

Tracking footprints of an experiment of selection in Iberian pigs

Carmen Rodríguez, Yolanda Núñez, Ana I Fernández, Almudena Fernández, Carmen Barragán and Luis Silió
 Department of Animal Breeding, INIA, Madrid, Spain

ABSTRACT: Genomic tools have been used for analyzing the changes originated by an experiment of selection performed on a strain of Iberian pigs. Both pedigree and SNP-based metrics provided similar estimates of the reduction of autosomal genetic diversity (*GD*) due to selection, equivalent to an 11% of the diversity of a group of contemporary unselected animals with a common demographic history. Different estimates of the loss of *GD* in SSCX were obtained with both metrics: 10 and 18%, respectively. This may be attributed to the low number of available SNPs. Traces of the performed selection have been identified in regions of chromosomes 1, 3, 4, 15, 17 and X, which include the *FAT1* QTL region and some genes related to fatness, as *TNC-C* and *PTGIS*. Further gene sequencing and association analysis should be carried out to obtain a better understanding of the causal polymorphisms.

Keywords:-Iberian pig;-Gene diversity;-Selection footprints

Introduction

An experiment of selection for lean growth in Iberian pigs was carried out for seven years by splitting the Torbiscal strain in two lines: one selected line (S) and one unselected line (C) maintained in a conservation programme (Silió et al, 1996). Selection was based on EBV for two early recorded traits: weight at 4 months of age and ultrasonic backfat depth adjusted to 40 kg of body weight. Genetic changes for these traits were estimated in both lines, being substantial the diminution of backfat thickness in the selected line (Figure 1). Correlated responses in other traits were observed in fattened progenies from both lines, with remarkable differences for two fatness related traits: backfat depth at 100 kg (S-C = -3.87 ± 0.48 mm) and intramuscular fat content at 160 kg (S-C = -1.83 ± 0.75 %).

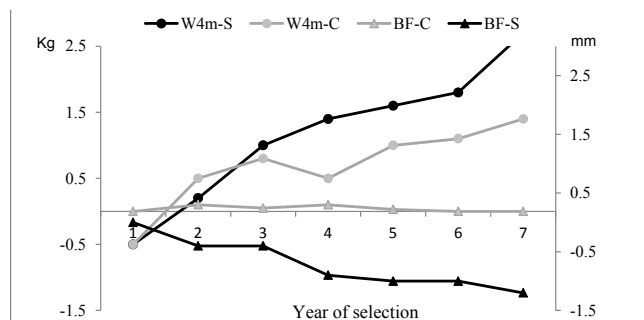


Figure 1. Genetic trends for weight at 4 months (W_{4m} , circles) and backfat thickness at 40 kg (BF, triangles) in the S and C lines of the Torbiscal Iberian pig strain

Recent advances in animal genomics, including the development of SNP genotyping arrays and more accurate genome assembly annotations of reference sequences, made it possible to get measures of genetic diversity (*GD*) and genetic differentiation (F_{ST}) for each SNP, and to examine their distribution at the genome, chromosome and SNP level. These resources have enabled several studies for identifying signatures on chromosomal regions that have been targets of the selection performed in domestic animals, such as breeds of dogs (Akey et al., 2010) or pigs (Wilkinson et al., 2013). Differences in SNP-based genetic diversity for specific chromosomal regions have also been detected between small populations that have recently diverged (Engelsma et al., 2012).

In the present study, both approaches based on SNP-metrics were applied to understand the changes caused by the quoted selection experiment in Iberian pigs. Our goals were: a) to compare the genetic diversity of three groups of animals with a common ancestral origin: the S line and two groups of the C line sampled at different periods, and b) to detect genomic regions associated with variation in fatness traits.

Materials and Methods

DNA samples, SNP genotyping and phenotypic records. DNA from 144 Iberian sows of the Torbiscal line were genotyped using the Porcine SNP60 BeadChip v.1 (Illumina). After filtering, only 27,097 SNPs were retained in the dataset and used in the analyses. These sows proceeded from three groups: a) 50 of the selected line (S), b) 51 contemporary sows of the conservation program (C1), and c) 43 sows of the conservation program born about 5.6 equivalent generations later (C2). Additionally, 184 progenies of S and C1 sows with available growth and carcass records were genotyped for the ALGA0025026 and MARC0002405 SNPs.

Statistical analysis. Pairwise kinship coefficients (f) were calculated for the animals of the three quoted groups from a complete pedigree of 1,819 animals (individual/sire/dam) born along 27 equivalent generations. Genetic diversity (*GD*) was calculated for each group and for all the animals from the corresponding pairwise kinship coefficients: $GD_{PEDIG} = 1 - f$. A similar SNP-based metrics of within population *GD* is the expected heterozygosity (Nei, 1973), that was calculated for the whole set of animals, for each group and at genome or chromosome level from the

allelic frequencies of each one of the corresponding m SNPs:

$$H_e = \frac{1}{m} \left(\sum_{j=1}^m \left(1 - \sum_{i=1}^2 p_{ij}^2 \right) \right)$$

Between group differences in H_e values were tested by bootstrapping, using 10,000 bootstrap samples created by repeated random sampling with replacement of the m loci.

As proposed by Akey et al (2010), the locus specific divergence in allele frequencies of each group was measured by unbiased estimates of pairwise F_{ST} . For each SNP and i group we calculated the statistics

$$d_i = \sum_{j \neq i} \frac{F_{ST}^{ij} - E[F_{ST}^{ij}]}{SD[F_{ST}^{ij}]},$$

where $E[F_{ST}^{ij}]$ and $SD[F_{ST}^{ij}]$ are the mean and standard deviation of the F_{ST} between i and j groups calculated from the 27,097 SNPs. For each group, d_i were averaged for sliding overlapping windows of ten successive SNPs. Candidate regions were defined as the 99.0th or 99.9th percentile of the genome-wide distribution of the averaged d_i values for the S line against the C1 and C2 groups. Gene content across candidate regions was determined using the BioMart tool from Ensembl dataset.

Candidate regions. Effects on fatness traits of some SNPs located on one of the candidate regions were estimated using animal models fitting additive and dominant effects. Analyses were performed using the Qxpak v.4 software (Pérez-Enciso and Misztal, 2004). The statistical significance of each SNP effects was tested comparing the full and reduced models by means of χ^2 approaches to the distribution of log-likelihood ratios (LR). The *TNC-C* gene, located in other candidate regions, was sequenced amplifying nine overlapping fragments (4,400 bp, 17 exons) of the mRNA sequence (GenBank accession: NM_214230) on muscle tissue samples from three Torbiscal Iberian pigs.

Results and Discussion

Genetic diversity. Mean values of pedigree and SNP-based metrics of GD of the three groups of sows are shown in Table 1. In spite of its different scale, both pedigree and marker measures indicate that the reduction of autosomal GD in the S line is equivalent to 11% of the diversity of the contemporary unselected C1 group. The reduction of the SSCX GD between both groups is equal to 10 or 18%, according to the pedigree or SNP metrics. The comparison between the two groups of the C line indicates lower but relevant differences in GD , with no overlapping bootstrap 95%CI. 4% decreases of autosomal GD in the C2 line with respect to the C1 group was estimated with both metrics, being the diminution of SSCX equivalent to 5 or 11%, considering pedigree or marker metrics respectively.

Table 1. Pedigree (GD_{PEDIG}) and SNP-based (GD_{SNP}) measures of genetic diversity in three related groups of Iberian sows for autosomes and chromosome X

	Group of Iberian sows			Total
	C1	C2	S	
N° of sows	51	43	50	144
GD_{PEDIG}				
Autosomes	0.825	0.791	0.734	0.820
SSC X	0.807	0.770	0.725	0.806
$GD_{SNP-AUTO}$				
Mean	0.339	0.327	0.303	0.337
95%CI	0.337 0.341	0.325 0.329	0.301 0.305	0.335 0.339
$GD_{SNP-SSCX}$				
Mean	0.350	0.310	0.286	0.335
95%CI	0.336 0.363	0.295 0.326	0.270 0.303	0.321 0.349

Marker-based metrics of GD allow for examining the diversity in each chromosome and the possible differences between groups at chromosome level. These differences, with their bootstrap 95%CI, are represented in the Figure 2 for the two main comparisons. The graphic mainly highlights the effects on GD at chromosome level of the performed selection (C1 vs S), and those of the genetic drift along an average number of 5.6 equivalent generations of the conservation program (C1 vs C2). A different heterozygosity for the SSCX may be expected because the effective size for sex-linked genes may differ from the calculated size for autosomes (Caballero, 1995). It occurs in the comparison between groups of the C line. However, the comparison between marker-based H_e measures for this and other chromosomes is difficult due to the variable number of available SNPs per chromosome.

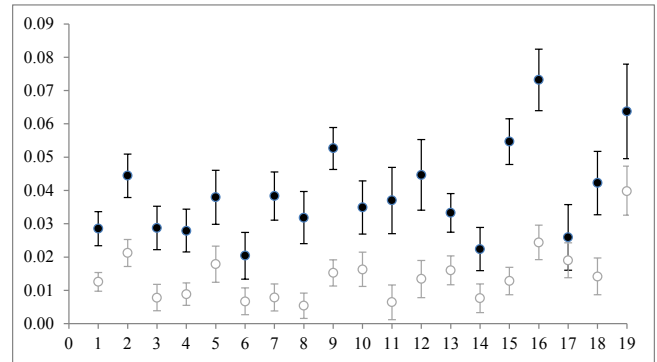


Figure 2. Differences in expected heterozygosity based on SNP data between Iberian sows groups C1 and S (black circles) and C1 and C2 (empty circles) over all 19 chromosomes, including 95%CI calculated by bootstrap

Signatures of selection. The genome-wide empirical distribution of d_i for the S line is shown in Figure 3. In total, 270 windows reached the 99.0th and only 27 the 99.9th percentile, respectively. The last ones may be grouped in nine regions of SSC 1, 3, 4, 15, 17 and X. Their location

according to the *Sscrofa10.2* annotation, number of SNPs and maximum d_i value within each region are presented in Table 2. Eight genes related to fatness traits have been identified in seven of these regions. The genome-wide distribution of d_i values only based on autosomal SNPs provided consistent results for the identified autosomal candidate regions.

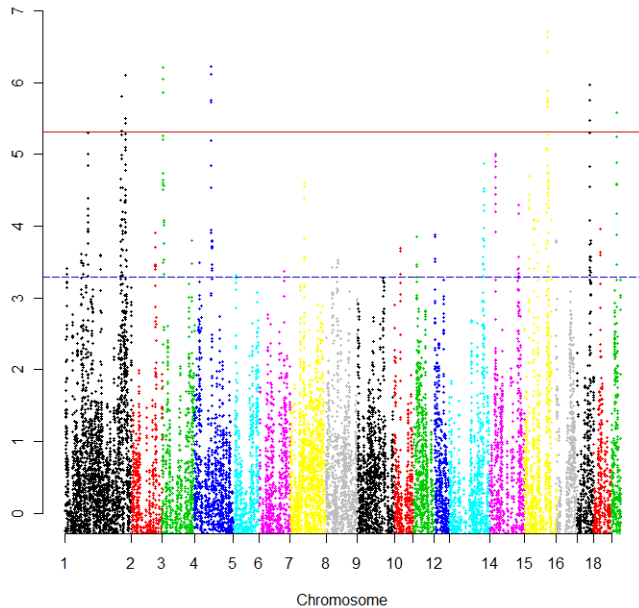


Figure 3. Genomic distribution of d_i values for the S line against the C1 and C2 groups. Dashed blue line and red line respectively denote the 99th and 99.9th percentiles.

Table 2. Genomic regions with outlier d_i windows for the S line against the unselected C1 and C2 groups

Chr	Region (Mb)	SNP	Max d_i	Candidate Gene
1	276.2-277.0	11	5.81	<i>SLC44A1</i>
1	286.5-286.9	10	5.50	<i>TNC-C</i>
1	288.7-289.5	11	6.10	<i>TLR4</i>
3	5.4-5.9	12	6.21	-
4	52.7-56.1	13	6.23	<i>CPNE3, CA2</i>
15	123.6-124.4	11	5.78	<i>ACADL</i>
15	125.2-126.0	15	6.71	-
17	57.1-57.5	12	5.97	<i>PTGIS</i>
X	65.7-83.7	13	5.58	<i>CITED1</i>

The results related to SSCX should be interpreted with caution. The candidate region found in SSCX is very large (18 Mb) and coincides with a recombination coldspot (Ma et al., 2010) overlapping with a run of homozygosity previously fixed to domestication (Rubin et al., 2012). Other candidate regions or genes present a remarkable interest. The region identified on SSC4 matches with the well-known porcine FAT1 QTL region (Andersson et al, 1994). Two SNPs (MARC0002405 and ALGA0025026) located within this region on 54.24 and 54.27 Mb, showed significant effects on fatness traits in the performed association analysis. Their respective additive effects on backfat depth at nine months of age represent a 3.42 and a 4.65% of the trait mean. Because of these SNPs are strongly correlated,

these effects cannot be considered independent. A good candidate for this region may be the *CA2* gene, related to fat accumulation in bovine skeletal muscle (Zhang et al., 2010).

Other interesting candidates for the regions on SSC1 and SSC17 are the *TNC-C* and the *PTGIS* genes, respectively. The tenascin C (*TNC-C*) is a member of the family of genes responsible for coding extracellular matrix proteins, and it is considered a promising candidate for porcine carcass and meat quality traits (Kayan et al., 2011). The sequence comparison across three Iberian pigs allowed us to detect 14 polymorphisms: six of them are non-synonymous. Some of these SNPs have been selected for genotyping and further association analyses. A similar procedure of sequencing and detection of polymorphisms in our material will be carried out for the *PTGIS* gene, which has been related to early growth and fatness traits in native Korean pigs (Li et al., 2011).

Conclusion

Our results provide a first assessment of the changes originated by a selection experiment for early growth and fatness in Iberian pigs. They profit of the common demographic history of the compared groups of animals. Both pedigree and SNP-based metrics provided similar estimates of the reduction of *GD* due to selection. Traces of the performed selection have been identified in some chromosomal regions and genes. Further gene sequencing and association analysis should be carried out to identify causal polymorphisms.

Acknowledgements

Financial support was provided by RTA2011-00113 grant. Technical assistance of Fabián García is acknowledged.

Literature Cited

- Andersson, L., Haley, C. S., Ellegren, H. et al. (1994). *Science* 263:1771-1774.
- Akey, J.M., Zhang, G., Zhang, K. et al. (2002). *Genom. Res.* 12: 1805-1814.
- Engelsma, K.A., Veerkamp, R.F., Calus, M.P.L. et al. (2012). *J. Anim. Breed. Genet.* 129: 195-205.
- Kayan, A., Cinar, M.U., Uddin, M.J. et al. (2011). *Meat Sci.* 89: 76-83.
- Li, X., Kim, S-W., Do, K-T. et al. (2011). *Mol. Biol. Rep.* 38: 3805-3820.
- Ma J., Ianuccelli, N., Duan, Y. et al. (2012). *BMC Genomics* 11: 159.
- Nei, M. (1973). *Proc. Natl. Acad. Sci. U.S.A.* 70: 3321-3323.
- Perez-Enciso, M., Misztal, I. (2011). *Bioinformatics* 20: 2792-2798.
- Rubin, C-J., Megens, H-J., Martínez-Barrio, A. et al. (2012). *Proc. Natl. Acad. Sci. U.S.A.* 109: 19529-19536
- Silió, L., Toro, M.A., Rodríguez, M.C., Rodríguez, J. (1996). *47th Ann. Meeting of the EAAP*, Lillehammer, Norway, 25-29 August.
- Wilkinson, S., Lu, Z.H., Meggens, H.J. et al. (2013). *PLOS Genetics* 9(4): e1003453.
- Zhang, Q., Li, H-G., Han, J-A. et al. (2010) *Meat Sci.* 86: 814-820.