Received: 19 June 2023 | Revised: 22 September 2023 | Accepted: 26 September 2023

DOI: 10.1111/jbg.12829

Journal of Animal Breeding and Genetics WILEY

ORIGINAL ARTICLE

Ascertaining the genetic background of the Celtic-Iberian pig strain: A signatures of selection approach

Katherine D. Arias[1] | **Hanboreum Lee**[2] | **Riccardo Bozzi**[3] | **Isabel Álvarez**[1] | **Juan Pablo Gutiérrez**[4] | **Iván Fernandez**[1] | **Juan Menéndez**[5] | **Albano Beja-Pereira**[6,7,8] | **Félix Goyache**[1]

[1]SERIDA-Deva, Gijón, Asturias, Spain

[2]Gyeongbuk Provincial College, Gyungbuk, Korea

[3]DAGRI, Università degli Studi di Firenze, Firenze, Italy

[4]Departamento de Producción Animal, Universidad Complutense de Madrid, Madrid, Spain

[5]ACGA, C/ Párroco José Fernández Teral 5A, Avilés, Asturias, Spain

[6]CIBIO-InBio, Universidade do Porto, Vairão, Portugal

[7]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal

[8]DGAOT, Faculty of Sciences, Universidade do Porto, Porto, Portugal

Correspondence
Félix Goyache, SERIDA-Deva, Camino de Rioseco 1225, 33394-Gijón, Asturias, Spain.
Email: fgoyache@serida.org

Funding information
AEI-ESF, Grant No. PRE2020-092905; AEI-FEDER, Grant No. PID2019-103951RB/AEI/10.13039/501100011033

Abstract

Celtic-Iberian pig breeds were majority in Spain and Portugal until the first half of the 20th century. In the 1990s, they were nearly extinct as a result of the introduction of foreign improved pig breeds. Despite its historical importance, the genetic background of the Celtic-Iberian pig strain is poorly documented. In this study, we have identified genomic regions that might contain signatures of selection peculiar of the Celtic-Iberian genetic lineage. A total of 153 DNA samples of Celtic-Iberian pigs (Spanish Gochu Asturcelta and Portuguese Bísara breeds), Iberian pigs (Spanish Iberian and Portuguese Alentejano breeds), Cinta Senese pig, Korean local pig and Cosmopolitan pig (Hampshire, Landrace and Large White individuals) were analysed. A pairwise-comparison approach was applied: the Gochu Asturcelta and the Bísara samples as test populations and the five other pig populations as reference populations. Three different statistics (XP-EHH, F_{ST} and ΔDAF) were computed on each comparison. Strict criteria were used to identify selection sweeps in order to reduce the noise brought on by the Gochu Asturcelta and Bísara breeds' severe population bottlenecks. Within test population, SNPs used to construct potential candidate genomic areas under selection were only considered if they were identified in four of ten two-by-two pairwise comparisons and in at least two of three statistics. Genomic regions under selection constructed within test population were subsequently overlapped to construct candidate regions under selection putatively unique to the Celtic-Iberian pig strain. These genomic regions were finally used for enrichment analyses. A total of 39 candidate regions, mainly located on SSC5 and SSC9 and covering 3130.5 kb, were identified and could be considered representative of the ancient genomic background of the Celtic-Iberian strain. Enrichment analysis allowed to identify a total of seven candidate genes (*NOL12*, *LGALS1*, *PDXP*, *SH3BP1*, *GGA1*, *WIF1*, and *LYPD6*). Other studies reported that the *WIF1* gene is associated with ear size, one of the characteristic traits of the Celtic-Iberian pig strain. The function of the other candidate genes could be related to reproduction, adaptation and

immunity traits, indirectly fitting with the rusticity of a non-improved pig strain traditionally exploited under semi-extensive conditions.

KEYWORDS

candidate genes, Celtic-Iberian pig strain, population bottleneck, rusticity, signatures of selection

1 | INTRODUCTION

Genomic variation results from a combination of evolutionary forces such as migration, drift, bottlenecks, environmental shifts, or changes in genetic interactions due to selection (Hohenlohe et al., 2010; Oleksyk et al., 2010; Sabeti et al., 2006). The distribution of selected loci is sensitive to the strength of selection as well as the landscape of environmental variation on which selection occurred (Hohenlohe et al., 2010; Lotterhos & Whitlock, 2014; Oleksyk et al., 2010). Natural selection can shape the between-populations differentiation within a species (Oleksyk et al., 2010). Populations subject to different environmental conditions will initially differ at a few key genomic sites, and the surrounding DNA will tend to accumulate differences due to linkage disequilibrium (Beaumont & Balding, 2004). Unlike classical population genetics, which focuses on allelic variation at a single site, methods aiming at the identification of sites under selection should consider the influence of their flanking areas (Hohenlohe et al., 2010).

Different tests are available to identify selection sweeps (Guo et al., 2009; Hohenlohe et al., 2010; Oleksyk et al., 2010; Randhawa et al., 2014). Selection causes deviations from the expected allelic frequency under neutrality resulting in signatures on the DNA sequence (Sabeti et al., 2006) that can be identified without needing to record phenotypic information on populations (Randhawa et al., 2015). However, even if nearly complete genomic sequence information is available, the identification of signatures of selection has limitations: it depends on the population structure and the assumptions underlying the models fitted to test the effects of either natural or artificial selection (Bovo et al., 2020; Hohenlohe et al., 2010; Randhawa et al., 2014; Schiavo et al., 2021). The sensitivity of the different methods to discriminate true sweeps caused by directional selection is a critical factor in the choice of selection tests (Randhawa et al., 2014). It frequently depends on the phenotypic and genetic divergence between the population under study and the population used as a reference (Randhawa et al., 2015). Combining the information provided by multiple selection tests and grouping more or less differentiated breeds are efficient strategies for identifying signatures of selection and identifying candidate genomic regions putatively important for the regulation of complex traits (Randhawa et al., 2015).

Although both archaeological evidence and genetic evidence suggest that the first domestic pigs arrived in Europe from the Near East about 8500 years ago (Larson et al., 2005), these ancestral genomes disappeared over 3000 years later as a result of interbreeding with local wild boars (Frantz et al., 2019). Within the continental European domestic pigs, two main mitochondrial DNA lineages, the so-called haplogroups A and C, were identified: the lineage A being representative of the Central and Northern Europe pig and the C lineage being mainly present in the Iberian domestic pig (Portuguese Alentejano and Spanish Iberian breeds) and in the Iberian and Maghrebian Wild Boar (Larson et al., 2005; Ramírez et al., 2009).

Historically, two different pig strains were bred in the Iberian Peninsula (Aparicio, 1944; Sotillo & Serrano, 1985): the Celtic and the Iberian pig breeds. The Iberian and Lusitano pig breeds are hypothesized to result from a particular domestication event occurred in the Iberian Peninsula, whereas Celtic-Iberian pigs would result from an ancient process of migration of Northern-Central European pigs, with population replacement, into the Iberian Peninsula (Aparicio, 1944; Sotillo & Serrano, 1985). Such hypotheses are consistent with genetic data (Menéndez, Goyache, et al., 2016).

In any case, recent population history may have erased the differences in genetic background between the two Iberian pig strains (Menéndez, Goyache, et al., 2016): (a) Celtic-Iberian pig breeds became nearly extinct at the end of the 20th century because of replacement by foreign improved breeds (Conde et al., 2010; Sotillo & Serrano, 1985); (b) although with little influence in the present Iberian pig population due to the strict control of crossbreeding (Rodríguez-Valdovinos et al., 2019), during the second half of the 20th century improved Cosmopolitan pigs introgressed into specific lines of the Iberian pig (Spotted Black Jabugo) due to unsupervised crosses (Alves et al., 2003); and (c) inappropriate management practices may have facilitated the introgression of Cosmopolitan breeds into Wild Boar populations in areas of Southern Spain (van Asch et al., 2012) and the Pyrenees.

While documenting the genetic background of local populations stands as a key concern within pig genomics (Dadousis et al., 2022; Muñoz et al., 2019; Schiavo et al., 2021), the understanding of the genomic profile of the Celtic-Iberian pig strain is still limited. The Iberian pig has received major scientific attention due to their importance as a source of quality food products (see, as examples, Herrero-Medrano et al., 2013; Lopez-Bote, 1998; Silió, 2000). However, this is not the case for Celtic-Iberian pig breeds. Recovery programmes for the Celtic-Iberian pig breeds were implemented only recently: the programme for the Portuguese Bísara breed started in 1996 by implementing two governmental-run conservation nuclei in Guimarães and Montalegre (Santos Silva et al., 2019). In 2002, a conservation programme for the Gochu Asturcelta was initiated using six founders (Menéndez et al., 2015; Menéndez, Álvarez, et al., 2016). Addressing this gap is important because Celtic-Iberian pigs were majority in Spain and Portugal until the 1950s (Aparicio, 1944; Santos Silva et al., 2019). Recovery programmes for the Celtic-Iberian pig breeds were implemented only recently: the programme for the Portuguese Bísara breed started in 1996 by implementing two governmental-run conservation nuclei in Guimarães and Montalegre (Santos Silva et al., 2019). In 2002, a conservation programme for the Gochu Asturcelta was initiated using six founders (Menéndez et al., 2015; Menéndez, Álvarez, et al., 2016). There is an increasing interest in characterizing the type and production traits of the Celtic-Iberian pig (Argamentería, de la Roza-Delgado, Cueto, Hidalgo, Tamargo, & Menéndez, 2012; Argamentería, de la Roza-Delgado, Cueto, Hidalgo, Tamargo, Rodríguez, et al., 2012; de la Roza-Delgado et al., 2022; Santos Silva et al., 2019). Individuals belonging to the Celtic-Iberian pig strain are animals showing different coat colour patterns (white, roan, or spotted) with long, non-compact, bodies, good skeletal development, relatively lower development of the rear musculature, and typically huge ears dropping to the sides of a well-developed head (Figure 1; Table S1; Argamentería, de la Roza-Delgado, Cueto, Hidalgo, Tamargo, & Menéndez, 2012; Santos Silva et al., 2019). Characterized as environmentally well adapted, these individuals exhibit notable prolificacy (farrowing rate ≥ 9) and a predisposition towards subcutaneous fat accumulation (Argamentería, de la Roza-Delgado, Cueto, Hidalgo, Tamargo, Rodríguez, et al., 2012; Santos Silva et al., 2019), highly efficient on the utilization of locally sourced feed resources, but presenting low growth rates (Santos Silva et al., 2019). Consequently, the enhancement of carcass yields keeping the established traditional extensive or semi-extensive husbandry systems presents a formidable challenge (de la Roza-Delgado et al., 2022; Santos Silva et al., 2019).

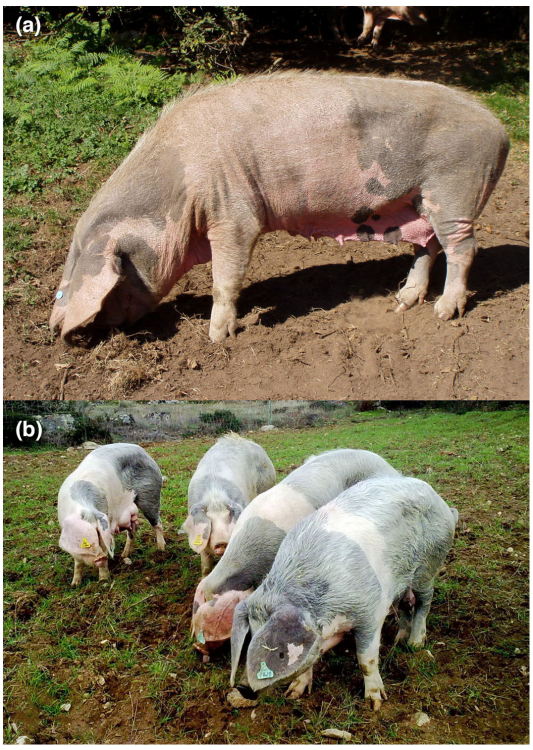

FIGURE 1 Pictures showing a typical Gochu Asturcelta sow (a) and typical Bísara sows (b).

The aim of this research was to enhance the understanding of the genetic background of the Celtic-Iberian pig strain through the utilization of more than one approach to identify signatures of selection. This involved identifying genomic regions that are potentially distinct within their genome. To achieve this, a series of multi-pairwise comparisons were conducted using samples from Iberian, Cosmopolitan, Wild Boar, and potentially unrelated Italian and Korean pig populations.

2 | MATERIALS AND METHODS

SERIDA is adhered to the Ethical Committee in Research of the University of Oviedo (Spain), which ensures that all research with biological agents follows Good Laboratory Practices and European and Spanish regulations on biosecurity under the Regulation of 13 February 2014 (BOPA no. 47 on 26 February 2014). Blood and hair root samples used in this project were collected by veterinary practitioners following standard procedures and relevant national guidelines to ensure appropriate animal care, with the permission and in the presence of the owners. For this reason, permission from the Ethical Committee in Research of the University of Oviedo was not required.

2.1 | Samples, genotyping, and population structuring analysis

A total of 153 DNA samples were analysed. Samples belonged to the following pig breeds: Gochu Asturcelta (20), Portuguese Bísara (19), Spanish Iberian (11), Portuguese Alentejano (7), Hampshire (5), Landrace (13), Large White (20), Wild Boar (11), Italian Cinta Senese (27), and Korean local (20). Wild Boar samples were obtained in different provinces of North-western Spain. A part of these samples was previously analysed for microsatellites and mitochondrial DNA markers (Menéndez, Goyache, et al., 2016). Since various studies reported little differentiation between Spanish Iberian and Portuguese Alentejano breeds (Dadousis et al., 2022; Muñoz et al., 2018, 2019), Iberian and Alentejano samples were grouped into the Iberian pig population (18 individuals). Following Menéndez, Goyache, et al. (2016), Hampshire, Landrace, and Large White samples were grouped into the Cosmopolitan (38 individuals) pig population.

The whole dataset was genotyped with the AxiomPorcineHDv1 of Affymetrix (658,692 SNPs). The software Axiom Analysis Suite v4.0.3 (Thermo Fisher Scientific, Waltham, MA) was used to create standard .ped and .map files. SNPs were mapped using the Sscrofa genome build 11.1 (Groenen et al., 2012). Only autosomal chromosomes with known positions were considered. SNP data were edited as follows: first, loci at which Mendelian errors were identified in a Gochu Asturcelta pedigree (Arias et al., 2022) were removed from all samples in the data set; and second, following the recommendations of Arias et al. (2022) for populations in which pedigrees are not available (here Bísara, Iberian, Cosmopolitan, Wild Boar, Cinta Senese, and Korean local), we fitted $F_{IS} > 0.9$ as a threshold to avoid the presence of null and partially null alleles due to technical issues. Compared with classical editing strategies using Minor Allele Frequency and deviation from Hardy–Weinberg proportions, this strategy has been proven to be useful to remove genotyping errors while keeping putatively informative loci and datasets size (Arias et al., 2022). Finally, a total of 458,384 SNPs with a minimum call rate of 0.97 were retained. Missing genotypes were imputed using BEAGLE v5.4 (Browning et al., 2018, 2021) using default parameters.

A cluster analysis was carried out using the program Admixture v1.23 (Alexander et al., 2009; Alexander & Lange, 2011), which calculates maximum-likelihood estimates of individual ancestries based on data provided by multiple loci. Analyses were conducted for $1 \leq K \leq 14$ being K the number of clusters given the data. The optimal number of clusters was determined by performing the cross-validation procedure included in the program Admixture. By default, the procedure partitions all the observed genotypes into five roughly equally sized folds for each K. Within K, each fold was used as a test set, while the other four were used for training. For each fold, prediction error is estimated by averaging the squares of the deviance residuals for the binomial model. Admixture barplots were constructed using the library ggplot2 of R (http://CRAN.R-project.org/).

The program PLINK v1.9 (Chang et al., 2015) was used to compute principal component analysis (PCA). Eigenvectors computed for each individual were used to construct dispersion plots using the library ggplot2 of R (http://CRAN.R-project.org/).

2.2 | Identification of selection sweeps

For the ascertainment of genomic regions under positive selection, SNP data were arranged into a set of population comparisons as follows: (a) the Gochu Asturcelta and the Bísara samples were sequentially used as test populations; and (b) the Cosmopolitan, Iberian, Spanish Wild Boar, Cinta Senese, and Korean local pig samples were sequentially used as reference populations for either test populations. Furthermore, within test population, ten different pairwise comparisons were constructed taking two-by-two the results obtained in each of the five populations' comparisons. Only polymorphic SNPs were used in each pairwise comparison (test populations vs. reference population) varying from 392,278 to 450,086 SNPs.

Three different statistics were computed on each pairwise comparison: XP-EHH (Sabeti et al., 2007), F_{ST} (Weir & Cockerham, 1984), and ΔDAF based on the change in frequency in the putatively derived allele (Grossman et al., 2010).

The XP-EHH test (Sabeti et al., 2007), across-population extended haplotype homozygosity, is a modification of the Extended Haplotype Homozygosity (EHH) statistics (Sabeti et al., 2002) with increased power to detect ongoing or nearly fixed signatures of selection (positive scores for the population under study). This test allows the use of small sample sizes and genetically similar test and reference populations (Álvarez et al., 2020; Sabeti et al., 2007). XP-EHH estimates were carried out using the program *selscan* v.1.0.40 (Szpiech & Hernandez, 2014), fitting the parameters recommended by the authors: maximum EHH extension in bp ("–max-extend" option) 1,000,000, maximum gap allowed between two SNPs in bp ("–max-gap" option) 200,000, EHH decay cut-off ("–cutoff" option) 0.05. The output results for each SNP were frequency-normalized over all chromosomes using the program *norm*, provided with *selscan*. This normalization was carried out using default parameters as well: number of frequency bins

("–bins" option) 100. SNPs with normalized values higher than 2 were considered under putative selection in the test population and used for further analyses.

The fixation index F_{ST} (Weir & Cockerham, 1984) is a classical statistic of genetic differentiation that can be used to identify loci in positive selection (Lotterhos & Whitlock, 2014). F_{ST} accounts for the variance in allele frequency between the test and the reference populations. At the loci level, extremely positive F_{ST} values mean divergent selection, strong positive selection, or random drift (Randhawa et al., 2014).

The ΔDAF test requires the definition of an ancestral allele and a derived allele (Grossman et al., 2010). Following Hsu et al. (2023), the major alleles (more frequent variants at a locus) were assigned as ancestral alleles. ΔDAF was computed as ΔDAF = DAF *test* – DAF *reference*, where DAF is the frequency of the derived allele in either the test, putatively selected, population or the reference, alternative non-selected, population (ancestral population in terms of Grossman et al., 2010). The normal distribution of the derived allele frequency (DAF) was estimated, and the corresponding ΔDAF values were transformed to Z scores (0,1).

Only loci having either F_{ST} or ΔDAF scores within the top 1% were considered under putative selection and used for further analyses (Kijas et al., 2012; Randhawa et al., 2015).

2.3 | Construction of genomic areas under selection

Chromosomal regions under potential selection were first constructed within test population (either Gochu Asturcelta or Bísara breeds) starting from the SNPs identified in three steps: first, SNPs identified as putatively under selection were only considered if: (a) they were identified in at least two out of three statistics (either XP-EHH, F_{ST}, or ΔDAF) and in four or more (out of ten) pairwise population comparisons within statistics (note that the five assessments carried out within test population gave 10 two-by-two pairwise population comparisons); second, using the *intersect* function of the software BedTools (Quinlan & Hall, 2010), the genomic areas 16 kb (roughly three-fold mean distance among SNPs) upstream and downstream the retained SNPs were overlapped with the genomic areas surrounding those SNPs under putative selection identified in at least four (out of ten) pairwise comparisons identified using one statistics only to ensure that the selection signal in the surroundings of the retained loci was correctly assessed; and third, the lower and upper bounds of the obtained overlaps were considered potential candidate regions under selection.

Finally, the potential candidate regions identified in either the Gochu Asturcelta or the Bísara breeds were overlapped using the *intersect* function of the software BedTools (Quinlan & Hall, 2010). The upper and lower bounds of the overlaps were considered candidate regions under selection in the Celtic-Iberian pig strain and used for enrichment analyses.

2.4 | Enrichment and annotation analyses

Protein-coding genes found within the candidate regions under selection were retrieved from the Ensembl Genes 91 database, based on the SusScrofa 11.1 porcine reference genome using the BioMart tool (Kinsella et al., 2011) and further considered candidate genes. Candidate genes were further processed through a gene network analysis using GeneMANIA (Franz et al., 2018; Mostafavi et al., 2008), and physical interaction networks with other genes were constructed (see http://pages.genemania.org/help/ for details). Data are collected from primary studies and BioGRID for genes having a full description in either the Ensembl Genes 91 or GenBank databases only.

3 | RESULTS

Figure 2 illustrates the dispersion of the individual genomes analysed according to the two first factors computed via PCA. These two factors accounted for a total of 40% of the genetic variance. The local Korean samples were separated from the other pig populations studied on the X-axis by PCA. Additionally, the Mediterranean pig samples (Iberian, Cinta Senese, and Iberian Wild Boar) were separated from the Celtic-Iberian and Cosmopolitan pig individuals on the Y-axis. On the positive values of the Y-axis, the samples of Bísara and Gochu Asturcelta could be distinguished. On the X-axis, PCA separated the Mediterranean pig samples (Iberian, Cinta Senese, and Iberian Wild Boar) from the Celtic-Iberian and Cosmopolitan pig individuals. PCA also allowed to separate, on the positive values of the Y-axis, the Gochu Asturcelta individuals from the Bísara and the Cosmopolitan samples and the Cinta Senese samples from the Iberian and Wild Boar samples on the negative values of the Y-axis.

Admixture analysis informed that the lowest cross-validation error was at $K = 9$ (Figure 3). However, differences between cross-validation errors for $K = 8$ and $K = 9$ were negligible (lower than 0.001). Each of the clusters inferred corresponds with one of the populations analysed ($K = 8$) except for the Cosmopolitan pig, in which the Hampshire and Landrace samples clustered separately of

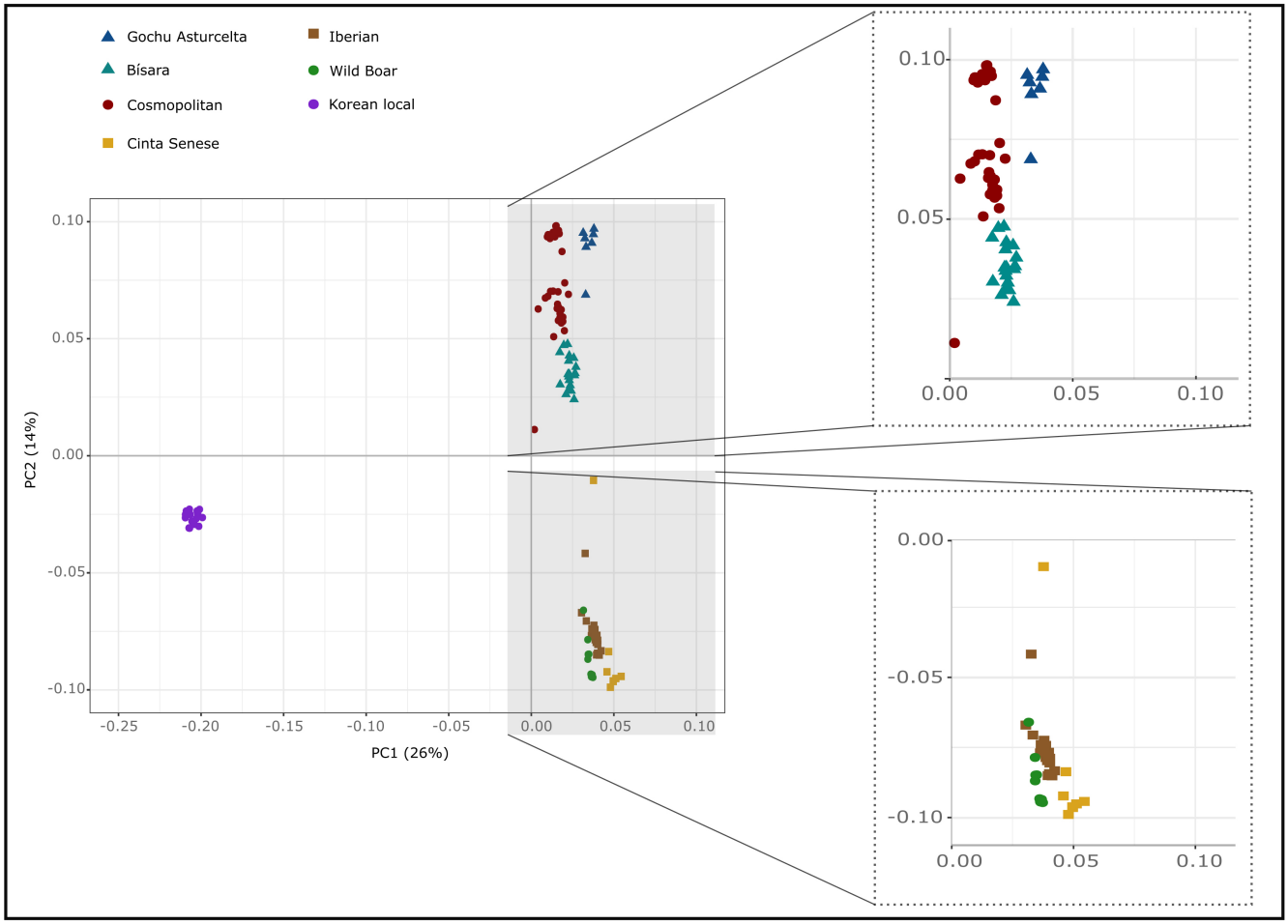

FIGURE 2 Dispersion plot summarizing the between-individuals genomic relationships constructed according to the two first factors computed via principal component analysis. On the X-axis, the first factors explained 26% of the total variability. On the Y-axis, the second factor explained 14% of the total variability. The Celtic-Iberian individuals are in triangles (Gochu Asturcelta individuals in blue and Bísara individuals in light blue), Cosmopolitan pig individuals are in red circles, Cinta Senese individuals are in yellow squares, Iberian pig individuals are in brown squares, Wild Boar individuals are in green circles, and Korean local pig (on the left of Y-axis) are in purple. The dispersion areas corresponding to the Celtic-Iberian and Cosmopolitan pigs (positive values on Y-axis) and Mediterranean pigs (negative values on Y-axis) were zoomed in to allow a better assessment of their differences.

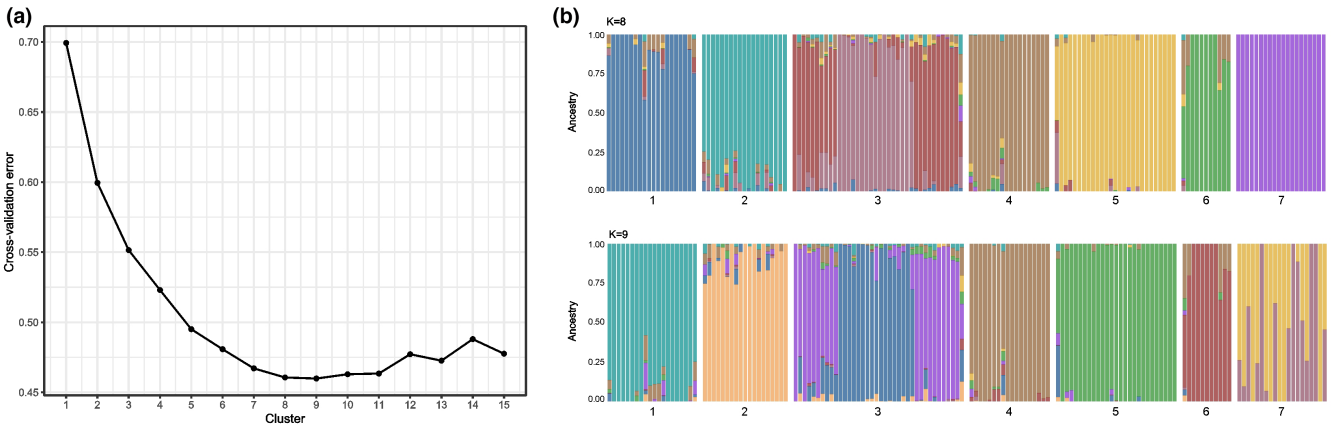

FIGURE 3 Plots summarizing admixture analyses. Plot a illustrates the cross-validation errors for each K tested. Plot b illustrates the individual ancestries estimated using Admixture v1.23 for $K=8$ and $K=9$. Numbers below the Admixture barplots mean the following: 1, Gochu Asturcelta; 2, Bísara breed; 3, Cosmopolitan; 4, Iberian pig; 5, Cinta Senese; 6, Wild Boar; and 7, Korean local.

the Large White individuals. However, the best fit with data was obtained for $K=9$ due to the identification of a hidden structure within the Korean local pig samples.

3.1 | SNPs under selection

The number of SNPs found under selection in at least four pairwise comparisons and one statistic in the Gochu Asturcelta and the Bísara pig breeds was 64,977 and 53,510, respectively. The full list of SNPs identified is available on request. Regarding the statistics assayed (not shown), XP-EHH yielded the higher number of SNPs under positive selection both in the Gochu Asturcelta (33,083) and in the Bísara (31,335) breeds, whereas ΔDAF yielded the lower number of SNPs under positive selection in both breeds (20,719 and 18,606, respectively). Performance of F_{ST} was intermediate although nearer to ΔDAF in both breeds.

In the Gochu Asturcelta breed, the comparison giving the higher number of SNPs under positive selection was that with Wild Boar (25,184 SNPs) and the lower that with the Italian Cinta Senese breed (11,783 SNPs). Within statistics assayed, up to 14,639, 5311, and 1179 SNPs were identified under selection in more than six, eight, and 10 pairwise comparisons, respectively (Table S2). Using two different reference populations (pairwise comparisons), the number of loci identified under positive selection varied from 1013, for the pair Iberian-Korean local pig breeds, to 6246, for the pair Cosmopolitan-Iberian pig. Up to 12,164 (18.7% of the loci retained) SNPs were identified in at least two of the tests performed and, therefore, were considered under putative selection.

In the Bísara breed, the higher number of SNPs under positive selection was obtained when compared with the Iberian population (15,418 SNPs) and the lower when compared with the Italian Cinta Senese breed (11,671 SNPs). Within statistics assayed, up to 32,146, 7468, and 2502 loci were identified under selection in more than six, eight, and 10 pairwise comparisons, respectively (Table S3). When two different reference populations were considered (pairwise comparisons), the number of loci identified under positive selection varied from 1980, for the pair Cosmopolitan-Cinta Senese pig, to 8887, for the pair Cinta Senese-Wild Boar. Up to 14,286 (26.7% of the loci retained) SNPs were identified in at least two of the tests performed and, therefore, were considered under putative selection.

3.2 | Construction of candidate regions

In the Gochu Asturcelta breed, the 16 kb genomic areas flanking the 12,164 SNPs retained allowed to construct a total of 1411 genomic regions under putative selection (covering a total of 45,268 kb across 17 out of the 18 porcine autosomes). The overlap of these 1411 regions with the areas surrounding those SNPs under putative selection identified in two or more pairwise comparisons using one statistic only allowed to construct of 5587 potential candidate regions under selection, covering 276,049 kb of the porcine genome, on 17 of the 18 porcine autosomes (Table 1; Table S2). Figure 4 illustrates the distribution, per porcine autosome, of the candidate regions under putative selection identified in the Gochu Asturcelta (in red). Most potential candidate regions identified in the Gochu Asturcelta genome (48% of the total) were located on SSC3 and SSC6. However, these regions gathered 26% of the total length only (Table 1). No potential candidate regions were identified on SSC18.

In the Bísara breed, a total of 1163 genomic regions under putative selection, covering a total of 37,311 kb across the 18 porcine autosomes (mainly on SSC5 and SSC10), were constructed. The overlap of these 1163 regions with the areas surrounding those SNPs under putative selection identified in two or more pairwise comparisons using one statistic only allowed to construct 2969 potential candidate regions under selection, covering 145,050.7 kb of the porcine genome, on the 18 porcine autosomes (Table 1; Table S3). Figure 4 illustrates the distribution, per porcine autosome, of the candidate regions under putative selection identified in the Bísara breed (in blue). Most potential candidate regions identified in the Bísara pig breed genome (36% of the total), gathering 36% of the total length, were located on SSC4 and SSC9.

Finally, 62 and 94 potential candidate regions identified in the Gochu Asturcelta (35 on SSC5, 5 on SSC7, 20 on SSC9, and 2 on SSC15) and the Bísara (60 on SSC5, 2 on SSC7, 30 on SSC9, and 2 on SSC15) pig breeds, respectively, overlapped (Figure 4). This allowed to identify a total of 39 genomic candidate regions under selection typical of the Celtic-Iberian pig strain (Table 1; Table S4). These typical candidate regions under selection covered a total of 3130.5 kb on four porcine autosomes only (SSC5, SSC7, SSC9, and SSC15), with SSC5 and SSC9 gathering 28 and 8 of these genomic areas, respectively.

3.3 | Enrichment and network analyses

Gene-annotation enrichment analysis allowed to identify a total of seven potential candidate porcine genes in the constructed candidate regions under selection (Table 2). Five of them (*NOL12*, nucleolar protein 12; *LGALS1*, galectin 1; *PDXP*, pyridoxal phosphatase; *SH3BP1*, SH3

TABLE 1 Description, per porcine autosome (SSC), of the potential candidate regions identified in both the Gochu Asturcelta and the Bísara pig breeds and of the candidate regions identified in the Celtic-Iberian pig strain.

	Potential candidate regions					
	Gochu Asturcelta		Bísara		Candidate regions	
SSC	*N*	Length	*N*	Length	*N*	Length
1	836	41181.0 [3.20; 64.1]	64	3000.6 [32.2; 63.4]		
2	4	168.8 [3.4; 48.6]	4	209.4 [38.3; 62.4]		
3	1287	64,238.8 [32.1; 64.1]	85	4011.0 [33.2; 63.9]		
4	165	8174.9 [33.1; 63.8]	375	18,478.6 [32.2; 64.2]		
5	192	9494.2 [32.7; 64.0]	279	13,544.2 [32.1; 64.1]	1475	2265.1 [48.1; 116.1]
6	1395	69,086.1 [32,2; 64.2]	71	3260.3 [32.7; 64.0]		
7	405	19,920.5 [32.8; 64.0]	68	3152.3 [33.9; 64.1]	4	187.81 [86,2; 101,6]
8	161	7763.3 [32.6; 64.0]	283	14,131.3 [33.4; 64.1]		
9	119	5865.6 [34.4; 63.9]	717	34,973.6 [32.3; 64.0]	1160	636.7 [70.1; 91.0]
10	9	435.3 [35.8; 60.2]	232	11,558.3 [32.3; 64.1]		
11	324	16,023.7 [32.8; 64.0]	28	1439.0 [35.8; 64.0]		
12	57	2662.3 [33,2; 63.9]	14	725.3 [42.0; 63.3]		
13	3	139,2 [34.6; 62.3]	73	3510.7 [33.2; 63.9]		
14	401	19,597.0 [32.9; 64.1]	47	2215.4 [33.0; 63.9]		
15	189	9253.4 [32.7; 64.2]	69	3149.3 [32.1; 63.9]	4	40.8 [40.8; 40,8]
16	5	269.6 [44.0; 60.2]	74	3480.2 [33.4; 62.9]		
17	35	1775.0 [35.7; 62.5]	463	23,101.9 [32.2; 64.1]		
18			23	1109.3 [32.1; 61.4]		
Totals	5587	276,049.0 [32.1; 64.2]	2969	145,050.7 [32.1; 64.2]	2643	3130.5 [40.8; 116,1]

Note: For each autosome, the number of chromosomal regions (*N*) and the total length (in kb) of the genome covered by these areas [minimum and maximum lengths in brackets] is given.

domain binding protein 1; *GGA1*, golgi associated, gamma adaptin ear containing, ARF binding protein 1) were located in the same locus on SSC5 spanning 76.1 kb. The other candidate genes fully identified were the *WIF1* (WNT inhibitory factor 1) gene on SSC5 and the *LYPD6* (LY6/PLAUR domain containing 6) gene on SSC15. Only gene ENSSSCG00000036198 on SSC9 had no description in the consulted databases.

Despite the close genomics location of most genes identified, enrichment analysis carried out using the DAVID Bioinformatics resources 6.8 (Huang et al., 2009), fitting classification stringency to "high" and enrichment thresholds to 1.3 (equivalent to $p < 0.05$), informed that they did not form functional clusters. This was confirmed via gene network analysis. Figure 5 illustrates that no links were reported among the seven candidate genes identified. However, several co-expression patterns were identified. The candidate genes formed networks with 19 other functionally complementary genes, four of them belonging to the Wnt genes family (*WNT11*, *WNT5A*, *WNT7A*, and *WNT1*, networking with the *WIF1* gene) (Table 2; Figure 5).

4 | DISCUSSION

The analyses carried out showed that the pig populations under study were organized according to geography and common ancestry (Figures 2 and 3). Our findings suggest that Celtic-Iberian pig breeds have a genetic association with Cosmopolitan pigs rather than the Iberian pig strain.

According to Muñoz et al. (2018), there is a distinct difference between the Bísara breed and south-western European pig breeds including Iberian and Alentejano pigs. Moreover, consistently with recent reports (Dadousis et al., 2022; Muñoz et al., 2018, 2019) and the classical theory on the origin of the strain, the Iberian domestic pig is genetically close to Iberian Wild Board (Figure 2) as well as with to the other Mediterranean pig breed analysed (Cinta Senese).

Both PCA and Admixture studies confirmed the distinct separation between the Bísara and Gochu Asturcelta pig breeds. This difference may result of the recent population bottlenecks and reproductive isolation to which these breeds were exposed. Breeding practices can also lead to population substructuring in very small animal populations, as previously determined by employing

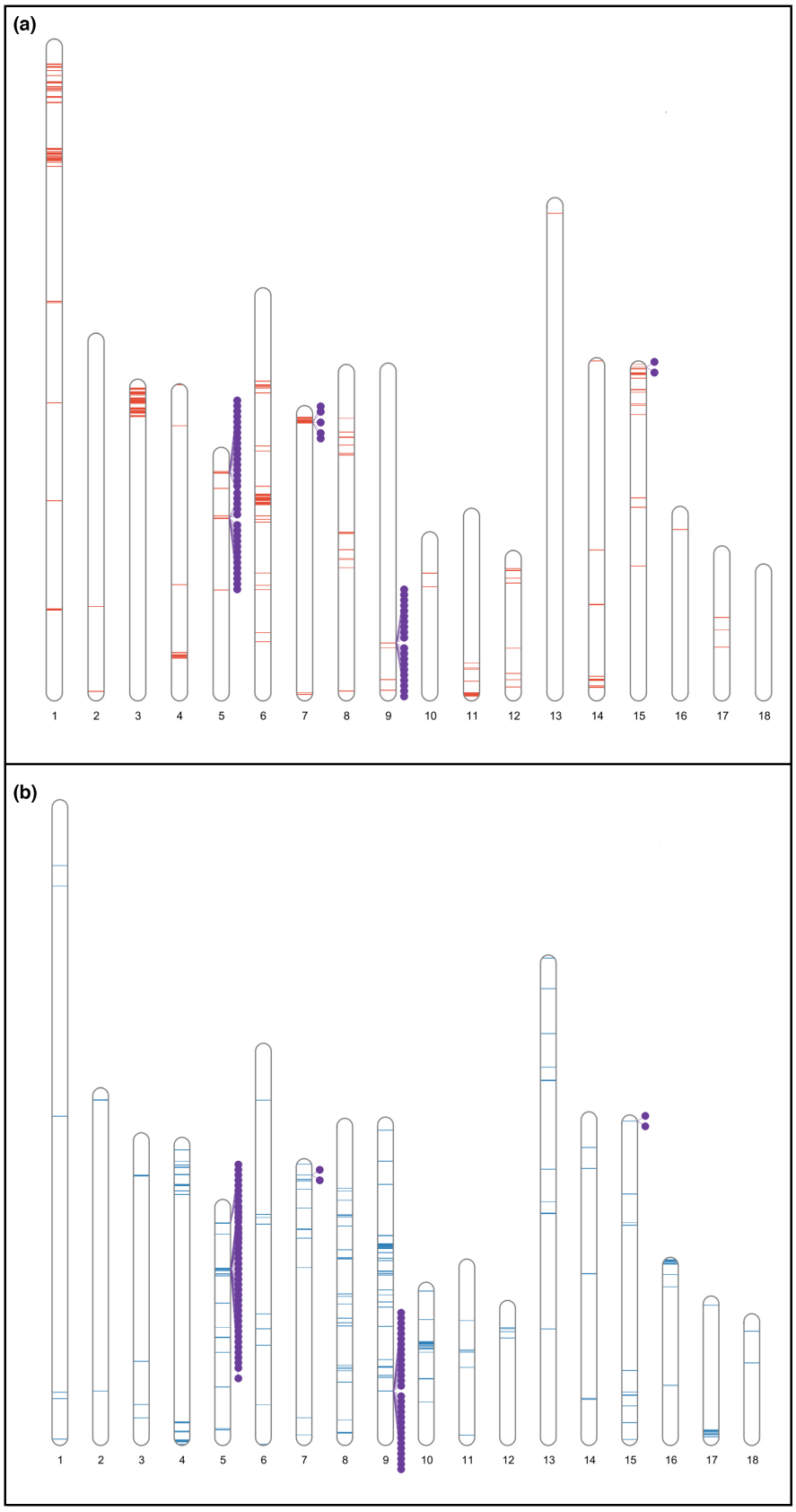

FIGURE 4 Ideogram illustrating, per porcine autosome, the distribution of the potential candidate regions identified in Gochu Asturcelta (Plot a) and the Bísara (Plot b) pig breeds. The 62 and 94 potential candidate regions identified in the Gochu Asturcelta and the Bísara pig breeds, respectively, on SSC5, SSC7, SSC9, and SSC15 (zoomed-in in purple circles) were finally used to construct candidate regions in the Celtic-Iberian pig strain.

microsatellites in the Gochu Asturcelta breed (Menéndez, Goyache, et al., 2016) and assessed in the current analysis for the Korean local pig breed (Figure 3).

Despite being considered both Celtic-Iberian strains, the differentiation between the Gochu Asturcelta and the Bísara breed might be justified by the different foreign

TABLE 2 Description of the candidate genes identified in the Celtic-Iberian pig strain.

Gene stable ID	SSC	Start (bp)	End (bp)	Name	Description	Physical interactions
ENSSSCG00000021663	5	10,137,362	10,144,140	*NOL12*	Nucleolar protein 12 [Source: HGNC Symbol;Acc:HGNC:28585]	*KIAA1191* (Putative monooxygenase p33MONOX)
ENSSSCG00000033854	5	10,149,068	10,156,023	*LGALS1*	Galectin 1 [Source: VGNC Symbol;Acc:VGNC:89691]	*CD2* (T-cell surface antigen CD2 precursor), *SUSD2* (sushi domain containing 2), *TRPV5* (transient receptor potential cation channel subfamily V member 5)
ENSSSCG00000034632	5	10,162,643	10,169,512	*PDXP*	Pyridoxal phosphatase [Source: VGNC Symbol;Acc:VGNC:91293]	*MED23* (mediator complex subunit 23)
ENSSSCG00000028531	5	10,172,667	10,187,906	*SH3BP1*	SH3 domain binding protein 1 [Source:HGNC Symbol;Acc:HGNC:10824]	*NLRP4* (NLR family pyrin domain containing 4)
ENSSSCG00000024895	5	10,190,367	10,213,477	*GGA1*	Golgi associated, gamma adaptin ear containing, ARF binding protein 1 [Source: VGNC Symbol;Acc:VGNC:88431]	*LRP3* (LDL receptor related protein 3), *SYNRG* (synergin gamma), *SORL1* (sortilin-related receptor 1), *F8* (coagulation factor VIII), *SYNRG* (Synergin Gamma), *AP1AR* (adaptor-related protein complex 1-associated regulatory protein)
ENSSSCG00000030998	5	29,450,442	29,543,913	*WIF1*	WNT inhibitory factor 1 [Source: VGNC Symbol;Acc:VGNC:94957]	*WNT11* (Wnt family member 11), *ICOS* (inducible T-cell costimulator), *WNT5A* (Wnt family member 5A), *WNT7A* (Wnt family member 7A), *CRADD* (CASP2 and RIPK1 domain containing adaptor with death domain), *WNT1* (Wnt family member 1)
ENSSSCG00000036198	9	116,327,344	116,330,248			
ENSSSCG00000023627	15	2,579,622	2,696,160	*LYPD6*	LY6/PLAUR domain containing 6 [Source:HGNC Symbol;Acc:HGNC:28751]	*CHRNA6* (cholinergic receptor nicotinic alpha 6 subunit), *SS18L2* (SS18 like 2)

Note: For each candidate gene, the identification (ID), porcine chromosome (SSC), start and end of the gene (in bp), and the name and description of the gene retrieved from the Ensembl Genes 91 database are detailed. If the description of the candidate gene was available, genes involved in functional networks as physical interactions with the candidate genes were listed as well.

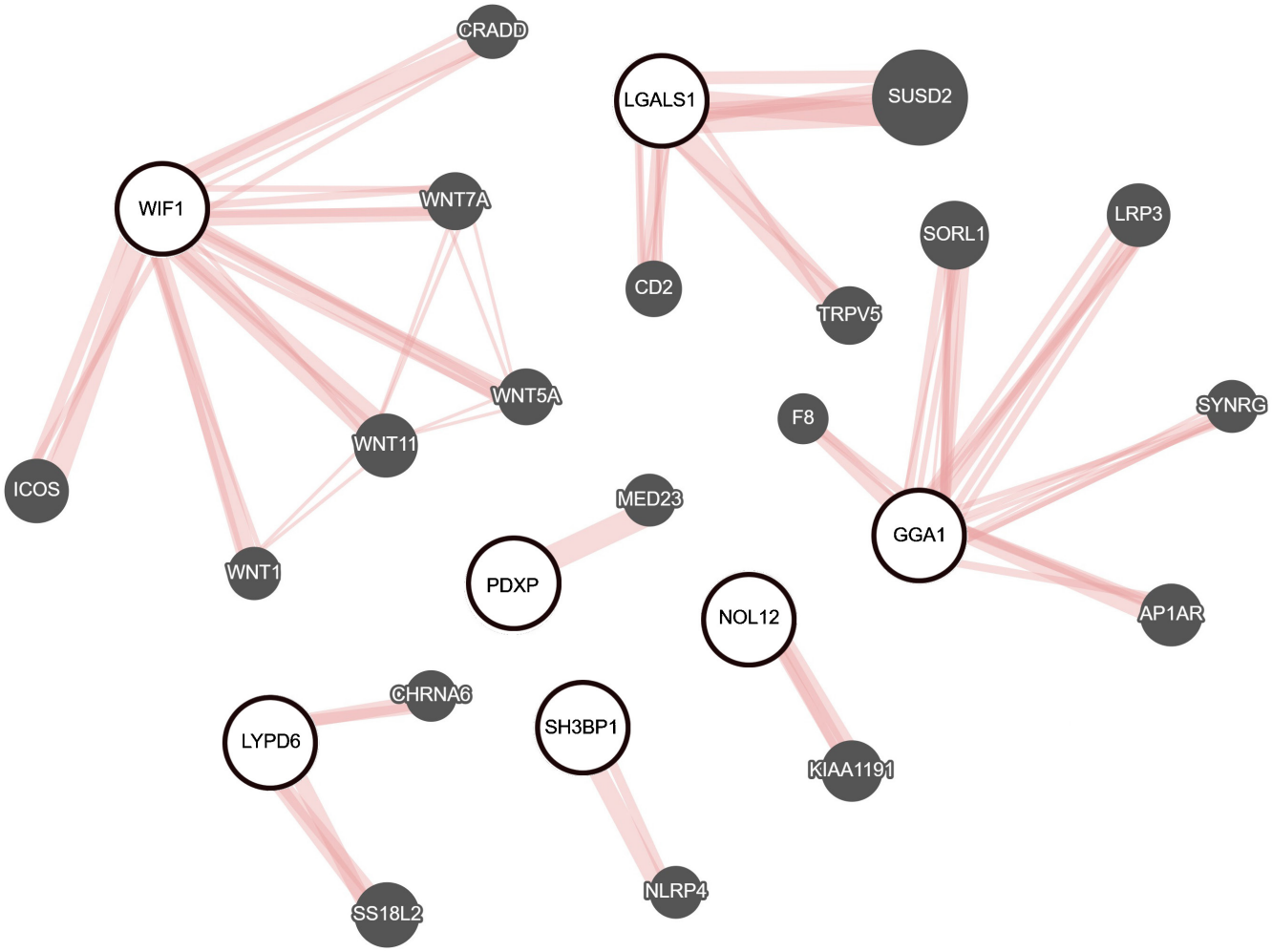

FIGURE 5 Graphical representation of the gene network analysis carried out using the software GeneMANIA. Candidate genes identified in this study are in open boxes. Links in light pink indicate the physical interaction networks (protein–protein interaction data assessed from publicly available protein interaction databases).

breeds, which introgressed the Bísara breed. Indeed, several studies have been finding moderate frequencies on alleles which their origin is well assigned to foreign breeds, such as the mutation responsible for the malignant hyperthermia the ryanodine receptor 1 gene (*RYR1*; Arg-615Cys), originated from the Belgian Pietrain (Beja-Pereira et al., 2001), the moderate frequency alleles in the *PPARD* gene (Muñoz et al., 2018) originated in Chinese breeds, and a close genetic proximity between the Cosmopolitan Landrace breed and the Bísara breed (Dadousis et al., 2022).

4.1 | Differences among selection signature approaches

Although the Gochu Asturcelta and Bísara breeds are thought to have common ancestors (Aparicio, 1944), the patterns of nucleotide variation caused by recent drift due to a population bottleneck may mimic those caused by selection pressures at specific loci (Hohenlohe et al., 2010; Oleksyk et al., 2010), and therefore, we have used a very conservative approach to identify genomic regions putatively specific to the Celtic-Iberian pig strain. Particularly, because of populations with significant levels of inbreeding, the identification of breed-specific characteristics may be difficult (Schiavo et al., 2021). Although both the Gochu Asturcelta and the Bísara breeds experienced strong population bottlenecks (Menéndez, Álvarez, et al., 2016), this issue has particular importance to the Gochu Asturcelta pig, a highly inbred population deriving from four founders only (Menéndez, Álvarez, et al., 2016). As a whole, the number of SNPs being under putative selection was 22% higher in the Gochu Asturcelta than in the Bísara breed (65,217 vs. 53,510) with XP-EHH and ΔDAF identifying 6% and 11% more SNPs, respectively, and F_{ST} yielding 28% more SNPs. This confirms previous reports, suggesting that F_{ST} may be more sensitive to false signatures of selection caused by drift due to the increase in the variance of

the F_{ST} distribution across loci (Colonna et al., 2014; Lotterhos & Whitlock, 2014; Megdiche et al., 2019; Randhawa et al., 2014).

Furthermore, replacement of Celtic-Iberian pig breeds by foreign improved pigs (Conde et al., 2010; Menéndez, Álvarez, et al., 2016; Sotillo & Serrano, 1985) has undoubtedly erased part of the ancient genomic background of the strain making it difficult to ascertain. In this scenario, applying the conservative approach used here is mandatory. In this respect, scanning data using different statistics could capture different aspects of the information content in data (Moscarelli et al., 2021; Randhawa et al., 2014). The three statistics used for analyses are expected to give complementary information: ΔDAF is expected to be precise in distinguishing the genomic sites with higher differentiation between populations (Colonna et al., 2014), whereas F_{ST} and XP-EHH signals peaked more narrowly around the causal variant, making them useful for spatial localization (Grossman et al., 2010). Therefore, we used strict criteria to select SNPs under putative selection (identified by two or more statistics and in at least four out of ten pairwise comparisons within breed) and later construct potential candidate areas under selection around these SNPs.

The same approaches as those used here, that is the combination of multi-pairwise comparisons between more or less differentiated populations with the combined use of different statistics expected to provide complementary information, have been performed to obtain reliable signals useful to identify the particular genomic background of Merino sheep (Megdiche et al., 2019), the genomic background of original Brown cattle (Moscarelli et al., 2021), and that of geographically distinct groups of Asian dogs (Hsu et al., 2023). Consistently with these previous reports, the genomic areas putatively typical of the Celtic-Iberian pig strain were limited in number and located on a few porcine chromosomes.

However, the multi-pairwise comparisons performed, using three different statistics giving complementary information, provide additional confidence to the results reported independently on both the recent breeding history of the two Celtic-Iberian pig breeds analysed and the fact that some pig populations used as reference were constructed by pooling different breeds (Iberian and Cosmopolitan) or have small sample size (Wild Boar).

4.2 | Biological importance of the candidate genes identified

Although their results should be interpreted with caution, selection signature analyses, while not reliant on phenotypic records, offer a mean to establish a connection between the genome and phenotype in cases where natural or human-induced selection result in the near-complete fixation of a specific allele (Megdiche et al., 2019). Furthermore, the evolutionary properties of the complex traits can be captured by exploring gene networks in which the candidate genes located in genomic regions under selection are involved (Randhawa et al., 2014).

This assertion finds validation in the current analysis: potential candidate regions identified in the Gochu Asturcelta harbour genes involved in ear anatomy (*WIF1* and *WNT1* genes on SSC5), as an example of this genotype–phenotype matching. The *WIF1* gene has previously been associated with ear size in pigs (Zhang et al., 2014), while the *WNT1* gene has been associated with ear malformation occurrences in the species (Hao et al., 2018). Huge ears dropping on the sides are the main phenotypic characteristic of Celtic-Iberian pigs. Remarkably, this feature served as one of the criteria employed by farmers during the recovery programmes of the Gochu Asturcelta and Bísara breeds (Menéndez et al., 2015; Menéndez, Álvarez, et al., 2016; Santos Silva et al., 2019).

Furthermore, the *WIF1* gene may also be involved in other typical characteristics of the Celtic-Iberian pig strain, such as skeletal development. This notion stems from the observation that the *WIF1* gene potentially influences bone development differences among European pig populations, acting as an inhibitor within the *Wnt* signalling pathway (Wilkinson et al., 2013). This pathway, represented in its network by genes like *WNT11*, *WNT5A*, *WNT7A*, and *WNT1*, potentially underscores the genetic regulation of skeletal growth in this strain. Another interesting finding worthy of noting is the *LYPD6* gene, on SSC15. It has been reported to be associated with total number of piglets born (Wang et al., 2022), which fits well with the good prolificacy of Celtic-Iberian pigs (from 9 to 14 piglets born; Argamentería, de la Roza-Delgado, Cueto, Hidalgo, Tamargo, Rodríguez, et al., 2012; Santos Silva et al., 2019).

The other candidate genes located on candidate genomic regions under positive selection may contribute to the rustic attributes of the Celtic-Iberian pig strain, a non-improved pig strain traditionally exploited under semi-extensive management, encompassing adaptation, vigour, and immunity. In this respect, the *GGA1* gene, involved in the regulation of acid metabolism and protein trafficking in liver (Kumar et al., 2010), may have importance in the ability of Celtic-Iberian pig for making use of different food resources due to its networking with genes such as *SORL1*, involved in the regulation of gut microbiota in pigs (Crespo-Piazuelo et al., 2019), and *AP1AR*, involved in the regulation of feeding behaviour in pig (Ding et al., 2017). Similarly, the *SH3BP1* gene, which contributes to the integrity of the skin and mucosae due to

its involvement in the involved epithelial junction formation (Elbediwy et al., 2012), is related to the gene *NLRP4*, which participates in the immune inflammatory response (Bouchier-Hayes et al., 2001). The other candidate genes identified are involved in basic metabolic pathways probably affecting different traits.

The *NOL12* gene encodes a multifunctional protein involved in many metabolic pathways of the cell, including RNA- and mitochondrial metabolism (Scott et al., 2017), and is related to the *KIAA1191* gene, which encodes a NAPDH-dependent oxidoreductase (Xu et al., 2022). The *PDXP* gene participates in vitamin B6 metabolism (Fonda, 1992) and networks with the *MED23* gene, which plays a critical role in the metabolic connection between the insulin signalling and the adipogenic transcriptional cascade (Wang et al., 2009). The *LGALS1* gene encodes galectin-1, which plays an important role during myoblast proliferation and differentiation (Qiu et al., 2008), and networking with genes such as *CD2*, involved in the immune response in pig (Cheng et al., 2022) or the *TRPV5* gene candidate gene for cold tolerance in Min pig (Zhang et al., 2023).

Finally, it is important to note that although a previous analysis focusing on adaptation in European pig breeds, Bovo et al. (2020) identified signatures of selection in the Bísara breed on SSC8 near the *KIT* gene in consistency with the spotted phenotype of the breed. The aim of the current analysis focuses on the Celtic-Iberian strain rather than on a particular breed. Although this coat colour patterning is predominant in the Bísara breed, the coat colour pattern showed a large variation in the first stages of the conservation programme of the Gochu Asturcelta pig, which included a proportion of both white and black individuals (Argamentería, de la Roza-Delgado, Cueto, Hidalgo, Tamargo, Rodríguez, et al., 2012).

5 | CONCLUSION

The Celtic-Iberian pig strain stands as a rarity in the European pig framework. Its historical importance justifies the implementation of recovery programmes and their genomic documentation. The current analysis suggests that both the Gochu Asturcelta and the Bísara breeds share genomic regions putatively representative of the ancient background of the strain. This alignment is reinforced by the involvement of the candidate genes linked to distinctive phenotypic traits of the strain, such as ear size and prolificacy, and the adaptation ability of a pig traditionally managed under semi-extensive conditions. While the recent breed history might introduce false-positive signatures of selection and the relatively low number of breeds and individuals tested may be an issue, the stringent criteria employed in pinpointing candidate genomic areas under selection, including a pairwise-comparison approach and diverse complementary statistical measures, bolster the confidence in the findings. Furthermore, the current research enriches our comprehension of rusticity (environmental adaptation) in pigs by providing a new set of genomic areas and candidate genes that can be targeted for further research aiming at the ascertainment of the genomic basis for pig adaptation within evolving environments.

AUTHOR CONTRIBUTIONS

FG, IA, and JM conceived and planned the project; KDA, FG, JPG, and IF did the data analyses; KDA and FG wrote the paper; LH, RB, ABP, IA, JM, and JPG undertook sampling and discussed and interpreted genetic data in the light of breeding and statistical evidence; KDA and IA did laboratory work. All authors gave final approval for publication.

ACKNOWLEDGEMENTS

This research was partially funded by AEI-FEDER, Grant No. PID2019-103951RB/AEI/10.13039/501100011033. KDA was funded by AEI-ESF, Grant No. PRE2020-092905. Authors are indebted to Dr Wei-Tse Hsu of The University of Sydney, Australia, for kindly providing the R code used to compute F_{ST} and ΔDAF scores. We would like to express our sincere gratitude to ACGA (Asociación de Criadores del Gochu Asturcelta; https://www.gochuasturcelta.org/) and Direção Geral de Alimentação e Veterinaria (https://www.dgav.pt) for providing pictures of typical individuals of the Gochu Asturcelta and Bísara breeds. The genotyping service was carried out at CEGEN-PRB3-ISCIII; it is supported by grant PT17/0019, of the PEI+D+i 2013-2016, funded by ISCIII and ERDF. Authors are indebted to Dr Gonçalo Carvalho (Thermo Fisher Scientific) for his willingness and kind contribution to the execution of this project.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The dataset used and analysed during the current study is available from the corresponding author on reasonable request.

ORCID

Katherine D. Arias https://orcid.org/0000-0001-6710-4208
Riccardo Bozzi https://orcid.org/0000-0001-8854-0834
Isabel Álvarez https://orcid.org/0000-0002-6052-8881
Juan Pablo Gutiérrez https://orcid.org/0000-0003-3142-2356

Albano Beja-Pereira https://orcid.org/0000-0002-1607-7382
Félix Goyache https://orcid.org/0000-0002-6872-1045

REFERENCES

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. https://doi.org/10.1186/1471-2105-12-246

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Álvarez, I., Fernández, I., Traoré, A., Pérez-Pardal, L., Menéndez-Arias, N. A., & Goyache, F. (2020). Genomic scan of selective sweeps in Djallonké (west African dwarf) sheep shed light on adaptation to harsh environments. *Scientific Reports*, *10*(1), 2824. https://doi.org/10.1038/s41598-020-59839-x

Alves, E., Óvilo, C., Rodríguez, M. C., & Silió, L. (2003). Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and other domestic and wild pig populations: Mitochondrial DNA sequence variation and phylogenetic relationships. *Animal Genetics*, *34*(5), 319–324. https://doi.org/10.1046/j.1365-2052.2003.01010.x

Aparicio, G. (1944). *Zootecnia especial: Etnología compendiada: descripción y estudio de razas nacionales y extranjeras* (3ª edición ed.). Imprenta Moderna.

Argamentería, A., de la Roza-Delgado, B., Cueto, M., Hidalgo, C., Tamargo, C., & Menéndez, J. (2012). Guía del Gochu Asturcelta. SERIDA. ISBN: 978-84-695-3048-1.

Argamentería, A., de la Roza-Delgado, B., Cueto, M., Hidalgo, C., Tamargo, C., Rodríguez, A., Fernández, A., Merino, M., & Menéndez, J. (2012). Manual del Gochu Asturcelta. SERIDA. ISBN: 978-84-695-3049-8.

Arias, K. D., Álvarez, I., Gutiérrez, J. P., Fernández, I., Menéndez, J., Menéndez-Arias, N. A., & Goyache, F. (2022). Understanding mendelian errors in SNP arrays data using a Gochu Asturcelta pig pedigree: Genomic alterations, family size and calling errors. *Scientific Reports*, *12*(1), 19686. https://doi.org/10.1038/s41598-022-24340-0

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, *13*(4), 969–980. https://doi.org/10.1111/j.1365-294X.2004.0225.x

Beja-Pereira, A., Bento, P., Ferrand, N., & Brenig, B. (2001). Genetic polymorphism of the 17th exon at the porcine *RYR1* locus: A new variant in a local Portuguese pig breed demonstrated by PCR-SSCP analysis. *Journal of Animal Breeding and Genetics*, *118*(4), 271–274. https://doi.org/10.1046/j.1439-0388.2001.00294.x

Bouchier-Hayes, L., Conroy, H., Egan, H., Adrain, C., Creagh, E. M., MacFarlane, M., & Martin, S. J. (2001). CARDINAL, a novel caspase recruitment domain protein, is an inhibitor of multiple NF-κB activation pathways. *Journal of Biological Chemistry*, *276*(47), 44069–44077. https://doi.org/10.1074/jbc.M107373200

Bovo, S., Ribani, A., Muñoz, M., Alves, E., Araujo, J. P., Bozzi, R., Čandek-Potokar, M., Charneca, R., Di Palma, F., Etherington, G., Fernández, A. I., García, F., García-Casco, J., Karolyi, D., Gallo, M., Margeta, V., Martins, J. M., Mercat, M. J., Moscatelli, G., … Fontanesi, L. (2020). Whole-genome sequencing of European autochthonous and commercial pig breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems. *Genetics Selection Evolution*, *52*(1), 33. https://doi.org/10.1186/s12711-020-00553-7

Browning, B. L., Tian, X., Zhou, Y., & Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics*, *108*(10), 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Cheng, J., Fernando, R., Cheng, H., Kachman, S. D., Lim, K., Harding, J. C. S., Dyck, M. K., Fortin, F., Plastow, G. S., Canada, P., & Dekkers, J. C. M. (2022). Genome-wide association study of disease resilience traits from a natural polymicrobial disease challenge model in pigs identifies the importance of the major histocompatibility complex region. *G3: Genes, Genomes, Genetics*, *12*(3), jkab441. https://doi.org/10.1093/g3journal/jkab441

Colonna, V., Ayub, Q., Chen, Y., Pagani, L., Luisi, P., Pybus, M., Garrison, E., Xue, Y., Tyler-Smith, C., & The 1000 Genomes Project Consortium. (2014). Human genomic regions with exceptionally high levels of population differentiation identified from 911 whole-genome sequences. *Genome Biology*, *15*(6), R88. https://doi.org/10.1186/gb-2014-15-6-r88

Conde, D., Cifuentes, M., & Fernández, L. (2010). La estación Pecuaria de Lugo (1932—1962). *Libro de Ponencias y Comunicaciones Del XVI Congreso Nacional y VII Iberoamericano de Historia de La Veterinaria*, 379–384. Imprenta de la Diputación Provincial de Córdoba.

Crespo-Piazuelo, D., Migura-Garcia, L., Estellé, J., Criado-Mesas, L., Revilla, M., Castelló, A., Muñoz, M., García-Casco, J. M., Fernández, A. I., Ballester, M., & Folch, J. M. (2019). Association between the pig genome and its gut microbiota composition. *Scientific Reports*, *9*(1), 8791. https://doi.org/10.1038/s41598-019-45066-6

Dadousis, C., Muñoz, M., Óvilo, C., Fabbri, M. C., Araújo, J. P., Bovo, S., Potokar, M. Č., Charneca, R., Crovetti, A., Gallo, M., García-Casco, J. M., Karolyi, D., Kušec, G., Martins, J. M., Mercat, M. J., Pugliese, C., Quintanilla, R., Radović, Č., Razmaite, V., … Bozzi, R. (2022). Admixture and breed traceability in European indigenous pig breeds and wild boar using genome-wide SNP data. *Scientific Reports*, *12*(1), 7346. https://doi.org/10.1038/s41598-022-10698-8

de la Roza-Delgado, B., Feito, I., Fuente-Maqueda, F., Modroño, S., Argamentería, A., & Ciordia, M. (2022). Influence of production system and feeds on performance, carcass traits and estimated energy balance of autochthonous Gochu Asturcelta pigs. *Spanish Journal of Agricultural Research*, *20*(3), e0604. https://doi.org/10.5424/sjar/2022203-18194

Ding, R., Quan, J., Yang, M., Wang, X., Zheng, E., Yang, H., Fu, D., Yang, Y., Yang, L., Li, Z., Liu, D., Cai, G., Wu, Z., & Yang, J. (2017). Genome-wide association analysis reveals genetic

loci and candidate genes for feeding behavior and eating efficiency in duroc boars. *PLoS One*, *12*(8), e0183244. https://doi.org/10.1371/journal.pone.0183244

Elbediwy, A., Zihni, C., Terry, S. J., Clark, P., Matter, K., & Balda, M. S. (2012). Epithelial junction formation requires confinement of Cdc42 activity by a novel SH3BP1 complex. *Journal of Cell Biology*, *198*(4), 677–693. https://doi.org/10.1083/jcb.201202094

Fonda, M. L. (1992). Purification and characterization of vitamin B6-phosphate phosphatase from human erythrocytes. *Journal of Biological Chemistry*, *267*(22), 15978–15983. https://doi.org/10.1016/S0021-9258(19)49630-0

Frantz, L. A. F., Haile, J., Lin, A. T., Scheu, A., Geörg, C., Benecke, N., Alexander, M., Linderholm, A., Mullin, V. E., Daly, K. G., Battista, V. M., Price, M., Gron, K. J., Alexandri, P., Arbogast, R.-M., Arbuckle, B., Bălăşescu, A., Barnett, R., Bartosiewicz, L., … Larson, G. (2019). Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. *Proceedings of the National Academy of Sciences*, *116*(35), 17231–17238. https://doi.org/10.1073/pnas.1901169116

Franz, M., Rodríguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G. D., & Morris, Q. (2018). GeneMANIA update 2018. *Nucleic Acids Research*, *46*, W60–W64. https://doi.org/10.1093/nar/gky311

Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D. M., Kim, H., Frantz, L. A. F., Caccamo, M., Ahn, H., Aken, B. L., Anselmo, A., Anthon, C., … Schook, L. B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, *491*(7424), 393–398. https://doi.org/10.1038/nature11622

Grossman, S. R., Shylakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., Lander, E. S., Schaffner, S. F., & Sabeti, P. C. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*, *327*(5967), 883–886. https://doi.org/10.1126/science.1183863

Guo, F., Dey, D. K., & Holsinger, K. E. (2009). A Bayesian hierarchical model for analysis of SNP diversity in multilocus, multipopulation samples. *Journal of the American Statistical Association*, *104*(485), 142–154. https://doi.org/10.1198/jasa.2009.0010

Hao, Q.-Q., Li, L., Chen, W., Jiang, Q.-Q., Ji, F., Sun, W., Wei, H., Guo, W.-W., & Yang, S.-M. (2018). Key genes and pathways associated with inner ear malformation in SOX10 $^{p.R109W}$ mutation pigs. *Frontiers in Molecular Neuroscience*, *11*, 181. https://doi.org/10.3389/fnmol.2018.00181

Herrero-Medrano, J., Megens, H.-J., Groenen, M. A. M., Ramis, G., Bosse, M., Pérez-Enciso, M., & Crooijmans, R. P. M. A. (2013). Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC Genetics*, *14*(1), 106. https://doi.org/10.1186/1471-2156-14-106

Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *International Journal of Plant Sciences*, *171*(9), 1059–1071. https://doi.org/10.1086/656306

Hsu, W.-T., Williamson, P., & Khatkar, M. S. (2023). Analysis of dog breed diversity using a composite selection index. *Scientific Reports*, *13*(1), 1674. https://doi.org/10.1038/s41598-023-28826-3

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B., & other members of the International Sheep Genomics Consortium. (2012). Genome-wide analysis of the World's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, *10*(2), e1001258. https://doi.org/10.1371/journal.pbio.1001258

Kinsella, R. J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., & Flicek, P. (2011). Ensembl BioMarts: A hub for data retrieval across taxonomic space. *Database*, *2011*, bar030. https://doi.org/10.1093/database/bar030

Kumar, C. G., Everts, R. E., Loor, J. J., & Lewin, H. A. (2010). Functional annotation of novel lineage-specific genes using co-expression and promoter analysis. *BMC Genomics*, *11*(1), 161. https://doi.org/10.1186/1471-2164-11-161

Larson, G., Dobney, K., Albarella, U., Fang, M., Smith, E. M., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Andersson, L., & Cooper, A. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, *307*(5715), 1618–1621. https://doi.org/10.1126/science.1106927

Lopez-Bote, C. J. (1998). Sustained utilization of the Iberian pig breed. *Meat Science*, *49*, 17–27. https://doi.org/10.1016/S0309-1740(98)90036-5

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, *23*(9), 2178–2192. https://doi.org/10.1111/mec.12725

Megdiche, S., Mastrangelo, S., Ben Hamouda, M., Lenstra, J. A., & Ciani, E. (2019). A combined multi-cohort approach reveals novel and known genome-wide signatures of selection for wool traits in Merino and Merino-derived sheep breeds. *Frontiers in Genetics*, *10*, 1025. https://doi.org/10.3389/fgene.2019.01025

Menéndez, J., Álvarez, I., Fernández, I., de la Roza-Delgado, B., & Goyache, F. (2015). Multiple paternity in domestic pigs under equally probable natural matings – A case study in the endangered Gochu Asturcelta pig breed. *Archives Animal Breeding*, *58*(1), 217–220. https://doi.org/10.5194/aab-58-217-2015

Menéndez, J., Álvarez, I., Fernández, I., & Goyache, F. (2016). Genealogical analysis of the Gochu Asturcelta pig breed: Insights for conservation. *Czech Journal of Animal Science*, *61*(3), 140–149. https://doi.org/10.17221/8787-CJAS

Menéndez, J., Goyache, F., Beja-Pereira, A., Fernández, I., Menéndez-Arias, N., Godinho, R., & Álvarez, I. (2016). Genetic characterisation of the endangered Gochu Asturcelta pig breed using microsatellite and mitochondrial markers: Insights for the composition of the Iberian native pig stock. *Livestock Science*, *187*, 162–167. https://doi.org/10.1016/j.livsci.2016.03.013

Moscarelli, A., Sardina, M. T., Cassandro, M., Ciani, E., Pilla, F., Senczuk, G., Portolano, B., & Mastrangelo, S. (2021).

Genome-wide assessment of diversity and differentiation between original and modern Brown cattle populations. *Animal Genetics*, *52*(1), 21–31. https://doi.org/10.1111/age.13019

Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., & Morris, Q. (2008). GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function. *Genome Biology*, *9*(Suppl 1), S4. https://doi.org/10.1186/gb-2008-9-s1-s4

Muñoz, M., Bozzi, R., García, F., Núñez, Y., Geraci, C., Crovetti, A., García-Casco, J., Alves, E., Škrlep, M., Charneca, R., Martins, J. M., Quintanilla, R., Tibau, J., Kušec, G., Djurkin-Kušec, I., Mercat, M. J., Riquet, J., Estellé, J., Zimmer, C., … Óvilo, C. (2018). Diversity across major and candidate genes in European local pig breeds. *PLoS One*, *13*(11), e0207475. https://doi.org/10.1371/journal.pone.0207475

Muñoz, M., Bozzi, R., García-Casco, J., Núñez, Y., Ribani, A., Franci, O., García, F., Škrlep, M., Schiavo, G., Bovo, S., Utzeri, V. J., Charneca, R., Martins, J. M., Quintanilla, R., Tibau, J., Margeta, V., Djurkin-Kušec, I., Mercat, M. J., Riquet, J., … Óvilo, C. (2019). Genomic diversity, linkage disequilibrium and signatures of selection in European local pig breeds assessed with a high density SNP chip. *Scientific Reports*, *19*(1), 13546. https://doi.org/10.1038/s41598-019-49830-6

Oleksyk, T. K., Smith, M. W., & O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *365*(1537), 185–205. https://doi.org/10.1098/rstb.2009.0219

Qiu, H., Zhao, S., Yu, M., Fan, B., & Liu, B. (2008). Expression patterns and subcellular localization of porcine (sus scrofa) lectin, galactose-binding, soluble 1 gene. *Acta Biochimica et Biophysica Sinica*, *40*(1), 85–90. https://doi.org/10.1111/j.1745-7270.2008.00373.x

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramírez, O., Ojeda, A., Tomas, A., Gallardo, D., Huang, L. S., Folch, J. M., Clop, A., Sanchez, A., Badaoui, B., Hanotte, O., Galman-Omitogun, O., Makuza, S. M., Soto, H., Cadillo, J., Kelly, L., Cho, I. C., Yeghoyan, S., Perez-Enciso, M., & Amills, M. (2009). Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. *Molecular Biology and Evolution*, *26*(9), 2061–2072. https://doi.org/10.1093/molbev/msp118

Randhawa, I. A. S., Khatkar, M., Thomson, P., & Raadsma, H. (2014). Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. *BMC Genetics*, *15*(1), 34. https://doi.org/10.1186/1471-2156-15-34

Randhawa, I. A. S., Khatkar, M. S., Thomson, P. C., & Raadsma, H. W. (2015). Composite selection signals for complex traits exemplified through bovine stature using multibreed cohorts of European and African *Bos taurus*. *G3: Genes, Genomes, Genetics*, *5*(7), 1391–1401. https://doi.org/10.1534/g3.115.017772

Rodríguez-Valdovinos, C., García-Casco, J., García-Ruiz, F., Nuñez-Moreno, Y., & Silió-López, L. (2019). A further look of the genetic origin and singularity of the Torbiscal Iberian pig line. *Spanish Journal of Agricultural Research*, *17*(1), e0402. https://doi.org/10.5424/sjar/2019171-13618

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. https://doi.org/10.1038/nature01140

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D., & Lander, E. S. (2006). Positive natural selection in the human lineage. *Science*, *312*(5780), 1614–1620. https://doi.org/10.1126/science.1124309

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., & Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913–918. https://doi.org/10.1038/nature06250

Santos Silva, J., Pedro Araújo, J., Orlando Cerqueira, J., Pires, P., Alves, C., & Batorek-Lukač, N. (2019). Bísaro Pig. In M. Candek-Potokar & R. M. Nieto Linan (Eds.), *European local pig breeds—Diversity and performance. A study of project TREASURE*. IntechOpen. https://doi.org/10.5772/intechopen.83759

Schiavo, G., Bovo, S., Muñoz, M., Ribani, A., Alves, E., Araújo, J. P., Bozzi, R., Čandek-Potokar, M., Charneca, R., Fernández, A. I., Gallo, M., García, F., Karolyi, D., Kušec, G., Martins, J. M., Mercat, M. J., Núñez, Y., Quintanilla, R., Radović, Č., … Fontanesi, L. (2021). Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds. *Animal Genetics*, *52*(2), 155–170. https://doi.org/10.1111/age.13045

Scott, D. D., Trahan, C., Zindy, P. J., Aguilar, L. C., Delubac, M. Y., Van Nostrand, E. L., Adivarahan, S., Wei, K. E., Yeo, G. W., Zenklusen, D., & Oeffinger, M. (2017). Nol12 is a multifunctional RNA binding protein at the nexus of RNA and DNA metabolism. *Nucleic Acids Research*, *45*(21), 12509–12528. https://doi.org/10.1093/nar/gkx963

Silió, L. (2000). Iberian pig breeding programme. In S. Galal, J. Boyazouglu, & K. Hammond (Eds.), *Developing breeding strategies for lower input animal production environments* (pp. 511–520). International Committee for Animal Recording.

Sotillo, J. L., & Serrano, V. (1985). *Producción animal: Etnología zootécnica*. Ediciones Tebar-Flores.

Szpiech, Z. A., & Hernandez, R. D. (2014). Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827. https://doi.org/10.1093/molbev/msu211

van Asch, B., Pereira, F., Santos, L. S., Carneiro, J., Santos, N., & Amorim, A. (2012). Mitochondrial lineages reveal intense gene flow between Iberian wild boars and south Iberian pig breeds: MtDNA lineages of Iberian wild boars and pigs. *Animal Genetics*, *43*(1), 35–41. https://doi.org/10.1111/j.1365-2052.2011.02222.x

Wang, W., Huang, L., Huang, Y., Yin, J., Berk, A. J., Friedman, J. M., & Wang, G. (2009). Mediator MED23 links insulin signaling to the adipogenesis transcription Cascade. *Developmental Cell*, *16*(5), 764–771. https://doi.org/10.1016/j.devcel.2009.04.006

Wang, X., Wang, L., Shi, L., Zhang, P., Li, Y., Li, M., Tian, J., Wang, L., & Zhao, F. (2022). GWAS of reproductive traits in large white pigs on Chip and Imputed whole-genome sequencing

data. *International Journal of Molecular Sciences*, *23*(21), 13338. https://doi.org/10.3390/ijms232113338

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358. https://doi.org/10.2307/2408641

Wilkinson, S., Lu, Z. H., Megens, H.-J., Archibald, A. L., Haley, C., Jackson, I. J., Groenen, M. A. M., Crooijmans, R. P. M. A., Ogden, R., & Wiener, P. (2013). Signatures of diversifying selection in European pig breeds. *PLoS Genetics*, *9*(4), e1003453. https://doi.org/10.1371/journal.pgen.1003453

Xu, Z., Sun, Y., Jiang, J., & Liu, P. (2022). The role of KIAA1191 in the necroptotic pathway of multiple myeloma. *Annals of Hematology*, *101*(2), 359–367. https://doi.org/10.1007/s00277-021-04710-w

Zhang, L., Liang, J., Luo, W., Liu, X., Yan, H., Zhao, K., Shi, H., Zhang, Y., Wang, L., & Wang, L. (2014). Genome-wide scan reveals LEMD3 and WIF1 on SSC5 as the candidates for porcine ear size. *PLoS One*, *9*(7), e102085. https://doi.org/10.1371/journal.pone.0102085

Zhang, W., Liu, L., Zhou, M., Su, S., Dong, L., Meng, X., Li, X., & Wang, C. (2023). Assessing population structure and signatures of selection in Wanbei pigs using whole genome resequencing data. *Animals*, *13*(1), 13. https://doi.org/10.3390/ani13010013

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Arias, K. D., Lee, H., Bozzi, R., Álvarez, I., Gutiérrez, J. P., Fernandez, I., Menéndez, J., Beja-Pereira, A., & Goyache, F. (2023). Ascertaining the genetic background of the Celtic-Iberian pig strain: A signatures of selection approach. *Journal of Animal Breeding and Genetics*, *00*, 1–17. https://doi.org/10.1111/jbg.12829