Published July 17, 2015

Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations[1]

**E.-S. Kim,* R. Ros-Freixedes,† R. N. Pena,†
T. J. Baas,* J. Estany,† and M. F. Rothschild*[2]**

*Department of Animal Science, Iowa State University, Ames; †Departament de Producció Animal, Universitat de Lleida–Agrotecnio Center, Lleida, Catalonia, Spain

ABSTRACT: Intramuscular fat (IMF) content is an important trait affecting the quality of pork. Two Duroc populations, one under positive selection for IMF and the other selected for decreased backfat but under stabilizing selection for IMF, were used to identify signatures of selection associated with IMF using 60,000 single-nucleotide polymorphism data. The effects of selection were analyzed between 2 lines or groups representing selected and control animals within each population using a discriminant analysis of principal components and Wright's fixation index (F_{ST}). Moreover, extended haplotype homozygosity-based approaches were used to examine the changes in haplotype frequency due to recent selection. Each statistical method identified 10–20 selection signatures. A few haplotype-based signatures of selection agreed with results from a genome-wide association study (GWAS), while F_{ST} measures showed a better agreement with GWAS results. Agreement of marker-trait associations and signatures of selection was limited, and further examination will be necessary to understand the effect of selection on IMF and why some regions identified by GWAS did not appear to respond to the selection practiced. The genes in 21 consensus selection signatures were examined. Several genes with an effect on overall fatness were identified, but further research is needed to assess whether or not some of them could have a specific effect on IMF.

Key words: Duroc, genes, genome-wide associations, intramuscular fat, selection signatures

J. Anim. Sci. 2015.93:3292–3302
doi:10.2527/jas2015-8879

INTRODUCTION

Intramuscular fat content (**IMF**) affects both the organoleptic quality and nutritional value of pork. There is an increasing interest in including this trait in selection schemes because of its influential role in determining the preference of meat (Fernandez et al., 1999a,b). Selection experiments for high levels of IMF have been performed in Duroc pigs (Suzuki et al., 2005b; Schwab et al., 2009). Also, several genome-wide association studies (**GWAS**) have been performed for IMF (Quintanilla et al., 2012; Rohrer et al., 2012; Nonneman et al., 2013). However, GWAS can generate some false positive associations, although sophisticated statistical tests have been proposed to reduce false positives. To tackle this problem, a complementary approach has been suggested: combining identification of signatures of selection with GWAS (Schwarzenbacher et al., 2012) to reveal genomic regions associated with a trait that has recently undergone selection.

The Wright's fixation index (F_{ST}; Wright, 1951) is calculated as a measure of population differentiation between 2 genetically divergent groups. Moreover, variations of the extended haplotype homozygosity (**EHH**) can be detected in regions associated with variation influencing fitness (Sabeti et al., 2002; Voight et al., 2006; Tang et al., 2007). Signatures of

[1]This research was supported in the United States by the College of Agriculture and Life Sciences, the Ensminger program, State of Iowa, and Hatch funding and in Spain by the MINECO-funded projects AGL2009-09779 and AGL2012-33529. Comments made by the editor and 2 reviewers are greatly appreciated.
[2]Corresponding author: mfrothsc@iastate.edu
Received January 5, 2015.
Accepted April 13, 2015.

selection between diversified pig breeds have been identified (Amaral et al., 2011; Rubin et al., 2012; Wilkinson et al., 2013), whereas no reports assessing signatures of selection in experimental livestock populations selected for a specific trait have been described. In this study, we used 2 independent Duroc populations, one selected for increased IMF and the other selected for decreased backfat thickness (**BT**), but under stabilizing selection for IMF, to identify genomic regions and signatures of selection for IMF with the aim to identify candidate regions and genes underlying genetic improvement for IMF. How the different statistical methods compared in finding the same genomic regions was also assessed.

MATERAL AND METHODS

Populations

Population A consisted of a total of 144 Duroc pigs sampled from the sixth generation of a selection experiment for IMF at Iowa State University (Schwab et al., 2009). In this experiment, one line was selected for increased IMF and a control line was randomly bred. Half of the animals ($n = 73$) were obtained from the selected line (referred to as High IMF line; with a mean 4.46% IMF in the loin, SD 1.80%), while the remaining 71 animals were randomly sampled from the control line that maintained average levels of IMF (referred to as Low IMF line; with mean 2.71%, SD 0.98%). Population B consisted of 138 Duroc barrows from a Spanish Duroc line (Ros-Freixedes et al., 2013). Animals from this population were sampled to represent 2 time periods. The first half of the sampled animals ($n = 66$) were born in 2002 and were used as a control group. Because IMF content in 2002 was considered near the optimum in Population B (3.58%, SD 1.21%), selection after 2002 was then aimed at maintaining IMF while decreasing BT (Ros-Freixedes et al., 2013). The other half of the sampled animals ($n = 72$), considered the selected group, were born in 2009 and had 1.9 mm less BT and similar IMF (−0.20%) than those born in 2002. Animals were chosen to be as unrelated as possible and representative of the whole population. Because selection for increased IMF in Population A also led to an unfavorable correlated response in BT (Schwab et al., 2009), Population B was used to compare those candidate regions also affecting BT.

SNP Genotypes

All pigs were genotyped using the PorcineSNP60 v2 Genotyping BeadChip (Illumina, CA). Data from both populations were analyzed using the same procedures. The PLINK software (Purcell et al., 2007) was used to filter out SNP with minor allele frequency (**MAF**) below 0.01 and genotyping rate below 0.90, and individuals with more than 10% missing genotypes. Unmapped SNP based on the current pig genome assembly *Sus scrofa* (**SSC**) build 10.2 were also excluded. Two additional samples were removed due to the high likelihood they were mislabeled as to which group they originated from. Remaining data comprised 41,012 SNP for 130 individuals in Population A and 135 individuals in Population B. Posteriorly, the Beagle software (Browning and Browning, 2007) was used to phase and impute the missing genotypes using all data combined (10 iterations) for the further analysis of signatures of selection.

Population Structure

A discriminant analysis of principal components (**DAPC**) was performed with SNP genotypes. In DAPC, the discriminant functions are linear combinations of variables (principal components) that optimize the separation of individuals using predefined groups (Jombart et al., 2010). For an analysis of admixture, membership probabilities of each individual for the 4 different groups were obtained based on the retained discriminant functions. Pairwise Wright's F_{ST} between lines within a population was computed for individual loci and then summarized with 20-SNP windows. The DAPC and F_{ST} analyses were performed using the *adegenet* R package (Jombart and Ahmed, 2011). For further examination of population structure, admixture (Alexander et al., 2009) was used under the assumption of no prior information of subpopulations (K). Ten runs were performed with different random seed numbers for each value of $1 \leq K \leq 12$, and cross validation errors were recorded to examine the proportion of admixture. Allosomes were excluded in these analyses.

Extended Haplotype Homozygosity-based Signatures of Selection

The *rehh* R package (Gautier and Vitalis, 2012) was used to compute the integrated haplotype score (**iHS**; Voight et al., 2006) of each of the lines/groups in Populations A and B. The iHS reveals time-independent signatures of selection in a population, which comprises signatures of selection for IMF and evidence of any other selection in each Duroc population. To compare the change in EHH of the selected line/group with respect to the control line/group, a standardized score of the ratio of EHH (Rsb) was calculated (Tang et al., 2007). A positive value is indicative of a higher level of EHH in the selected line/group

compared to the control line/group, whereas a negative value represents decreased homozygosity due to selection. Both analyses were performed with the default parameters of the *rehh* package. Allosomes were excluded in these analyses.

Genome-wide Association Study

In Population A, GWAS was performed to detect the additive genetic effect of SNP across the genome. The generalized linear mixed model used was $y = \mu + s + \beta G + u + e$, where y is the log-transformed IMF of an individual, μ is the mean, s is the sex, β is a vector of additive genetic effects, G is an indicator variable for the additive genetic effects of an individual, u is the polygenic effect, and e is the vector of individual error terms. The random effect u was assumed to be distributed as $u \sim N(0, A\sigma_u^2)$, where A is the individual genomic relationship between pigs based on the whole genomic similarity and σ_u^2 the additive genetic variance. The R package *rrBLUP* (Endelman, 2011) was used for this analysis, and thresholds were decided by the analytical method proposed by Lander and Kruglyak (1995). Given the sample size of this experiment, some readers may consider all these regions as only suggestive.

Comparison between Statistics

The Pearson's correlation coefficients among the scores in each locus for the different statistics were calculated. Candidate regions for each statistic were summarized and compared to detect consensus candidate regions. Candidate regions for F_{ST} were defined as those regions including at least 1 SNP with a signal in the top 1% ($F_{ST} > 0.3$) and at least 10 other SNP in the top 5% ($F_{ST} > 0.2$) within 5 Mb. Candidate regions for iHS and Rsb were defined as those including at least 1 SNP with a signal of iHS > 3 or |Rsb| > 3 and 10 other SNP with iHS > 2 or |Rsb| > 2 within 3 Mb. Finally, candidate regions for GWAS were defined as those including at least 1 SNP with an association of $-\log_{10} p > 3$ and 5 other SNP with $-\log_{10} p > 2$ within 3 Mb. Overlapping regions were merged into one. Genes in some consensus candidate regions were retrieved from Ensembl (EMBL-EBI) using *Sus scrofa* genome assembly 10.2, and their function was checked with Enrichr (Chen et al., 2013).

RESULTS

Population Structure

Populations A and B were clearly separated by the first principal component in the DAPC analysis (Fig. 1A). Using the second principal component, the selected line in Population A was completely separated from the control line, while the clusters for 2002-born and 2009-born groups of population B completely overlapped. Principal components 1 and 2 accounted for 93.3% and 6.0% of total variation, respectively, and no additional principal component separated groups in Population B. The analysis of the population structure of the 4 subgroups confirmed that there was no genetic admixture between Populations A and B and that the 2 lines in Population A had differentiated (Fig. 1B). The results for Population B reflected the common ancestry of the 2 groups, which represented 2 time points of the same genetic line. The best number of subpopulations appeared to be 4 (the 2 lines in Population A and 2 samples in Population B) based on a combination of the value of cross-validation error and known subpopulations (Figs. S1, S2). When comparing animals based on the whole SNP data, the selected and control lines were potentially different in Population A (mean $F_{ST} = 0.062$), whereas the 2009 group did not substantially differ from the 2002 group in Population B (mean $F_{ST} = 0.010$).

Signatures of Selection in Population A

In Population A (Fig. 2A), several candidate regions were found using the top 1% and 5% F_{ST} thresholds of 0.368 and 0.232, respectively, and these regions are summarized in Table 1. In particular, multiple signals with peak $F_{ST} > 0.5$ were found on 10 regions. A total of 134 SNP surpassed the |iHS| > 3 threshold in the High IMF line (Fig. 2B) and 110 SNP in the Low IMF line (Fig. 2C). Seven and 8 regions appeared to have been selected in each line, respectively (Table 1). The iHS between the High and Low IMF lines were almost uncorrelated (0.07; Table 2), indicating distinctive signatures of selection in the 2 lines despite similar starting points. Regarding Rsb, 226 loci were identified in Population A (Fig. 2D). In summary, 16 regions responded to selection for high IMF as seen by the increased haplotype homozygosity (positive Rsb of the High with respect to the Low IMF line), whereas levels of Rsb were negative at 4 regions (Table 1).

Signatures of Selection in Population B Compared to Population A

The maximum value of F_{ST} was 0.197 (Fig. 3A), which supported the weak evidence of differential selection between 2002 and 2009 groups in Population B, and less than 100 SNP were bound to selection detected by iHS (> 3) in each group (Fig. 3B, 3C). In contrast to Population A, both groups in Population B shared almost the same pattern of iHS across the

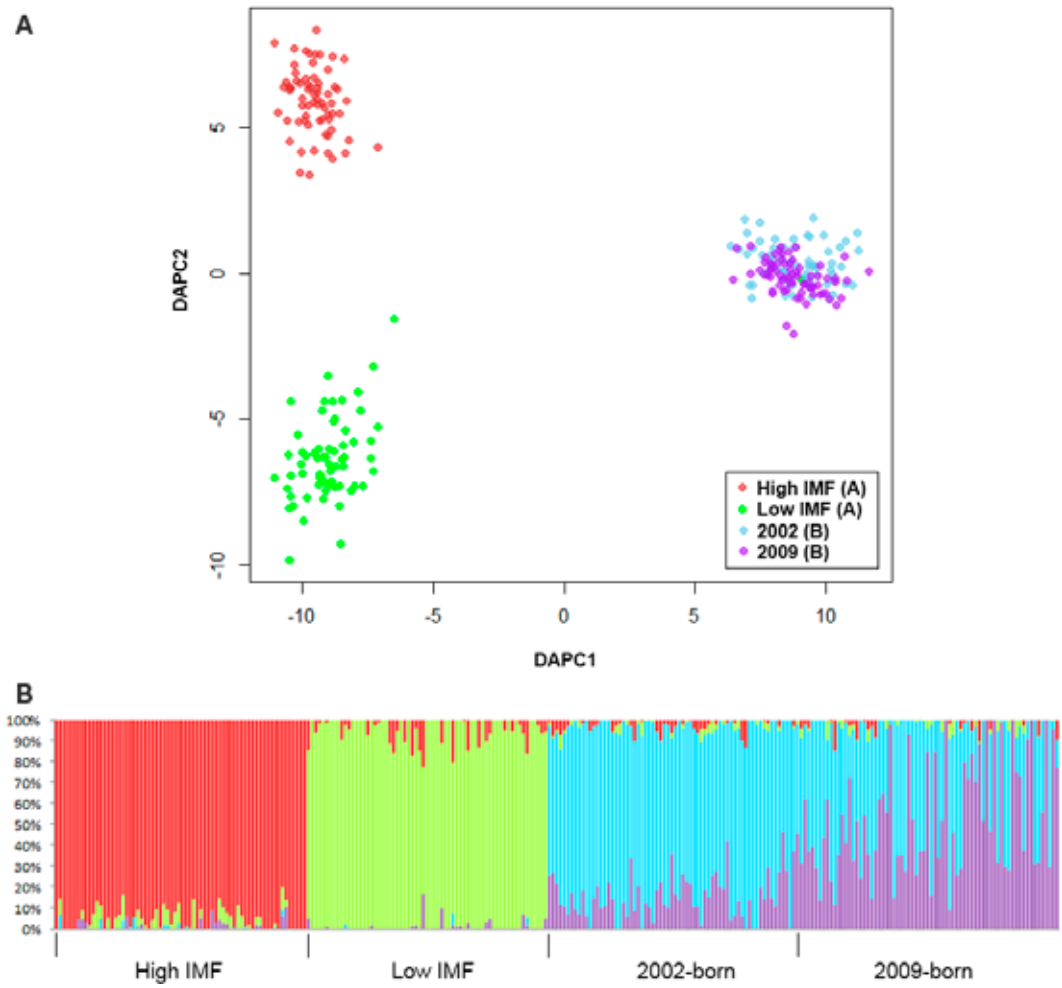

Figure 1. Analysis of population structure in the 2 Duroc populations by means of a discriminant analysis of principal components (A) and admixture analysis (B), K = 4. The probability of an individual to each group or line is indicated by different colors.

genome (r = 0.50; Table 2), reflecting common signatures of selection in both groups rather than differentiation between them. Using Rsb, several candidate regions were revealed (|Rsb| > 3) across the genome in Population B (Fig. 3D). It can be noted that the same number of positive and negative Rsb were detected (Table 3), which suggests no strong directional selection for a single trait in Population B during the 2002–2009 period. The iHS in both groups of Population B was moderately correlated (r = 0.17–0.19) to that of the Low IMF line in Population A (Table 2). Consistently, we identified some common signatures between both populations A and B for these statistics, such as the region between 210 and 230 Mb on SSC 1 (Tables 1 and 3). Thus, we focused on the analysis of Population A to search for regions involved in the recent selection for IMF.

GWAS and Consensus Signatures of Selection

Since strong signals of selection signatures were identified in Population A, GWAS was also performed for a better understanding of the effect of selection (Fig. 2E, S3). Fifteen regions were considered candidate regions explaining variation in IMF (Table 1). Of these, 10 regions located on SSC 1, 4, 6, 7, 9, 11, 13, and 15 comprised a considerable number of signals and were supported by signatures of selection, in particular, F_{ST} (Table 1). In fact, there was a moderate correlation between the GWAS associations and the F_{ST} values (0.33; Table 2). On the other hand, some regions associated with IMF were supported by haplotype-based selection signatures with moderate significance levels, but many candidate regions strongly supported by iHS or Rsb were not directly found to be associated with IMF as measured by GWAS (Table 1).

We studied the genes in the candidate regions supported by at least F_{ST} or GWAS and any additional methods in Population A. This included a total of 16 candidate regions (Table 1, in bold). Genes in 6 regions found in Population B for Rsb that coincided with signatures of selection in Population A were also examined (Table 3, in bold). Common candidate regions in both populations could be less likely to be

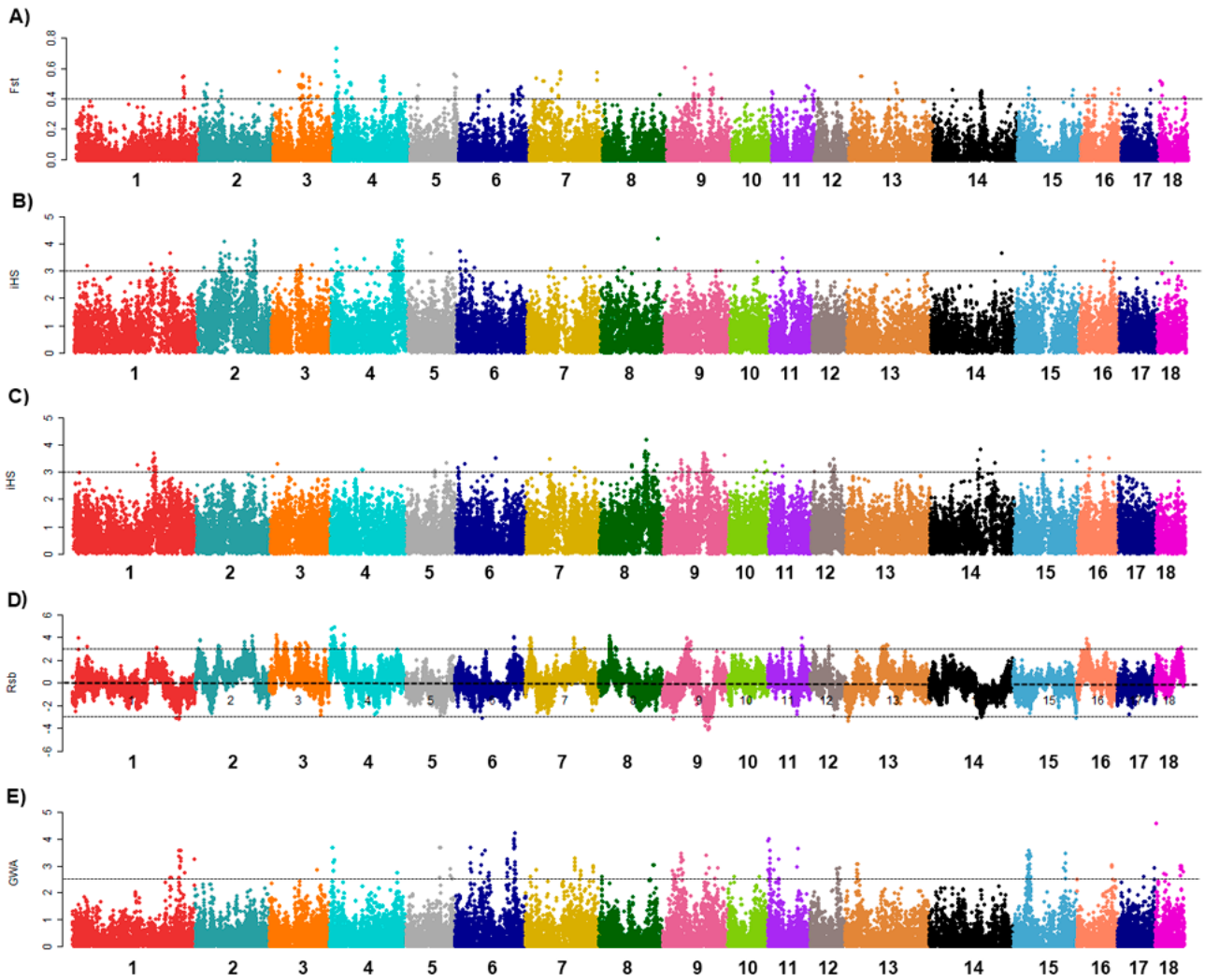

Figure 2. Genome-wide signatures of selection and associations in Population A. (A) F_{ST}; (B) iHS in High IMF line; (C) iHS in Low IMF line; (D) Rsb (High/Low IMF line); (E) GWAS ($-\log_{10}$p). For F_{ST}, iHS, Rsb, and GWAS, threshold values of 0.4, 3.0, 3.0, and 2.5, respectively, are plotted.

associated to IMF than other signals. On the other hand, they may reflect the positive genetic correlation between IMF and BT (Suzuki et al., 2005a; Solanes et al., 2009). Significant signatures of selection for BT in Population B were only supported by Rsb. A total of 1,118 genes were found in the examined candidate regions, from which 148 genes were found to have diverse functions in adipogenesis and lipid metabolism, as well as in muscle development or integration of energy metabolism, as summarized in Table 4.

DISCUSSION

By identifying SNP that have highly divergent F_{ST} values, several regions associated with additive effects on traits like residual beef yield, feed intake, or intramuscular fatness were found in Australian beef cattle (Barendse et al., 2009). In our study, a considerable correlation between F_{ST} and GWAS signals was also found. This could be expected because both methods depend on the differences in allele frequency of SNP and when 2 selection lines that have been divergently selected for a specific trait are compared, similar results would be expected for these 2 analyses. It has been noted that outliers in F_{ST} analyses often reflect genetic drift as well as selection, but those are hard to distinguish. Less than half of the regions in our study with high F_{ST} (> 0.4) appeared to match those most associated with additive values of IMF. In addition, a few GWAS-detected regions were independent of the regions with high levels of F_{ST}. These discrepancies led us to further investigate selection signatures to interpret the results from GWAS and F_{ST}.

In livestock, selection of superior animals for a particular phenotype will increase the frequency of haplotypes harboring the preferred alleles under selection (Kim et al., 2013). The frequency of a selected allele will be eventually be fixed if selection is practiced effectively. Compared to relatively long-term selection (e.g., 20–30 generations), the recent selection

Table 1. Selection signatures in Population A[1]

		Maximum (position, Mb)				
			iHS			
SSC	Region (Mb)[2]	F_{ST}	High IMF	Low IMF	Rsb[3]	GWAS[4]
1	202.5–229.5		3.25 (204.3)	3.67 (215.7)		
	256.7–261.4		3.65 (257.6)			
	281.2–285.5	**0.55 (283.5)**			**-3.18 (281.9)**	**3.57 (284.3)**
2	8.6–13.0	0.49 (12.9)				
	41.9–48.5	**0.45 (42.0)**	**3.66 (43.1)**		**3.28 (43.1)**	
	77.4–96.3		3.60 (81.8)			
	117.4–119.1				3.55 (118.4)	
	130.0–136.0				4.07 (133.1)	
3	14.7–18.7				4.15 (15.6)	
	27.9–34.9				3.39 (30.9)	
	61.2–66.8	**0.49 (65.9)**			**3.11 (62.0)**	
	68.5–72.2	**0.56 (71.5)**			**3.44 (71.7)**	
	88.2–91.4	0.54 (91.1)				
4	**2.6–12.8**	**0.73 (5.0)**			**4.67 (3.3)**	**3.68 (5.0)**
	19.5–22.9				4.20 (21.6)	
	92.4–101.8	0.55 (100.9)				
	119.1–135.0	**0.43 (128.4)**	**4.11 (128.6)**			
5	81.6–83.0					3.66 (81.8)
	102.8–105.7	0.55 (105.4)				
6	30.2–30.9					3.66 (30.3)
	40.1–43.1	0.42 (41.2)				
	62.6–65.4					3.42 (64.3)
	112.7–119.9					**3.25 (119.4)**
	134.1–138.3				**4.04 (134.2)**	**4.22 (136.2)**
7	8.1–12.0				3.97 (9.5)	
	51.9–53.7	0.57 (52.9)				
	89.3–92.7				**3.93 (90.0)**	**3.27 (91.5)**
	121.9–123.6	0.57 (122.6)				
8	19.3–22.6				3.86 (19.9)	
	110.4–118.2			4.17 (114.2)		
9	**29.9–38.1**	**0.61 (32.1)**		**3.42 (30.0)**		**3.47 (32.0)**
	39.1–53.3			3.18 (45.8)		
	92.2–123.6	**0.47 (117.0)**		**3.50 (99.7)**	**-4.12 (116.6)**	**3.38 (108.4)**
11	**0.3–4.4**	**0.45 (1.0)**				**4.01 (1.1)**
	21.1–22.8					3.25 (21.3)
	25.1–26.1		3.47 (25.3)			
	73.9–76.2				3.89 (74.9)	
	84.7–87.3	0.45 (87.3)				
12	7.7–11.1	0.40 (7.8)				
	42.6–44.5			3.47 (42.9)		
13	5.7–9.8				−3.42 (5.7)	
	17.9–21.5	**0.54 (19.9)**				**3.06 (19.9)**
	69.1–83.6				3.35 (83.3)	
14	**80.9–92.3**	**0.44 (81.9)**		**3.44 (81.2)**		
	92.9–97.9				−3.07 (93.3)	
15	**28.4–35.6**	**0.36 (31.6)**				**3.55 (31.4)**
	61.9–64.2			3.74 (62.6)		
	132.9–135.3					3.45 (134.5)
16	**20.9–30.5**	**0.46 (27.7)**			**3.85 (21.5)**	
	75.2–78.1		3.3 (77.3)			

[1]Bold indicates consensus regions examined for genes.
[2]Overlapping regions were merged into one.
[3]Rsb of High IMF line respect to Low IMF line.
[4]Significance level of associations ($-\log_{10} p$).

Table 2. Correlations among signatures of selection[1]

Statistics by population and method			Population A				Population B		
		F_{ST}	iHS High IMF	iHS Low IMF	Rsb[a]	GWAS[b]	F_{ST}	iHS 2009	iHS 2002
Population A	iHS High IMF	0.03							
	iHS Low IMF	0.02	0.07						
	Rsb[a]	0.01	0.12	−0.05					
	GWAS[b]	**0.33**	−0.01	0.03	0.01				
Population B	F_{ST}	0.00	0.00	−0.02	0.03	0.01			
	iHS 2009	0.01	0.02	**0.19**	−0.03	−0.02	−0.02		
	iHS 2002	0.02	0.03	**0.17**	−0.01	−0.02	0.05	**0.50**	
	Rsb[a]	0.04	0.03	0.02	−0.01	−0.03	0.06	−0.08	0.07

[1]Bold indicates correlations greater than 0.15.

[a]Rsb of High IMF line with respect to Low IMF line for Population A and of the 2009 group with respect to the 2002 group for Population B.

[b]Significance level of associations in Population A ($-\log_{10} p$).

of a haplotype may not substantially affect the change of the frequency of a selected allele in only a few generations. Despite the low correlations of F_{ST} with iHS and Rsb, several identified regions were supported by 2 or more of these approaches. However, inconsistencies with the results of GWAS produced only a limited number of candidate regions. The size of population in our study may be insufficient to detect small effects of marker-trait associations in GWAS, and a distinctive feature of EHH may produce markedly different results. In principle, single marker-trait associations are not necessarily consistent with selection responses, and unselected SNP could be considered a target of future selection if the association can be confirmed.

Variation of IMF and other fat-related traits has been explained by the overall effect of numerous loci in pigs (Hernández-Sánchez et al., 2013). Standard methods for detecting selective sweeps would have little power in the case of polygenic traits, even with strong selection for a trait because the response to selection

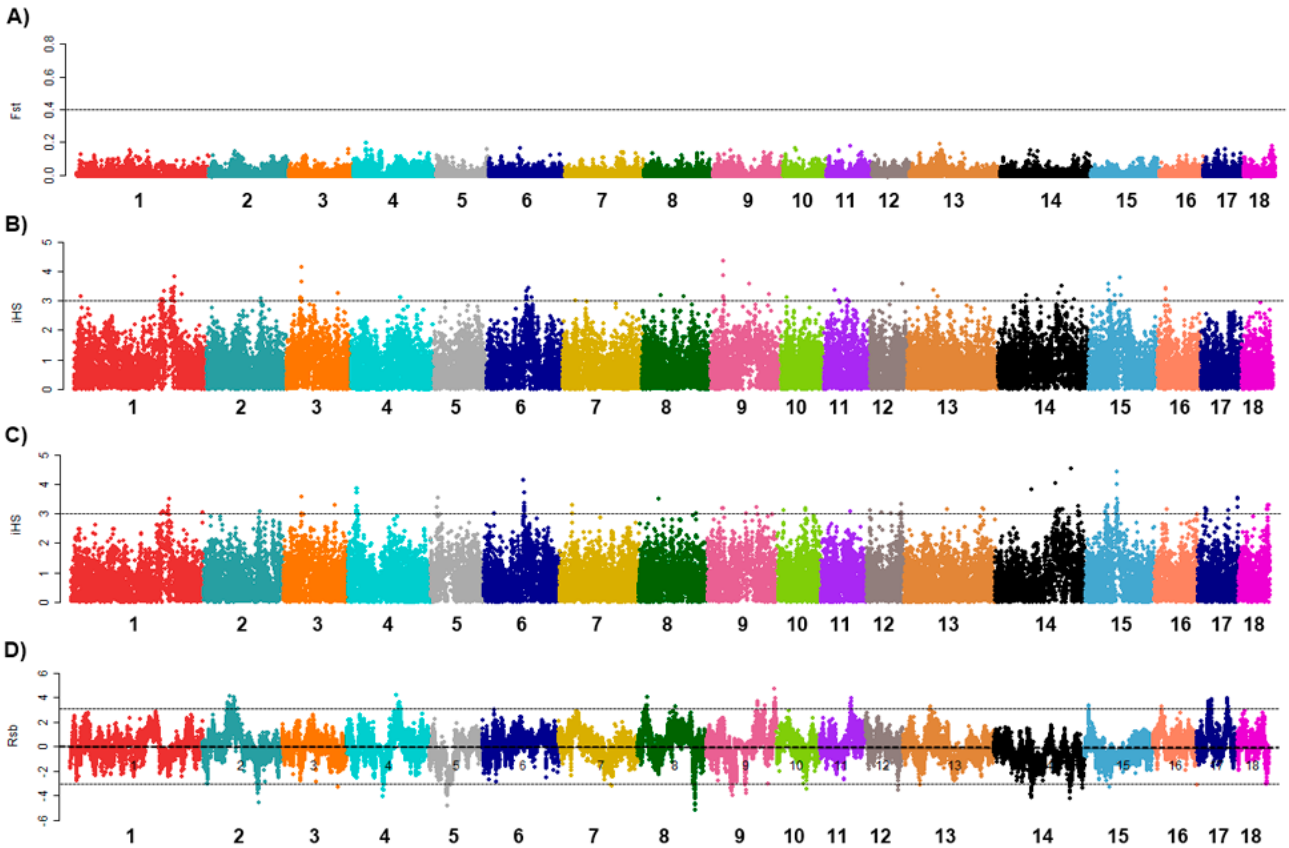

Figure 3. Genome-wide signatures of selection in Population B. (A) F_{ST}; (B) iHS (2009-born); (C) iHS (2002-born); (D) Rsb (2009/2002). For F_{ST}, iHS, and Rsb, threshold values of 0.4, 3.0, and 3.0, respectively, are plotted.

Table 3. Selection signatures (iHS and Rsb) in Population B

		Maximum (position, Mb)			
		iHS			
SSC	Region	2009	2002	Rsb[a]	Selection signatures in Population A[b]
1	213.1–227.5	2.63 (215.7)	3.03 (215.7)		iHS(High IMF, Low IMF)
	241.9–250.6	3.48 (247.5)	3.81 (249.5)		
2	43.3–69.1			4.06 (46.2)	F_{ST}, iHS(High IMF), Rsb(+)
	121.1–123.4			−4.57 (122.6)	
3	25.3–27.9		4.12 (26.8)		
4	11.8–14.7	3.85 (12.6)			F_{ST}, Rsb(+), GWAS
	63.9–68.5			−4.04 (63.9)	
	93.4–99.6			4.20 (94.9)	F_{ST}
5	32.3–54.6			−4.79 (34.5)	
6	80.7–86.5	4.15 (82.0)	3.33 (82.0)		
7	89.9–92.5			−3.17 (91.4)	Rsb(+), GWAS
8	13.7–16.6			4.04 (164.5)	
	127.5–131.7			−4.76 (130.9)	
9	19.9–20.8		4.35 (20.3)		
	44.6–47.3			−3.94 (45.8)	iHS(Low IMF)
	117.9–122.3			3.70 (120.6)	F_{ST}, iHS(Low IMF), Rsb(−), GWAS
	133.9–135.7			3.33 (135.1)	
	149.5–153.2			4.67 (150.2)	
11	65.8–67.5			3.92 (66.6)	
14	59.5–60.7			−4.14 (60.3)	
15	36.2–41.6	3.29 (40.0)	3.57 (37.4)		
	61.9–65.9	4.43 (62.6)	3.78 (62.6)		iHS(Low IMF)
17	20.9–28.3			3.81 (27.3)	
	52.2–55.2			3.95 (52.8)	
18	53.8–57.6	3.3 (55.9)		−3.06 (55.5)	

[a]Rsb of the 2009 group with respect to the 2002 group.

[b]Symbols + (High homozygosity in High IMF line) and – (High homozygosity in Low IMF line) indicate the sign of Rsb in Population A.

would be generated by modest allele frequency shifts at many loci that are already polymorphic (Pritchard et al., 2010). It has been demonstrated that the effect of polygenic adaptation using a simplified model with a trait and an alternative approach will be applicable using birth years of animals (Decker et al., 2012).

Using DAPC, 2 clusters were found that matched the 2 lines in Population A, which represented the consequences of selection for IMF in the whole genome. Thus, we focused most of our analysis on Population A because of a weak differentiation of the groups in Population B. All methods measuring diversity and selection signatures showed that there was no substantial evidence of strong recent selection in Population B. Genetic progress in Population B could have been limited by the need to restrain the genetic response of IMF during selection for reduced BT to offset the positive genetic correlation between these 2 traits (Ros-Freixedes et al., 2013), as well as the simultaneous inclusion of some other traits in the selection objectives of the line. In most selection programs, a combination of traits is generally considered in livestock breeding (Van Vleck et al., 1986), but the impact of multitrait selection seems to be limited with respect to changing polymorphic genotypes in the whole genome. Conversely, the most recent selection program has maintained the desired phenotype selected for and fixed before 2002 in Population B resulting in the same levels of allele or haplotype frequency for the past 7 yr.

Results from iHS represent the selection signatures observed. We obtained similar selection signatures in both groups of Population B from the different time points, which may reflect selection performed before 2002 as well as the recent selection for reduced BT and some other traits from after 2002. Moreover, some iHS results in Population B suggest common signatures of selection in Population A. For example, the region from 210 to 230 Mb on SSC 1 is a candidate region in Population B (2002-born group), which overlapped with loci with high values of iHS in both lines of Population A. Nonetheless, iHS results in Population B showed a low positive correlation with selection signatures in the Low IMF line of Population A.

Table 4. Functional candidate genes in the consensus selection signatures as retrieved from databases[a] integrated in the Enrichr gene analysis tool

Population	SSC	Region	Genes and functions[b]
A	1	202.5–229.5	*CER2*[4], *ADFP*[2], *BMP4*[8], *DDHD1*[6], *ERO1L*[2], *HIF1A*[2], *HMGA2*[2,3], *PLAA*[6], *PTPLAD2*[4]*, *SGPP1*[4], *SIX1*[8], *SIX4*[8], *TEK*[6], *TMEM30B*[7]
A	1	281.2–285.5	*GNG10*[9], *LPAR1*[3,6], *PTGR1*[6]*, *UGCG*[4,6]
A,B	2	43.3–48.5	*ABCC8*[9], *CSRP3*[8], *CYP2R1*[4], *KCNJ11*[9], *MYOD1*[8], *PIK3C2A*[4,6], *TPH1*[3]
B	2	48.5–69.1	*ADM*[4,8], *AMPD3*[9], *ARF1*[4], *ARNTL*[2,3,8], *CACNA1A*[9], *CALR*[8], *CRTC1*[3], *DKK3*[4], *FAR1*[1,6]*, *GNG12*[9], *JAK3*[1], *LPAR2*[6], *NDUFA13*[9], *NDUFB7*[9], *PRKACA*[6,9]*, *SIN3B*[8], *SLC27A1*[4,6,7]*, *TECR*[4,6]*, *UBA52*[9], *WNT3A*[2,8]
A	2	77.4–96.3	*ABCA7*[3,7], *ACOT12*[6]*, *ARSB*[3,6], *ARSK*[6], *ATP5D*[9], *CMYA5*[6,8], *COL4A3BP*[4,5,7], *F2R*[6], *FGFR4*[4,6], *FLT4*[3], *GAMT*[3], *GFPT2*[9], *GPX4*[6]*, *HEXB*[3,6], *HMGCR*[4,6,8], *HOMER1*[8], *LTC4S*[4,6]*, *MAML1*[8], *MAPK9*[1,4,6,7]*, *NDUFS7*[9], *PDE8B*[4,6], *PPAP2C*[4,6]*, *PRELID1*[7], *PTBP1*[8], *STK11*[1,4,6,9]*, *TCF3*[8]
A	3	61.2–66.8	*ST3GAL5*[4,6], *SUCLG1*[9]
A	3	68.5–72.2	*DOK1*[3]
A	4	2.6–12.8	*MYC*[3,9], *OC90*[6]
A	4	92.4–93.4	*ALDH9A1*[6]*, *RXRG*[1,2,8]
A,B	4	93.4–99.6	*APOA2*[4,5,6,7], *CASQ1*[8], *CRP*[6], *DDR2*[3], *FCER1A*[4,6]*, *HSD17B7*[4,6], *IGSF8*[8], *NDUFS2*[9], *PIGM*[6], *RGS4*[3], *RGS5*[3], *SDHC*[9], *USF1*[5]
A	4	99.6–101.8	*NTRK1*[6]
A	4	119.1–135	*ABCA4*[7], *ABCD3*[6]*, *AGL*[9], *CD53*[8], *CEPT1*[4], *CSF1*[2], *SLC30A7*[3], *SORT1*[6,8], *VAV3*[6]
A	6	112.7–119.9	*PIK3C3*[4,6], *PRKACB*[6,8]*
A	6	134.1–138.3	*ANGPTL3*[5,6]*, *DOCK7*[3], *JAK1*[1], *LEPR*[1,3,5,9], *LEPROT*[1]
A,B	7	89.9–92.5	*CHD2*[3,8], *ST8SIA2*[4,6]
A	9	29.9–38.1	*MTMR2*[4,6]
B	9	44.6–47.3	*BCO2*[6], *DLAT*[6,8]*, *DRD2*[3,6,7]*, *IL18*[3], *ZBTB16*[2]
A	9	92.2–117.9	*CD36*[1,7]*, *CROT*[6,7]*, *HDAC9*[8], *ITGB8*[6], *MEOX2*[8], *PIK3CG*[1,4,6]*, *SEMA3C*[8]
A,B	9	117.9–122.3	*DLD*[6,9]*, *EZH2*[8]
A	11	0.3–4.4	*FGF9*[8], *GTF3A*[2]
A	13	17.9–21.5	*GPD1L*[4,6]*, *OSBPL10*[7], *TGFBR2*[3,8]
A	14	80.9–92.3	*PLA2G12B*[5,6]*, *PLAU*[3], *SAMD8*[4,6]
B	14	131.7–135.2	*ADRA2A*[6,9], *TCF7L2*[2,4,6]*
A	15	28.4–35.6	*BIN1*[8], *DBI*[6]*, *TFCP2L1*[4], *TSN*[3]
A	16	20.9–30.5	*FGF10*[2,3,8], *GHR*[6]*, *HMGCS1*[4,6], *LIFR*[2], *NIPBL*[2,3], *OXCT1*[2,6], *PLCXD3*[6], *PRLR*[3,4], *RICTOR*[3]

[a]The Gene Ontology project, MGI Mammalian Phenotype Ontology, KEGG Pathway Database, WikiPathways, and Reactome Pathway Database.

[b]1: Adipocytokine signaling pathway; 2: Adipogenesis; 3: Related to abnormal adipose tissue; 4: Lipid biosynthesis; 5: Lipid homeostasis; 6: Lipid metabolism and catabolism; 7: Lipid transport; 8: Muscle development; 9: Integration of energy metabolism; *: Related to fatty acids or triglycerides.

Using the analysis of Rsb, some regions seemed to respond to recent artificial selection in Population B. Rsb revealed considerable differences in haplotype homozygosity between the 2 groups, but the same amount of positive and negative signals was found, in contrast with the mostly positive Rsb signals found in Population A. The standardized score of Rsb depends on the distribution of the ratio of EHH between 2 groups in a population, suggesting that the values of Rsb or iHS may not be directly comparable to the results from another population or other studies. Thus, the results from Rsb may mislead or incorrectly estimate the strength of selection unless supporting methods and additional populations are available to help interpret them.

Previous studies have reported quantitative trait loci (**QTL**) for IMF in several breeds (Hu et al., 2013). Comparison with QTL detected in previous studies showed some agreement with 18 of our consensus selection signatures, but candidate regions on SSC 3, 9 (92.2–123.6 Mb), 13, 14, 15, and 16 were exclusively identified in our study. Because the High IMF line was selected for IMF without restrictions, BT also increased in this line (Schwab et al., 2009). The regions and genes detected here (Table 4) have affected overall fatness rather than only specifically IMF. Using Rsb we found a selection signature that was supported by the results of GWAS in SSC 6 at 134.1–138.3 Mb. This region contains the leptin receptor (*LEPR*) gene (at 135.4 Mb). Plenty of QTL related to fatness were previously found in this region. A nonsynonymous polymorphism in this gene has been associated with increased feed intake and, as a consequence, with overall fatness, affecting both BT and IMF (Óvilo et al., 2005; Galve et al., 2012; Uemoto et al., 2012). Thus, the selection signature found in this region could indicate that selection for IMF also modified the haplotype frequencies at loci with nonspecific associations to all adipose depots. Association between this gene and IMF also has been reported in Population B (Ros-Freixedes et al., 2014), but no Rsb selection signature was detected in this population, maybe because the restriction applied on IMF in its selection objective limited the changes on genes affecting overall fatness.

Accumulation of fat in the intramuscular depots takes place late in the growth period of the animal. In muscle, fat is stored in 2 cell types: intramuscular adipocytes (about 80%) and in lipid droplets in the myocyte cytoplasm (up to 20%). Intramuscular adipocytes are morphologically and functionally different from adipocytes of other fat depots, but exclusive biomarkers have not yet been found. It is thus not surprising that the candidate genes underlying the regions affected by selection have general functions in adipocyte differen-

tiation, fat transport and metabolism, or distribution of the energy balance. The shared functionality of the fat depots translates in a positive correlation between, for instance, BT and IMF. Indeed, in commercial selection schemes aimed at increasing IMF (without restricting BT), the correlated increase of BT is one of the main disadvantages. Among the functional candidate genes that we identified in the selection signatures for Population A, the expression of gene *HMGCR* (SSC 2 at 86.0 Mb) has already been positively associated with increased IMF and with other fatness and lipid traits (Cánovas et al., 2010b). The retinoid X receptor (**RXR**) proteins are involved in the adipocytokine signaling pathway and play a role in the regulation of preadipocyte differentiation, lipid metabolism, and fatty acid catabolism (Brandebourg and Hu, 2005). Some results showed that gene *RXRG* (SSC 4 at 93.0 Mb) was downregulated in pigs with high IMF (Cánovas et al., 2010a), and differential expression of this gene was observed in pigs fed diets using oleic acid or carbohydrates as the energy source (Óvilo et al., 2014). The gene *SLC27A1* (SSC 2, 59.8 Mb), detected in a selection signature of Population B, was found to have greater expression in muscles than in liver or subcutaneous fat (Gallardo et al., 2013). Whether changes of haplotype homozygosity at these or any other loci could affect IMF and BT differentially should be further assessed.

Conclusions

Genomic signatures of artificial selection for IMF in Duroc pigs were identified by examining the differences of allele frequency and haplotype homozygosity between selected and control lines/groups in 2 populations. Selection signatures were analyzed using 4 different methods and GWAS. Substantial changes of genetic background were identified in a line selected for high IMF. In contrast, a population selected for multiple traits while under stabilizing selection for IMF showed little evidence of genetic change for IMF using DAPC, F_{ST}, and EHH analyses. Identifying the regions involved in selection for IMF will be useful to find potential candidate genes underlying genetic improvement. Despite dozens of signals generated in all, 21 consensus signatures of selection were examined. Genes in these regions are likely to have a general effect on overall fatness and further research is needed to assess whether some of them affect IMF specifically. The results from our study provide some insight of the relationships between selection signatures and marker-trait associations. Agreement of marker-trait associations and signatures of selection was limited and further examination will be necessary to understand the effect of selection on this trait and why some regions identified by GWAS did not appear to have responded to the selection practiced. When a measured phenotype is not available, F_{ST} will be a relatively useful method to infer regions affecting a trait in populations that have undergone strong or divergent selection.

LITERATURE CITED

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Amaral, A. J., L. Ferretti, H. J. Megens, R. P. Crooijmans, H. Nie, and S. E. Ramos-Onsins. 2011. Genome-wide footprints of pig domestication an selection revealed through massive parallel sequencing of pool DNA. PLoS ONE 6:e14782.

Barendse, W., B. E. Harrison, R. J. Bunch, M. B. Thomas, and L. B. Turner. 2009. Genome wide signature of positive selection: The comparison of independent samples and the identification of regions associated to traits. BMC Genomics 10:178.

Brandebourg, T. D., and C. Y. Hu. 2005. Regulation of differentiating pig preadipocytes by retinoic acid. J. Anim. Sci. 83:98–107.

Browning, S. R., and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097.

Cánovas, A., R. Quintanilla, M. Amills, and R. N. Pena. 2010a. Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. BMC Genomics 11:372.

Cánovas, A., R. Quintanilla, D. Gallardo, I. Díaz, J. L. Noguera, O. Ramírez, and R. N. Pena. 2010b. Functional and association studies on the pig HMGCR gene, a cholesterol-synthesis limiting enzyme. Animal 4:224–233.

Chen, E. Y., C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, and A. Ma'ayan. 2013. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14:129.

Decker, J. E., D. A. Vasco, S. D. McKay, M. C. McClure, M. M. Rolf, J. Kim, S. L. Northcutt, S. Bauck, B. W. Woodward, R. D. Schnabel, and J. F. Taylor. 2012. A novel analytical method, Birth Date Selection Mapping, detects response of the Angus (*Bos taurus*) genome to selection on complex traits. BMC Genomics 13:606.

Endelman, J. B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Gen. 4:250–255.

Fernandez, X., G. Monin, A. Talmant, J. Mourot, and B. Lebret. 1999a. Influence of intramuscular fat content on the quality of pig meat: 1. Composition of the lipid fraction and sensory characteristics of m. longissimus lumborum. Meat Sci. 53:59–65.

Fernandez, X., G. Monin, A. Talmant, J. Mourot, and B. Lebret. 1999b. Influence of intramuscular fat content on the quality of pig meat: 2. Consumer acceptability of m. longissimus lumborum. Meat Sci. 53:67–72.

Gallardo, D., M. Amills, R. Quintanilla, and R. N. Pena. 2013. Mapping and tissue mRNA expression analysis of the pig solute carrier 27A (SLC27A) multigene family. Gene 515:220–223.

Galve, A., C. Burgos, L. Silió, L. Varona, C. Rodríguez, C. Ovilo, and P. López-Buesa. 2012. The effects of leptin receptor (LEPR) and melanocortin-4 receptor (MC4R) polymorphisms on fat content, fat distribution and fat composition in a Duroc × Landrace/Large White cross. Livest. Sci. 145:145–152.

Gautier, M., and R. Vitalis. 2012. rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28:1176–1177.

Hernández-Sánchez, J., M. Amills, R. N. Pena, A. Mercadé, A. Manunza, and R. Quintanilla. 2013. Genomic architecture of heritability and genetic correlations for intramuscular and back fat contents in Duroc pigs. J. Anim. Sci. 91:623–632.

Hu, Z. L., C. A. Park, X. L. Wu, and J. M. Reecy. 2013. Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. 41:D871–D879.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genet. 11:94.

Kim, E. S., J. B. Cole, H. Huson, G. R. Wiggans, C. P. Van Tassell, B. A. Crooker, G. Liu, Y. Da, and T. S. Sonstegard. 2013. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. PLoS ONE 8:e80813.

Lander, E., and L. Kruglyak. 1995. Genetic dissection of complex traits: Guideleines for interpreting and reporting linkage results. Nat. Genet. 11:241–247.

Nonneman, D. J., S. D. Shackelford, D. A. King, T. L. Wheeler, R. T. Wiedmann, W. M. Snelling, and G. Rohrer. 2013. A genome-wide association of meat quality traits and tenderness in swine. J. Anim. Sci. 91:4043–4050.

Óvilo, C., R. Benítez, A. Fernández, B. Isabel, Y. Núñez, A. I. Fernández, C. Rodríguez, A. Daza, L. Silió, and C. López-Bote. 2014. Dietary energy source largely affects tissue fatty acid composition but has minor influence on gene transcription in Iberian pigs. J. Anim. Sci. 92:939–954.

Óvilo, C., A. Fernández, J. L. Noguera, C. Barragán, R. Letón, C. Rodríguez, A. Mercadé, E. Alves, J. M. Folch, L. Varona, and M. Toro. 2005. Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross. Genet. Res. 85:57–67.

Pritchard, J. K., J. K. Pickrell, and G. Coop. 2010. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 20:208–215.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, D. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

Quintanilla, R., R. N. Pena, D. Gallardo, A. Cánovas, O. Ramírez, I. Díaz, J. L. Noguera, and M. Amills. 2012. Porcine intramuscular fat content and composition are regulated by quantitative trait loci with muscle-specific effects. J. Anim. Sci. 89:2963–2971.

Rohrer, G. A., D. J. Nonneman, R. K. Miller, H. Zerby, and S. J. Moeller. 2012. Association of single nucleotide polymorphism (SNP) markers in candidate genes and QTL regions with pork quality traits in commercial pigs. Meat Sci. 92:511–518.

Ros-Freixedes, R., S. Gol, R. N. Pena, M. Tor, J. C. M. Dekkers, and J. Estany. 2014. Genome-wide study for intramuscular fat content and composition in Duroc pigs. Proc. 10th World Cong. Genet. Appl. Livest. Prod., Vancouver, BC, Canada.

Ros-Freixedes, R., J. Reixach, L. Bosch, M. Tor, and J. Estany. 2013. Response to selection for decreased backfat thickness at restrained intramuscular fat content in Duroc pigs. J. Anim. Sci. 91:3514–3521.

Rubin, C. J., H. J. Megens, A. Martinez, K. Barrio, S. Maqbool, D. Sayyab, C. Schwochow, Ö. Wang, P. Carlborg, C. B. Jern, A. L. Jørgensen, M. Archibald, M. A. Fredholm, M. A. Groenen, and L. Andersson. 2012. Strong signatures of selection in the domestic pig genome. Proc. Natl. Acad. Sci. U. S. A. 109:19529–19536.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald, H. C. Ackerman, S. J. Campbell, D. Altshuler, R. Cooper, D. Kwiatkowski, R. Ward, and E. S. Lander. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837.

Schwab, C. R., T. J. Baas, K. J. Stalder, and D. Nettleton. 2009. Results from six generations of selection for intramuscular fat in Duroc swine using real-time ultrasound. I. Direct and correlated phenotypic responses to selection. J. Anim. Sci. 87:2774–2780.

Schwarzenbacher, H., M. Dolezal, K. Flisikowski, F. Seefried, C. Wurmser, C. Schlötterer, and R. Fries. 2012. Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle. BMC Genomics 13:48.

Solanes, F. X., J. Reixach, M. Tor, J. Tibau, and J. Estany. 2009. Genetic correlations and expected response for intramuscular fat content in a Duroc pig line. Livest. Sci. 123:63–69.

Suzuki, K., M. Irie, H. Kadowaki, T. Shibata, M. Kumagai, and A. Nishida. 2005a. Genetic parameter estimates of meat quality traits in Duroc pigs selected for average daily gain, longissimus muscle area, backfat thickness, and intramuscular fat content. J. Anim. Sci. 83:2058–2065.

Suzuki, K., H. Kadowaki, T. Shibata, H. Uchida, and A. Nishida. 2005b. Selection for daily gain, loin-eye area, backfat thickness and intramuscular fat based on desired gains over seven generations of Duroc pigs. Livest. Prod. Sci. 97:193–202.

Tang, K., K. R. Thornton, and M. A. Stoneking. 2007. New approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 5:e171.

Uemoto, Y., T. Kikuchi, H. Nakano, S. Sato, T. Shibata, H. Kadowaki, K. Katoh, E. Kobayashi, and K. Suzuki. 2012. Effects of porcine leptin receptor gene polymorphisms on backfat thickness, fat area ratios by image analysis, and serum leptin concentrations in a Duroc purebred population. Anim. Sci. J. 83:375–385.

Van Vleck, L. D., E. J. Pollak, and E. A. B. Oltenacu. 1986. Genetics for the animal sciences. W.H. Freeman and Co., New York. p. 391.

Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:e72.

Wilkinson, S., Z. H. Lu, H. J. Megens, A. L. Archibald, C. Haley, I. J. Jackson, M. A. Groenen, R. P. Crooijmans, R. Ogden, and P. Wiener. 2013. Signatures of diversifying selection in European pig breeds. PLoS Genet. 9:e1003453.

Wright, S. 1951. The genetical structure of populations. Ann. Eugen. 15:323–354.