u}_i^{\,2} * 2p_iq_i$, where p_i and q_i are the allele frequencies of SNP_i for the two alleles. SNP effects were blocked into sliding windows of 10 SNPs and genetic variances within each window (n = 34,959) were summed for each polynomial regression coefficient. The 5 largest 10-SNP sliding window variances were extended by 1 Megabase (Mb) in both directions to locate potential candidate genes.

Results and Discussion

The estimated heritability ranged from 0.03 to 0.40 and 0.06 to 0.24 for DBW_{AVG} and DFI_{ADJ} across age and the average heritability within each age class is outlined in Table 1. The range of heritability estimates is in line with previous results for both traits (Haraldsen et al. (2009); Schnyder et al. (2002); Wetten et al. (2012)). The genetic correlation across the age classes were moderate to highly positive ($r_g \geq 0.50$) throughout the growth curve, but become moderate to largely negative ($r_g \leq -0.50$) for the beginning and end of the growth curve. Therefore genes that have an effect on the early part of the trajectory are not consistently affecting the trait later in the trajectory. Due to this, locating genes that affect a trait across time using a single measure to characterize the entire trajectory may not be advantageous and furthermore the estimates are biased due to heterogeneity of effect across ages.

Table 1. Estimated heritability by age (days) class for adjusted daily feed intake (kg) and daily weight (kg).

	Age Class[2]						
Trait[1]	1	2	3	4	5	6	7
DBW_{Avg}	0.36	0.39	0.29	0.22	0.16	0.09	0.04
DFI_{Adj}	0.15	0.08	0.08	0.09	0.08	0.07	0.07

[1] DBW_{Avg} refers to average daily body weight and DFI_{Adj} refers to adjusted daily feed intake measurements.
[2] Age was blocked into 7 classes with approximately 14 days in each class.

Table 2. Top 5 largest 10 SNP sliding window variance for adjusted daily feed intake (kg).

Polynomial	Chromosome	Location (Mb[1])	Candidate Genes
β_0	1	168	*SOCS6, DOK6*
	1	256	*GNCT1*
	4	127	*AMY2B*
	9	145	*LPGAT1, PPP2R5A*
	13	6	
β_1	2	79	*OBSCN*
	5	2	*SULT4A1*
	9	1	*TUB*
	9	145	*LPGAT1, PPP2R5A*
	12	56	*PIK3R6*
β_2	1	256	*GNCT1*
	6	55	*GALP*
	12	56	*PIK3R6*
	13	6	
	15	2	

[1] Location refers to the start position in Megabases of the 10-SNP sliding window derived from swine genome build 10.2.

The 5 largest 10-SNP sliding window variances along with potential candidate genes for the intercept (β_0), slope (β_1), and curvature (β_2) are outlined in Table 2 and Table 3 for DFI_{adj} and DBW_{Avg}, respectively. The top 2.5% (n=875) sliding windows that were in close proximity to the top 5 10-SNP sliding windows were included in the estimate of proportion of variance explained (i.e. subset divided by total 10-SNP sliding windows) by the top 5 10-SNP sliding windows. Sliding windows were utilized to account for the variability in LD across the genome and to remove the chance of splitting a window in two, as is the case with using predefined window lengths. For DFI_{adj} the percent of variance explained by the top 5 10-SNP sliding windows were 0.021, 0.024, and 0.022 for the intercept, linear and quadratic regression coefficient. The window located on SSC1 (168 Mb) for the DFI_{adj} intercept coefficient was also found to be

significantly associated with ADFI by Jiao et al. (Submitted). A large portion of the regions had candidate genes within the enlarged 3 Mb window and in particular functions related to mucosal gastrointestinal tract homestatis (*GNCT1*; Yoshihisa et al. (2012)), stimulation of feed intake from the hypothalamus (*GALP*; Kuramochi et al. (2006)), and late onset of obesity in humans (TUB; Snieder et al. (2008)). For DBW_{Avg} the percent of variance explained by the top 5 10-SNP sliding windows were 0.047, 0.030, and 0.030 for the intercept, linear and quadratic regression coefficient. A region on SSC1 (270 Mb) harbors the *TGFBR1*, which has been previously found to be associated with ADG and multiple carcass traits (Chen et al. (2012)). Lastly, a region on SSC8 (46 Mb) has the gene *CPE*, which has been shown to be associated with obesity and insulin regulation (Cool et al. (1997)).

Table 3. Top 5 largest 10 SNP sliding window variance for daily weight (kg).

Polynomial	Chromosome	Location (Mb[1])	Candidate Genes
β_0	1	285	*SAL1*
	9	3	*DNHD1*
	9	8	*STARD10*
	17	7	*ASAH1*
	17	13	*SLC20A2*
β_1	6	120	
	9	3	*DNHD1*
	9	8	*STARD10*
	17	7	*ASAH1*
	17	13	*SLC20A2*
β_2	1	270	*TGFBR1*
	2	145	*SMAD5*
	7	127	
	8	46	*CPE, SC4MOL*
	8	64	

[1] Location refers to the start position in Megabases of the 10-SNP sliding window derived from swine genome build 10.2.

Conclusion

It has been shown that the use of genomic information to estimate an animal's growth curve is possible. Furthermore the effect of a SNP is transient throughout the trajectory and different weighting schemes can be utilized to alter growth and feed intake curves across the trajectory. Future research will involve combining DBW_{AVG} and DFI_{ADJ} along with carcass traits in order generate a gene network for the combination of growth, feed intake and carcass traits.

Literature Cited

Aguilar, I., Misztal, I., Johnson, D. L., et al. (2010) J. Dairy Sci. 93:743-752.

Casey, D. S. (2003) PhD Dissertation. Iowa State University, Ames.

Casey, D. S., Stern, H. S., & Dekkers, J. C. M. (2005) J. Anim. Sci. 83:969-982.

Chen, K., Hawken, R., Flickinger G. H., et al. (2012) Animal Biotechnology 23: 43–63.

Cool, D. R., Normant, E., Shen, F., et al. (1997) Cell 88 73-83.

Haraldsen, M., Ødegård, J., Olsen, D., et al. (2009) Animal 3:475-481.

Jiao, S., Maltecca, C., Gray, K. A. et al. (2014) J. Anim. Sci. (Submitted).

Kuramochi, M., Onaka, T., Kohno, et al. (2006) Endocrinology 147:1744–1752.

Misztal, I. (2008) J. Anim. Breed. Genet. 125:363–370.

Snieder, H., Wang, X., Shiri-Sverdlov, R., et al. (2008) Diabetologia 51:54–61.

Schnyder, U., Hofer, A., Labroue, et al. (2002) Genet. Sel. Evol. 34:61-81.

Tetens, J., Seidenspinner, T., Buttchereit, et al. (2012) Animal Genetics 44:1-8.

VanRaden .P.M. (2008) J. Dairy Sci. 91:4414-4423.

Wang, H., Misztal, I., Aguilar, I., et al. (2012). Genetics Research Cambridge 94:73-84.

Wetten, M., Ødegård, J., Vangen, et al. (2012) Animal 6: 433-439.

Wolc ,A., Arango, J., Fulton, et al. (2013) Poultry Sci. 92:1486-91.

Yoshihisa N., Hamajima, Y., Komori, et al. (2012) Int J Otolaryngol 2012: Article ID 438609.

Zumbach, B., Misztal, I., Chen, et al. (2010) J. Anim. Breed. Genet. 127:93-99.