Instituto Universitario de Investigación
Biocomputación y Física de Sistemas Complejos
Universidad Zaragoza

CIENCIAS ZARAGOZA

1542 Universidad Zaragoza

Final Master´s Project in

Quantitative Biotechnology

Genomic differentiation and detection of signatures of selection between strains of Iberian pigs.

Author

Inés Alonso Jáuregui

Director

Luis Varona Aguado

2018

INDEX

ABSTRACT

The Iberian pig is a racial entity native to the Iberian Peninsula with great adipogenic capacity that involves an extremely good meat quality. Nevertheless, the Iberian pig breed is composed by several differentiate strains involving a very large genetic diversity. The genomic differentiation between strains can be due to genetic drift or because of selection and adaptation processes that may have left relevant signals within the genomic structure of the populations.

We have used the genotypes from 349 individuals from pure and crossbred populations with the Illumina porcine SNP60 BeadChip provided by the IBEROMICS grant (CDTI-IDI-20170304 and CICYT-CGL-2016-80155). The porcine SNP60 BeadChip included around 55,000 SNP markers evenly spaced within the porcine genome. With this information, three alternative procedures for the detection of signatures of selection were applied. The first (F_{ST}) was based on the differentiation between populations, the second (*ROH*) on the reduction of local variation and the third (*nSL*) on the extension of the linkage disequilibrium. Later on, the genes present within the genomic regions associated with signatures of selection were identified using a genomic browser (www.ensembl.org). Finally, gene enrichment procedures were used to identify the metabolic pathways associated with selection or differentiation between the Iberian strains.

The metabolic pathways identified with each procedure were different as they were associated with selection events at different evolutionary times. The older ones were identified with the *ROH* method and they corresponded to aminoacid transportation, immune response, neurotransmission, cellular structure process or very primary biological processes. The F_{ST} procedure identified pathways associated with morphological development, lipid metabolism, immune response and Na-K equilibrium. Finally, the most recent signals were hormonal and metabolic regulation, cell proliferation and cellular response as they were identified by the *nSL* procedure. In addition, the genes located in the selected genomic regions were also discussed.

1. BACKGROUND AND OBJECTIVES

1.1. Signatures of selection

The processes of selection, adaptation and differentiation suffered by populations throughout their evolution can leave detectable signals in the structure of the genome. The ability to identify these signals clarifies the evolutionary reconstruction of these populations, and provides a potential identification of genes of interest, either because of their association with characters of economic importance, or due to specific adaptation processes. These signals in the genome or "signatures of selection" are defined as regions of the genome that contain functionally important sequence variants, and which, therefore, are or have been subjected to natural or artificial selection processes that have left special patterns in the DNA structure (Qanbari and Simianer, 2014; Utsunomiya *et al.*, 2015).

All procedures for detecting selection signals are based on the postulates of the Theory of Neutral Evolution (Kimura, 1983). This theory states that most of the variability within species is not caused by natural selection, but by genetic drift of mutant alleles that are selectively neutral or close to neutrality and proposes that (Kimura, 1989; Walsh and Lynch, 2014):

- Many of the mutations that arise in a population are deleterious and, consequently, are quickly eliminated.
- A very small proportion of the mutations that arise are advantageous and are rapidly fixed in the population.
- Most of the mutations that are observed are selectively neutral and, consequently, are those that contribute to the polymorphism and divergence between populations.

In the literature, there a huge number of procedures to detect “signatures of selection”, that are applicable to a long term or to a more recent selection scenarios. In particular, procedures that study a more recent evolutionary framework can be classified according to Oleksyk *et al.* (2010) in the following categories:

a) Differentiation between populations. The starting point of these methods implied that populations differ by drift processes. However, if there is a region that, in addition, has been affected by selective forces, it will be reflected in a greater divergence among the populations in that specific region. Among the methods described in this category, the most traditional and used is the Fixation Index (F_{ST}), described by Wright (1943). This method assumed that an original population was split into several subpopulations and it calculated the reduction of the heterozygosity in the subpopulations, with respect to the expected heterozygosity as calculated with the average allelic frequency. However, it assumed that all populations had the same effective size and they derive independently for the same ancestral population (Qanbari and Simianer, 2014). Moreover, there are other more recent methods that are based on a Bayesian modeling of the allelic frequencies, such as *Bayescan* (Foll and Gaggiotti, 2008) and *Selestim* (Vitalis *et al.*, 2014). Nevertheless, empirical results (González-Rodríguez *et al*., 2016) using SNP data from Spanish beef cattle populations indicated that the results of these Bayesian approaches lead to the same conclusions that the simple F_{ST} approach, as its assumptions are robust to minor modifications.

b) Analysis of the reduction of local variation. The second group of methods were based on the reduction of genetic variation in specific areas of the genome and thus avoiding the need for information from more than one population. Several methods have been proposed for this approach such a simple measure of average heterozygosity within genomic regions (Oleksyk *et al.*, 2008) or the identification of Runs of Homozygosity -*ROH*- (McQuillan *et al.*, 2008). Runs of Homozygosity are genomic regions where each individual is homozygous for all polymorphisms within them. The procedure was initially described as a measure of inbreeding but a simple count of the number of individuals that are homozygous for each SNP allows to identify genomic regions with a reduction of genetic variation that are subject to be recognized as "signatures of selection".

c) Extension of linkage disequilibrium. The length and frequency of the haplotypes associated with the favored allele depend on the intensity of the selective process and the frequency of recombination, so that the presence of long haplotypes with high frequency suggest the presence of a recent selection signature, since it can be interpreted that the selected allele has increased its frequency quickly in recent generations.

There are several methods to quantity this phenomenon, such as the *EHH* (Sabeti *et al.*, 2002) or modifications of this procedure such as *iHS* (Voight *et al.*, 2006) or *nSL* (Ferrer-Admetlla *et al.*, 2014). The main advantage of this last procedure is that it avoids the use of a previously calculated genetic map, which will need a very large sample to be accurately estimated.

d) Modifications of the frequency spectrum. The starting hypothesis was that the appearance of a new mutation, or derived allele, occurs very infrequently, and is usually eliminated by fast ancestral allele fixation. On the contrary, if this new mutation has a selective advantage, or is affected by the drag effect with some other advantageous mutation, its frequency will increase. Among the method of this category, it is worth to mention the classical *Tajima* (Tajima, 1989), *Fu and Li* (Fu and Li, 1993) and *Fay and Wu* (Fay and Wu, 2000) methods. Nevertheless, all these methods are defined for full sequence information and they cannot produce accurate results with sparse genotypes such as provided by SNP genotyping devices. Moreover, recent results (González-Rodríguez *et al.*, 2016) suggested that their results are similar to the ones provided by the reduction of local variation methods.

1.2. The Iberian pig

The pork production sector in Spain occupies the fourth place of importance worldwide, after China, USA and Germany. At European level, Spain is in second place in production with 17.5% of the tons produced. The Spanish pig sector has a fundamental importance in the economy of our country that 12.7% of the Total Agricultural Income. In this sense, the number of pigs slaughtered in Spain (MAPAMA, 2017) was more than 17 million individuals, of which more than 3 million were Iberian pigs. The Iberian pig is a racial entity native to the Iberian Peninsula with great adipogenic capacity that involves an extremely good meat quality (Benítez *et al.*, 2018).
The production of Iberian pigs, both in extensive ("de bellota", linked to the “dehesa”) and in intensive conditions, (“de cebo"), reached its peak in 2007 leading to a situation of oversupply and imbalance of the market, which led to a decrease in prices (Coordinadora de Organizaciones de Agricultores y Ganaderos, 2016). In later years, the recovery of production has favored the restoration of the prices of Iberian animals, which recorded historical figures in the 2013/2014 campaign. In fact, in 2017 the

Iberian pig (and crosses) census was over 3 million individuals, continuing the increase of the previous two years (MAPAMA, 2017).

The geographical distribution of the Iberian pig has traditionally been limited to the southwest of the Iberian Peninsula, linked to the "dehesa". As it can be seen in Figure 2, more than 50% of the total heads of Iberian pigs in Spain are located in Andalucía and Extremadura (MAPAMA, 2015).

CCAA	Total reproductores		Total animales		Total	Nº Ganaderías
	Hembras	*Machos*	*Hembras*	*Machos*		
ANDALUCÍA	124.441	3.556	161.322	18.301	179.623	1.650
ARAGÓN	0	0	4	0	4	1
CANTABRIA	0	0	0	0	0	0
CASTILLA LA MANCHA	22.953	100	27.147	2.141	29.288	75
CASTILLA LEÓN	99.220	614	122.855	3.579	126.434	660
CATALUÑA	7.265	3	10.136	3	10.139	5
CEUTA	0	0	0	0	0	0
COMUNITAT VALENCIANA	1.145	0	1.295	0	1.295	2
EXTREMADURA	205.238	2.565	254.963	30.110	285.073	2.454
GALICIA	0	0	0	0	0	0
ILLES BALEARS	0	0	0	0	0	0
ISLAS CANARIAS	0	0	0	0	0	0
LA RIOJA	0	0	0	0	0	0
MADRID	21	3	32	4	36	1
MELILLA	0	0	0	0	0	0
MURCIA	20.667	15	29.148	15	29.163	4
NAVARRA	0	0	0	0	0	0
PAÍS VASCO	0	0	0	0	0	0
PRINCIPADO DE ASTURIAS	0	0	0	0	0	0
Totales	480.950	6.856	606.902	54.153	661.055	4.852

Figure 1. Distribution of the number of pure Iberian reproductive individuals, total number of animal and number of farms in Spain, distributed by Autonomous Communities. (Source: MAPAMA, 2015).

The population structure of the Iberian pig is made up of numerous strains that have differentiated thanks to the processes of genetic drift or selection and adaptation after reproductive isolation. In fact, the Spanish Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente considers two varieties (Entrepelado and Retinto) as *Razas Autóctona de Fomento* and another three (Lampiño, Manchado de Jabugo and Torbiscal) as *Razas Autóctonas en Peligro de Extinción.* In fact, several studies (Martínez *et al.*, 2000, Fabuel *et al.*, 2004) have observed that the genetic variability

between Iberian strains is higher than that observed between varieties of commercial white pig (Laval *et al.*, 2000).

The varieties of the Iberian breed object of study in this Master's Thesis are Retinto, Entrepelado and Torbiscal (see Figure 2). The first one has an evolutionary tendency of the population in expansion. The variety Entrepelado is a fixed hybrid of the cross between Lampiño and Retinto, showing something earlier and less greasy than Lampiño, but without reaching the levels of the Retinto. Unlike the previous two, the population of the Torbiscal variety has an evolutionary tendency in recession. They are animals of greater height, very resistant and with greater prolificacy (BOE, 2007; MAPAMA, 2015).

Entrepelado Retinto Torsbical

Figure 2. Varieties of Iberian pig object of this Master thesis: Entrepelado, Retinto and Torbiscal (Source: MAPAMA, 2015).

Nevertheless, there are still no studies of regional genetic diversity that allow the identification of the genes associated to the differentiation processes between Iberian pig strains carried out with high density genotypes and with a relevant number of individuals. Therefore, the objective of this work is to locate the genomic regions associated with signatures of selection due to adaptation or differentiation between three of the strains of greater implantation (Entrepelado, Retinto and Torbiscal).

1.3. Objectives

The objectives of this Master Thesis are to study the genomic differentiation between three strains of Iberian pig (Retinto, Entrepelado and Torbiscal) and to identify the genomic regions and metabolic pathways associated with signatures of selection using three alternative methods:

1). Measures of the differentiation between populations
2). Analysis of the reduction of local variation
3). A procedure based on the extension of the linkage disequilibrium.

2. METHODOLOGY

2.1. Material

The biological material used in this Master Thesis consist of the genotypes with the PorcineSNP60 v2GenotypingBeadChip of Illumina (*www.illumina.com*) for 349 individuals of the Iberian pig breed. The PorcineSNP60 v2GenotypingBeadChipwas composed by 64,232 SNPs. This dataset was obtained within the scope of two research grants (CDTI-IDI-20170304 and CICYT-CGL-2016-80155). They were distributed in the following way, 21 individuals belong to the purebred Entrepelado population, 50 to the Retinto and 78 to the Torbiscal. In addition, 25 individuals were crosses between Entrepelado and Retinto, 37 between Entrepelado and Torbiscal and 138 between Torbiscal and Retinto. Moreover, the pedigree (individual, sire, dam) of the genotyped individuals were also used.

2.2. Filtering

The result of genotyping with the PorcineSNP60 v2 Genotyping BeadChip were filtered with the software *PLINK* (Purcell *et al.*, 2007). The criteria of filtering were:

- SNP call rate over 0.95: 61,565 SNP were kept.
- Individual call rate over 0.95: All individuals were retained.
- Known location within the autosomal chromosomes: 52,458 SNP were kept.
- Minor Allele Frequency over 0.01: A total of 31,180 were finally used.

2.3. Imputation of Haplotype Phase

After the filtering step, the software *FImpute* (Sargolzaei *et al.*, 2014) was used to impute the paternal and maternal haplotype phases of the genotyped individual. The procedure of haplotype phasing is illustrated in Figure 3.

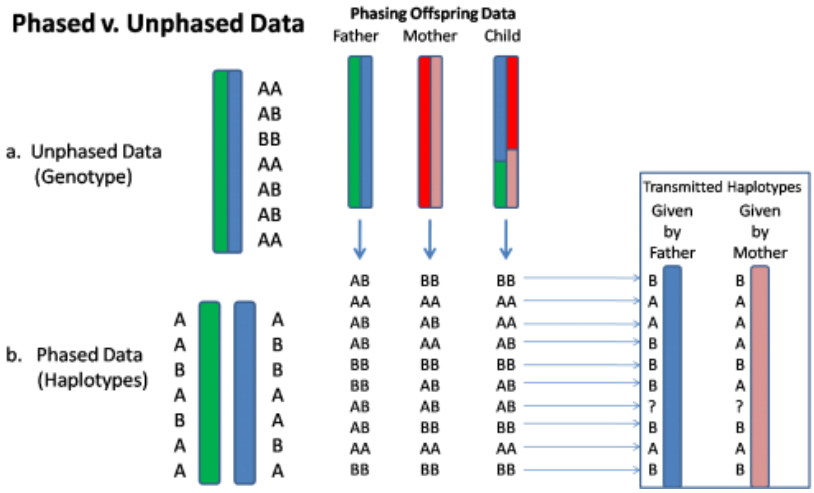

Figure 3.Phased v. Unphased Data. In the phase data, the allele provided by the father and the one provided by the mother can be known (source: http://www.chromosomechronicles.com).

The software *FImpute* uses familiar and population information to infer the haplotype phases. After the execution of the software, a total of 47 distinct haplotype phases of the Entrepelado population, 67 of the Retinto and 123 of the Torbiscal population were identified.

2.4. Genomic differentiation between populations.

The first group of procedures for the detection of selection and adaptation signatures makes use of the usual methodologies for the analysis of genetic diversity, but restricted to specific regions of the genome. The starting point of these methods implies that populations diverge by drift processes. However, if there is a region that, in addition, has been affected by selective forces, it will be reflected in a greater divergence among the populations in that specific region. Among the procedures that analyze the processes of differentiation between populations we have used the F_{ST} statistic, defined by Wright (1943). It calculates the reduction in heterozygosity observed in the subpopulations with respect to the expected heterozygosity under the average allele frequency.

$$F_{ST} = \frac{Expected\ heterozygosity\ (EH) - Observed\ heterozygosity\ (OH)}{Expected\ heterozygosity\ (EH)}$$

F_{ST} is directly related to the variance in allele frequency among populations and, conversely, to the degree of resemblance among individuals within populations. If F_{ST} is small, it means that the allele frequencies within each population are similar; if it is large, it means that the allele frequencies are different (Kent *et al.*, 2009).In order to calculate the F_{ST} for each SNP marker, the allelic frequencies (p_{ij} and q_{ij}=1-p_{ij}) of the i^{th} population (Entrepelado, Retinto or Torbical) and j^{th} SNP marker were calculated from the phased haplotypes provided by *FImpute*. Afterwards the expected and observed heterozygosity for the j^{th} SNP were calculated as:

$$OH_j = \frac{\sum_{i=1,3} 2p_{ij}\,q_{ij}}{3} \qquad EH_j = 2\bar{p}_j\,\bar{q}_j$$

Where $\bar{p}_j = \frac{p_{1j}+p_{2j}+p_{3j}}{3}$ and $\bar{q}_j = \frac{q_{1j}+q_{2j}+q_{3j}}{3}$.

Further, and in order to reduce the uncertainty caused by sampling at each SNP, we have calculated sliding windows of 5, 10 and 20 SNP and centered at each SNP. This procedure is illustrated in Figure 4.

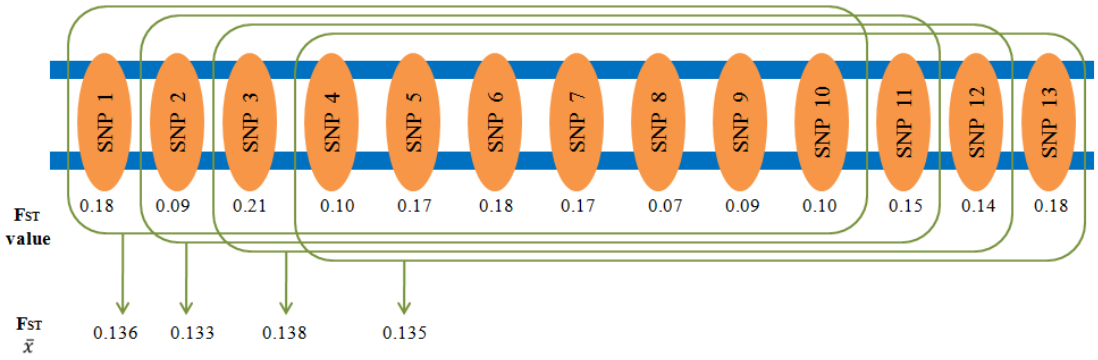

Figure 4. Sliding window of 10 SNP. Each green box represents a group of 10 SNPs, which is called sliding window. In this way, the mean values of F_{ST} are obtained for each group of 10 SNPs along the genome.

2.5. Reduction of local variation

The methods based on the reduction of local variation start from the hypothesis that those regions of the genome with a high frequency of homozygosis and large in size correspond to animals that have been subjected to selection. In this study, we have used the calculation of the runs of homozygosity (*ROH*) for the purebred genotyped individuals (21 Entrepelado, 50 Retinto and 78 Torbiscal). *ROH* (Ceballos *et al.*, 2018) are defined as segments of the genome where one individual is homozygote (See Figure 5).

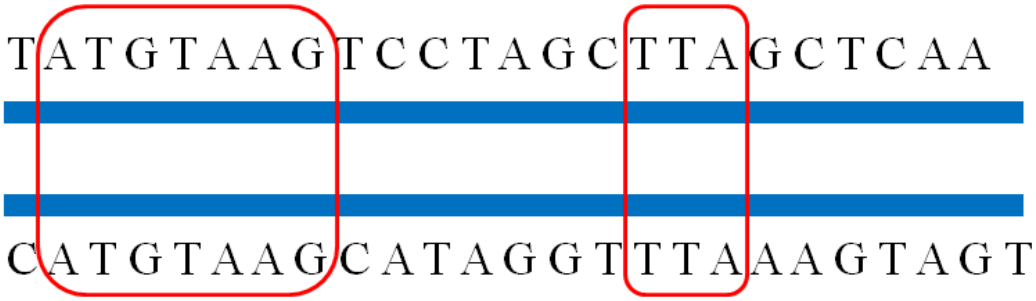

Figure 5. Runs of homozygosity. The red boxes indicate the runs of homozygosity, that is, those regions where an individual is homozygote.

The ROH can be defined according the number of SNP that contains or their size in bp. After an exploratory analysis, the threshold for a definition of an *ROH* here used was to contain at least 45 SNP and the software *detectRUNS* (Biscarini *et al.*, 2018) in R was used. *ROH* were calculated using the *consecutiveRuns*() function and in order to identify the genomic regions with a reduced local variation the function *SNPinRUNS()* was used to compute the percentage of individuals within population that have each SNP included within a *ROH*.

2.6. Extension of the linkage disequilibrium

The third approach that was used in this study is based on the extension of the haplotype homozygosity or linkage disequilibrium, implemented by the software *Selscan* (Szpiech and Hernández, 2014). Among the procedures available in this software we used the *nSL* (Ferrer-Admetlla *et al.*, 2014) method. It is based on the identification of long haplotypes with high frequency (see Figure 6).

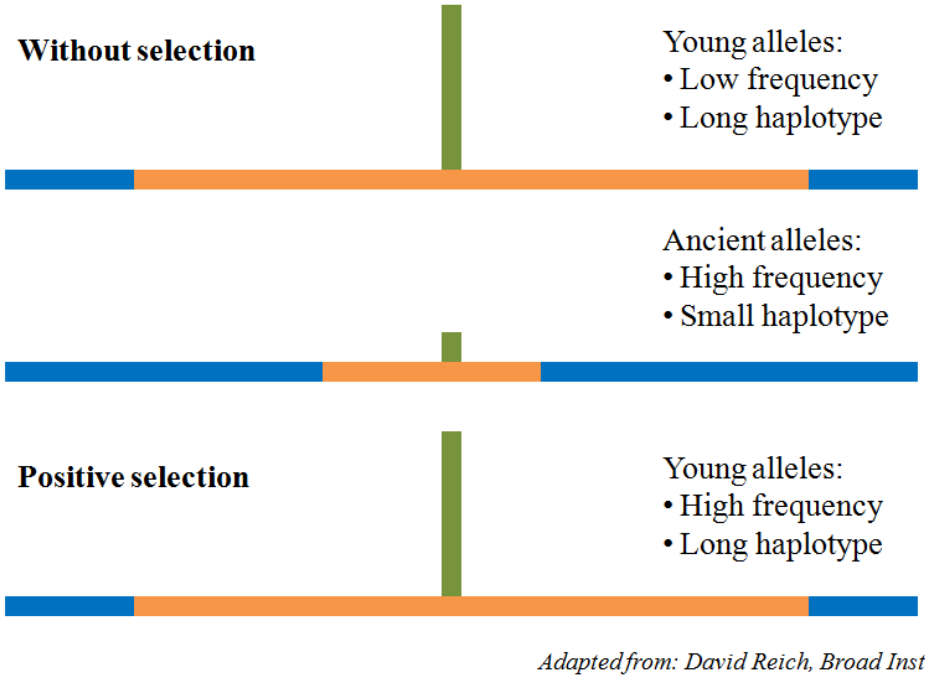

Figure 6. Different type of selection according to the frequency and length of the haplotypes (Source: http://slideplayer.com/slide/4639927/)

The hypothesis is that a new allele that is favored with selection increases rapidly its allelic frequency and it is accompanied by a long haplotype because it has few opportunities to recombine. If alleles are not affected by selection, the expectation is that alleles with low frequency were associated with longer haplotypes whereas alleles with high frequency present a smaller haplotype surrounding them.

2.7. Identification of candidate genes and metabolic pathways.

From the results of the above described procedures, the genomic regions associated with relevant signals of selection were analyzed with a two non-exclusive procedures.

1. The genes located within one Megabase around the SNP whose signals were over the 99.9% percentile for each of the procedures were identified with the *Biomart* tool from the genomic browser *Ensembl* (www.ensembl.org; Flicek *et al.*, 2014).

2. The genomic regions with a signal over the 95% percentile were identified and an enrichment analysis of the GO-*Gene Ontology*- terms with the genes located within them was performed with the software *GOrilla* (http://cbl-gorilla.cs.technion.ac.il/; Eden *et al.*, 2009).

3. RESULTS

3.1. Differentiation between populations

The average results of the F_{ST} procedure were 0.069 with a standard deviation of 0.060. However, when the results were plotted with respect to their genomic location, the results were noisy (Figure 7), because the amount of information provided by each particular SNP is low.

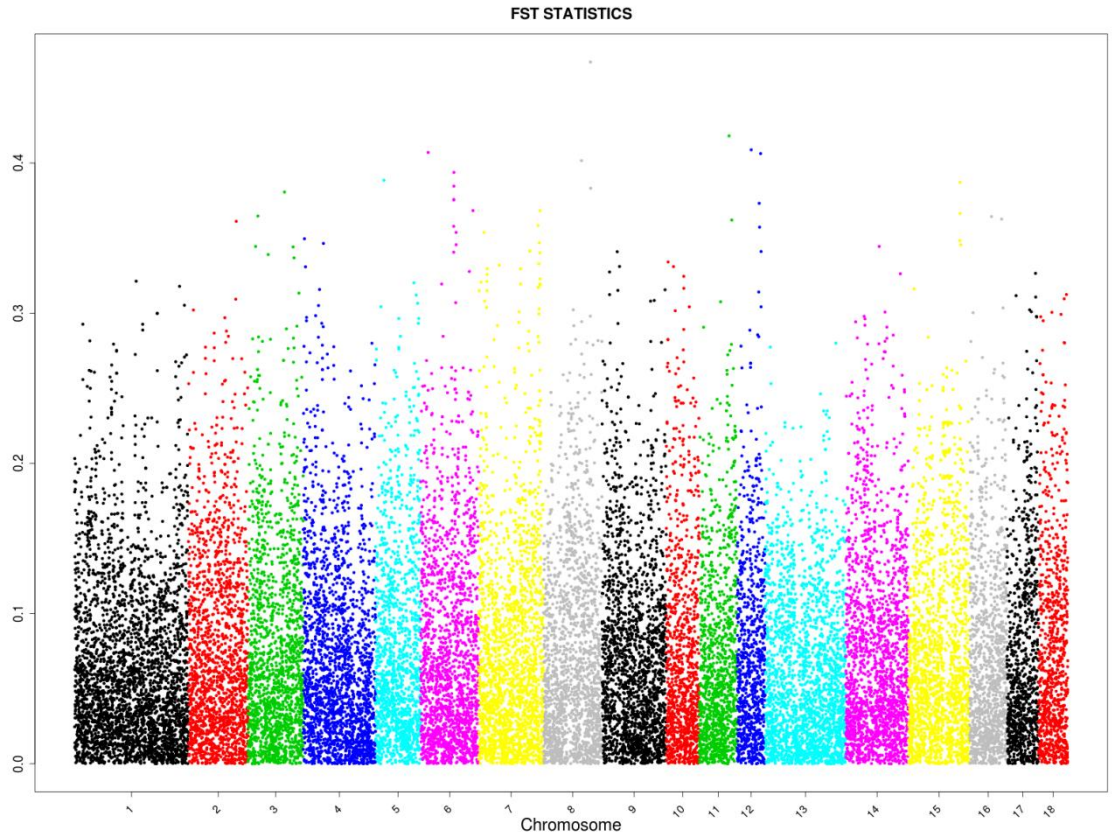

Figure 7.Genomic scan of the F_{ST} statistic for each SNP.

Thus, the results of the F_{ST} statistic per SNP were averaged in sliding windows of 5, 10 and 20 SNP. Afterwards, the standard deviations for the averaged F_{ST} results were 0.037, 0.030 and 0.024, respectively. The results are presented in Figure 8.

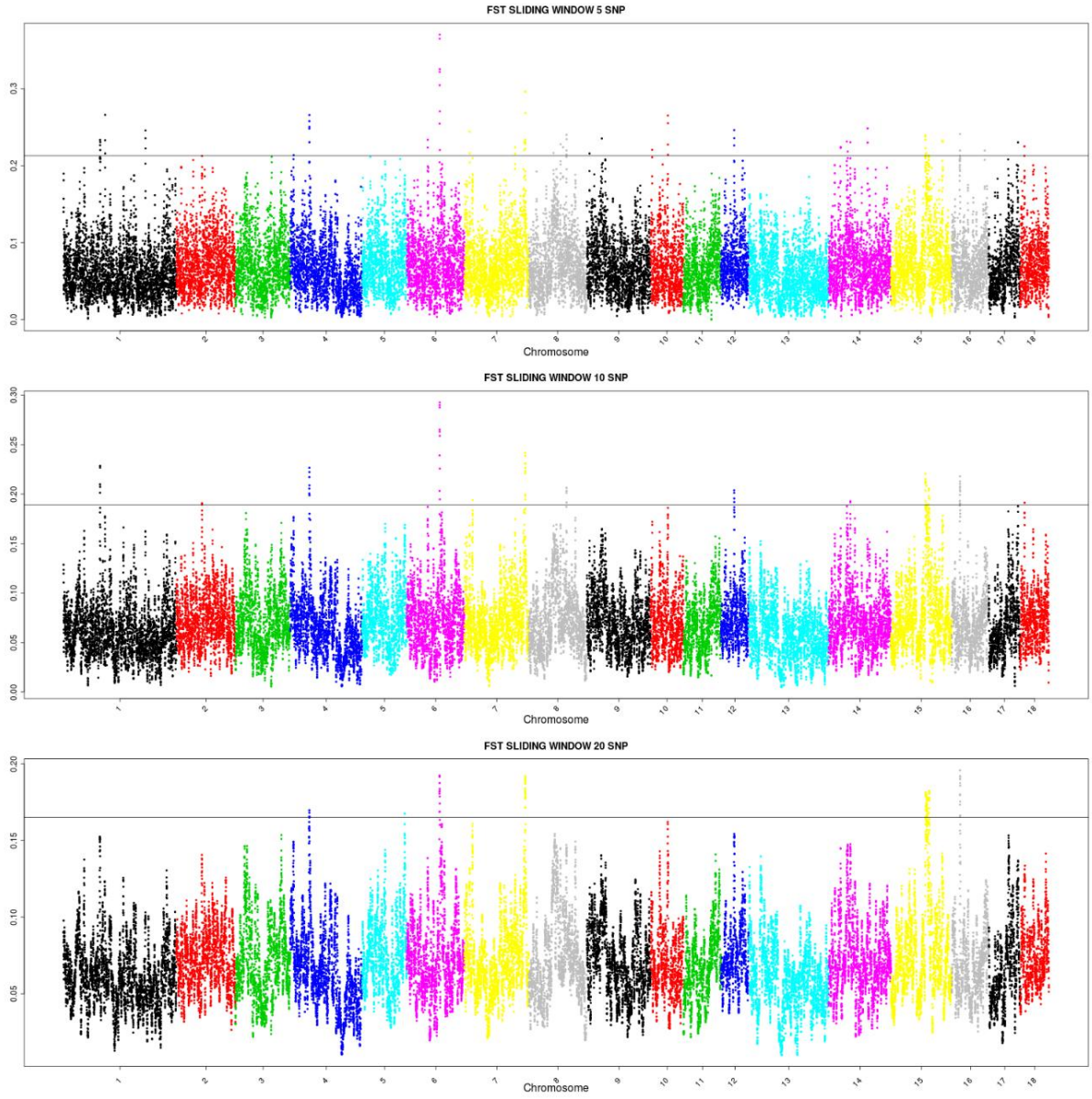

Figure 8. Genomic scan of the average F_{ST} statistic in sliding windows of 5, 10 and 20 SNP.

In addition, the F_{ST} statistic can be also computed between pairs of population, in order to discriminate the specific genomic regions involve in the differentiation between each pair. So, in Figures 9, 10 and 11, the plots of genomic differentiation computed in sliding windows of 20 SNP and between each pair of populations (Entrepelado-Retinto, Entrepelado-Torbiscal and Retinto-Torbiscal) are presented.

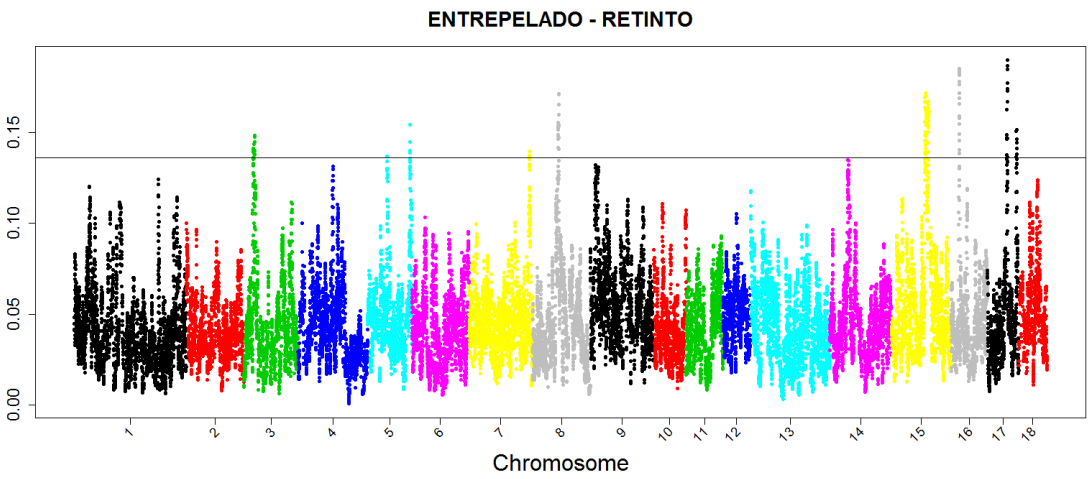

Figure 9. Genomic scan of the average F_{ST} statistic in sliding windows of 20 SNP and between Entrepelado-Retinto.

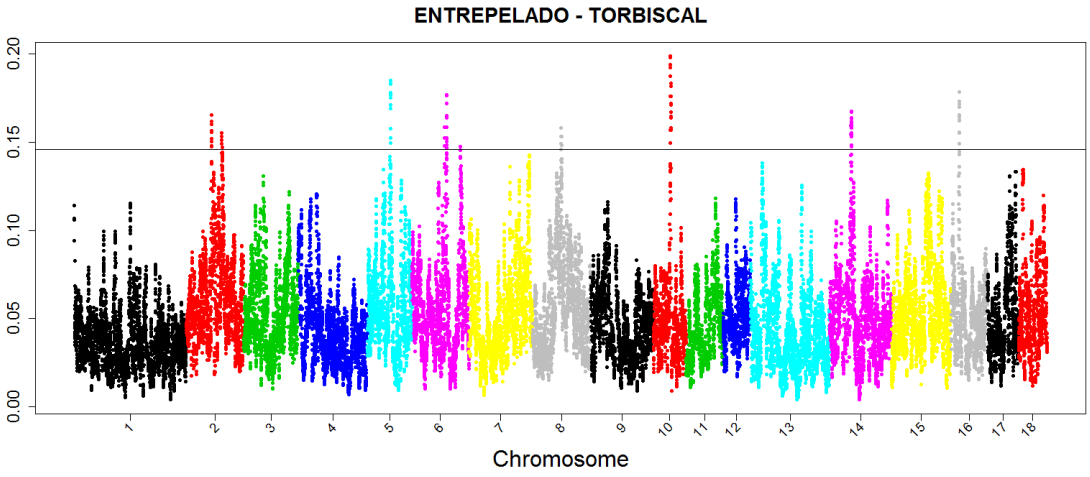

Figure 10. Genomic scan of the average F_{ST} statistic in sliding windows of 20 SNP and between Entrepelado-Torbiscal.

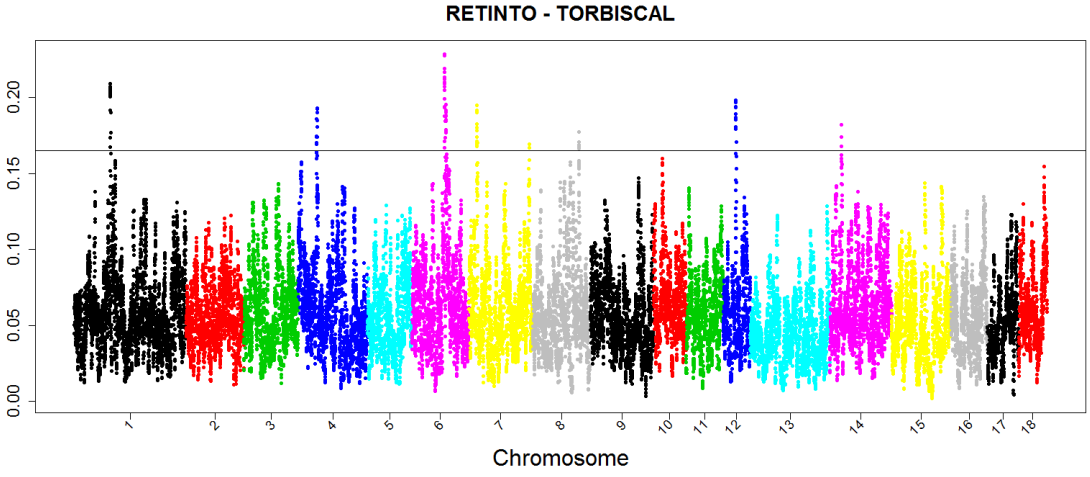

Figure 11. Genomic scan of the average F_{ST} statistic in sliding windows of 20 SNP and between Retinto-Torbiscal

The genomic regions over the 99.9% percentile were selected and the genes there located were identified using the *Ensembl* database (www.emsembl.org). The genes obtained are presented in Table 1.

Population	Chr.	Start	End	GENES
Entrepelado - Retinto	8	59780416	60060727	-
	15	80154402	80584372	CIR1, SCRN3, GPR155, WIPF1
	15	82011689	82043862	EVX1, HOXD1, HOXD3, HOXD4, HOXD8, HOXD9, HOXD10, HOXD12, HOXD13
	15	87043862	87157562	ITGA4, CERKL, NEUROD1
	16	19195924	19410017	TARS, ADAMTS12
	17	41992643	42480736	ADIG, SLC32A1, ACTR5, PPP1R16B, DHX35, FAM83D
Entrepelado - Torbiscal	5	59526012	60545861	ETV6, LRP6
	10	36213710	37943981	-
	14	44986228	45015971	MN1, TTC28, PITPNB
	16	19189083	19252487	TARS, ADAMTS12
Retinto-Torbiscal	1	77096908	77472299	REV3L, TRAF3IP2, FYN
	6	104376948	105370752	ADCYAP1, YES1
	12	28551479	28781834	-

Table 1. Genes obtained by the F_{ST} method by regions, chromosome and population.

In addition, in a second approach, the genomic regions with a F_{ST} over the 95% percentile were also identified. They comprised up to 1442 genes that were used to perform and enrichment analysis with the *GOrilla* software (http://cbl-gorilla.cs.technion.ac.il/) in order to detect the enriched GO terms for biological processes. The enriched terms with a p-value lower than 0.001 are presented in Table 2.

GO term	Description	P-value	FDR q-value
GO:0001501	skeletal system development	2.1E-6	1.56E-2
GO:0002028	regulation of sodium ion transport	1.38E-4	5.13E-1
GO:0019372	lipoxygenase pathway	2.09E-4	5.17E-1
GO:0051122	hepoxilin biosynthetic process	2.09E-4	3.87E-1
GO:0051121	hepoxilin metabolic process	2.09E-4	3.1E-1
GO:0048305	immunoglobulin secretion	5.78E-4	7.14E-1
GO:0048598	embryonic morphogenesis	6.56E-4	6.95E-1
GO:0001764	neuron migration	7.58E-4	7.03E-1
GO:0042733	embryonic digit morphogenesis	9.86E-4	8.13E-1

Table 2. Enriched GO terms (p-value < 0.001) for biological processes using the genes obtained from the F_{ST} analysis (see Table 1)

3.2. Reduction of local variation

The results of reduction of local variation were studied using Runs of Homozygosity (Ceballos *et al.*, 2018). This procedure was used to approximate the degree and antiquity of inbreeding. Nevertheless, it may be also applied to study the amount of local genetic variation by calculating the percentage of individuals that have each SNP included within a ROH. Here, that percentage of individuals has been calculated for each population (Entrepelado, Retinto and Torbiscal) and for each SNP. The results are presented in Figures 12, 13 and 14, respectively.

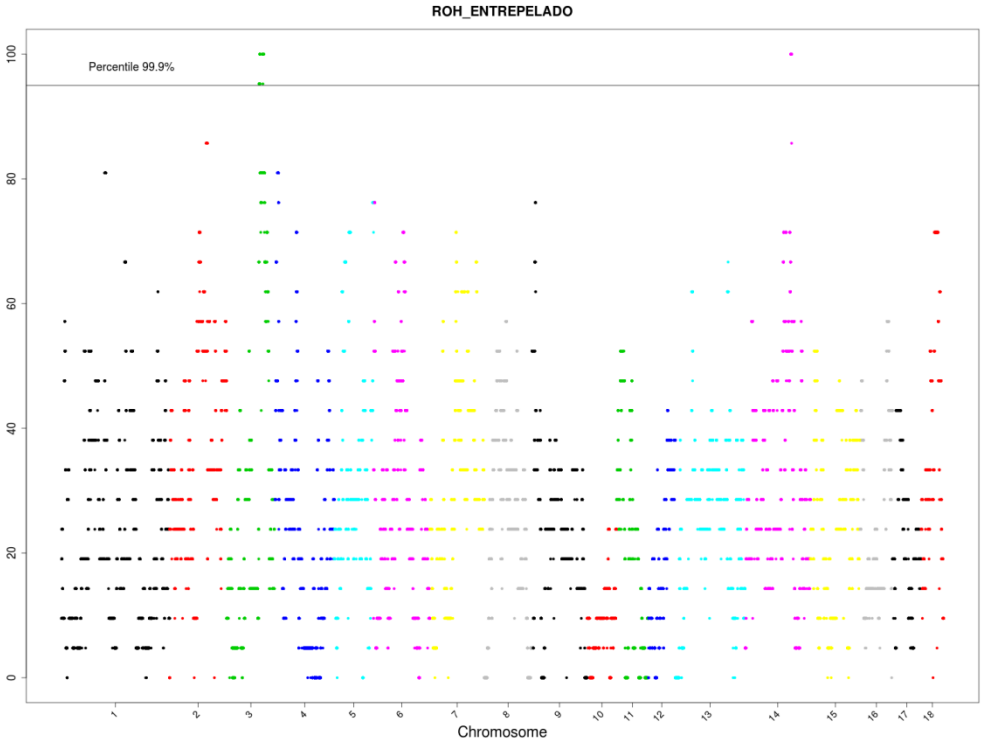

Figure 12. Percentage of individual with each SNP included within a ROH in the Entrepelado population.

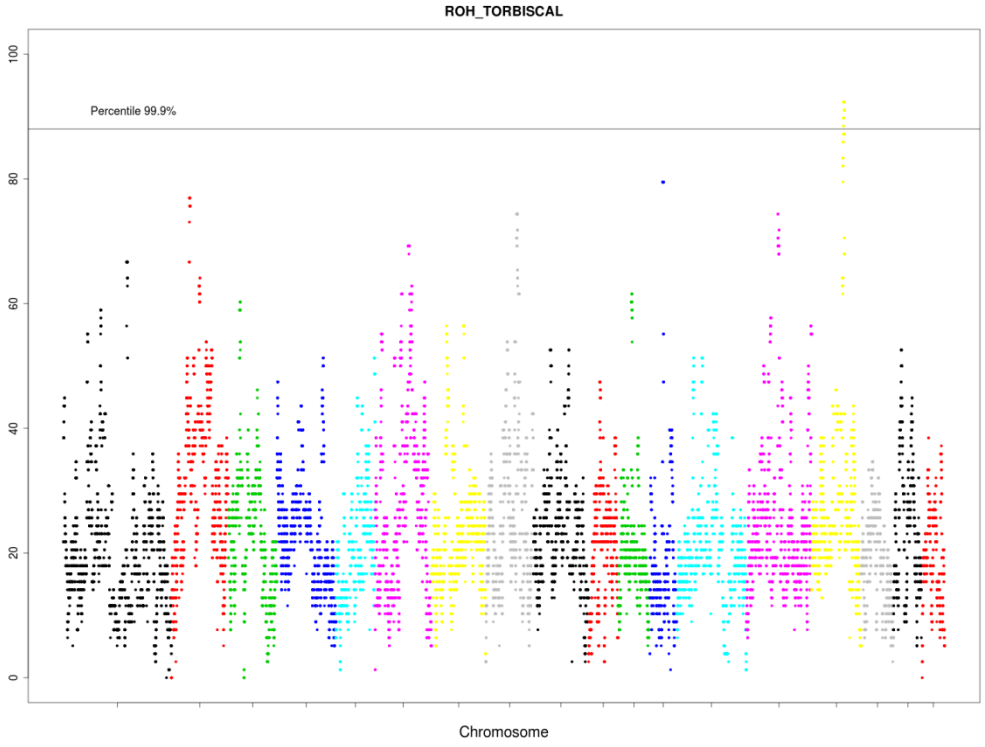

Figure 13.Percentage of individual with each SNP included within a ROH in the Torbiscal population.

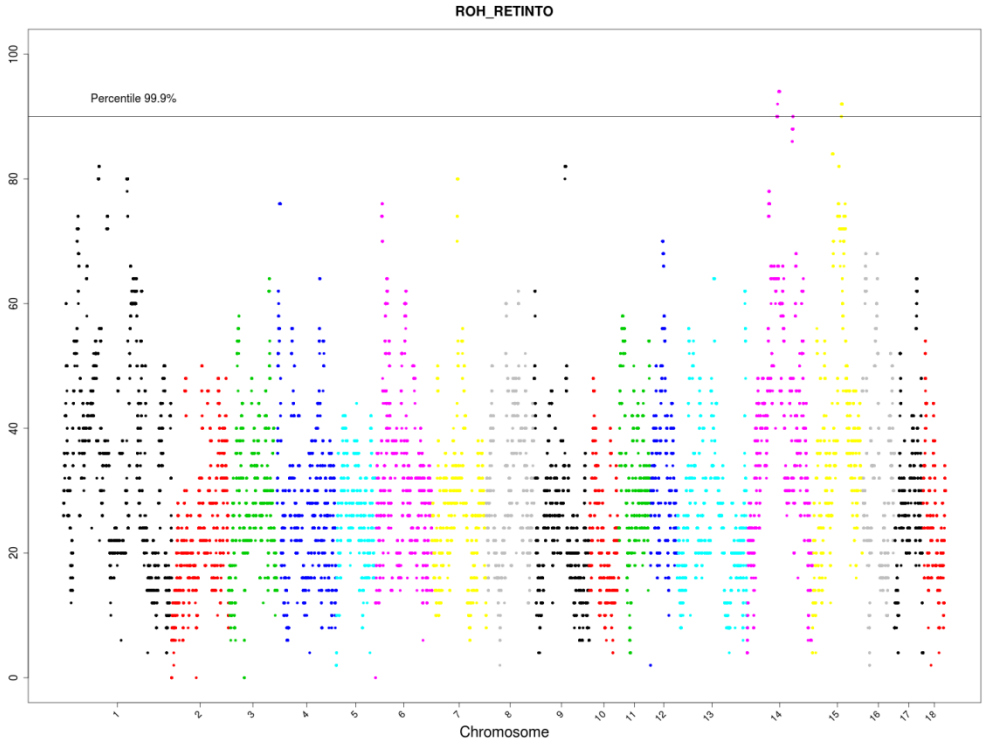

Figure 14.Percentage of individual with each SNP included within a ROH in the Retinto population.

The average (and standard deviation) percentage of individuals with an each specific SNP included within a *ROH* were 25.87 (17.09), 30.39 (14.65) and 23.35 (11.53) for Entrepelado, Retinto and Torbiscal. However, as it can be observed in the above presented figures, there are some genomic regions with a higher reduction of genetic variation expressed by a much higher percentage of individuals included within a *ROH* for some specific SNP. In order to identify the genomic regions associated with a more reduced level of genetic variation we identify the genes present at the genomic regions with a percentage of individuals with SNP over the 99.9% percentile. The results are presented in Table 3.

Population	Chr.	Start	End	GENES
Entrepelado	3	94599687	97330822	CAMKMT, PREPL, SLC3A1, PPM1B, LRPPRC
	3	103664734	106098985	FEZ2, CRIM1, FAM98A, LTBP1
	14	99022505	101003904	SGMS1, MINPP1, PAPSS2, ATAD1, KLLN, RNLS, LIPK, ANKRD22, STAMBPL1, ACTA2, FAS
Torbiscal	15	91695047	96458175	ITGAV, FAM171B, ZSWIM2, CALCRL, TFPI, GULP1, COL3A1, COL5A2, WDR75, SLC40A1, ANKAR, PMS1, OSGEPL1, MSTN, INPP1, MFSD5, NAB1, GLS, STAT1, STAT4, MYO1B
Retinto	14	63934099	69888948	RHOBTB1, TMEM26, CABCOCO1, ARID5B, RTKN2, ZNF365, EGR2, ADO, NRBF2, JMJD1C, REEP3, CTNNA3
	14	99701595	100011521	ATAD1, KLLN
	15	82043862	83890342	EVX1, HOXD1, HOXD3, HOXD4, HOXD8, HOXD9, HOXD10, HOXD12, HOXD13

Table 3. Genes of each of the regions obtained in each chromosome by the ROH procedure.

As in the previous chapter, the genomic regions over the 95% percentile were also identified. They comprise 1831 genes and the results of enrichment with *GOrilla* are presented in Table 4.

GO term	Description	P-value	FDR q-value
GO:0010043	response to zinc ion	2.67E-4	1E0
GO:0050851	antigen receptor-mediated signaling pathway	3.34E-4	1E0
GO:0046928	regulation of neurotransmitter secretion	3.6E-4	9.5E-1
GO:0016926	protein desumoylation	5.46E-4	1E0
GO:0090230	regulation of centromere complex assembly	5.46E-4	8.64E-1
GO:0090234	regulation of kinetochore assembly	5.46E-4	7.2E-1
GO:0051588	regulation of neurotransmitter transport	5.58E-4	6.3E-1
GO:0032328	alanine transport	8.85E-4	8.75E-1
GO:0015808	L-alanine transport	8.85E-4	7.78E-1
GO:0015824	proline transport	8.85E-4	7E-1
GO:0035524	proline transmembrane transport	8.85E-4	6.36E-1
GO:0002250	adaptive immune response	9.51E-4	6.26E-1

Table 4. Enriched GO terms (p-value < 0.001) for biological processes using the genes obtained from the ROH analysis (see Table 3)

3.3. Extension of the linkage disequilibrium

The third approach was the *nSL* procedure (Ferrer-Admetlla *et al.*, 2014) that was calculated for each population separately with the haplotype phases of the founder individuals. The results for each population are presented in Figures 15, 16 and 17.

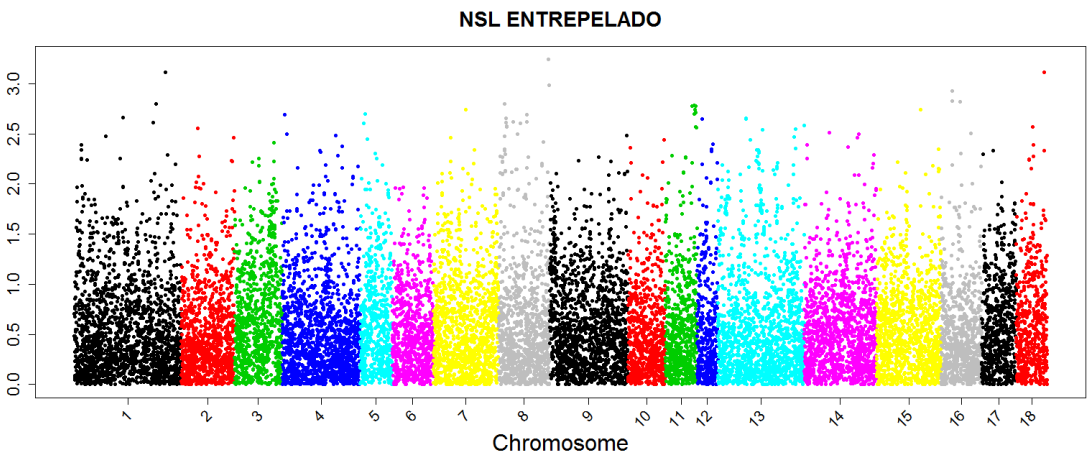

Figure 15. Genomic scan of nSL values for each one of the chromosomes in the Entrepelado population.

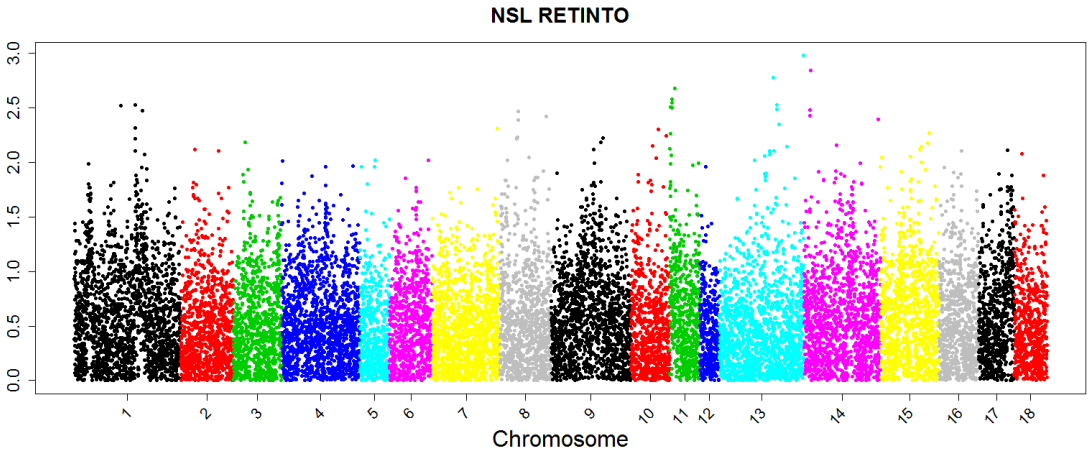

Figure 16. Genomic scan ofnSL values for each one of the chromosomes in the Retinto population.

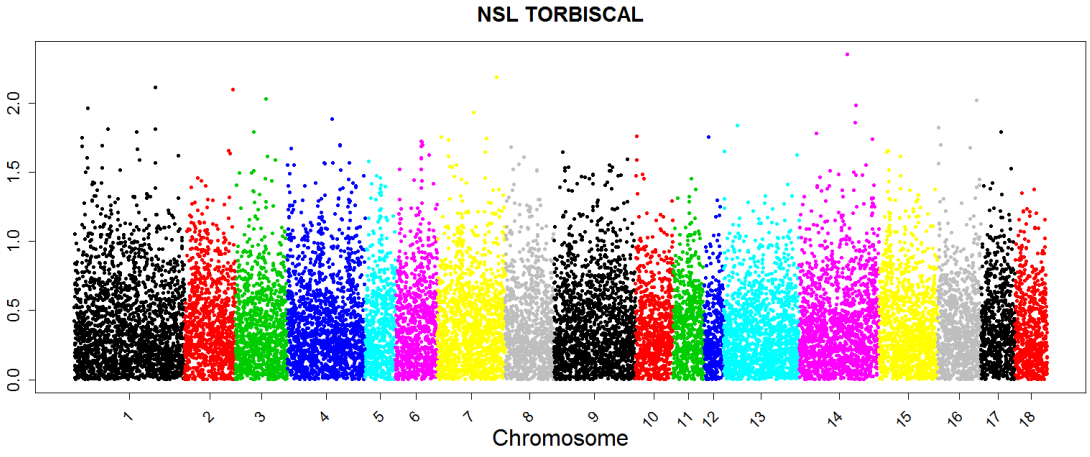

Figure 17.Genomic scan of nSL values for each one of the chromosomes in the Torbiscal population.

As it can be observed the results were noisy and there is not any clear signal at any chromosome and population. Nevertheless, there are some genomic regions with a denser frequency of positive signals, such as in SSC1 for the Retinto population. Thus, the *nSL* estimated were averaged in sliding windows of 20 SNP as it was performed previously for the F_{ST} analysis. Afterwards, the results were clearer and they are presented in Figures 18, 19 and 20.

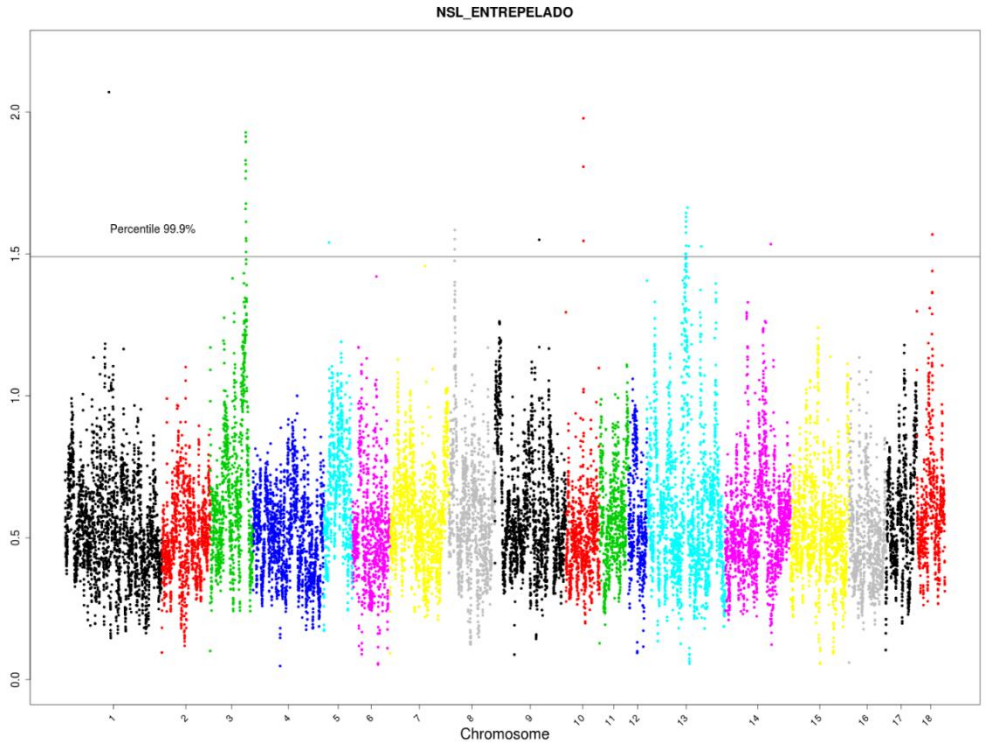

Figure 18. Genomic Scan of nSL values computed in sliding windows of 20 SNP for Entrepelado population

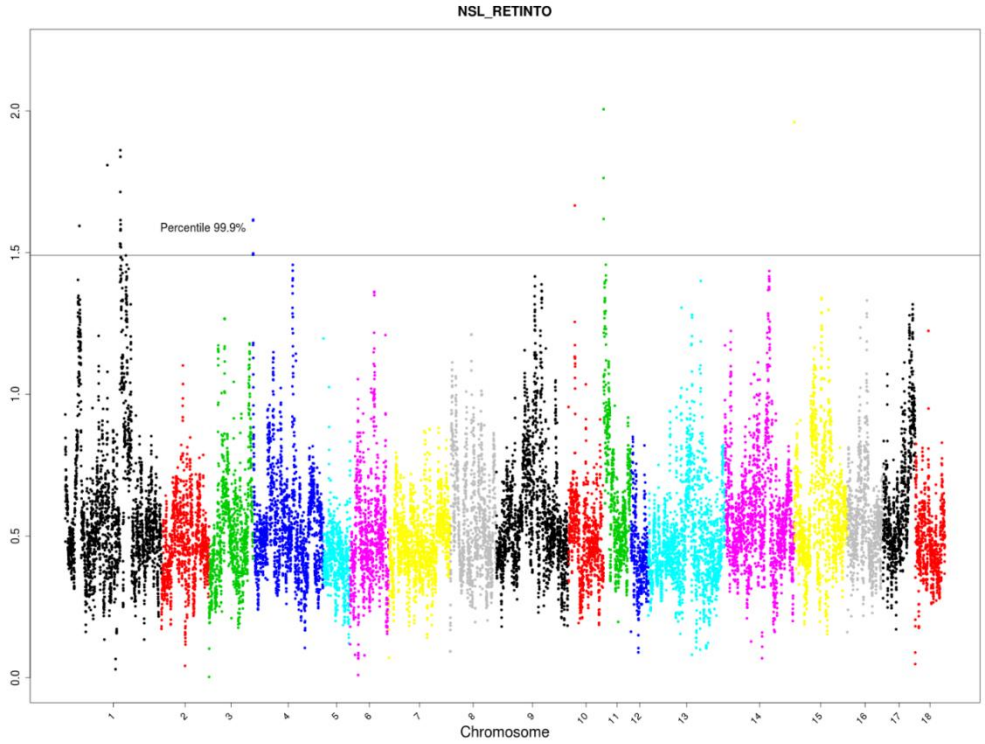

Figure 19. Genomic scan of nSL values computed in sliding windows of 20 SNP for Retinto population

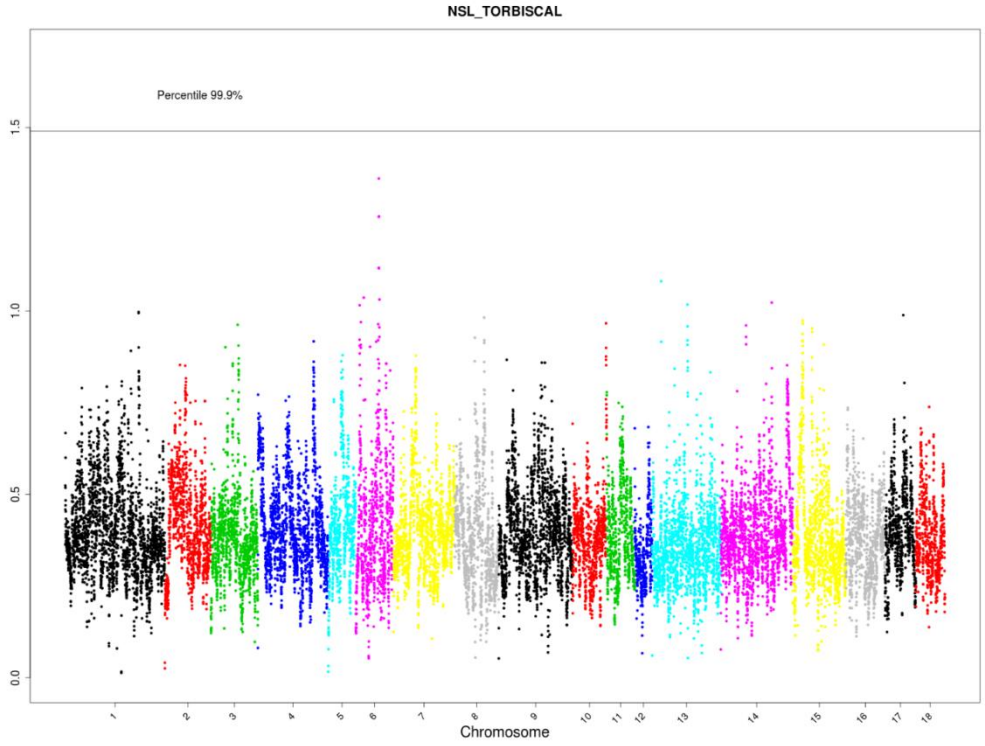

Figure 20. Genomic Scan of nSL values computed in sliding windows of 20 SNP for Torbiscal population

As in the two previous chapters, we identify the genes located in genomic regions over the 99.9% percentile that are presented in Table 5.

Population	Chr	Start	End	GENES
Entrepelado	3	114228362	115145014	ITSN2, SF3B6, PFN4, TP53I3, WDCP, MFSD2B, ATAD2B, UBXN2A, KLHL29
	8	15205723	15277271	PACRGL, KCNIP4
	10	35940169	36161775	-
	13	65976067	66668301	BRPF1, OGG1, CAMK1, ARPC4, TADA3, RPUSD3, CRELD1, GHRL, ATP2B2, FANCD2OS, SEC13, EMC3, JAGN1, PRRT3, CIDEC, FANCD2, IL17RC
Retinto	1	164854328	166608531	IQCH, MAP2K5, SKOR1, PIAS1, CALML4, CLN6, ITGA11, FEM1B, CORO2B
	4	6711580	7029908	ZFAT
	11	5532122	5814466	PAN3, FLT1

Table 5. Genes of each region obtained in each chromosome by the nSL procedure.

Finally, the enrichment analysis with the genes located in the genomic region with signals over the 95% were also performed and the results are present in Table 6.

GO term	Description	P-value	FDR q-value
GO:0043124	negative regulation of I-kappaB kinase/NF-kappaB signaling	5.11E-5	5.12E-1
GO:0043401	steroid hormone mediated signaling pathway	8.66E-5	4.34E-1
GO:0009755	hormone-mediated signaling pathway	2.06E-4	6.88E-1
GO:0060334	regulation of interferon-gamma-mediated signaling pathway	2.42E-4	6.07E-1
GO:0060330	regulation of response to interferon-gamma	2.42E-4	4.86E-1
GO:0046328	regulation of JNK cascade	3.82E-4	6.38E-1
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	4.43E-4	6.33E-1
GO:0060759	regulation of response to cytokine stimulus	4.57E-4	5.72E-1
GO:0071277	cellular response to calcium ion	6.23E-4	6.94E-1
GO:2001242	regulation of intrinsic apoptotic signaling pathway	6.51E-4	6.52E-1
GO:0060442	branching involved in prostate gland morphogenesis	8.41E-4	7.65E-1
GO:0021924	cell proliferation in external granule layer	9.11E-4	7.6E-1
GO:0021930	cerebellar granule cell precursor proliferation	9.11E-4	7.02E-1
GO:0021534	cell proliferation in hindbrain	9.11E-4	6.52E-1

Table 6. Enriched GO terms (p-value < 0.001) for biological processes using the genes obtained from the nSL analysis (see Table 5)

4. DISCUSSION

In this study, three alternative procedures for the detection of signatures of selection have been developed, and their results were summarized with two different approaches. In the first approach, the genes located at the genome regions with a signal over the 99.9% percentile were explicitly identified. Moreover, the second identified the genes located in genomic regions over the 95% percentile and they were used to carry out an enrichment analysis with the GO terms for biological processes. The aim of this second approach was to identify the most affected biological processes by the genes located at the genomic regions with moderate and high signatures of selection with each procedure. The results provided by each method were different as can be observed in Figure 21 and if the list of GO terms presented in Tables 2, 4 and 6 were compared.

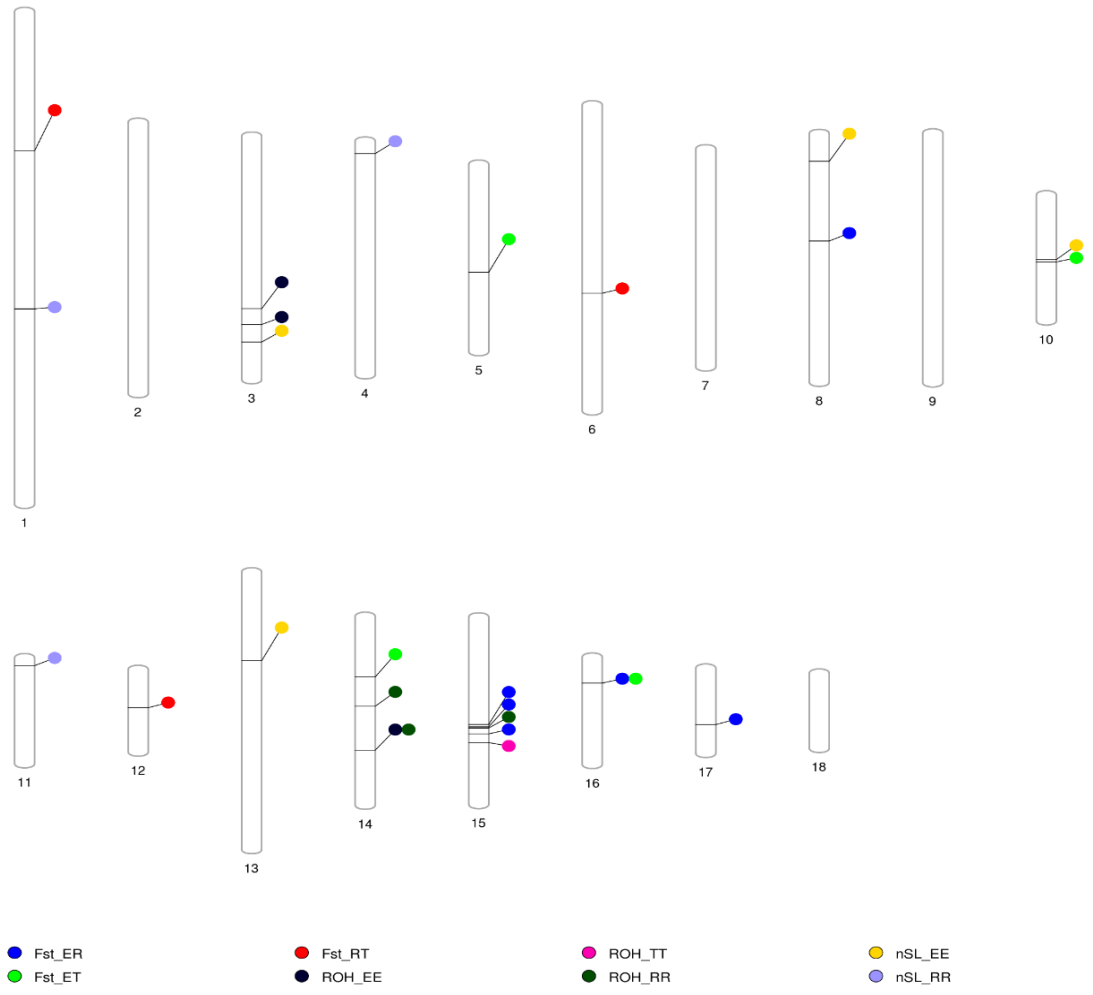

Figure 21. Genomic regions obtained by the F_{ST}, ROH and nSL methods at each porcine chromosome.

The difference of the results between methods was expected as it was also observed by González-Rodríguez *et al.* (2016) using a bovine dataset. These authors argued, following Sabeti *et al.* (2006), that each method or group of provided signatures generated by selection events at different moments of the evolutionary history of the populations. In fact, the methods based on the reduction of the genetic variation, such as *ROH*, detect the signals of very old selection processes. The methods based on the differentiation of the populations reflect the selection or adaptation processes associated with the generation of the populations or breeds, and the procedures based on the extension of linkage disequilibrium, such as the *nSL*, were able to detect very recent or current selection events.

According to the results above presented, the older selection events are those identified by the *ROH* procedure. Therefore, if the enriched GO terms with this procedure were observed, it is possible to detect several groups of pathways. The first is associated with aminoacid transportation (*alanine transport*, *L-alanine transport*, *proline transport*, *proline transmembrane transport-*, the second is related with immune response (*antigen receptor-mediated signaling pathway*, *adaptive immune response*), the third is associated with neurotransmission (*regulation of neurotransmitter secretion, regulation of neurotransmitter transport*), the fourth is related with cellular structure processes (*regulation of centromere complex assembly, regulation of kinetochore assembly*) and, finally, the last one reflected very primary biological processes (*response to zinc ion*and*protein desumoylation).*

These results were reinforced by the identification of some genes within the detected genomic regions. So, in the first group of pathways (*alanine transport*, *L-alanine transport*, *proline transport*, *proline transmembrane transport),* the gene SLC3A1 -*Solute Carrier Family 3 Member 1-* (Nagamori *et al.,* 2016)can be highlighted as involved in the amino acid transportation. The pathways of the second group (*antigen receptor-mediated signaling pathway*, *adaptive immune response*) makes sense becausepig populations havebeen selected or adapted to the presence of several pathogenic threats. An example that supports it is the case of the KLLN -*killin*- gene that intervenes in the process of cellular apoptosis (Cho and Liang, 2008) or the STAT1 -*signal transducer and activator of transcription 1*- that plays an important role in macrophage maturation (Coccia *et al.,* 1999). In addition, there are several studies in the

literature that located QTL -*Quantitative Trait Loci*- related with resistance to diseases within the same genomic regions (Uthe *et al.*, 2011).

The third group of pathways (*regulation of neurotransmitter secretion, regulation of neurotransmitter transport*) is related with the process of neurotransmission. Some interesting genes related with that pathways are the ATAD1 -*ATPase Family, AAA Domain Containing 1*-, whose mutation can lead to the development of encephalopathies (Piard *et al.*, 2018), the NAB1-*signal transducer and activator of transcription 1*- and the EGR2 -*early growth response 2*- that are essential for nervous system myelination (Le *et al.*, 2005; Jones *et al.*, 2007), the EVX1-*even-skipped homeobox 1*- that is involved in the development of the spinal cord (Moran-Rivard *et al.*, 2001), the CRIM1 -*cysteine rich transmembrane BMP regulator 1* - associated with the development of the central nervous system (Kolle *et al.*, 2000) and the MYO1B -*myosin IB* - that encode a motor protein with a relevant function in the neuronal development (Iuliano *et al.*,2018).The presence of this kind of pathways can be related with very old processes of selection associated with the pig domestication, understood as a very strong selection event.

The fourth group is related with basic cellular structural processes (*regulation of centromere complex assembly, regulation of kinetochore assembly*) among the genes identifies, it is worth noting that the gen FEZ2 -*fasciculation and elongation protein zeta 2*- is necessary for elongation within axon bundles (Bloom and Horvitz, 1997) and the gen ITGAV -*Integrin Subunit Alpha V*-encodes membrane proteins that function in cell surface adhesion and signaling (Humphries *et al.*, 2006).Nevertheless, we were not able to identify any gene clearly related with the last GO terms (*response to zinc ion* and *protein desumoylation*).

In a more recent evolutionary time, the divergence between the studied populations had left signatures of selection that can be identified by the F_{ST} approach. First, if we look at the GO biological processes four groups of processes can be identified. They were related with morphological development (*skeletal system development, embryonic morphogenesis, neuron migration* and *embryonic digit morphogenesis),* lipid metabolism *(lipoxygenase pathway, hepoxilin metabolic process* and *hepoxilin biosynthetic process),* immune response (*immunoglobulin secretion*) and another related with the Na-K equilibrium (*regulation of sodium ion transport*). The presence of

biological processes related with morphological development is expected due to the mechanism of breed of variety formation itself. The formation of the varieties was developed by the farmers by crossing individuals of similar morphology that, after several generations, constitute differentiated populations. Moreover, the difference between populations in lipid metabolism confirmed the results of Ibáñez-Escriche *et al.*, (2016) that detected relevant differences in the fatty acid composition of the same analyzed Iberian pig varieties.

Moreover, the presence of the first group of pathways (*skeletal system development, embryonic morphogenesis, neuron migration* and *embryonic digit morphogenesis*) in reinforced by the presence of some relevant genes within those genomic regions. For instance, there is a family of HOXD -*Homeobox protein*- genes that encode a family of transcription factor that play a crucial role in morphogenesis (Myers, 2008), and that were also identified with the ROH procedure. In addition,the MN1 -*Trasncriptional activator MN1*- is involved in the development of craniofacial traits (Pallares *et al.*, 2015) or the NEUROD1 -*Neurogenic differentiation 1*- that is transcription factor involved in neurogenesis (Hsieh, 2012) and that have a key role in insulin expression and β cell proliferation (Cerf, 2006). In addition, there some interesting studies in the literature that located QTL within these genomic regions related with morphological traits (Rohrer *et al.*, 2015).

The relevance of the second group of pathways (*lipoxygenase pathway, hepoxilin metabolic process* and *hepoxilin biosynthetic process)* is confirmed by the presence of a large number of QTL related with fatness or meat quality, such as the ones described by Fontanesi *et al.*, (2012) or Zhang *et al.*, (2016) within the same genomic regions. In addition, some interesting genes, such as ADCYAP1 -*Adenylate Cyclase Activating Polypeptide 1*- which is a member of the glucagon superfamily of hormones that have important roles in growth, metabolism and immune response (Sherwood *et al.*, 2000) or ADIG -*Adipogenin*- that is involved in adipocyte differentiation (Hong *et al.*,2005) also strengthened this hypothesis.

Finally, the pathway associated with immune response (*immunoglobulin secretion*) is confirmed by the presence of the genes TRAF3IP2 -*TRAF3 interacting protein*- plays a central role in innate immunity in response to pathogens (Wu *et al.*, 2013),ITGA4 -*Integrin, alpha 4*- that is also involved in the immunity processes (Dedrick, 2007) and

associated jointly with the CERKL -*ceramide kinase like*- to the Monocyte-Lymphocyte ratio (Lin *et al.,* 2017) and the WIPF1 -*WAS/WASL interacting protein family member 1* - that is involved in T and B lymphocytes activation (Anton *et al.,* 2002).

Regarding the *nSL* procedure, the most recent selection processes can be identified. It is important to mention that we were able to identify relevant signals at the Entrepelado and Retinto populations, whereas there is not highly relevant ones in the Torbiscal population. These results are coherent with the current genetic status of the Torbiscal population as the less developed Iberian strain with a reduction of census during the last years.

In addition, the enrichment analysis with the *nSL* results identified several groups of pathways related to biological processes, such as hormonal and metabolic regulation (*steroid hormone mediated signaling pathway* and *hormone-mediated signaling pathway,negative regulation of I-kappaB kinase/NF-kappaB signaling, regulation of interferon-gamma-mediated signaling pathway, regulation of response to interferon-gamma, regulation of JNK cascade, regulation of response to cytokine stimulus* and *regulation of intrinsic apoptotic signaling pathway*), cell proliferation (*cell proliferation in external granule layer, cerebellar granule cell precursor proliferation* and *cell proliferation in hindbrain*)and cellular response (*homophilic cell adhesion via plasma membrane adhesion molecules, cellular response to calcium ion* and *branching involved in prostate gland morphogenesis*).

In the first group of hormonal and metabolic regulation, the GHRL -*appetite-regulating hormone precursor*-or *Ghrelin* gene is outstanding for its relevance in the regulation of appetite and consequently with its relationship in fat accumulation, of particular relevance in Iberian pigs due to the seasonality of acorns, the main source of their feeding (Fontanesi *et al.,* 2012).In addition, the CIDEC - *cell death activator CIDE-3*- is also involved in the regulation of lipid deposition (Li *et al.,* 2018) and several QTL related with fatness has been also detected in the selected genomic regions (Fan *et al.,* 2009).

In the second group of cell proliferation, among the genes located in these genomic regions, some of them are clearly associated with this kind or processes, such as the FLT1 -*fms related tyrosine kinase 1* - gene that regulates cell migration (Lee *et al.,* 2011), the OGG1 - *8-oxoguanine DNA glycosylase*- that is associated with reparation of

the DNA (Rosenquist *et al.,* 1997), the PAN3 - Poly(A)-Nuclease Deadenylation Complex Subunit 3 - related with RNA processing (Wolf *et al.,* 2014) and the SKOR1 - *SKI family transcriptional corepressor 1* - that is a transcriptional co-repressor (Mizuhara *et al.,* 2005). In addition, the ITGA11 -*integrin subunit alpha 11*- plays a role in the organization of the extracellular matrix (Velling *et al.,* 1999).

Finally, in the last group related to cellular response, several genes can be highlighted. The IQCH -*IQ motif containing H*- and CALML4 -*calmodulin-like 4*-can be mentioned due to their relation with the cellular response to calcium ion through binding calmodulin (Bähler *et al.,* 2002). Other genes related to the cellular response to calcium are CLN6 -*ceroid-lipofuscinosis neuronal protein 6* - and CAMK1- *calcium/calmodulin-dependent protein kinase type 1*-. Moreover, CRELD1 -*cysteine rich with EGF like domains 1* – gen acts as a regulator of calcineurin/NFATc1 signaling (Mass *et al.,* 2014). Finally, the KCNIP4 -*potassium voltage-gated channel interacting protein 4*- is also involved in the calcium-potassium equilibrium (Morohasi *et al.,* 2002).

5. CONCLUSIONS

- Each of the three methods provides different signatures of selection, indicating different evolutionary times. The F_{ST}, *ROH* and *nSL* values indicate differentiation between populations, very old selection processes and recent selection processes, respectively.

- The difference of F_{ST} values respect to the three populations is small, so no clear differences are observed between them. Although there are some regions throughout the genome with more divergence than others.

- The differentiation between populations is mainly focused on the selection by physical features related to their morphology development and lipid metabolism.

- The old selection processes will focus on very primary processes such as aminoacids transportation or structural processes, as well as on resistance to diseases (immune response).

- The absence of recent signals for the Torbiscal population suggest absence or small recent selection with respect to the other populations.

6. BIBLIOGRAPHY

Anton, I.M., de la Fuente, M.A., Sims, T.N., Freeman, S., Ramesh, N., Hartwig, J.H., Dustin, M.L., Geha, R.S. (2002). WIP deficiency reveals a differential role for WIP and the actin cytoskeleton in T and B cell activation. *Inmmunity*, **16**:193-204.

Bähler, M. and Rhoads, A. (2002). Calmodulin signaling via the IQ motif. *FEBS Letters*, **513**:107-113.

Benítez, R., Fernández, A., Isabel, B., Núñez, Y., De Mercado, E., Gómez-Izquierdo, E., García-Casco, J., López-Bpte, C., Óvilo, C. (2018). Modulatory Effects of Breed, Feeding Status, and Diet on Adipogenic, Lipogenic, and Lipolytic Gene Expression in Growing Iberian and Duroc Pigs. *Int J Mol Sci*, **19**(1):22.

Biscarini, F., Cozzi, P., Gaspa, G., Marras, G. (2018). Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes.

Bloom, L. and Horvitz, R. (1997). The *Caenorhabditis elegans* gene *unc-76* and its human homologs define a new gene family involved in axonal outgrowth and fasciculation.*Proc. Natl, Acad. Sci*., **94**:3414-3419.

Boletín Oficial del Estado (BOE) (2007). *ORDEN APA/3376/2007, de 12 de noviembre, por la que se aprueba el Reglamento del Libro Genealógico de la Raza Porcina Ibérica* https://www.boe.es/boe/dias/2007/11/22/pdfs/A47908-47911.pdf

Ceballos, F.C., Joshi, P.K., Clark, D.W, Ramsay, M., Wilson, J.F., (2018). Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet.*, **19**(4):220-234.

Cerf, M.E. (2006). Transcription factors regulating β-cell function. *European Journal of Endocrinology*, **155**:671-679.

Cho, Y.J., and Liang, P. (2008). Killin is a p53-regulated nuclear inhibitor of DNA synthesis. *Proc Natl Acad Sci U S A*, **105**(14):5396-401.

Coccia, E.M., Del Russo, N., Stellacci, E., Testa, U., Marziali, G., Battistini, A. (1999). STAT1 activation during monocyte to macrophage maturation: role of adhesion molecules. *Int Immunol*, **11**:1075-1083.

Coordinadora de Organizaciones de Agricultores y Ganaderos, 2016

Dedrick, R.L. 2007. Understanding gene expression patterns in Inmune-Mediated Disorders. *Journal of Inmunotoxicology*, **4**:201-207.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z. (2009). GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists, *BMC Bioinformatics*, **10**:48.

Fabuel, E.C., Barragan, L., Silio, M.C., Rodriguez, M., Toro, M.A. (2004) Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity*, **93**:104-113.

Fan, B., Onteru, S..K., Nikkilä, M.T., Stalder, K.J, Rothschild, M.F. (2009). Identification of genetic markers associated with fatness and leg weakness traits in the pig, *Animal genetics*, **40**(6):967-70.

Fay, J.C. and Wu, C.I.(2000). Hitchhiking under positive Darwinian selection. *Genetics*, **155**:1405-1413.

Ferrer-Admetlla, A., Liang, M., Korneliussen T., Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*, **31**(5):1275-1291.

Flicek,P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., García, C., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kähäri, A., Keenan, S., Kulesha, E., Martin, F.J., Maurel, T., McLaren, S.M., Murphy, D.N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H.S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S.J., Vullo, A., Wilder, S.P., Wilson, M., Zadissa, A., Aken, B.L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D.R., Searle, S. Ensembl2014**.***Nucleic Acids Research*, **42** Database issue:D749-D755

Foll, M. and Gaggiotti, O.E. (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**:977-993

Fontanesi, J., Galimberti, G., Calò, D.G., ronza, R., Martelli, P.L., Scotti, E., Colombo, M., Schiavo, G., Casadio, R., Buttazzoni, L., Russo, V. (2012) Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian Large White pigs using a selective genotyping approach. *J Anim Sci.*, **90**(8):2450-64.

Fu, Y.X. and Li, W.H. (1993). Statistical tests of neutrality of mutations. *Genetics*, **133**:693–709.

González-Rodríguez, A., Munilla, S., Mouresan, E., Cañas-Álvarez, J., Díaz, C., Piedrafita, J, Altarriba, J., Baro, J., Molina, A., Varona, L. (2016). On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. *Genetics Selection Evolution,* **48** (1):81.

Hong, Y.H., Hishikawa, D., Miyahara, H.(2005). Up-regulation of adipogenin, an adipocyte plasma transmembrane protein, during adipogenesis. *Mol Cell Biochem*., **276**:133-141.

Hsieh, J. (2012). Orchestrating transcriptional control of adult neurogenesis. *Genes and Development.*, **26**:1010-1021.

Humphries, J.D., Byron, A., Humphries, M.J. (2006). Integrin ligand at a glance. *Journal of Cell Science*, **119**:3901-3903.

Ibáñez-Escriche, N., Magallón, E., Gonzalez, E., Tejeda, J.F., Noguera, J.L. (2016). Genetic parameters and crossbreeding effects of fat deposition and fatty acid profiles in Iberian pig lines. *Journal of animal science*, **94**(1), 28-37.

Iuliano, O., Yoshimura, A.,Prospéri, M.T., Martin, R., Knölker, H.J., Coudrier, E. (2018). Myosin 1b promotes axon formation by regulating actin wave propagation and growth cone dynamics.*Journal of Cell Biology*, **217**(6):2033-2046.

Jones, E.A., Jang, S.W., Mager, G.M., Chang, L-W., Srinivasan, R., Gokey, N.G., Ward, R.M., Nagarajan, R., Svaren, J. (2007). Interactions of Sox10 and Egr2 in myelin gene regulation. *Neuron Glia Biol*, **3**:377–387.

Kent, E.H and Bruce S.W.(2009). Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Genetics*, **10**:639-650.

Kimura, M., (1983). The Neutral Theory of Molecular Evolution, Cambridge University Press, ReinoUnido.

Kimura, M., (1989). The neutral theory of molecular evolution and the world view of the neutralists. *Japanese Journal of Genetics*, **64**(4):315-34.

Kolle, G., Georgas, K., Holmes, G.P., Little, M.H., Yamada, T. (2000). *CRIM1***,** a novel gene encoding a cysteine-rich repeat protein, is developmentally regulated and implicated in vertebrate CNS development and organogénesis.*Mechanisms of Development.*, **90**:181-193.

Laval, G.N., Iannuccelli, C., Legault, D., Milan, M.A.M., Groenen, E., Giuffra, L., Andersson, P.H., Nissen, C.B., Jargensen, P., Beeckmann, H., Geldermann, J.L., Foulley, C., Chevalet, Ollivier, L. (2000) Genetic diversity of eleven European pig breeds. *Genetic Selection and Evolution*, **32:**187-203.

Le, N., Nagarajan, R., Wang, J.Y., Svaren, J., LaPash, C., Araki, T., Schmidt, R.E., Milbrandt, J. (2005). Nab proteins are essential for peripheral nervous system myelination. *Nature neuroscience*, **8**:932-940.

Lee, H.K., Chauchan, S.K., Kay, E., Reza, D. (2011). Flt-1 regulates vascular endothelial cell migration via a protein tyrosine kinase-7–dependent pathway. *Blood*, **117**(21):5762-5771.

Li, Y., Kang, H., Chu, Y., Jin, Y., Zhang, L., Yang, R., Zhang, Z., Zhao, S., Zhou, L. (2018). Cidec differentially regulates lipid deposition and secretion through two tissue-specific isoforms. *Gene*, **641**:265-271.

Lin, B.D., Willemsen, G., Fedko, I.O., Jansen, R., Penninx, B., de Geus, E., Kluft, C., Hottenga, J., Boomsma, D.I. (2017). *Twin research and human genetics*, **20**:97-107

Martínez, A.M., Delgado, J.V., Rodero, A., Vega-Pla, J.L. (2000) Genetic structure of the Iberian pig breed using microsatellites. *Animal Genetics*, **31**:295-301.

Mass, E., Wachten, D., Aschenbrenner, A.C., Voelzmann, A., Hoch, M. (2014). Murine Creld1 controls cardiac development through activation of calcineurin/NFATc1 signaling. *Dev. Cell*., **28**(6):711-26.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M.G., Wright, A.F., Campbell, H., Wilson, J.F. (2008). Runs of Homozygosity in European Populations. *American Journal of Human Genetics*, **83**:359-372.

Ministerio de Agricultura y Pesca, Alimentación y Medio (2017). "INFORME TRIMESTRAL INDICADORES DE PORCINO" Subdirección General de Productos Ganaderos, Dirección General de Producciones y Mercados Agrarios.

Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente, 2015. Caracterización del sector porcino español.

Mizuhara, E., Nakatani, T., Minaki, Y., Sakamoto, T., Ono, Y. (2005) Corl1, a novel neuronal lineage-specific transcriptional corepressor for the homeodomain transcription factor Lbx1. *J Biol Chem.*, **280**:3645-3655.

Moran-Rivard, L., Kagawa, T., Saueressig, H., Gross, M.K., Burrill, J., Goulding, M.(2001). Evx1 is a postmitotic determinant of v0 interneuron identity in the spinal cord. *Neuron,* **29**:385-399.

Morohasi, Y., Hatano, N., Ohya, S., Takikawa, R., Watabiki, T., Takasugi, N., Imaizumi, Y., Tomita, T., Iwatsubo, T. (2002). Molecular cloning and characterization of CALP/KChIP4, a novel EF-hand protein interacting with presenilin 2 and voltage-gated potassium channel subunit Kv4. *J. Biol. Chem.*, **277**:14965-14975.

Myers, P. Z. (2008). Hox Genes in Development: The Hox Code. *Nature Education*, **1**:2

Nagamori, S., Wiriyasermkul, P., Guarch, M.E., Okuyama, H., Nakagomi, S., Tadagaki, K., Nishinaka, Y., Bodoy, S., Takafuji, K., Okuda, S. (2016). Novel cystine transporter in renal proximal tubule identified as a missing partner of cystinuria-related plasma membrane protein rBAT/SLC3A1. *Proc. Natl. Acad. Sci*., **113:**775-780

Oleksyk, T.K., Smith, M.W. and O'Brien, S.J. (2010). Genome-wide scans for footprints of natural selection. Philosophical transactions of the Royal Society of London. Series B, *Biological sciences*, **365**:185-205.

Oleksyk, T.K., Zhao, K., De la Vega, F.M., Gilbert, D.A., O´Brien, S.J., Smith M.W. (2008). Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations. *PLoS One*, **3**(3):e1712.

Pallares, L. F., Carbonetto, P., Gopalakrishnan, S., Parker, C. S., Ackert-Bicknell et al. 2015. Mapping of craniofacial traits in outbred mice identifies major developmental genes involved in shape determination. *Plos Genet*., **11**: e1005607.

Piard, J.,Essien, G.K., Harms, F.L., Abalde-Atristain, L., Amram, D., Chang, M., Chen, R., Alawi, M., Salpietro, V., Chung, M., Houlden, H., Verloes, A., Dawson, T.M., Dawson, V.L., Maldergem, L.V., Kutsche, K. (2018) A homozygous ATAD1 mutation impairs postsynaptic AMPA receptor trafficking and causes a lethal encephalopathy. *Brain*, **141**:651-661.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A. R., Bender, D., Maller, J., Sklar, P, de Bakker, P. I. W., Daly, M. J., Sham, P. C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, **81**:559-575.

Qanbari, S. and Simianer, H.(2014). Mapping signatures of positive selection in the genome of livestock. *Livestock Science*, **116**:133–143.

Rohrer, G.A., Nonneman, D.J., Wiedmann, R.T., Schneider, J.F. (2015). A study of vertebra number in pigs confirms the association of vertnin and reveals additional QTL. *BMC genetics*, **16**(1):129.

Rosenquist, T.A., Zharkov, D.O., Grollman, A.P. (1997). Cloning and characterization of a mammalian 8-oxoguanine DNA glycosylase. *PNAS*., **14**:7429-7434.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S.(2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**(6909):832–837.

Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O. (2006). Positive natural selection in human lineage. *Science*, **312**:1614-1620.

Sargolzaei, M., Chesnais, J. P., Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, **15**:478.

Sherwood, N. M., Krueckl, S. L., McRory, J. E. (2000). The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. *Endocr. Rev.*, **21**:619-670.

Szpiech, Z.A. and Hernandez, R.D. (2014). Selscan: an efficient multi-threaded program to calculate EHH-based scans for positive selection. *Molecular Biology and Evolution*, **31**:2824-2827.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**(3):585–595.

Uthe, J.J., Bearson, S.M., Qu, L., Dekkers, J.C., Nettleton, D., Rodriguez, Y., O'Connor, A.M., McKean, J.D., Tuggle, C.K. (2011). Integrating comparative expression profiling data and association of SNPs with Salmonella shedding for improved food safety and porcine disease resistance". *Anim Genet*, **42**(5):521-34.

Utsunomiya, Y.T., Pérez-O'Brien, A.M., Sonstegard, T.S., Sölkner, J., Garcia, J.F. (2015). Genomic data as the "hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome. *Frontiers in Genetics*, **6**:36.

Velling, T., Kusche-Gullberg, M., Sejersen, T., Gullberg, D. (1999) cDNA Cloning and Chromosomal Localization of Human α_{11} Integrin**.** A COLLAGEN-BINDING, I DOMAIN-CONTAINING, β_1-ASSOCIATED INTEGRIN α-CHAIN PRESENT IN MUSCLE TISSUES. *J. Biol. Chem.*, **274**:25735-25742.

Vitalis, R., Gautier, M., Dawson, K.J., Beaumont, M.A. (2014) Detecting and measuring selection from gene frequency data. *Genetics*, **196**: 799-817

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K.(2006). A map of recent positive selection in the human genome. *PLOS Biology*, **4**(3):0446-0458.

Walsh, B. and Lynch, M. (2014). Volume 2: Evolution and selection of quantitative traits. http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html

Wolf, J., Valkov, E., Allen, M.D., Meineke, B., Gordiyenko, Y., McLaughlin, S.H., Olsen, T.M., Robinson, C.V., Bycroft, M., Stewart, M., Passmore, L.A. (2014) Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. *EMBO J.*, **33**:1514-1526.

Wright, S. (1943). Isolation by Distance. *Genetics*, **28**(2):114-138.

Wu, B., Gong, J., Yuan, S., Zhang, Y., Wei, T. (2013). Patterns of evolutionary selection pressure in the immune signaling protein TRAF3IP2 in mammals. *Gene,* **531**:403-410.

Zhang, W., Yang, B., Zhang, J., Cui, L., Ma, J., Chen, C., Ai, H., Xiao, S., Ren, J., Huang, L. (2016). Genome-wide association studies for fatty acid metabolic traits in five divergent pig populations, *Scientific reports*, **6**: 24718.

ANNEX 1 – NAMES OF GENES

SYMBOL	NAME
ACTA2	Actin, aortic smooth muscle
ACTR5	ARP5 actin related protein 5 homolog
ADAMTS12	ADAM metallopeptidase with thrombospondin
ADCYAP1	pituitary adenylate cyclase-activating polypeptide precursor
ADIG	Adipogenin
ADO	2-aminoethanethiol dioxygenase
ANKAR	ankyrin and armadillo repeat containing
ANKRD22	ankyrin repeat domain 22
ARID5B	AT-rich interaction domain 5B
ARPC4	actin-related protein 2/3 complex subunit 4
ATAD1	ATPase family, AAA domain containing 1
ATAD2B	ATPase family, AAA domain containing 2B
ATP2B2	ATPase plasma membrane Ca2+ transporting 2
BRPF1	bromodomain and PHD finger containing 1
CABCOCO1	ciliary associated calcium binding coiled-coil 1
CALCRL	calcitonin gene-related peptide type 1 receptor precurso
CALML4	calmodulin-like 4
CAMK1	calcium/calmodulin-dependent protein kinase type 1
CAMKMT	calmodulin-lysine N-methyltransferase
CERKL	ceramide kinase like
CIDEC	cell death activator CIDE-3
CIR1	corepressor interacting with RBPJ 1
CLN6	ceroid-lipofuscinosis neuronal protein 6
COL3A1	collagen alpha-1(III) chain precursor
COL5A2	collagen alpha-2(V) chain precursor
CORO2B	coronin 2B
CRELD1	cysteine rich with EGF like domains 1
CRIM1	cysteine rich transmembrane BMP regulator 1
CTNNA3	catenin alpha 3
DHX35	DEAH-box helicase 35
EGR2	early growth response 2
EMC3	ER membrane protein complex subunit 3
ETV6	ETS variant 6
EVX1	even-skipped homeobox 1
FAM171B	family with sequence similarity 171 member B
FAM83D	family with sequence similarity 83 member D
FAM98A	family with sequence similarity 98 member A
FANCD2	Fanconi anemia complementation group D2
FANCD2OS	FANCD2 opposite strand
FAS	tumor necrosis factor receptor superfamily member 6 precursor
FEM1B	fem-1 homolog B

FEZ2	fasciculation and elongation protein zeta 2
FLT1	fms related tyrosine kinase 1
FYN	Tyrosine-protein kinase Fyn
GHRL	appetite-regulating hormone precursor
GLS	glutaminase
GPR155	G protein-coupled receptor 155
GULP1	GULP, engulfment adaptor PTB domain containing 1
HOXD1	homeobox D1
HOXD10	homeobox D10
HOXD12	homeobox D12
HOXD13	homeobox D13
HOXD3	homeobox D3
HOXD4	homeobox D4
HOXD8	homeobox D8
HOXD9	homeobox D9
IL17RC	interleukin 17 receptor C
INPP1	inositol polyphosphate 1-phosphatase
IQCH	IQ motif containing H
ITGA11	integrin subunit alpha 11
ITGA4	integrin subunit alpha 4
ITGAV	integrin alpha-V precursor
ITSN2	intersectin 2
JAGN1	jagunal homolog 1
JMJD1C	jumonji domain containing 1C
KCNIP4	potassium voltage-gated channel interacting protein 4
KLHL29	kelch like family member 29
KLLN	killin, p53 regulated DNA replication inhibitor
LIPK	lipase family member K
LRP6	LDL receptor related protein 6
LRPPRC	leucine rich pentatricopeptide repeat containing
LTBP1	latent transforming growth factor beta binding protein 1
MAP2K5	mitogen-activated protein kinase kinase 5
MFSD2B	major facilitator superfamily domain containing 2B
MFSD5	major facilitator superfamily domain containing 5
MINPP1	multiple inositol-polyphosphate phosphatase 1
MN1	MN1 proto-oncogene, transcriptional regulator
MSTN	Growth/differentiation factor 8
MYO1B	myosin IB
NAB1	NGFI-A binding protein 1
NEUROD1	neuronal differentiation 1
NRBF2	nuclear receptor binding factor 2
OGG1	8-oxoguanine DNA glycosylase
OSGEPL1	O-sialoglycoprotein endopeptidase like 1

PACRGL	parkin coregulated like
PAN3	Poly(A)-Nuclease Deadenylation Complex Subunit 3
PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2
PFN4	profilin family member 4
PIAS1	protein inhibitor of activated STAT 1
PITPNB	Phosphatidylinositol Transfer Protein Beta
PMS1	PMS1 homolog 1, mismatch repair system component
PPM1B	protein phosphatase, Mg2+/Mn2+ dependent 1B
PPP1R16B	protein phosphatase 1 regulatory subunit 16B
PREPL	prolyl endopeptidase like
PRRT3	proline rich transmembrane protein 3
REEP3	receptor accessory protein 3
REV3L	REV3 like, DNA directed polymerase zeta catalytic subunit
RHOBTB1	Rho related BTB domain containing 1
RNLS	renalase
RPUSD3	RNA pseudouridylate synthase domain containing 3
RTKN2	rhotekin 2
SCRN3	secernin 3
SEC13	SEC13 homolog, nuclear pore and COPII coat complex component
SF3B6	splicing factor 3b subunit 6
SGMS1	phosphatidylcholine:ceramide cholinephosphotransferase 1
SKOR1	SKI family transcriptional corepressor 1
SLC32A1	solute carrier family 32 member 1
SLC3A1	Solute carrier family 3 member 1
SLC40A1	solute carrier family 40 member 1
STAMBPL1	AMSH-like protease
STAT1	signal transducer and activator of transcription 1
STAT4	signal transducer and activator of transcription 4
TADA3	transcriptional adapter 3
TFPI	tissue factor pathway inhibitor precursor
TMEM26	transmembrane protein 26
TP53I3	tumor protein p53 inducible protein 3
TRAF3IP2	TRAF3 interacting protein 2
TARS	threonyl-tRNA synthetase
TTC28	tetratricopeptide repeat domain 28
UBXN2A	UBX domain protein 2A
WDCP	WD repeat and coiled coil containing
WDR75	WD repeat domain 75
WIPF1	WAS/WASL interacting protein family member 1
YES1	YES proto-oncogene 1, Src family tyrosine kinase
ZFAT	zinc finger and AT-hook domain containing
ZNF365	zinc finger protein 365
ZSWIM2	zinc finger SWIM-type containing 2

ANNEX 2 - ABBREVIATIONS INDEX

A: Adenine

C: Cytosine

Chr: Chromosome

DNA: Deoxyribonucleic acid

EH: Expected Heterozygosity

EHH: Extended Haplotype Homozygosity

FDR: False Discovery Rate

F_{ST}: Fixation Index

G: Guanine

GO: Gene Ontology

iHS: Integrated Haplotype Score

MAF: Minor Allele Frequency

MAPAMA: Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente

nSL: Number of Segregating sites by Length

OH: Observed Heterozygosity

QTL: Quantitative Trait Locus

ROH: Runs of Homozygosity

SNP: Single Nucleotide Polymorphism

SSC: Susscrofa chromosome.

T: Thymine