Detection Of Imprinting In Two Commercial Pig Populations

Albart Coster[*], Ole Madsen[*], Henri Heuven[*], Johan van Arendonk[*] and Henk Bovenhuis[*]

Introduction

Genomic imprinting is an epigenetic phenomenon where the level of expression of alleles depends on their parental origin. On the molecular level, genomic imprinting is established by differential methylation of particular chromosomal regions which results in a parent-of-origin dependent expression of RNA (Feil and Berger, 2007). On the phenotypic level, imprinting is manifested through a contrast between the two heterozygote classes that exist for a genotype (A/a and a/A) (Hager *et al.*, 2009).

Genomic imprinting has been found in viviparous mammals and in seeded plants (Feil and Berger, 2007). In mammals, imprinting has mainly been studied in human and mice, and to a lower account in other species as pig, cattle, and sheep (Imprinted gene catalog: `http://igc.otago.ac.nz/`). Imprinted genes are organized in clusters and play key roles in growth and development (Feil and Berger, 2007). In pigs, the imprinted gene Igf2 was found associated to muscle mass and backfat thickness (de Koning *et al.*, 2000; Van Laere *et al.*, 1998)

A preliminary study at our group of a F2 Meishan x commercial cross indicated imprinting effects on reproductive traits. The objective of this research was to assess the degree of imprinting affecting reproductive traits in two commercial pig populations.

Material and methods

Composition of the populations. In this study, data of sows from two purebred lines from breeding companies Hypor (LW1) and Topigs (LW2) were used. Both lines belonged to the large-white breed. To enable accurate inference of allele origin, a sow was only selected when her father was available for genotyping and when more than nine of her paternal halfsibs were available for genotyping. Ancestors of the selected sows were selected when available, this included the fathers of the selected sows.

Phenotypes considered in this study were the total number of piglets born (TB) and the number of piglets born alive (LB). Records of the first four parities were used in the analyses, records of later parities were discarded from the analyses together with records defined as outliers. There were records of approximately 4000 sows in the data of line LW1, of which 490 were

[*]Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH,Wageningen, The Netherlands

genotyped. There were records of approximately 3000 sows in the data of line LW2, of which 973 were genotyped.

Composition of the marker panel. Fifteen chromosomal regions were selected which were orthologous to regions in human and mouse in which imprinting was detected. An Illumina 384-plex golden gate assay was used for genotyping. The 384 markers were distributed over the fifteen regions; the number of markers per region was a function of the size of the region and of the number of imprinted genes detected in that regions in human and mouse, with a minimum of 20 markers per region.

`Mendelsoft` (de Givry *et al.*, 2005) was used to identify and correct mendelian inconsistencies in the genotype data. Subsequently, `cvmhaplo` (Albers *et al.*, 2006) was used to infer allele origin. Due to the large scale of the data, `cvmhaplo` was run separately for each company on a sliding window of six consecutive markers.

Statistical analyses. The markers were analyzed in an animal model that included an additive, a dominance and an imprinting effect for each marker separately. The following model was fitted to the data in `ASReml` (Gilmour *et al.*, 2002):

$$\begin{aligned} \mathbf{y} =& \mathbf{Xb} + \mathbf{Qq} + \mathbf{Za} + \mathbf{Z}_{pe}\mathbf{pe} + \mathbf{Mv} + \mathbf{e} \\ & \mathbf{a} \sim \mathbf{N}(0, \mathbf{A}\sigma_a^2); \quad \mathbf{pe} \sim \mathbf{N}(0, \mathbf{I}\sigma_{pe}^2); \\ & \mathbf{v} \sim \mathbf{N}(0, \mathbf{I}\sigma_m^2); \quad \mathbf{e} \sim \mathbf{N}(0, \mathbf{I}\sigma_e^2), \end{aligned} \tag{1}$$

where **y** is the vector of phenotypic observations, **X** is an incidence matrix for the fixed effects, **b** is an unknown vector of fixed effects, **Q** is the incidence matrix of a marker which is explained in the next paragraph, **q** is an unknown vector of marker effects, **Z** is an incidence matrix for the random animal effects, **a** is an unknown vector of breeding values, $\mathbf{Z}_{pe}$ is an incidence matrix of permanent environment effects, **pe** is an unknown vector of permanent environment effects, **M** is a an incidence matrix of maternal effects, **v** is an unknown vector of maternal effects, and **e** is the residual of the model. Maternal effects were included in the model because these can be confounded with imprinting effects (Hager *et al.*, 2008). Fixed effects included in the model were a class effect accounting for the breed of the litter, the class effect accounting for parity of the sow and the class effect accounting for the combination of farm, year and season. The model was fitted independently for each marker in each combination of company and trait.

Incidence matrix **Q** relates the genotype of the sows to the corresponding elements of **q** , which correspond to the additive, dominance and imprinting effect of a marker. Matrix **Q** was build as $\mathbf{Q} = \mathbf{GS}$, where **G** is matrix denoting to which of the four genotype classes (00,01,10,11) each genotype belongs and **S** the contrast matrix used by Hager *et al.* (2008):

$$\mathbf{S} = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{A} & \mathbf{D} & \mathbf{I} \end{array} \\ \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 0 \end{bmatrix} \end{array}.$$

Table 1: Regions in which an effect with a FDR below 0.10 was found, letters indicate whether this was an additive (A), dominance (D), or imprinting (I) effect.

Company	Trait	4	7	8	10	12	13
LW1							
	LB						D
	TB			D		I	D
LW2							
	LB				D	A	
	TB	A	A		D	A	A

The first column of **S** corresponds to the additive effect, the second column of **S** corresponds to the dominance effect and the third column of **S** corresponds to the imprinting effect. The four rows of **S** correspond to the four genotype classes.

Incremental F-ratios were calculated for the three effects corresponding to each marker, where the marker was included as the last fixed effect in the model. Dominance was included after additive and imprinting was included after dominance, corresponding the the order of the columns of **Q** .

After fitting the models, false discovery rate (FDR) was calculated for each significance value using R-package `qvalue` (Dabney *et al.*, 2009). False discovery rate was calculated within each combination of line, trait, and genetic effect. Tests with a false discovery rate below 0.1 were considered as significant.

Results and discussion

Analyses without marker effects were performed to estimate variance components for the random effects in Model 1. Heritability of both traits was 0.10 in the LW1 data. In the LW2 data, h^2 of TB was 0.16 and h^2 of LB was 0.12. The standard error of h^2 was lower than 0.03 in the four analyses.

Effects with a significant FDR were located in regions 4, 7, 8, 10, 12, and 13 (Table 1). Most significant effects were found in the LW2 data, which is not surprising because most of the genotyped sows belonged to this line. A significant imprinting effect was found in region 12 for line LW1.

The imprinting effect in region 12 was due to a single marker (Figure 1). The FDR of this effect was 0.075. Estimated variance of this effect was 0.10; approximately 14% of σ_a^2. The sign of this effect indicated that a maternal allele led to a greater TB than a paternal allele.

The estimated variances of the additive effects with a significant FDR, expressed as percentage of σ_a^2, ranged between 10% and 1%. The estimated variance of the dominance effects with a significant FDR, expressed as percentage of σ_a^2 ranged between 32% and 4%.

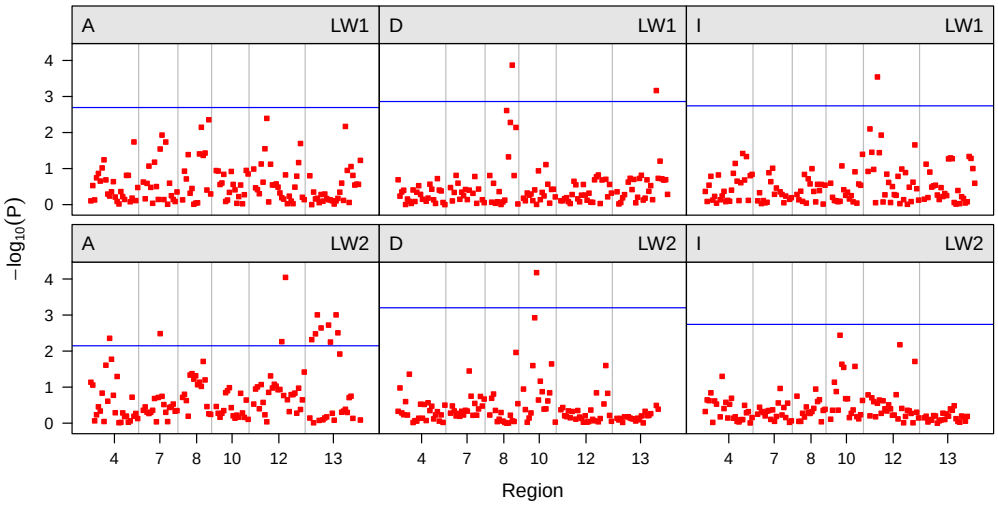

Figure 1: Plot of -$\log_{10}$(P) of marker effects in genomic regions 4, 7, 8, 10, 12, and 13. The horizontal lines are 10% FDR cutoff when this could be determined.

Acknowledgments

We acknowledge Technologiestichting STW for the financial support of this work. We acknowledge Hypor and Topigs for providing the data on which this work is based.

References

Albers, C. A., Leisink, M., and Kappen, H. J. (2006). *BMC Bioinformatics*, 7:S1.

Dabney, A., Storey, J. D., and with assistance from Gregory R. Warnes (2009). *qvalue: Q-value estimation for false discovery rate control*. R package version 1.18.0.

de Givry, S., Palhiere, I., Vitezica, Z. *et al.* (2005). In *Proceedings of WCB05 Workshop on Constraint Based Methods for Bioinformatics*. page 47.

de Koning, D. J., Rattink, A., Harlizius, B. *et al.* (2000). *Proceedings of the National Academy of Sciences*, 97:7947.

Feil, R. and Berger, F. (2007). *Trends in Genetics*, 23:192 – 199.

Gilmour, A. R., Cullis, B. R., Welham, S. J. *et al.* (2002). *ASREML Reference Manual*. NSW Agriculture, Locked Bag, Orange, NSW 2800, Australia, 2 edition.

Hager, R., Cheverud, J. M., and Wolf, J. B. (2008). *Genetics*, 178:1755–1762.

Hager, R., Cheverud, J. M., Wolf, J. B. *et al.* (2009). *Evolution*, 63:1118–1128.

Van Laere, A., Nguyen, M., Braunschweig, M. *et al.* (1998). *Hippocampus*, 8:244–261.