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

Genome-Wide Association Study for Intramuscular Fat Content and Composition in Duroc Pigs

R. Ros-Freixedes[1], **S. Gol**[1], **R. N. Pena**[1], **M. Tor**[1], **J. C. M. Dekkers**[2] **and J. Estany**[1]
[1]University of Lleida – Agrotecnio Center, Spain, [2]Iowa State University, USA.

ABSTRACT: Intramuscular fat (IMF) content and fatty acid composition affect the organoleptic quality and nutritional value of pork and, thus, there is increasing interest in including these traits in selection schemes. In a genome-wide association study based on 138 purebred Duroc barrows we detected only weak associations with IMF content. In contrast, strong associations were found between two chromosomal regions and IMF composition, which co-localized with the *SCD* (SSC 14) and the *LEPR* (SSC 6) genes. The *SCD* gene is responsible for the biosynthesis of oleic acid from stearic acid and affects saturated and monounsaturated fat content. The association of *LEPR* with IMF composition is, at least in part, a consequence of its effect on IMF content, resulting in higher saturated and lower polyunsaturated fatty acids. Markers in these two genes could be useful to genetically improve intramuscular fatty acid composition.
Keywords: fatty acids; genome-wide association study; intramuscular fat

Introduction

Intramuscular fat (IMF) content and fatty acid composition affect both organoleptic quality and nutritional value of pork and, thus, there is increasing interest in including these traits in the selection objectives for pig lines for quality pork markets. Both traits have been shown to display substantial genetic variation (Ros-Freixedes et al., 2012), with evidence of associations with single nucleotide polymorphisms (SNPs). In particular, several SNPs in the promoter region of the stearoyl-CoA desaturase (*SCD*) gene have been associated with desaturase activity in Duroc pigs (Uemoto et al., 2012a; Estany et al., 2014), resulting in greater content of monounsaturated fatty acids (MUFA) but not IMF. The objectives of this study were to use genome-wide association study (GWAS) techniques to validate the role of *SCD* genetic variants on IMF composition, and to detect other candidate genes affecting IMF content and composition. The potential use of these markers for marker-assisted selection is discussed.

Materials and Methods

Animals and data. We genotyped 138 purebred Duroc barrows from the line described in Ros-Freixedes et al. (2012) using the PorcineSNP60 v2 Genotyping BeadChip (Illumina, CA). Animals were chosen to be as unrelated as possible and representative of the whole population. Half of the animals (n=66) were born in 2002-2003, and the other half (n=72) in 2009-2010. Animals were raised in 6 batches (3 in each period) under commercial conditions and fed ad libitum with a pelleted finishing diet from 160 days of age until slaughter. Animals were slaughtered in the same commercial abattoir at ~210 days of age (~125 kg of body weight). Immediately following slaughter, a sample of subcutaneous fat (SF) at the level of the third and fourth ribs was collected. After chilling for about 24 h at 2ºC, samples of the muscles *gluteus medius* (GM) and *longissimus dorsi* at the same location as the SF sample (LD) were also collected. The IMF content and fatty acid composition of the samples were determined in duplicate by quantitative determination of the individual fatty acids by gas chromatography. The IMF content was calculated as the sum of each individual fatty acid expressed as triglyceride equivalents and expressed as percentage of fresh sample (mean in GM: 5.4%, SD 2.1). Fatty acid contents were expressed as the percentage relative to total fatty acids, including total saturated (SFA; 38.7%, SD 2.0), MUFA (48.5%, SD 2.0), and polyunsaturated (PUFA; 12.8%, SD 2.0) fatty acids, and individual oleic acid (C18:1; 44.1%, SD 2.0). The desaturation ratios of oleic to stearic acid (C18:1/C18:0) (3.6, SD 0.5) and SFA/PUFA (3.1, SD 0.6) were calculated.

High-density SNP data quality control. The PLINK software (Purcell et al., 2007) was used to filter out SNPs with minor allele frequency below 0.05 and genotyping rate below 0.95, and individuals with more than 10% missing genotypes. Unmapped SNPs based on the current pig genome assembly *Sus scrofa* (SSC) Build 10.2 were also excluded. The remaining data comprised 135 individuals and 36,432 SNPs.

Genome-wide association study. GWAS for the described traits were performed using GenSel (Fernando and Garrick, 2009). For each trait we used the Bayes B approach with π=0.997 (where π is the prior proportion of SNPs considered to have no effect on the trait; i.e., our model fitted ~110 SNPs per iteration), with a Markov chain of 750,000 iterations and a burn-in of 250,000. Apart from the additive SNP effects, which were fitted as random effects, the model included batch as a systematic fixed effect and age at slaughter as a covariate. The posterior means and posterior samples of the effects of all SNPs within 1-Mb windows were collectively used to predict the genomic merit of the window and the proportion of total genetic variance that the window accounted for, respectively. Windows that accounted for at least 1% of total genetic variance were considered as candidate regions, including contiguous 1-Mb windows that accounted for at least 0.25%.

Genomic prediction. We used the animals born in 2002-2003 as training data to estimate the SNP effects and then to predict the genomic estimated breeding values (GEBV) of the animals born in 2009-2010. The effect of each SNP was estimated using the same procedure as for GWAS but with π=0.9985. The correlation between GEBV and the adjusted phenotypic values was used as a measure of the prediction accuracy. Results on GEBV were compared to those obtained using only SNPs from the two regions that explained the greatest amount of variance and with pedigree-based estimated breeding values (P-EBV). The P-EBV and heritabilities of the traits were obtained as described in Ros-Freixedes et al. (2012) using all available records since 2002 but excluding those from the animals in the 2009-born testing set (n=1,132).

Results and Discussion

Genome-wide association study. The regions associated with IMF content and composition in GM are given in Table 1 and the genes located in each of these regions in Table 2. A region in SSC 14 at 120.1-124.0 Mb was found to be associated with SFA, MUFA, C18:1, and the desaturation index C18:1/C18:0. This region corresponds to the location of the *SCD* gene, thereby confirming the association reported in the same population by Estany et al. (2014) in which a three-SNP haplotype in the promoter region of the *SCD* gene affected C18:1/C18:0. The SCD is the rate-limiting enzyme for the biosynthesis of MUFA C18:1 from SFA C18:0 and therefore, although other genes involved in lipid metabolism are located in this region, *SCD* remains the strongest candidate gene for this association. The percentages of genetic variance explained by this region for each trait were close to those obtained by Estany et al. (2014) using pedigree-based estimates and available records. Interestingly, the same region was also found to be associated with these traits in LD and SF.

Table 1. Candidate regions for intramuscular fat content and composition in gluteus medius muscle and percentage of genetic variance accounted for by trait.

Trait[1]	Chr	Position (Mb)	% σ_a^2
SFA	6	135.1-137.0	9.0
	7	94.3-96.0	1.2
	12	24.0-24.9	1.0
	14	120.1-124.0	9.8
MUFA	2	23.0-25.0	1.0
	3	13.2-15.0	1.1
	6	68.0-71.0	1.0
	14	118.0-124.0	11.4
C18:1	6	99.2-102.0	1.0
	13	46.0-50.0	1.9
	14	120.1-124.0	3.7
	18	17.0-21.0	1.2
PUFA	1	306.0-306.9	1.9
	6	135.1-136.0	29.8
C18:1/C18:0	1	309.0-309.9	1.3
	3	13.2-15.0	2.5
	13	210.1-212.8	1.1
	14	120.1-124.0	25.9
SFA/PUFA	6	135.1-136.0	34.5
IMF	3	3.0-3.9	1.4
	17	47.0-48.0	1.7

[1]SFA = saturated fatty acids; MUFA = monounsaturated fatty acids; PUFA = polyunsaturated fatty acids; C18:1 = oleic acid; C18:0 = stearic acid; IMF = intramuscular fat content.

Table 2. Candidate genes within identified regions.

Chr	Position (Mb)	Genes[1]
1	306.0-306.9	*CEL*
1	309.0-309.9	*LCN9**, *KCNG2**
2	23.0-25.0	-
3	3.0-3.9	*FOXK1**
3	13.2-15.0	-
6	68.0-71.0	-
6	99.2-102.0	*GATA6*, *NPC1*, *OSBPL1A*
6	135.1-136.0	*LEPR*, *LEPROT*
7	94.3-96.0	*AKAP5*
12	24.0-24.9	*STAT5A*, *GIP**, *MED1**, *OSBPL7**, *NGFR**, *STARD3**
13	46.0-50.0	*ACOX2***
13	210.1-212.8	-
14	120.1-124.0	*ELOVL3*, *GOT1*, *NKX2-3*, *SCD*
17	47.0-48.0	*ADIG*, *LPIN3***
18	17.0-21.0	*MEST*, *LEP**

[1]Only genes involved in lipid metabolism are shown.
* (**) Within 1 Mb (2 Mb) from the described region.

A region in SSC 6 at 135.1-137.0 Mb was found to be associated with SFA and PUFA and, as a result, with SFA/PUFA. This region includes the leptin receptor (*LEPR*) and the *LEPROT* genes. *LEPROT* encodes a protein that negatively regulates the expression of the leptin receptor in the cell surface, decreasing the response to leptin. A polymorphism in exon 14 of *LEPR* has already been found to be associated with increased feed intake and fatness, measured both in terms of SF and IMF (Óvilo et al., 2005; Uemoto et al., 2012b). Moreover, several quantitative trait loci (QTL) for IMF have been reported in this region (animalgenome.org/QTLdb/pig.html). Although we did not find an association of this region with IMF, the percentage of genetic variance accounted for this region decreased to 3.6% and 13.4% for SFA and PUFA, respectively, when IMF content was included in the model. This indicates that, at least in part, the associations of the SNPs in this region with SFA and PUFA are an indirect effect of increased IMF. It is well known that the endogenous synthesis of SFA and MUFA increases with IMF content, which leads PUFA to proportionally decrease. Again, the associations of the *LEPR* (*LEPROT*) region with SFA and PUFA in GM were replicated in LD and SF. Moreover, the leptin gene (*LEP*) is located only 0.2 Mb downstream from the 17.0-21.0 Mb region on SSC 18 that was associated with C18:1.

Other associations were found for IMF content and composition in GM, but they were all weak and in general were not confirmed in LD or SF. For example, for IMF, we found an interesting association with a region in SSC 17 at 47.0-48.0 Mb, which includes the adipogenin (*ADIG*) gene, which is related to adipogenesis. However, this association only accounted for 1.7% of the genetic variance for IMF in

GM and was not found for LD, for which the strongest association was on SSC 11 at 19.3-21.0 Mb.

Table 3. Accuracy of genomic (GEBV) and pedigree-based (P-EBV) estimated breeding values and heritability (h^2) of the traits.

	GEBV[2]				Pedigree[3]	
Trait[1]	36k	*SCD*	*LEPR*	*SCD*+ *LEPR*	P-EBV	h^2
SFA	0.31	0.23	0.24	0.31	0.47	0.40
MUFA	0.28	0.26	0.36	0.30	0.48	0.50
C18:1	0.21	0.22	-0.35	0.24	0.42	0.45
PUFA	0.07	0.12	0.49	0.51	0.70	0.59
C18:1/C18:0	0.40	0.40	-0.05	0.39	0.26	0.44
SFA/PUFA	0.12	-0.06	0.43	0.41	0.69	0.54
IMF	0.11	0.14	0.19	0.24	0.52	0.51

[1]See abbreviations in Table 1.
[2]36k: using all 36,432 SNPs; *SCD*: using only 3 SNPs at the *SCD* promoter; *LEPR*: using only ASGA0089937 and ASGA0093565; *SCD+LEPR*: using SNPs at both *SCD* and *LEPR* loci.
[3]Using the full pedigree and 1,132 records.

The small sample size forced us to use a high π (few SNPs fitted per iteration), which may explain why only very strong associations were detected. For the same reason, the genetic variance that they explained may be overestimated. However, despite using bigger sample sizes, other reported GWAS experiments for fatty acid composition in Duroc did not reach much different results. Uemoto et al. (2012c), using a low-density linkage map in purebred Duroc, found only significant associations on SSC 14 at ~90-115 Mb for C18:1 and C18:0 in both LD and SF, and for the melting point of SF, which is a trait related to the unsaturation degree of fat. None of their other suggestive QTL for IMF composition matched those found here. Yang et al. (2013), using a Duroc × Erhualian F_2 cross, gave rather similar results, with the *SCD* locus being the only reported QTL for major fatty acids in IMF. No coincident regions were found with other reported GWAS experiments using Iberian and Landrace (Ramayo-Caldas et al., 2012; Muñoz et al., 2013), including the *SCD* locus. This is consistent with previous analyses showing no segregation of *SCD* genetic variants in these breeds (Estany et al., 2014). However, it is interesting to note that some of the experiments cited above generally found more QTL for minor than for major fatty acids (like C18:1). When performing GWAS for fatty acid composition it is important to be aware of the compositional nature of these data (Ros-Freixedes and Estany, 2013), i.e., when one fatty acid percentage increases, at least one other fatty acid has to decrease, which can cause spurious results. Because of this, associations found for minor fatty acids, which are more variable and more influenced by the relative amount of major fatty acids, can be particularly meaningless.

Genomic prediction. Accuracies of GEBV based on 36,432 SNPs were ~0.40 for C18:1/C18:0, ~0.30 for SFA and PUFA, ~0.20 for C18:1, and ~0.10 for PUFA, SFA/PUFA, and IMF. However, it is worth noting that for C18:1/C18:0, MUFA, and C18:1, the same prediction accuracies were obtained when using for prediction only the three SNPs reported by Estany et al. (2014) at the promoter region of the *SCD* gene. A combination of SNPs at the *SCD* promoter and at the *LEPR* (*LEPROT*) locus (ASGA0089937 and ASGA0093565) raised accuracies to ~0.50 for PUFA, ~0.40 for SFA/PUFA, and ~0.25 for IMF. However, accuracies of GEBV for IMF and C18:1 were always lower than those of P-EBV.

These results should be extrapolated with caution. Here we used a very small training set (n=65) to predict the GEBV of a very small testing set of pigs born 7 years later (n=70). In contrast, P-EBV were predicted using records collected in several batches throughout the period. Whether markers in the *SCD* and *LEPR* loci are useful for improving fatty acid composition or not should be evaluated against the cost of routine phenotyping for these traits. Overall, these results show that GWAS on small amount of data is able to capture the effect of genes with relative high effects segregating at intermediate frequencies (minor allele frequencies at *SCD* and *LEPR* loci were 0.42-0.45). Because pigs in the training and the predicted sets were separated by a 7-year span these results also confirm that the effect of the SNP at *SCD* and *LEPR* loci is consistent across generations.

Conclusion

The GWAS analysis allowed us to confirm the segregation of SNPs affecting fatty acid composition in the *SCD* locus in the investigated Duroc line. Another set of SNPs was identified that segregated at the *LEPR* (*LEPROT*) region, which had not been previously detected in this population. Markers in these two loci could be useful to genetically improve intramuscular fatty acid composition.

Literature Cited

Estany, J., Ros-Freixedes, R., Tor, M. et al. (2014). PLoS One, 9:e86177.

Fernando, R. L., and Garrick, D. J. (2009). Animal Breeding and Genetics, Iowa State University, Ames, IA. Available at: http://bigs.ansci.iastate.edu/bigsgui

Óvilo, C., Fernández, A., Noguera, J. L. et al. (2005). Genet. Res. 85:57-67.

Muñoz, M., Rodríguez, M. C., Alves, E. et al. (2013). BMC Genomics 14:845.

Purcell, S., Neale, B., Todd-Brown, K. et al. (2007). Am. J. Hum. Genet. 81:559-575.

Ramayo-Caldas, Y., Mercadé, A., Castelló, A. et al. (2012). J. Anim.Sci. 90:2883-2893.

Ros-Freixedes, R., and Estany, J. (2013). J. Agr. Biol. Envir. St. Doi: 10.1007/s13253-013-0162-x

Ros-Freixedes, R., Reixach, J., Tor, M. et al. (2012). J. Anim. Sci. 90:4230-4238.

Uemoto, Y., Nakano, H., Kikuchi, T., et al. (2012a). Anim. Genet. 43:225-228.

Uemoto, Y., Kikuchi, T., Nakano, H. et al. (2012b). Anim. Sci. J. 83:375-385.

Uemoto, Y., Soma, Y., Sato, S. et al. (2012c). Anim. Genet. 43:27-34.

Yang, B., Zhang, W., Zhang, Z. et al. (2013). PLoS One 8:e65554.