



RESEARCH ARTICLE

OPEN ACCESS

Genomic differentiation among varieties of Iberian pig

Inés Alonso (Alonso, I)¹, Noelia Ibáñez-Escriche (Ibáñez-Escriche, N)², José L. Noguera (Noguera, JL)³, Joaquim Casellas (Casellas, J)⁴, Melani Martín de Hijas-Villalba (Martín de Hijas-Villalba, M)⁴, María J. Gracia-Santana (Gracia-Santana, MJ)⁵ and Luis Varona (Varona, L)¹

¹ Universidad de Zaragoza, Instituto Agroalimentario de Aragón (IA2), 50013 Zaragoza, Spain. ² Universitat Politècnica de València, Instituto Universitario de Ciencia y Tecnología Animal, 46022 Valencia, Spain. ³ Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Genètica i Millora Animal, Av. Alcalde Rovira Roure 191, 25198 Lleida, Spain. ⁴ Universitat Autònoma de Barcelona, Dept Ciència Animal i dels Aliments, 08193 Bellaterra (Barcelona), Spain. ⁵ Programa de Mejora Genética "CASTUA". INGA FOOD S.A. (Nutreco group), Av. de a Rúa, 2, 06200 Almedralejo (Badajoz), Spain.

Abstract

Aim of study: The objective of this study was to identify the autosomal genomic regions associated with genetic differentiation between three commercial strains of Iberian pig.

Area of study: Extremadura (Spain).

Material and methods: We used the Porcine v2 BeadChip to genotype 349 individuals from three varieties of Iberian pig (EE, Entrepelado; RR, Retinto; and TT, Torbiscal) and their crosses. After standard filtering of the Single Nucleotide Polymorphism (SNP) markers, 47, 67, and 123 haplotypic phases from EE, RR, and TT origins were identified. The allelic frequencies of 31,180 SNP markers were used to calculate the fixation index (F_{ST}) that were averaged in sliding windows of 2Mb.

Main results: The results confirmed the greater genetic closeness of the EE and RR varieties, and we were able to identify several genomic regions with a divergence greater than expected. The genes present in those genomic regions were used to perform an Overrepresentation Enrichment Analysis (ORA) for the Gene Ontology (GO) terms for biological process. The ORA indicated that several groups of biological processes were overrepresented: a large group involving morphogenesis and development, and others associated with neurogenesis, cellular responses, or metabolic processes. These results were reinforced by the presence of some genes within the genomic regions that had the highest genomic differentiation.

Research highlights: The genomic differentiation among varieties of the Iberian pig is heterogeneous along the genome. The genomic regions with the highest differentiation contain an overrepresentation of genes related with morphogenesis and development, neurogenesis, cellular responses and metabolic processes.

Additional key words: *Sus scrofa*; single nucleotide polymorphism; founder haplotypes; candidate genes; gene ontology.

Abbreviations used: EE (Entrepelado); F_{ST} (Fixation Index); GO (Gene Ontology); MDS (Multidimensional Scaling Analysis); ORA (Overrepresentation Enrichment Analysis); QTL (Quantitative Trait Loci); RR (Retinto) SNP (Single Nucleotide Polymorphism); TT (Torbiscal).

Authors' contributions: Conceived, designed and performed the experiments: NIE, JLN, JC, MMHV, MJGS and LV. Analyzed the data: IA and LV. Wrote the paper: LV. All authors read and approved the final manuscript.

Citation: Alonso, I; Ibáñez-Escriche, N; Noguera, JL; Casellas, J; Martín de Hijas-Villalba, M; Gracia-Santana, MJ; Varona, L (2020). Genomic differentiation among varieties of Iberian pig. Spanish Journal of Agricultural Research, Volume 18, Issue 1, e0401. <https://doi.org/10.5424/sjar/2020181-15411>

Supplementary material (Tables S1, S2 and S3) accompanies the paper on SJAR's website

Received: 05 Jul 2019. **Accepted:** 26 Feb 2020.

Copyright © 2020 INIA. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC-by 4.0) License.

Funding Agencies/Institutions	Project/Grant
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Spain	RTA2012-00054-C02-01
Ministry of Science, Innovation and Universities, Spain	CGL2016-80155-R; IDI-20170304 (CDTI)

Competing interests: The authors have declared that no competing interests exist.

Correspondence should be addressed to Luis Varona: lvarona@unizar.es

Introduction

The Iberian pig is a native breed from the Iberian Peninsula which has high adipogenic capability and meat of excellent quality (Ventanas *et al.*, 2005). It is the largest extant population of the Mediterranean-type pig and, traditionally, its geographical distribution has been limited to the southwest of the Iberian Peninsula. The population structure of the Iberian pig comprises several varieties that have diverged because of genetic drift, selection, and adaptation. Some authors have reported that genetic variability among Iberian varieties is as high as it is among commercial breeds of white pig (Laval *et al.*, 2000; Martínez *et al.*, 2000; Fabuel *et al.*, 2004).

Genomic diversity between Iberian pig varieties is not expected to be homogeneous throughout the genome. In fact, it is plausible that some genomic regions exhibit higher divergence because of selection or adaptation processes (Qanbari & Simianer, 2014). As far as we know, few studies (Herrero-Medrano *et al.*, 2013; Silió *et al.*, 2016) have studied the genomic diversity of the Iberian population using high-density genotypes, and none of them has analyzed the genomic differentiation between the Iberian varieties throughout of the genome. Therefore, the objective of this study was to identify the genomic regions that had the highest degree of differentiation among three of the most widely used Iberian pig varieties (EE, Entrepelado; RR, Retinto; and TT, Torbiscal). In addition, the study identified candidate genes and the biological processes associated with the differentiation between varieties.

Material and methods

Genotypes from 349 individuals with the Porcine-SNP60 v2 BeadChip (Illumina Inc., San Diego, USA) were used. The data set included purebred Entrepelado (EE, n=21 individuals), Retinto (RR, n=50) and Torbiscal (TT, n=78). In addition, there were individuals from all of the reciprocal crosses: Entrepelado × Retinto (ER) and Retinto × Entrepelado (RE, n=25), Entrepelado × Torbiscal (ET) and Torbiscal × Entrepelado (TE, n=37), and Retinto × Torbiscal (RT) and Torbiscal × Retinto (TR, n=138). All analyzed individuals were descendants from 44 sires (12 EE, 14 RR and 18 TT) and 139 dams (24 EE, 42 RR and 73 TT) with an average progeny (\pm standard deviation) of 7.93 ± 8.01 and 2.51 ± 8.01 , respectively. Genotypes were filtered using *PLINK* (Purcell *et al.*, 2007). The criteria of selection included a call rate higher than 95% at the individual and single nucleotide polymorphism (SNP) levels and a minor allelic frequency (MAF) > 0.01 for autosomal SNPs. Subject to those criteria, 31,180 of 61,565 SNP's were included

in the analysis, which covered an autosomal genome of 22.61 Mb that had a density of one SNP marker per 7251.38 bp. In first place, we tried to corroborate the divergence between varieties by using a multidimensional scaling analysis (MDS) with the *cmdscale* () command from R package *stats* (R Core Team, 2019), and a maximum likelihood estimation of individual ancestries using the ADMIXTURE software (Alexander *et al.*, 2009) with the assumption of three populations.

In a second step, we performed an imputation process of missing alleles and the reconstruction of the haplotype phases for the genotyped individuals and their parents with the *FImpute* software (Sargolzaei *et al.*, 2014). First, a haplotype library built from reference individuals (24 sires and 30 dams with more than 4 progeny) was generated using the *save_hap_lib* option of the *FImpute* software. Second, we used this haplotype library to reconstruct the founder haplotypic phases of each Iberian pig variety using ad-hoc software written in FORTRAN90 and we were able to identify 47, 67 and 123 different haplotypes for EE, RR and TT, respectively. Third, these haplotypes were used to calculate allelic frequencies for each population and SNP by counting the alleles and dividing by the number of haplotypes. These allelic frequencies were used to calculate the Weir-Cockerham (Weir & Cockerham, 1984) estimator of the fixation index (F_{STi}) (Wright, 1951) for each (*ith*) SNP among the three populations and for each specific pair of populations, Entrepelado-Retinto ($F_{ST(ER)_i}$), Entrepelado-Torbiscal ($F_{ST(ET)_i}$) and Retinto-Torbiscal ($F_{ST(RT)_i}$). Finally, the single SNP F_{ST} -statistics were averaged in sliding windows of 1, 2 and 3 Mb centered at each SNP.

The genes located within the genomic regions associated with an average F_{ST} greater than the 95th and 99.9th percentiles were identified using the *Biomart* Tool (Smedley *et al.*, 2015) with the *Sus scrofa* 11.1 genome map. The genes within the genomic regions that had an average F_{ST} greater than the 95th percentile were used in an overrepresentation enrichment analysis (ORA) with the gene ontology (GO) terms for biological process for *Homo sapiens* and *Sus scrofa* with the *WEB-based Gene Set Analysis Toolkit* (Wang *et al.*, 2017; www.webgestalt.org) and the complete genome as a reference set.

Results and discussion

Genomic divergence between populations

The results of the MDS and the estimation of individual ancestries with maximum likelihood are presented in Figures 1 and 2, respectively. Both approach-

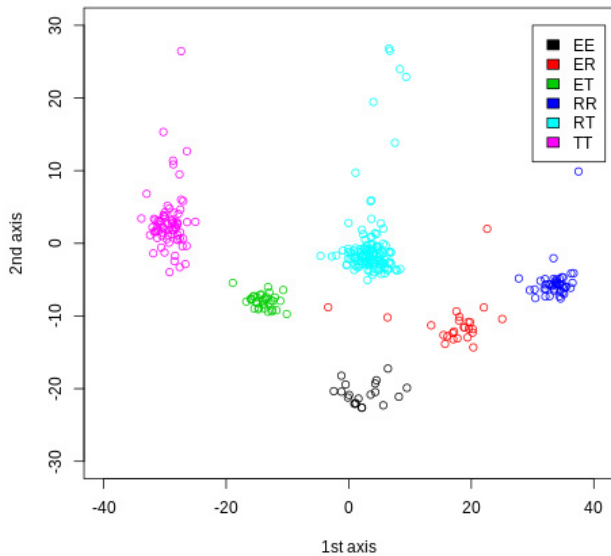


Figure 1. Multidimensional scaling (MDS) plot between individual samples of the EE (Entrepelado), ER (Entrepelado × Retinto), ET (Entrepelado × Torbiscal), RR (Retinto), RT (Retinto × Torbiscal) and TT (Torbiscal) populations.

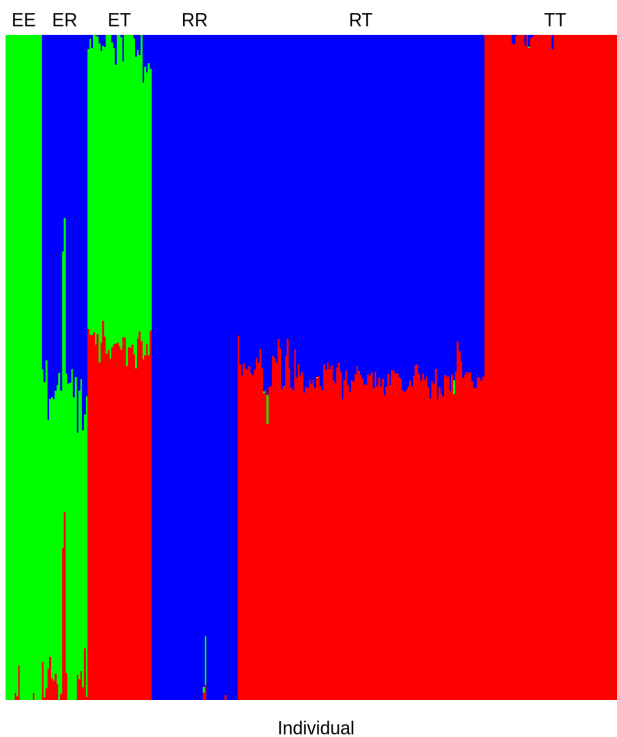


Figure 2. Inferred admixture proportions under the assumption of three populations for the individual samples of the EE (Entrepelado), ER (Entrepelado × Retinto), ET (Entrepelado × Torbiscal), RR (Retinto), RT (Retinto × Torbiscal) and TT (Torbiscal) populations.

es indicate that the genomic data clearly identified the genetic structure of the populations. The representation of the MDS with the first two dimensions distributed the individuals in six groups that correspond to the pure (EE, RR and TT) and crossbred (ER, ET and RT) populations. Graphically, the crossbred were located

between the purebred populations that generated the cross. In addition, the estimation of individual ancestries after the assumption of three populations distributed the purebred individuals into three different ancestries and assigned crossbred individuals with approximately half ancestry from each parental population.

Average \pm standard deviation of the F_{ST} results between the three varieties of Iberian pig was 0.069 ± 0.060 , which is lower than the estimate reported by Fabuel *et al.* (2004), who used a set of 36 microsatellites ($F_{ST}=0.129$). However, they are not directly comparable since microsatellites have a higher mutation rate and, therefore, F_{ST} estimates are not in the same scale. In addition, Fabuel *et al.* (2004) included five varieties (Entrepelado, Retinto, Lampiño, Torbiscal, and Guadyervas), and the Guadyervas population had the greatest genetic distance from the other populations. Therefore, a greater estimate of F_{ST} was expected in their analysis. We did not include Guadyervas because it is almost entirely restricted to conservation programs. The mean \pm standard deviation of F_{ST} statistics between pairs of populations was 0.045 ± 0.055 between EE and RR, 0.049 ± 0.059 between EE and TT, and 0.057 ± 0.072 between RR and TT. The results of paired t-tests were highly significant ($p < 1e-8$) for all comparisons. The results confirmed the closeness between Entrepelado and Retinto, as it was previously reported by Fabuel *et al.* (2004).

The results of the genomic scans through the porcine autosomal genome of the single SNP F_{ST} statistic for all populations and after averaging them in sliding windows of 1, 2, and 3 Mb centered at each SNP are presented in Figure 3. The distribution along the genome of F_{ST} statistics calculated with a single SNP was extremely noisy. Therefore, it was not possible to extract a clear pattern of the genomic differentiation between populations and confirmed the need to average estimates of the F_{ST} in wider genomic regions. The number of markers included in each window was 18.37 ± 6.63 , 33.77 ± 11.36 , and 48.97 ± 15.68 SNP for sliding windows of 1, 2, and 3 Mb, respectively. The results from the genomic scan with sliding windows of 1 Mb were somewhat noisy, and the results based on sliding windows of 2 Mb and 3 Mb were very similar. Therefore, to achieve a compromise between noise reduction and the genomic size of the windows, we decided to explore the results based on 2 Mb sliding windows.

The genomic scans for each pair of populations (EE and RR, EE and TT and RR and TT) with sliding windows of 2 Mb are presented in Figure 4. The distributions of the F_{ST} estimates along the autosomal chromosomes were

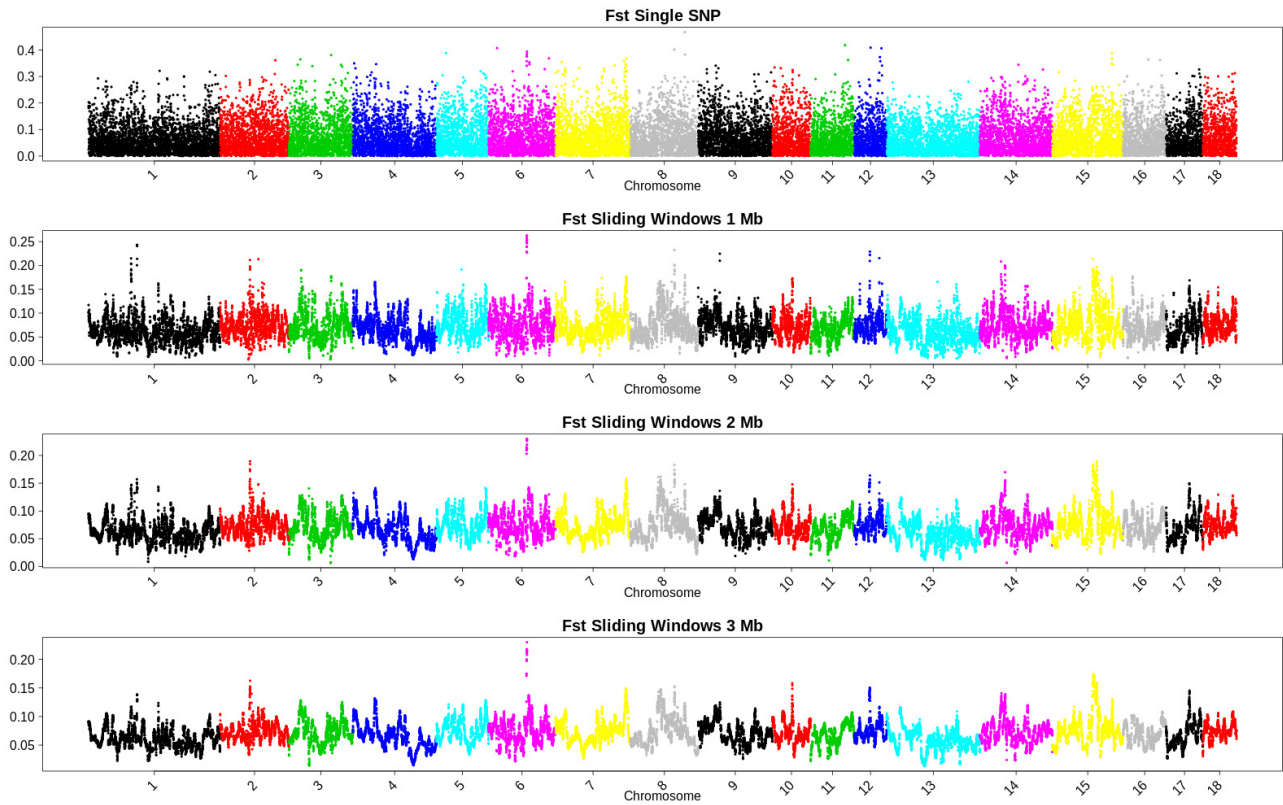


Figure 3. Genomic scan for a single SNP F_{ST} among three varieties of Iberian pig (Entrepelado, Torbiscal and Retinto) and for the average F_{ST} in sliding windows of 1, 2, and 3 Mb, centered at each SNP.

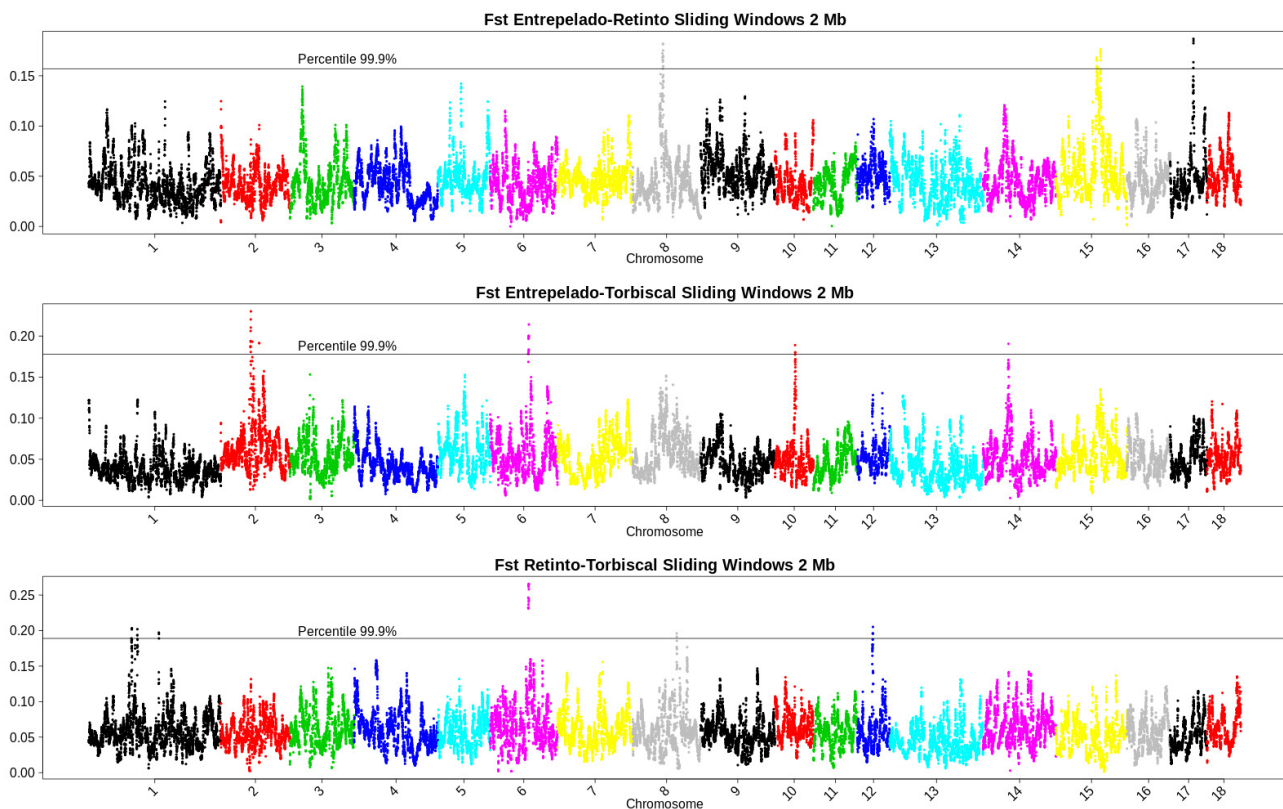


Figure 4. Genomic scan for average F_{ST} in sliding windows of 2 Mb, centered at each SNP, between Entrepelado and Retinto, Entrepelado and Torbiscal, and Retinto and Torbiscal populations of Iberian pig.

clearly different among the three genomic scans. However, we were able to detect some degree of similarity since the correlations between the F_{ST} estimates obtained from EE and RR and EE and TT, EE and RR and RR and TT and EE and TT and RR and TT were 0.218 ($p < 0.001$), 0.214 ($p < 0.001$), 0.317 ($p < 0.001$), respectively.

Biological processes and candidate genes between Entrepelado and Retinto

The genomic regions that had an average $F_{ST(ER)}$ greater than the 95th percentile (0.082) contained 651 genes, which were used in an ORA for the GO for biological process. Among them, 569 and 157 genes were annotated to functional categories in the *Homo sapiens* and *Sus scrofa* databases, respectively. The enriched GO terms that had a p -value < 0.0001 are presented in Table 1. We identified up to 15 and 3 terms within the human and the porcine databases, respectively. All of the GO terms identified with the *Homo sapiens* reference (*anterior/posterior pattern specification*, *skeletal system development*, *limb morphogenesis*, *appendage morphogenesis*, *pattern specification process*, *regionalization*, *embryonic skeletal system development*, *appendage development*, *limb development*, *skeletal system morphogenesis*, *embryo development*, *embryonic limb morphogenesis* and *embryonic appendage morphogenesis*) and two identified with the *Sus scrofa* (*skeletal system development* and *anatomical structure morphogenesis*) were associated with embryogenesis and morphogenesis. In addition, a biological process

associated with regulation of gene expression was significant with the *Sus scrofa* database.

Those results are reinforced by the genes within the genomic regions greater than the 99.9th percentile (0.157) which were located at SSC8 (56983392-60232132 bp), SSC15 (80515676-86990138 bp) and SSC17 (42130987-42480736 bp) as presented in the Table S1 [suppl.]. Among them is a family of *HOXD* (*Homeobox protein*) genes that encode a family of transcription factors that play a crucial role in morphogenesis (Myers, 2008), jointly, with the tightly linked *EVX2* (*Even-Skipped Homeobox 2*) (Hérault *et al.*, 1997), and the *SP9* (*Sp9 Transcription Factor*) (Kawakami *et al.*, 2004). In addition, the *NEUROD1* (*Neurogenic differentiation 1*) is a transcription factor involved in regulatory networks in embryonic stem cells (Marchand *et al.*, 2009), the *OLAI* (*Obg Like ATPase 1*) gene plays a role on the attachment of cells to the extracellular matrix (Jeyabal *et al.*, 2014), and the *ATF2* (*Activating transcription factor-2*) which has been found to affect skeletal growth (Vale-Cruz *et al.*, 2008). Some other interesting genes located within them that are related with morphogenesis are the *CHN1* (*Chimerin 1*) and the *PRKRA* (*protein kinase, interferon inducible double stranded RNA dependent activator*). The *CHN1* is mostly expressed in the brain and it is associated with the early development of the nervous system (Lim *et al.*, 1992) and *PRKRA* has been related with the development of the cerebellum (Yong *et al.*, 2015). An additional evidence of the relationship of those genomic regions with the embryological development is that, in pigs, they have been associated with QTL for teat numbers in SSC8 (Velardo *et al.*, 2016), number of mummies

Table 1. Enriched gene ontology (GO) terms for biological processes ($p < 0.0001$) with genes located within the genomic regions that have an average F_{ST} over the 95th percentile between Entrepelado and Retinto Iberian pig populations based on the *Homo sapiens* and *Sus scrofa* databases.

Homo sapiens	p-value
anterior/posterior pattern specification (<i>Homo sapiens</i>)	1.73e-08
skeletal system development (<i>Homo sapiens</i>)	1.50e-07
limb morphogenesis (<i>Homo sapiens</i>)	1.93e-06
appendage morphogenesis (<i>Homo sapiens</i>)	1.93e-06
pattern specification process (<i>Homo sapiens</i>)	2.52e-06
regionalization(<i>Homo sapiens</i>)	2.53e-06
embryonic skeletal system development (<i>Homo sapiens</i>)	3.10e-06
appendage development (<i>Homo sapiens</i>)	3.76e-06
limb development (<i>Homo sapiens</i>)	3.76e-06
skeletal system morphogenesis (<i>Homo sapiens</i>)	5.83e-06
embryo development (<i>Homo sapiens</i>)	7.30e-06
embryonic limb morphogenesis (<i>Homo sapiens</i>)	2.27e-05
embryonic appendage morphogenesis (<i>Homo sapiens</i>)	2.27e-05
embryonic organ morphogenesis (<i>Homo sapiens</i>)	2.27e-05
embryonic skeletal system morphogenesis (<i>Homo sapiens</i>)	3.87e-05
skeletal system development (<i>Sus scrofa</i>)	5.09e-05
anatomical structure morphogenesis (<i>Sus scrofa</i>)	6.38e-05
regulation of gene expression (<i>Sus scrofa</i>)	6.54e-05

(Onteru *et al.*, 2012) and stillbirths (Schneider *et al.*, 2015) in SSC15.

Moreover, some of those genes such as the *HOXD* family are also associated with the regulation of gene expression, but it is noteworthy that the *ACRT5* (*ARP5 Actin Related Protein 5 Homolog*) is involved in the INO80 complex that contributes to transcription, DNA repair, and DNA replication (Conaway & Conaway, 2009). Therefore, they can be related with the biological process last identified (*regulation of gene expression*) in the *Sus scrofa* database.

Biological processes and candidate genes between Entrepelado and Retinto

The genomic regions that were within the 95th percentile (0.092) of the average $F_{ST}(ET)$ in sliding windows of 2 Mb contained 886 genes, of which 784 and 228 were annotated in the *Homo Sapiens* and *Sus scrofa* databases, respectively. The results of the ORA with these genes are presented in Table 2. One of the enriched GO terms for biological processes was also associated with morphogenesis (*animal organ morphogenesis*) and the other terms were associated with the global metabolism of the individuals through metabolic or catabolic processes (*collagen catabolic process*, *multicellular organismal catabolic process*, and *multicellular organism metabolic process*), regulation of hormone secretion (*negative regulation of hormone secretion*), or the ubiquitination of proteins (*positive regulation of ubiquitin-protein transferase activity*). Furthermore, six genomic regions were detected that had an average $F_{ST}(ET)$ greater than the 99.9th percentile (0.178) and were at SSC2 (60177754-61303696, 71588075-71588075 and 92958395-93175114 bp), SSC6 (104376948-105459825 bp), SSC10 (36887963-37186755 bp), and SSC14 (45509383-45509383 bp). A search of the porcine QTL database (<https://www.animalgenome.org/QTLdb/pig/>) indicated that these regions have been associated with copying behavior

(Ponsuksili *et al.*, 2015) and hemoglobin content (Zhang *et al.*, 2014) in SSC2, intramuscular fat content (Cepica *et al.*, 2012) and vertebra number (Rohrer *et al.*, 2015) in SSC6 and the ratio to non-productive days in SSC14 (Onteru *et al.*, 2011). In addition, some of the genes within or near those genomic regions (see Table S2 [suppl.]) are associated with the same biological processes highlighted above, and are good candidates for had being affected by selection or adaptation; e.g., *INSR* (*insulin receptor*) or *ADCYAP1* (*Adenylate Cyclase Activating Polypeptide 1*). The *INSR* gene has been proposed as a candidate gene for intramuscular fat content by Cepica *et al.* (2012) while *ADCYAP1* is a member of the glucagon superfamily of hormones that are involved in growth, metabolism, and immune response (Sherwood *et al.*, 2000). In addition, *NWD1* (*NACHT and WD Repeat Domain Containing 1*) modulates androgen receptor signaling (Correa *et al.*, 2014), and *MNI* (*Transcriptional activator MNI*) is involved in the development of craniofacial traits (Pallares *et al.*, 2015). Further, another interesting gene is the *CD209* (*dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin*, *DC-SIGN*), that functions as an important pattern recognition receptor (*PRR*) in immune defense and plays a role in the immune modulation during pathogen infection (Soilleux *et al.*, 2002).

Biological processes and candidate genes between Retinto and Torbiscal

The genomic regions that had an average $F_{ST}(RT)$ within the 95th percentile (0.101) in sliding windows of 2 Mb contained 596 genes, of which 530 and 146 were annotated in *Homo sapiens* and *Sus scrofa* databases, respectively. The results of the ORA analyses identified statistically significant ($p < 0.0001$) GO terms when crossed with the *Homo sapiens* database, only (Table 3). There were up to four GO terms associated with development of the neuronal system (*generation*

Table 2. Enriched gene ontology (GO) terms for biological processes ($p < 0.0001$) with genes located within the genomic regions that have an average F_{ST} over the 95th percentile between Entrepelado and Torbiscal Iberian pig populations based on the *Homo sapiens* and *Sus scrofa* databases.

Homo sapiens	p-value
collagen catabolic process (<i>Homo sapiens</i>)	8.69e-06
multicellular organismal catabolic process (<i>Homo sapiens</i>)	1.97e-05
multicellular organism metabolic process (<i>Homo sapiens</i>)	7.94e-05
animal organ morphogenesis (<i>Homo sapiens</i>)	8.69e-05
negative regulation of hormone secretion (<i>Sus scrofa</i>)	7.22e-05
positive regulation of ubiquitin-protein transferase activity (<i>Sus scrofa</i>)	8.81e-05

Table 3. Enriched gene ontology (GO) terms for biological processes ($p < 0.0001$) with genes located within the genomic regions that have an average F_{ST} over the 95th percentile between Retinto and Torbiscal Iberian pig populations based on the *Homo sapiens* and *Sus scrofa* databases.

Homo sapiens	p-value
regulation of multicellular organismal development (<i>Homo sapiens</i>)	1.13e-05
generation of neurons (<i>Homo sapiens</i>)	2.03e-05
Neurogenesis (<i>Homo sapiens</i>)	3.21e-05
regulation of response to stress (<i>Homo sapiens</i>)	3.66e-05
neuron differentiation (<i>Homo sapiens</i>)	4.04e-05
neuron projection development (<i>Homo sapiens</i>)	4.29e-05
regulation of microtubule motor activity (<i>Homo sapiens</i>)	6.20e-05
cellular response to lipopolysaccharide (<i>Homo sapiens</i>)	6.95e-05

of neurons, neurogenesis, neuron differentiation and neuron projection development), and another term was associated with the general development of the organism (regulation of multicellular organismal development). Furthermore, two GO terms were associated with the responses of the organism to stress (regulation of response to stress) and lipopolysaccharides (cellular response to lipopolysaccharide), and another was associated with microtubule activity (regulation of microtubule motor activity). Some of the genes within the genomic regions that had an average F_{ST} that was in the 99.9th percentile (0.189) confirmed these results (see Table S3 [suppl.]). They were located at SSC1 (77056462-77472299, 90748703-90896710, and 141989498-142042139 bp), SSC6 (104376948-105459825 bp), SSC8 (89447995-89679243 bp), and SSC12 (28731262-28831639 bp). The genomic regions of SSC6 was also identified within the most divergent genomic regions between EE and TT, which strengthens the argument that, among other, *INSR* and *ADCYAP1* genes may be good candidates to have been affected by selection or adaptation processes that have influenced the genetic configuration of the TT population. Moreover, other genes of note include *FYN* (*FYN Proto-Oncogene, Src Family Tyrosine Kinase*), which is involved in the early stages of neurogenesis (Yagi *et al.*, 1994), *CDK19* (*Cyclin Dependent Kinase 19*), a regulator of the p53 network for cellular response to stress (Audetat *et al.*, 2017), and *TRAF3IP2* (*TRAF3 interacting protein*), which plays a central role in innate immunity in response to pathogens (Wu *et al.*, 2013). In addition, *COL12A1* (*Collagen, type XII, alpha-1*) is involved in bone formation (Izu *et al.*, 2011) and *CA10* (*Carbonic anhydrase-related protein CA10*) has been identified as an important neurexin ligand (Sterky *et al.*, 2017). Finally, it must be remarked that the genomic regions of SSC1 (77056462-77472299) were associated by Fontanesi *et al.* (2017) with a large QTL for intramuscular fat and that the genomic regions of SSC6 and SSC12 has been related with the genetic variation in vertebra numbers (Rohrer *et al.*, 2015).

Final remarks

The main conclusion of this study is that the processes of differentiation among Iberian pig varieties have heterogeneously affected the autosomal genome. A first approximation has identified potential candidate genes, most of which are associated with morphogenesis, neuronal development, regulation of metabolism, or cellular response to stressors. The presence of those candidate genes is coherent with the recent evolution of the Iberian pig populations, they have evolved through adaptation to harsh environmental conditions. Furthermore, producers have subjected the Iberian pig to “empirical” selection in which adipogenic capacity and morphological traits have played an important role. Nevertheless, further research must be done to confirm these results.

Acknowledgements

The authors gratefully acknowledge INGA FOOD S.A. (Almendralejo, Spain) and its technicians (E. Magallón, J. P. Rosas, L. Muñoz, P. Díaz, D. Iniesta, and M. Ramos) and S. Negro (IRTA), for their cooperation and technical support.

References

- Alexander DH, Novembre J, Lange K, 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19: 1655-1664. <https://doi.org/10.1101/gr.094052.109>
- Audetat KA, Galbraith MD, Odell AT, Lee T, Pandey A, Espinosa JM, Dowell RD, Taatjes D J, 2017. A kinase-independent role for cyclin-dependent kinase 19 in p53 response. *Mol Cell Biol* 37: e00626-16. <https://doi.org/10.1128/MCB.00626-16>
- Cepica S, Ovilo, C, Masopust M, Knoll A, Fernández A, López A, Rohrer GA, Nonneman D, 2012. Four genes located on a SSC2 meat quality QTL region are associated with different meat quality traits in Landrace x Chinese-European

- crossbred population. *Anim Genet* 43: 333-336. <https://doi.org/10.1111/j.1365-2052.2011.02252.x>
- Conaway RC, Conaway JW, 2009. The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem Sci* 34: 71-77. <https://doi.org/10.1016/j.tibs.2008.10.010>
- Correa RG, Krajewska M, Ware CF, Gerlic M, Reed JC, 2014. The NLR-related protein NWD1 is associated with prostate cancer and modulates androgen receptor signaling. *Oncotarget* 30: 1666-1682. <https://doi.org/10.1158/1538-7445.AM2014-LB-74>
- Fabuel EC, Barragán C, Silió L, Rodríguez MC, Toro MA, 2004. Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity* 93: 104-113. <https://doi.org/10.1038/sj.hdy.6800488>
- Fontanesi L, Schiavo G, Galimberti G, Bovo S, Russo V, Gallo M, Buttazzoni L, 2017. A genome-wide association study for a proxy of intermuscular fat level in the Italian Large White breed identifies genomic regions affecting an important quality parameter for dry-cured hams. *Anim Genet* 48: 459-465. <https://doi.org/10.1111/age.12542>
- Hérault Y, Hraba-Renevey S, van der Hoeven F, Duboule D, 1997. Function of the *Evx-2* gene in the morphogenesis of vertebrate limbs. *EMBO J* 15: 6727-6738. <https://doi.org/10.1002/j.1460-2075.1996.tb01062.x>
- Herrero-Medrano JM, Megens HJ, Groenen MAM, Ramis G, Bosse M, Pérez-Enciso M, Crooijmans RPMA, 2013. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC Genet* 14: 106. <https://doi.org/10.1186/1471-2156-14-106>
- Izu Y, Sun M, Zwolanek D, Veit G, Williams V, Cha B, Jepsen KJ, Koch M, Birk DE, 2011. Type XII collagen regulates osteoblast polarity and communication during bone formation. *J Cell Biol* 193: 1115-1130. <https://doi.org/10.1083/jcb.201010010>
- Jeyabal PVS, Rubio V, Chen H, Zhang J, Shi ZZ, 2014. Regulation of cell-matrix adhesion by OLA1, the Olg-like ATPase 1. *Biochem Biophys Res Commun* 444: 568-574. <https://doi.org/10.1016/j.bbrc.2014.01.099>
- Kawakami Y, Rodríguez-Esteban C, Matsui T, Rodríguez-León J, Kato S, Izpisua-Belmonte JC, 2004. Sp8 and Sp9, two closely related buttonhead-like transcription factors, regulate *Fgf8* expression and limb outgrowth in vertebrate embryos. *Development* 131: 4763-4774. <https://doi.org/10.1242/dev.01331>
- Laval G, Iannucelli N, Legault C, Milan D, Groenen MAM, Giuffra E, Andersson L, Nissen PH, Jorgensen CB, Beckmann P *et al.*, 2000. Genetic diversity of eleven European pig breeds. *Genet Sel Evol* 32: 187-203. <https://doi.org/10.1186/1297-9686-32-2-187>
- Lim HH, Michael GJ, Smith P, Lim L, Hall C, 1992. Developmental regulation and neuronal expression of the mRNA of rat n-chimaerin, a p21rac GAP:cDNA sequence. *Biochem J* 287: 415-422. <https://doi.org/10.1042/bj2870415>
- Marchand M, Schroeder IS, Markossian S, Skoudy A, Nègre D, Cosset FL, Real P, Kaiser C, Wobus AM, Savarier P, 2009. Mouse ES cells over-expressing the transcription factor *NeuroD1* show increased differentiation towards endocrine lineages and insulin-expressing cells. *Int J Dev Biol* 53: 569-578. <https://doi.org/10.1387/ijdb.092856mm>
- Martínez AM, Delgado JV, Rodero A, Vega-Pla JL, 2000. Genetic structure of the Iberian pig breed using microsatellites. *Anim Genet* 31: 295-301. <https://doi.org/10.1046/j.1365-2052.2000.00645.x>
- Myers P, 2008. Hox genes in development: the HOX code. *Nature Education* 1: 2.
- Onteru SK, Fan B, Nikkilä MT, Garrick DJ, Stalder KJ, Rothschild MF, 2011. Whole-genome association analyses for lifetime reproductive traits in pig. *J Anim Sci* 89: 988-995. <https://doi.org/10.2527/jas.2010-3236>
- Onteru SK, Fan B, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF, 2012. A whole-genome association study for pig reproductive traits. *Anim Genet* 43: 18-26. <https://doi.org/10.1111/j.1365-2052.2011.02213.x>
- Pallares LF, Carbonetto P, Gopalakrishnan S, Parker CC, Ackert-Bicknell CL, Palmer AA, Tautz D, 2015. Mapping of craniofacial traits in outbred mice identifies major developmental genes involved in shape determination. *Plos Genet* 11: e1005607. <https://doi.org/10.1371/journal.pgen.1005607>
- Ponsuksili S, Zebunke M, Murani E, Trakooljul N, Krieter J, Puppe B, Schwerin M, Wimmers K, 2015. Integrated genome-wide association and hypothalamus eQTL studies indicate a link between the circadian rhythm-related gene *PER1* and coping behavior. *Sci Rep* 5: 16264. <https://doi.org/10.1038/srep16264>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC, 2007. PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am J Human Genet* 81: 559-575. <https://doi.org/10.1086/519795>
- Qanbari S, Simianer H, 2014. Mapping signatures of positive selection in the genome of livestock. *Livest Sci* 166: 133-143. <https://doi.org/10.1016/j.livsci.2014.05.003>
- R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
- Rohrer GA, Nonneman DJ, Wiedmann RT, Schneider JF, 2015. A study of vertebra number in pigs confirms the association of *vertnin* and reveals additional QTL. *BMC Genet* 16: 129. <https://doi.org/10.1186/s12863-015-0286-9>
- Sargolzaei M, Chesnais JP, Schenkel FS, 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genom* 15: 478. <https://doi.org/10.1186/1471-2164-15-478>
- Schneider JF, Miles JR, Brown-Brandl TM, Nienaber JA, Rohrer GA, Vallet JL, 2015. Genomewide association analysis for average birth interval and stillbirth in swine. *J Anim Sci* 93: 529-540. <https://doi.org/10.2527/jas.2014-7899>
- Sherwood NM, Krueckl SL, McRory JE, 2000. The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. *Endocr Rev* 21: 619-670. <https://doi.org/10.1210/edrv.21.6.0414>
- Silió L, Barragan C, Fernández AI, García-Casco J, Rodríguez MC, 2016. Assessing effective population size, coancestry and inbreeding effects on litter size using the

- pedigree and SNP data in closed lines of the Iberian pig breed. *J Anim Breed Genet* 133: 145-154. <https://doi.org/10.1111/jbg.12168>
- Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, et al., 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucl Acids Res* 43: W589-W598. <https://doi.org/10.1093/nar/gkv350>
- Soilleux EJ, Morris LS, Leslie G, Chehimi J, Luo Q, Levrony E, Trowsdale J, Montaner LJ, Doms RW, Weissman D, Coleman N, Lee B., 2002. Constitutive and induced expression of DC-SIGN on dendritic cell and macrophage subpopulations in situ and in vitro. *J Leukoc Biol* 71: 445-457.
- Sterky FH, Trotter JH, Lee S, Recktenwald CV, Du X, Zhou B, Zhou P, Schwenk J, Fakler B, Südhof TC, 2017. Carbonic anhydrase-related protein CA10 is an evolutionary conserved pan-neurexin ligand. *Proc Natl Acad Sci* 114: E1253-E1262. <https://doi.org/10.1073/pnas.1621321114>
- Vale-Cruz DS, Ma Q, Syme J, LuValle PA, 2008. Activating transcription factor-2 affects skeletal growth by modulating pRb gene expression. *Mech Dev* 125: 843-856. <https://doi.org/10.1016/j.mod.2008.06.010>
- Velardo LL, Silva FF, Lopes MS, Madsen O, Bastiaansen JW, Knol EF, Kelly M, Varona L, Lopes PS, Guimarães SEF. 2016. Revealing new candidate genes for reproductive traits in pigs: combining Bayesian GWAS and functional pathways. *Genet Sel Evol* 48: 9. <https://doi.org/10.1186/s12711-016-0189-x>
- Ventanas S, Ventanas J, Ruiz J, Estévez M, 2005. Iberian pigs for the development of high-quality cured products. In: *Recent Res Devel Agricultural & Food Chem*; SG Pandalai (Ed.) 6: 27-53.
- Wang J, Vasaikar S, Shi Z, Greer M, Zhang B, 2017. Web-Gestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucl Acids Res* 45: W130-W137. <https://doi.org/10.1093/nar/gkx356>
- Weir WS, Cockerham CC, 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38: 1358-1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Wright S, 1951. The genetical structure of populations. *Ann Eugenics* 15: 323-354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>
- 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. <https://doi.org/10.1016/j.gene.2013.08.074>
- Yagi T, Shigetani Y, Furuta Y, Nada S, Okado N, Ikawa Y, Aizawa S, 1994. Fyn expression during early neurogenesis in mouse embryos. *Oncogene* 9: 2433-2440.
- Yong Y, Meng Y, Ding H, Fan Z, Tang Y, Zhou C, Luo J, Ke ZJ, 2015. PACT/RAX regulates the migration of cerebellar granule neurons in the developing cerebellum. *Sci Rep* 5: 7961. <https://doi.org/10.1038/srep07961>
- Zhang F, Zhang Z, Yan X, Chen H, Zhang W, Hong Y, Huang L, 2014. Genome-wide association studies for hematological traits in Chinese Sutai pigs. *BMC Genet* 15:41. <https://doi.org/10.1186/1471-2156-15-41>