Received: 13 June 2023 | Accepted: 15 June 2023

DOI: 10.1111/age.13344

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

Genome-wide population structure, homozygosity, and heterozygosity patterns of Nero Siciliano pig in the framework of Italian and cosmopolitan breeds

Salvatore Bordonaro[1] | **Giorgio Chessari**[1] | **Salvatore Mastrangelo**[2] | **Gabriele Senczuk**[3] | **Stefania Chessa**[4] | **Bianca Castiglioni**[5] | **Serena Tumino**[1] | **Donata Marletta**[1] | **Andrea Criscione**[1]

[1]Dipartimento di Agricoltura, Alimentazione e Ambiente, Università di Catania, Catania, Italy

[2]Dipartimento Scienze Agrarie, Alimentari e Forestali, Università di Palermo, Palermo, Italy

[3]Dipartimento di Agricoltura, Ambiente e Alimenti, Università del Molise, Campobasso, Italy

[4]Dipartimento di Scienze Veterinarie, Università di Torino, Torino, Italy

[5]Istituto di Biologia e Biotecnologia Agraria, National Research Council, Lodi, Italy

Correspondence
Serena Tumino, Dipartimento di Agricoltura, Alimentazione e Ambiente, Università di Catania, 95123, Catania, Italy.
Email: serena.tumino@unict.it

Funding information
Part of this research was funded by the project "QUALIGEN"; Linea 2 - Piano di Incentivi per la Ricerca di Ateneo 2020/2022; P.I. Giuseppe Luciano.

Abstract

Analysis of genomic data is becoming more and more common for the effective management of livestock breeding programmes, even in the case of local populations. In this work, the genome-wide data of Nero Siciliano pig breed were compared to that of wild boar, Italian local and cosmopolitan breeds to investigate its genetic structure, and runs of homozygosity (ROH) and heterozygosity patterns. The Nero Siciliano has been reported to have the highest rate of genetic diversity among the Italian breeds, and a genetic variability comparable to that of the cosmopolitan breeds. Analyses of genomic structure and relationships underlined its proximity to wild boar, and an internal substructure probably linked to different family lines. The breed showed a low value of inbreeding estimated from ROH, and the highest diversity index among the Italian breeds, even if lower than that of the cosmopolitans. Four ROH islands in three chromosomes (SSC8, SSC11, and SSC14) and one heterozygosity-rich region (SSC1) were identified in Nero Siciliano, highlighting genomic regions related to productive QTL. Across breeds, SSC8 and SSC14 were the chromosomes with most ROH islands, with Mora Romagnola and wild boar showing the highest level of autozygosity. Chromosomes SSC2, SSC6, SSC8 and SSC13 showed the majority of runs of heterozygosity regions, mainly found in the cosmopolitan pig breeds, which reported several genes associated with health-related QTL. The outlined results can help to better identify the genomic profile of this local breed in order to plan matings, maintain adequate internal diversity and exploit the production system.

KEYWORDS
autochthonous pigs, genetic diversity, runs of homozygosity/heterozygosity, SNPs

Salvatore Bordonaro and Giorgio Chessari contributed equally to this work.

INTRODUCTION

The widespread use of cosmopolitan and highly productive breeds has resulted in severe biodiversity loss, threatening the integrity of many local populations. However, native breeds often reared in marginal areas are gaining interest nowadays (Candek-Potokar & Nieto Linan, 2019). Local breeds not only represent a cultural and genetic heritage but are also a fundamental genetic resource thanks to their strong disease resistance and their adaptation to harsh environments and therefore are essential for the sustainable exploitation of natural resources. These breeds require careful genomic characterisation to highlight their genetic diversity and provide efficient means for conservation strategies and breeding programmes (Bovo, Ribani, Munoz, Alves, Araujo, Bozzi, Candek-Potokar, et al., 2020; Franci & Pugliese, 2007). In Italy, conservation programmes for local pig breeds started in the 1980s to avoid the complete extinction of these genetic resources. Currently, the National Association of Pig Breeders (ANAS) has officially recognised and manages eight Italian autochthonous pig breeds. Among these, the Nero Siciliano pig breed represents the Sicilian heritage since the Greek and Carthaginian dominations (7th–6th century BC). In the 19th century the population size of Nero Siciliano decreased dramatically due to widespread breeding of highly prolific and productive English pig breeds (Large Black and Berkshire) and also due to extensive deforestation which reduced its breeding area. (Franci & Pugliese, 2007; Zumbo et al., 2020). In 2001, Nero Siciliano was officially recognised as a breed, and then, national and regional projects dealt with its safeguarding and exploitation. Nowadays, about 6500 animals are recorded in the Herd Book, of which ~800 are sows and boars, while the rest are animals reared for producing fresh or processed meat. According to the breed standard, Nero Siciliano must have a black coat with long bristles in the dorsal–lumbar region, long black or partially white head with a straight front-nasal profile (ANAS, 2022). The breed has always been reared in the northeast of Sicily in semi-extensive or extensive systems, exploiting marginal woody lands through eco-compatible practices. In such a context, the breed has acquired a specific adaptation to difficult breeding conditions, including high disease resistance, good prolificacy and good adaptation to poor-quality feeding (Franci & Pugliese, 2007; Guastella et al., 2010; Zumbo et al., 2020).

In livestock research, BeadChip arrays are the most widely used genome-wide tool for many topics including genomic characterisation and structure, selection signature studies and the detection of QTL (Ramos et al., 2009). Similarly to other species, genome-wide analyses in pig have been carried out at local, European and global levels, to estimate the rate of genetic diversity, population structure and patterns of differentiation (e.g. D'Alessandro et al., 2019; Herrero-Medrano et al., 2014; Iacolina et al., 2016; Muñoz et al., 2019; Schiavo, Bovo, Tinarelli, et al., 2020; Yang et al., 2017).

In the current study, we compared the SNP data of Nero Siciliano with those of the other Italian local and cosmopolitan pig breeds, with the aim to understand the genetic variability, population structure, runs of homozygosity (ROH) and heterozygosity (ROHet) patterns. The genetic characterisation and understanding of genome structure are fundamental steps for the development of conservation programmes and for the enhancement of the production system of this local resource.

MATERIALS AND METHODS

DNA sampling

In total, 80 Nero Siciliano pigs, from 16 different farms, were selected for genotyping. No pedigree data were available except for a few animals. Pigs were chosen based on their phenotypic profile and the information supplied by farmers to collect animals as much unrelated as possible and try to avoid admixed or cross-species animals. Individual peripheral blood (10 mL) was collected in Vacutainer tubes containing EDTA at pH 8.0 as an anticoagulant. Authorised personnel, avoiding any suffering to animals, performed the sampling procedures following Directive 2010/63 /EU. DNA was extracted from blood using the commercial Illustra blood genomic Prep Mini Spin kit (GE Healthcare).

Genotyping, quality control and data handling

All animals were genotyped with the Illumina PorcineSNP60 v2 BeadChip (Illumina). The PiHat test was performed on Nero Siciliano pigs (NSIC = 80) using PLINK ver. 1.9 (Chang et al., 2015). Based on the within-breed average relatedness, no individual was removed. The SNP data of NSIC were then merged with the genotypic data retrieved from previous studies (Iacolina et al., 2016; Yang et al., 2017) also obtained using the PorcineSNP60 BeadChip. The final dataset consisted of 281 individuals belonging to five local breeds, four cosmopolitan breeds, and a sample of wild boar. In detail, the dataset was assembled as follows (Table 1): Calabrese (ITCA = 15), Casertana (ITCT = 15), Cinta Senese (ITCS = 13), Mora Romagnola (ITMR = 9), Italian Wild Boar (WBIT = 15), Duroc (PDUR = 30), Landrace (PLDR = 30), Large White (PLWT = 30), Pietrain (PPIT = 30) and another sample of Nero Siciliano (ITNS = 15) which was included for comparison with the experimental sample (NSIC). The software PLINK ver. 1.9 (Chang et al., 2015) was used to perform filtering and quality control. Chromosomal coordinates for each SNP were obtained using the latest version of the assembled pig genome (*Sscrofa 11.1*). Quality control on genotype

TABLE 1 Population's full name (breed), acronymous (code), number of individuals (size), type of classification (type) and related reference (ref) of the pig populations taken into analyses.

Breed	Code	Size	Type	References
Calabrese	ITCA	15	Autochthonous	Yang et al. (2017)
Casertana	ITCT	15	Autochthonous	Yang et al. (2017)
Cinta Senese	ITCS	13	Autochthonous	Yang et al. (2017)
Duroc	PDUR	30	Cosmopolitan	Yang et al. (2017)
Italian Wild Boar	WBIT	15	Wild	Iacolina et al. (2016)
Landrace	PLDR	30	Cosmopolitan	Yang et al. (2017)
Large White	PLWT	30	Cosmopolitan	Yang et al. (2017)
Mora Romagnola	ITMR	9	Autochthonous	Yang et al. (2017)
Nero Siciliano	NSIC	80	Autochthonous	This study
Nero Siciliano	ITNS	15	Autochthonous	Yang et al. (2017)
Pietrain	PPIT	30	Cosmopolitan	Yang et al. (2017)
Sus verrucosus	SVV	8	Ancestral	Yang et al. (2017)

data was performed using the following criteria: a minor allele frequency ≥0.02, a genotype call rate for an SNP ≥0.95 and an individual call rate ≥0.90, which resulted in 39 752 SNPs and 281 pigs.

Genetic diversity indices

PLINK ver. 1.9 (Chang et al., 2015) was used to estimate within-population genetic diversity coefficients, namely observed (H_O) and expected heterozygosity (H_E), inbreeding coefficient relative to each population (F_{IS}), and the minor allele frequencies (*MAF*). Contemporary effective population size (*cNe*) was estimated using NEESTIMATOR ver. 2.1 (Do et al., 2014), using the random mating option within the linkage disequilibrium method proposed by Waples and Do (2010). Historical effective population size (*hNe*) was also calculated using the GONE software (https://github.com/esrud/GONE) by default setting options (Santiago et al., 2020).

Genetic relationship and population structure

To explore the genetic relationships and the population structure, the genotype data were filtered to remove SNPs in high linkage disequilibrium ($r^2 > 0.2$) by using the *--indep-pairwise 50 10 0.2* function in PLINK ver. 1.9 (Chang et al., 2015), thus generating a pruned dataset of 15 239 SNPs. PLINK ver. 1.9 was also used to perform multidimensional scaling (MDS) analysis, based on the pairwise identity-by-state distances between individuals. The genetic structure was estimated by the software ADMIXTURE ver. 1.3.0 (Alexander et al., 2009), using the model-based clustering algorithm to range from $K=2$ to $K=25$ and a tenfold cross-validation procedure. The estimated ancestral matrices were then plotted using the R package *BITE* ver. 1.2.0008 (Milanesi et al., 2017) under the open-source programming environment for statistical analysis R (R Development Core Team, 2020). A Neighbor-Net was constructed on the basis of pairwise Reynolds' genetic distances inferred by ARLEQUIN ver. 3.5.2.2 (Excoffier & Lischer, 2010) and plotted using SPLITSTREE4 ver. 4.14.8 (Huson & Bryant, 2006). F_{ST} pairwise distances were calculated by ARLEQUIN ver. 3.5.2.2 (Excoffier & Lischer, 2010). Finally, to reconstruct genomic relationships among all populations and to investigate the presence of patterns of population splitting and migration events, a maximum likelihood tree has been generated using the software TREEMIX (Pickrell & Pritchard, 2012). For this analysis, a sample of *Sus verrucosus* ($n=8$) was used as an outgroup (Yang et al., 2017). Five independent iterations were performed allowing migration events to range between 1 and 10, while the covariance matrix was estimated using 500 contiguous SNPs per block. The most supported number of migration edges was assessed using the Evanno method as implemented in the R package *OptM* (Fitak, 2021).

ROH and ROHet detection

ROHs and ROHetwere detected through the R package *detectRUNS* ver. 0.9.6 (Biscarini et al., 2018), using both SNP windows (or Sliding Windows) and consecutive SNPs (or Consecutive Runs) methods.

In accordance with the literature (Ruan et al., 2022; Schiavo et al., 2021), the following parameters have been set in the sliding windows (SW) method for ROH and ROHet detection: (i) the minimum number of SNPs in a run was 15; (ii) no missing or opposite genotypes were allowed in the window; (iii) sliding window of 50 SNPs for ROH and of 15 SNPs for ROHet; (iv) the minimum density of one SNP every 100 kb; (v) the threshold to call an SNP within a run was set to 0.05; (vi) the maximum gap between consecutive SNPs was set to 1 Mb; (vii) the minimum run length was set to 1 Mb.

As regard the consecutive runs (CR) method the following setting was used for ROH and ROHet detection: (i) the minimum number of consecutive SNPs included

in a run was 15; (ii) the number of missing or opposite genotypes were set to zero; (iii) the maximum gap between consecutive SNPs was set to 1 Mb; (iv) the minimum run length was set to 1 Mb.

The mean number of ROH/ROHet per breed, the average length of ROH/ROHet per chromosome within each breed, as well as the sum of ROH/ROHet per breed, were calculated. In addition, the total length of the genome covered by ROH and ROHet was divided by the total autosomal genome length covered by the SNP array (~2.25 Gb) to evaluate the individual genomic inbreeding coefficient (F_{ROH}) and coefficient of diversity (D_{ROHet}), respectively. Pearson correlation was calculated for the aforementioned parameters obtained with the two methods.

Highly homozygous (ROH islands) and heterozygous (ROHet islands) genomic regions were identified as follows: the SNP-within-ROH/ROHet incidences per population were transformed into standard normal z-scores. Based on z-scores, *p*-values were calculated and the top 0.5% of SNPs were selected. The markers with *p*-value ≥0.995 identified with both methods (Sliding Windows and Consecutive Runs) were considered to identify ROH and ROHet islands. Finally, we defined ROH and ROHet islands those regions identified by overlapping markers obtained with the two methods of detection.

The genomic coordinates of ROH and ROHet islands were examined using the Ensemble browser for the pig genome, according to the assembly *Sscrofa 11.1* (https://www.ensembl.org/index.html), in order to retrieve annotated gene lists. The Pig Quantitative Trait Locus Database (Pig QTLdb) (https://www.animalgenome.org/cgi-bin/QTLdb/SS/index) was used to search for possible associations between markers and reported QTL in pig species, as well as to clarify the gene's identity and functions.

Gene ontology (GO) and the enrichment analysis of annotated genes were conducted using the open-source Database for Annotation, Visualization, and Integrated Discovery ver. 2021 package (https://david-d.ncifcrf.gov) (Huang da et al., 2009). For the GO terms and Kyoto Encyclopedia of Genes and Genomes pathway analysis, the level of significance was set as $p<0.05$. GO terms outcomes corresponded to the highest specificity. Corrections for multiple testing were made by applying the Bonferroni test.

RESULTS

Genetic diversity indices

The summary statistics of genetic diversity are reported in Table 2. PLDR, PLWT and PPIT showed the highest values of H_E and H_O, followed by NSIC, while ITMR reported the lowest values. High F_{IS} values were observed in ITCS, WBIT, and particularly in ITMR. Lower F_{IS} values were found for the three cosmopolitan breeds and NSIC, whereas intermediate values were reported for ITNS, ITCA, ITCT, and PDUR. Among the autochthonous breeds, ITMR showed the lowest *MAF* value and NSIC the highest; similar *MAF* values were reported among the cosmopolitan breeds. The contemporary (*cNe*) effective population size is reported in Table 2. NSIC had a *cNe*=34.3. The cosmopolitan breeds showed the highest values, ranging from 30.4 of PPIT to 50.7 of PLDR, whereas Italian local breeds and WBIT had the lowest values ranging from 2.5 of ITCT to 20.1 of ITCS. The variation of historical effective population size (*hNe*) over 100 generations is shown in Figure S1 as a logarithmic trend. Values at the 1st generation resembled those of *cNe*, then increased significantly starting from the 4th generation, and tended to stabilise at ~19th generation. A notable rise in the curve was observed at the 31st generation for WBIT, and at the 71st generation for all the cosmopolitan breeds and NSIC.

TABLE 2 Genetic diversity indices for the analysed pig populations.

Breed	H_O±SD	H_E±SD	F_{IS}±SD	*MAF*±SD	*cNe*
NSIC	0.309±0.146	0.325±0.147	0.167±0.103	0.241±0.143	34.3
ITNS	0.288±0.207	0.283±0.177	0.222±0.082	0.211±0.158	6.0
ITCA	0.296±0.245	0.255±0.198	0.203±0.127	0.194±0.169	2.8
ITCS	0.212±0.201	0.212±0.188	0.430±0.123	0.156±0.160	20.1
ITCT	0.281±0.206	0.288±0.183	0.241±0.225	0.217±0.161	2.5
ITMR	0.161±0.240	0.133±0.187	0.565±0.060	0.099±0.150	8.8
WBIT	0.197±0.197	0.212±0.197	0.470±0.068	0.158±0.164	14.8
PDUR	0.269±0.179	0.291±0.180	0.273±0.049	0.220±0.161	45.6
PLDR	0.323±0.157	0.347±0.150	0.130±0.069	0.265±0.146	50.7
PLWT	0.315±0.150	0.352±0.146	0.150±0.064	0.269±0.143	47.4
PPIT	0.329±0.171	0.335±0.157	0.113±0.052	0.254±0.149	30.4

Note: Observed (H_O) and expected heterozygosity (H_E), population inbreeding coefficient (F_{IS}), average minor allele frequency (*MAF*), contemporary effective population size (*cNe*), and standard deviation (SD). For the full definition of breeds, see Table 1.

Population structure

According to the first two components of inferred variance (C1 = 12.84% and C2 = 12.29%), the MDS analysis (Figure 1) highlighted a clear separation between cosmopolitan and autochthonous breeds. In particular, PLDR, PLWT, and PPIT clustered as a cohesive group, while PDUR separated from the rest of the sample. Component C2 showed a gradient of variation from left to right, which, almost without discontinuity, identified first WBIT and NSIC, then the other admixed black pigs, then the PDUR and finally the cluster of other cosmopolitan breeds (PLDR, PLWT, and PPIT). The C1 component did not discriminate the WBIT from the local breeds. ITMR was detached from the rest of the local breeds, which were not separated from each other. In fact, NSIC, ITCA, and ITCS partially overlapped, whereas a total overlapping was observed between NSIC and ITNS, as expected. Within the NSIC sample, a group of 12 animals was close to WBIT.

In the admixture analysis, the distribution of the cross-validation errors of the inferred clusters (K = 25) suggested K = 13 as the most likely number of ancestries (Figure S2). The circle plot (Figure 2), which refers to the genomic structure of the 11 populations, showed the most representative ancestral clusters from K = 2 to K = 13. At $K = 2$, the genomic pattern of populations reproduced the evidence of the first component of MDS by separating the PDUR breed from the other populations. At $K = 3$, three ancestral clusters were distinguished, represented respectively by PDUR, the rest of the cosmopolitan breeds, and the local breeds together with WBIT. At $K = 4$, ITMR clustered separately from the other Italian local breeds. Further clusters were formed by PLDR and PPIT ($K = 5$), followed by PLWT and WBIT ($K = 7$). NSIC and ITNS shared a common ancestry with the other local breeds until $K = 7$, whereas at $K = 9$ they formed their own cluster. Moreover, from $K = 9$ to $K = 13$, NSIC showed a heterogeneous genetic pattern with evidence of a sub-structure. In general, PDUR, PPIT, and PLDR reported the highest genomic divergence among the cosmopolitan breeds, whereas ITCA, ITCS, and, particularly, ITMR highlighted the lowest rate of admixture within the local breeds. WBIT evidenced high internal homogeneity and showed its genomic influence on autochthonous breeds (from $K = 3$ to $K = 5$) despite its divergent origin.

The Neighbor-Net based on pairwise Reynolds' genetic distances (Figure 3) showed a complementary picture of the genetic relationships among the breeds. Consistent with the MDS plot, the figure revealed two main clusters composed of NSIC, ITNS, WBIT, ITCA, and ITCS on one side, and PLDR, PLWT, PPIT, PDUR, ITMR, and ITCT on the other side. Within the first group, WBIT was closer to ITNS and NSIC than to the other breeds. Within the second group, ITMR departed from the same phylogenetic node with PDUR and highlighted the longest branch of the network; ITCT was in an intermediate position between the cluster formed by PLDR, PLWT, and PPIT and the ITMR-PDUR group. The F_{ST} pairwise distances (Table S1) highlighted the highest values for the comparisons ITMR vs WBIT (0.453), vs ITCS (0.451) and vs ITCA (0.422), while the lowest value was 0.048 between NSIC and ITNS. NSIC showed the lowest average distance (0.132) and relatively low values towards ITCT (0.116), PLDR (0.128) and PLWT (0.130).

The TreeMix dendrogram (Figure 4), basically confirmed the previous analyses showing that all Italian local breeds clustered together, except the ITCA. However, the ITCA was part of a group placed in a basal position with respect to the Italian breeds, which also included PDUR and PLDR breeds. The Italian Wild Boar fell within the Italian local breeds, while the other cosmopolitan breeds were in a basal position. The Evanno method indicated eight as the most supported number of migration events. Two migration edges were shown from the base of the outgroup (*Sus verrucosus*) to PLDR and PDUR, another migration event was highlighted from the base of the branch including all breeds to ITCT. WBIT showed two migration edges towards ITCA and the node including ITCS

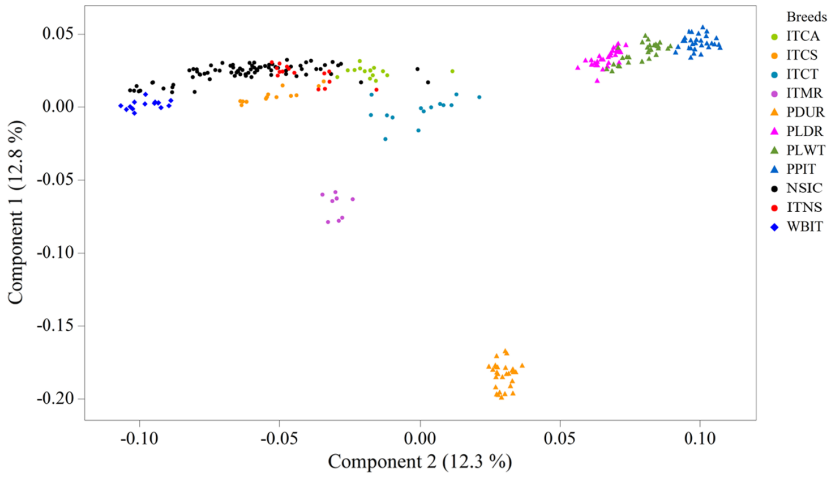

FIGURE 1 Genetic relationship based on the multidimensional scaling analysis. For a full definition of the breeds, see Table 1.

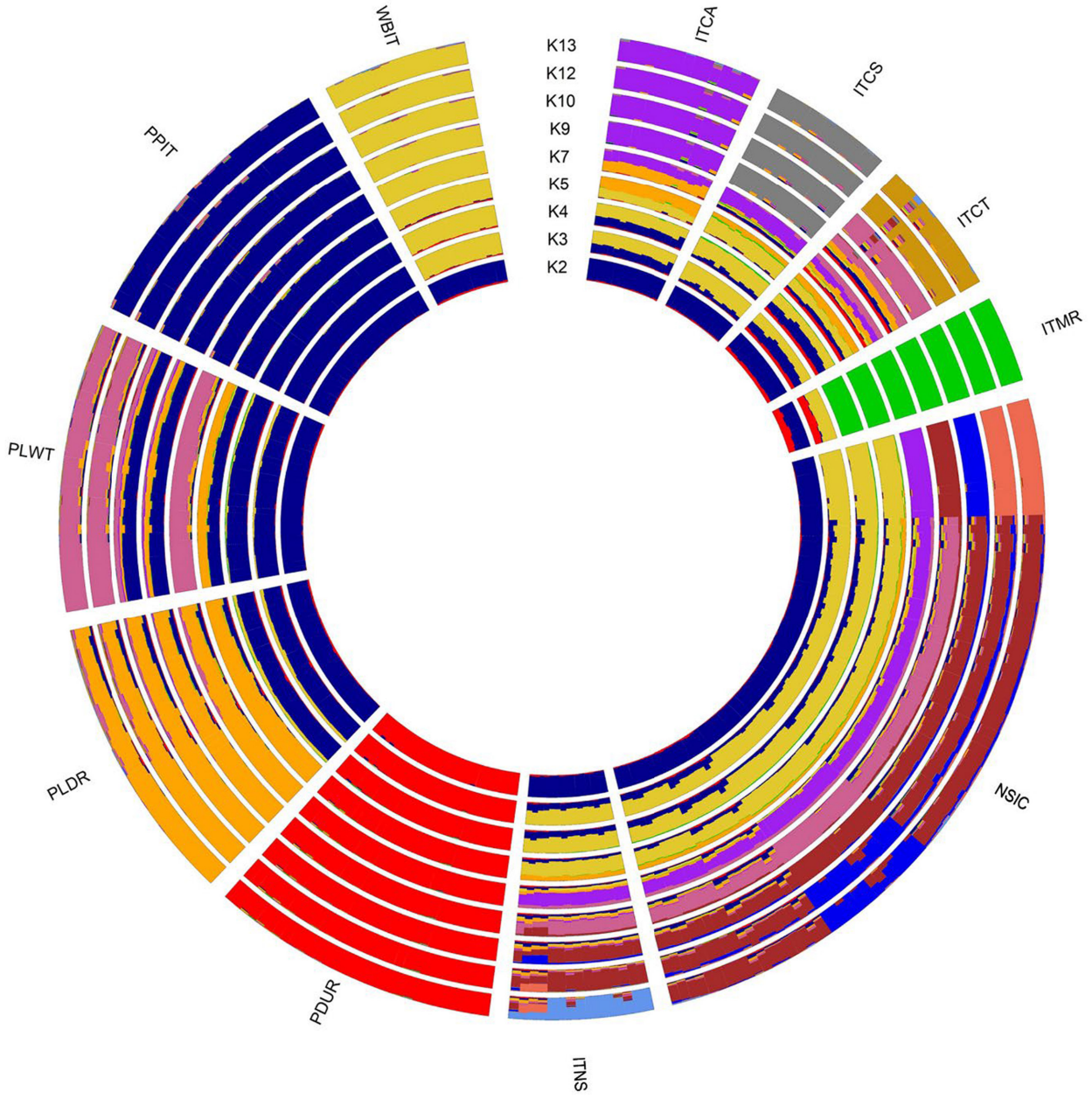

FIGURE 2 Circle plot showing ancestral clusters (K) inferred by ADMIXTURE analysis of the 10 pig breeds. For a full definition of the breeds, see Table 1.

and ITMR. The strongest migration weight was estimated from ITMR to PDUR while the weakest signal was detected from PDUR to ITCT. NSIC branched together with ITNS, ITCT, and WBIT and reported no migration events, whereas a weak migration edge was shown from the base of ITCA to ITNS.

ROH and ROHet detection

Table 3 summarises the mean number of ROH and ROHet per breed (N_{ROH}, N_{ROHet}), the mean ROH and ROHet length expressed in Mb (L_{ROH}, L_{ROHet}), the mean genome length covered by ROH and ROHet in Mb (S_{ROH}, S_{ROHet}), the inbreeding coefficient estimated from ROH (F_{ROH}), and the diversity coefficient estimated from ROHet (D_{ROHet}), obtained with SW and CR methods of detection.

All the individuals reported ROH identified by SW and CR methods: only one pig (ITCA) lacked homozygous segments when investigated by SW (data not shown). Within each method, the parameter values varied between breeds. The Pearson correlation coefficient of each parameter inferred with the two different methods was close to one for S_{ROH} and F_{ROH} ($r = 0.98$), whereas lower for N_{ROH} and L_{ROH} ($r = 0.72$ and $r = 0.75$, respectively). ITCS, ITMR, and PDUR always showed high values of N_{ROH}, whereas WBIT reported the

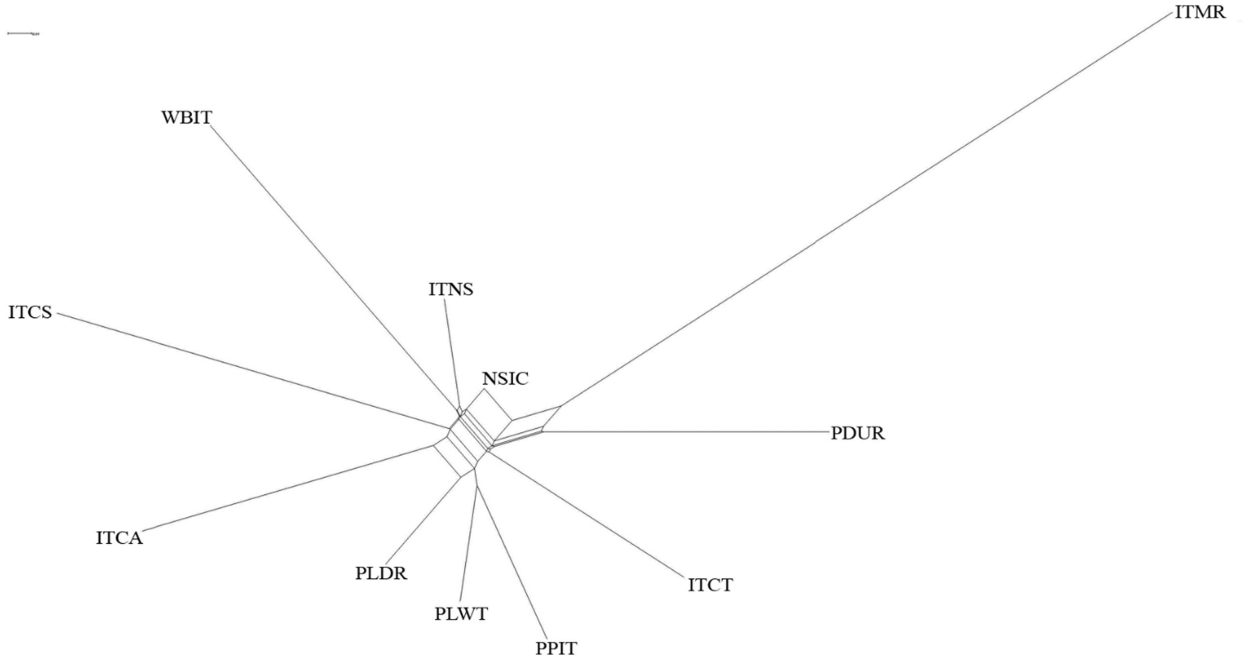

FIGURE 3 Neighbor-Net based on Reynolds' pairwise genetic distances among the 10 pig breeds. For a full definition of the breeds, see Table 1.

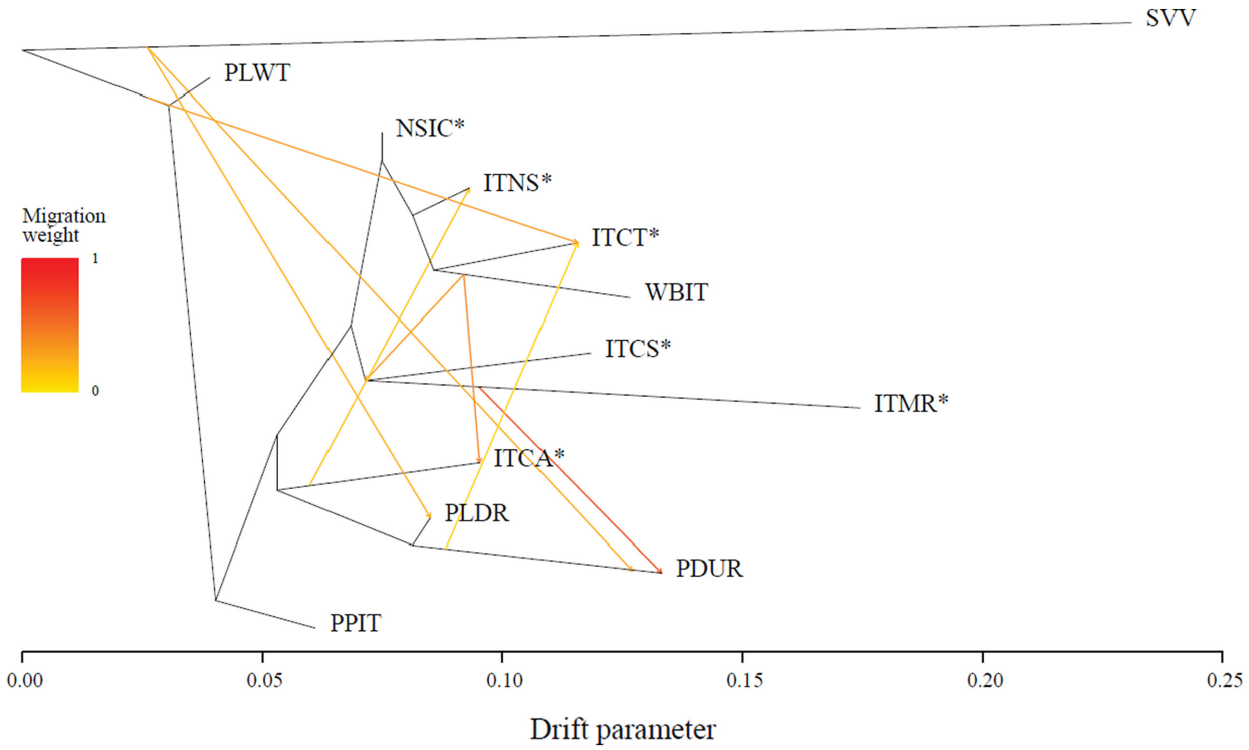

FIGURE 4 Maximum likelihood tree inferred using the software TreeMix with coloured arrows indicating the strength and direction of the migration events.

highest mean number of ROH for CR and intermediate value for SW. NSIC (SW = 20.39, CR = 127.55) and the rest of the local breeds reported relatively low N_{ROH}, while PLDR, PLWT and PPIT showed moderate values. As regards L_{ROH}, it reported slightly different rankings between the two methods. ITCT and ITMR always showed the highest mean length values, NSIC and ITNS reported high length for SW (9.6 Mb; 10.0 Mb) and the lowest for CR (3.2 Mb), ITCS and ITCA had relatively higher values for SW than CR. Cosmopolitan breeds reported the lowest L_{ROH} values (SW = 6.7–7.3 Mb; CR = 3–4.2 Mb). S_{ROH} showed the

TABLE 3 Descriptive statistics for runs of homozygosity (ROH) and runs of heterozygosity (ROHet) analysis through sliding windows (SW) and consecutive runs (CR) methods on the analysed populations.

		ROH			ROHet	
		SW	CR		SW	CR
Breed	Index	Mean±SD	Mean±SD	Index	Mean±SD	Mean±SD
NSIC	N_{ROH}	20.39±13.77	127.55±14.30	N_{ROHet}	2.25±1.48	2.79±1.74
	L_{ROH}	9.6±2.7	3.2±0.6	L_{ROHet}	0.92±0.61	1.12±0.69
	S_{ROH}	200.5±204.0	394.3±194.7	S_{ROHet}	2.97±1.92	4.23±2.74
	F_{ROH}	0.09±0.09	0.17±0.09	D_{ROHet}	0.0013±0.0009	0.0019±0.0012
ITNS	N_{ROH}	21.47±13.42	140.93±12.75	N_{ROHet}	1.07±1.16	1.40±1.40
	L_{ROH}	10.0±2.4	3.2±0.6	L_{ROHet}	0.49±0.63	0.76±0.88
	S_{ROH}	216.8±254.3	431.6±227.4	S_{ROHet}	1.31±1.49	2.04±2.14
	F_{ROH}	0.10±0.11	0.19±0.10	D_{ROHet}	0.0006±0.0007	0.0009±0.0010
ITCA	N_{ROH}	38.73±26.53	144.93±48.62	N_{ROHet}	2.07±1.39	2.13±1.46
	L_{ROH}	7.7±1.3	3.6±0.6	L_{ROHet}	0.48±0.63	0.51±0.68
	S_{ROH}	300.4±226.5	509.8±235.2	S_{ROHet}	2.60±1.89	2.78±2.09
	F_{ROH}	0.13±0.10	0.23±0.10	D_{ROHet}	0.0012±0.0008	0.0012±0.0009
ITCS	N_{ROH}	66.77±16.51	194.62±54.70	N_{ROHet}	0.85±0.99	1.08±1.12
	L_{ROH}	8.6±2.7	4.3±0.9	L_{ROHet}	0.45±0.58	0.83±0.88
	S_{ROH}	600.4±303.7	855.9±258.1	S_{ROHet}	1.00±1.20	1.53±1.62
	F_{ROH}	0.27±0.13	0.38±0.11	D_{ROHet}	0.0004±0.0005	0.0007±0.0007
ITCT	N_{ROH}	31.33±24.93	95.27±33.59	N_{ROHet}	2.73±2.12	2.93±2.34
	L_{ROH}	14.8±3.9	6.7±2.0	L_{ROHet}	0.58±0.69	0.71±0.79
	S_{ROH}	523.5±446.7	665.0±441.3	S_{ROHet}	3.74±2.82	4.19±3.27
	F_{ROH}	0.23±0.20	0.29±0.20	D_{ROHet}	0.0017±0.0013	0.0019 ±0.0015
ITMR	N_{ROH}	88.78±33.62	189.78±113.54	N_{ROHet}	1.44±0.73	1.56±0.53
	L_{ROH}	10.9±3.1	6.4±1.7	L_{ROHet}	0.32±0.63	0.40±0.69
	S_{ROH}	1037.2±337.7	1274.2±185.9	S_{ROHet}	2.09±1.18	2.26±0.92
	F_{ROH}	0.46±0.15	0.56±0.08	D_{ROHet}	0.0009±0.0005	0.0010±0.0004
WBIT	N_{ROH}	53.27±17.87	225.53±25.14	N_{ROHet}	0.13±0.52	0.13±0.52
	L_{ROH}	6.7±1.5	3.0±0.4	L_{ROHet}	0.08±0.00	0.08±0.00
	S_{ROH}	374.0±162.3	701.1±173.1	S_{ROHet}	0.18±0.00	0.18±0.00
	F_{ROH}	0.17±0.07	0.30±0.1	D_{ROHet}	0.0001±0.0000	0.0001±0.0000
PDUR	N_{ROH}	77.87±8.23	206.63±12.00	N_{ROHet}	2.97±1.73	3.80±1.77
	L_{ROH}	6.9±1.6	3.9±0.8	L_{ROHet}	0.79±0.66	1.12±0.84
	S_{ROH}	568.4±97.8	849.3±99.7	S_{ROHet}	3.87±2.41	6.16±2.97
	F_{ROH}	0.25±0.04	0.38±0.04	D_{ROHet}	0.0017±0.0011	0.0027±0.0013
PLDR	N_{ROH}	54.10±12.50	158.63±15.26	N_{ROHet}	3.83±2.15	4.77±2.39
	L_{ROH}	6.7±1.5	3.7±0.6	L_{ROHet}	0.90±0.67	1.20±0.75
	S_{ROH}	380.0±130.5	605.6±136.7	S_{ROHet}	5.23±3.06	7.48±4.03
	F_{ROH}	0.17±0.06	0.27±0.06	D_{ROHet}	0.0023±0.0014	0.0033±0.0018
PLWT	N_{ROH}	58.47±10.23	164.53±14.21	N_{ROHet}	3.97±1.73	5.77±2.64
	L_{ROH}	7.1±1.7	3.8±0.8	L_{ROHet}	0.93±0.61	1.20±0.75
	S_{ROH}	436.0±123.4	661.3±124.8	S_{ROHet}	5.29±2.35	10.07±5.02
	F_{ROH}	0.19±0.05	0.29±0.06	D_{ROHet}	0.0023±0.0010	0.0045±0.0022

TABLE 3 (Continued)

		ROH			ROHet	
		SW	CR		SW	CR
Breed	Index	Mean±SD	Mean±SD	Index	Mean±SD	Mean±SD
PPIT	N_{ROH}	54.23±9.54	142.87±13.32	N_{ROHet}	5.27±2.12	6.53±2.54
	L_{ROH}	7.3±2.0	4.2±1.2	L_{ROHet}	0.93±0.61	1.16±0.75
	S_{ROH}	416.3±87.6	614.3±86.7	S_{ROHet}	7.23±2.94	10.69±4.31
	F_{ROH}	0.18±0.04	0.27±0.04	D_{ROHet}	0.0032±0.0013	0.0047±0.0019

Note: Parameters show mean values over individuals and chromosomes of the number of detected ROH (N_{ROH}) and ROHet (N_{ROHet}), of the length of ROH (L_{ROH}) and ROHet (L_{ROHet}) in Mb, of the sum of ROH (S_{ROH}) and ROHet (S_{ROHet}) in Mb, and of the coefficients of inbreeding (F_{ROH}) and diversity (D_{ROHet}), and the respective standard deviation (SD). For the full definition of breeds, see Table 1.

highest values for ITMR (SW = 1037.2, CR = 1274.2), and the lowest for NSIC (SW = 200.5, CR = 394.3). F_{ROH} highlighted low values of inbreeding for NSIC and ITNS (SW = 0.09–0.10; CR = 0.17–0.19), the highest values for ITMR (SW = 0.46; CR = 0.56), and relatively moderate inbreeding for PLDR, PLWT, and PPIT.

Not all the individuals reported segments of heterozygosity: the within-breed percentages of ROHet were very similar between SW and CR, varying from 7% in WBIT to 93–96% in NSIC, to 93–100% in cosmopolitan breeds (data not shown). The mean values of the parameters obtained with the two detection methods provided overlapping breed rankings. The Pearson correlation coefficients were higher than 0.96 for the four calculated parameters. PPIT showed the highest N_{ROHet} (SW = 5.27; CR = 6.53), followed by the other cosmopolitan breeds, then ITCT and NSIC (SW = 2.25; CR = 2.79), whereas WBIT had the lowest number of heterozygosity segments (0.13 for SW and CR). The mean length of heterozygosity segments (L_{ROHet}) reported the highest values for cosmopolitan breeds and NSIC (SW = 0.92; CR = 1.12). WBIT showed the lowest values (0.08 for SW and CR), whereas the other Italian local breeds had intermediate values. S_{ROHet} showed PPIT as the breed with the highest values (SW = 7.23; CR = 10.69), followed by all the other cosmopolitan breeds. ITCT and NSIC (SW = 2.97; CR = 4.23) had the highest values among the Italian breeds. WBIT was the sample with the lowest S_{ROHet} values detected with both methods.

The diversity index estimated from ROHet (D_{ROHet}) reported the highest values in cosmopolitan breeds with PPIT leading the ranking (SW = 0.0032; CR = 0.0047), followed by ITCT and NSIC, whereas WBIT showed the lowest internal diversity (0.0001 for SW and CR).

ROH and ROHet islands

Further analyses were performed on ROH and ROHet patterns to investigate autozygous and heterozygous segments that recurred within breeds. Significant markers (p-value ≥0.995) that overlapped between the two detection methods identified ROH and ROHet islands. Moreover, only those islands with a percentage of recurrence within breed ≥15% were taken into consideration.

Table S2 reports the genomic coordinates of each ROH island per breed, the number of included markers, and the annotated genes and QTL traits. The highest number of ROH islands was identified in ITMR: 13 islands of homozygosity containing 713 SNPs. Four ROH islands were detected in NSIC, counting 80 SNPs distributed in three chromosomes (SSC8, SSC11, and SCC14). The ROH island in SSC8 was shared with WBIT, the one in SSC14 was in common with both ITNS and WBIT. The ROH island in SSC14 was partially shared with ITNS, ITCA and ITCT.

The search on the PigQTLdb of SNP-within-ROH islands revealed 255 different markers associated with 321 different QTL (Table S2). The highest number of markers associated with QTL was detected in ITMR (146), while NSIC reported 25 SNPs linked to 10 different traits.

Table S3 reports the genomic coordinates of ROHet islands per breed, the number of included SNPs, and the annotated genes and QTL traits. NSIC showed only one rich heterozygosity island in SSC1, comprising 22 SNPs. Cosmopolitan breeds and ITCT had the highest number of ROHet islands, whereas ITNS and WBIT did not have any. The search on the PigQTLdb database of SNP-within-ROHet islands revealed 203 different markers associated with 252 different QTL (Table S3). NSIC showed 15 SNPs associated with five different traits and did not share any markers with the other breeds. The highest number of markers associated with QTL was detected in PPIT (63).

Tables S4 and S5 report the biological processes (BPs), the cellular components (CCs), the molecular functions and the Kyoto Encyclopedia of Genes and Genomes pathways of the genes within ROH and ROHet islands, respectively. Within ROH islands, the enrichment analysis showed outcomes for most breeds, except for ITCT, PDUR, PLDR, and PLWT.

The investigation of ROH-islands genes highlighted common BPs among NSIC, ITNS, and WBIT. They concern mechanisms for cytokine production involved in immune response (GO:0061081, GO:0061082), mechanisms involved in the regulation of

cell communication and cycle modulation (GO:0007346, GO:0009966, GO:0010564, GO:0050678, GO:1903047), and mechanisms related to the DNA metabolic process (GO:0051052). NSIC and ITNS also shared CC GO-terms (GO:0031981, GO:0044428, GO:0070013) related to the nucleus and nuclear lumen, whereas NSIC and WBIT shared the intracellular non-membrane-bounded organelle (GO:0043232). The only molecular function in NSIC and WBIT was the pyrophosphatase activity (GO:0016462), that is the catalysis of the hydrolysis of a pyrophosphate bond between two phosphate groups in order to generate energy.

Gene enrichment analysis on ROHet-islands genes outlined common GO-terms among cosmopolitan breeds. PDUR, PLWT, and PLDR shared two BPs (GO: 0000491, GO: 0070199), while PPIT shared one BP (GO: 0000491) and one CC (GO: 0044451) with PLDR. The BPs GO:0000491 (snoRNA activity) and GO:0070199 (mechanism that brings proteins to a specific position on a chromosome), involve the genes *RUVBL2* and *PIH1D1*, which, as part of the *R2TP* complex, regulate and suppress the RNA synthesis of paramyxovirus (Katoh et al., 2019).

DISCUSSION

High-throughput SNP arrays have greatly facilitated the study of genetic structure in livestock species, but while much effort has been devoted to commercial breeds, local populations are generally understudied (Beynon et al., 2015). In this study, the Nero Siciliano pig was investigated in comparison with other Italian and cosmopolitan breeds in order to understand its genetic structure and conservation status.

Genetic diversity indices

The genetic diversity indices, which are key parameters in the genetic management of local populations, showed a fair degree of variability in the Nero Siciliano and placed it in an intermediate position between the other local Italian pig breeds and the cosmopolitans. The expected heterozygosity was generally comparable to or higher than the observed heterozygosity, except for Calabrese and Mora Romagnola, for which an isolate-breaking effect can be supposed. In Nero Siciliano the values of H_O and H_E were close to those reported by D'Alessandro et al. (2019) and lower than those reported by Muñoz et al. (2019) in a sample of 50 individuals (H_O=0.342; H_E=0.360). The values of the inbreeding coefficient reflected the ranking of heterozygosity and were probably linked to the size of the registered breeds' individuals, particularly in Mora Romagnola, which includes about 377 breeding animals (ANAS). The Nero Siciliano showed relatively low inbreeding values (F_{IS}), higher than those reported for the same breed in previous studies (D'Alessandro et al., 2009; Muñoz et al., 2019) and comparable to those of the cosmopolitan breeds. Keeping several breeding animals, which ensures an adequate level of variability from generation to generation, is the basis of conservation plans (Meuwissen, 2009). As expected, *cNe* values are particularly low in the local breeds, except for Nero Siciliano whose *cNe* approaches those found in cosmopolitan breeds. The trend of *hNe* showed an expected increase that was particularly remarkable for wild boar, Nero Siciliano, and Large White.

Population structure

Nero Siciliano is reared in semi-extensive or extensive systems, exploits marginal woodlands, and has probably undergone introgressions by the wild boar present in the same breeding area (Iacolina et al., 2016). This breed has been officially recognised only in the recent past and for a long time it has not been subjected to any rational selection plan. The results regarding its genetic relationships and structure, investigated here with different statistics, converged towards a common direction. MDS analysis, model-based clustering, measurement of population differentiation, Neighbor-Net and Treemix highlighted the admixture between Nero Siciliano, the other four Italian black breeds and the wild boar. Tinarelli et al. (2021) reported the wild-type E^+ *MC1R* allele in 108 Nero Siciliano and five other local Italian breeds, thus indicating gene flow between the domestic pig and wild boar. Mascheroni (1927) reported the progressive decreases in the census of Nero Siciliano due to crossbreeding with Large White and Large Black breeds, and occasionally with Casertana boars. Although in the last 20 years the Italian local pig breeds have benefited from conservation programmes that limited and then precluded crossbreeding, the influence of gene flow on their genomic structure is still detectable. The Mora Romagnola showed a close relationship with the Duroc breed, as already reported by previous studies (Bovo, Ribani, Munoz, Alves, Araujo, Bozzi, Candek-Potokar, et al., 2020; Bovo, Ribani, Munoz, Alves, Araujo, Bozzi, Charneca, et al., 2020; Tinarelli et al., 2021) and confirmed by historical reports. Nero Siciliano showed low levels of admixture with Pietrain and even lower with Large White, not evenly distributed across the whole sample. Indeed, it highlighted an internal substructure probably linked to different family lines and past admixture events (Muñoz et al., 2018; Russo et al., 2004). This finding is also confirmed by the genomic structure observed in ITNS. Furthermore, Nero Siciliano showed its relative proximity to the other local black breeds through the genomic relationship that all have with wild boar, except for the Mora Romagnola. The evidence of genomic heterogeneity in the Nero Siciliano together with its improved management in the last twenty years explains the relatively low inbreeding detected.

The maximum likelihood tree, which explored the traces of admixture and gene flow of populations, underlined the general dichotomy between cosmopolitan and local breeds. The Nero Siciliano, as to be expected, was found close to the ITNS and to some of the Italian breeds that branched along with the wild boar. Although the Calabrese breed highlighted a migration event toward ITNS, it parted away from the rest of the local breeds and branched close to cosmopolitan breeds. This genetic relationship might be the result of the reported crossbreeding with Yorkshire, which is a breed derived from Large White (Bigi & Zanon, 2008). The observed migration events confirmed the past crossbreeding between local and cosmopolitan breeds, such as the influence of Duroc on Mora Romagnola and Large White on Casertana (Bigi & Zanon, 2008). Finally, the marked migration events highlighted from the wild boar to several domestic pigs might reflect gene exchanges not just limited to past events (Giuffra et al., 2000; Goedbloed et al., 2013; Tinarelli et al., 2021).

ROH and ROHet patterns

Consecutive stretches of homozygous and heterozygous SNP genotypes are both a consequence of the selection pressure that shapes the structure of livestock genome. Previous studies reported the distribution of ROHet or ROH and the function of genes within them to investigate the population characteristics in livestock species (e.g., Biscarini et al., 2020; Bizarria Dos Santos et al., 2021; Chen et al., 2022). The occurrence of ROHet avoids the deleterious effects of continuous homozygous genotypes and favours the so-called heterozygote advantage in immune-related genes as well as in productive and reproductive traits (Chen et al., 2022; Hedrick, 2015; Ruan et al., 2022; Sanglard et al., 2021). On the contrary, the study of ROH distribution has been a helpful tool to detect regions under selection (e.g. Schiavo, Bovo, Bertolini, et al., 2020; Szmatola et al., 2020; Xie et al., 2019).

Sliding window and consecutive runs are the methods of choice to detect stretches of consecutive ROH and ROHet in livestock species (e.g. Bizarria Dos Santos et al., 2021; Dixit et al., 2020; Mulim et al., 2022; Selli et al., 2021). Here, we investigated both approaches (Bizarria Dos Santos et al., 2021) and then outlined the results of the overlapping regions in order to identify and characterise the ROHet and ROH patterns, as well as to reveal the related fixed islands harbouring the candidate genes linked to specific traits.

The detection of autozygosity using the consecutive runs method outlined a mean N_{ROH} per breed that was two to six times higher than that reported by the sliding window approach, and this was reflected in the S_{ROH} values. Consequently, the values of F_{ROH} were higher when detected by CR than SW. On the contrary, the L_{ROH} was 2–3 times higher in SW, particularly in local breeds, whereas the values of cosmopolitan breeds were closer between the two methods.

Compared to the other local and cosmopolitan breeds, Nero Siciliano showed a low level of autozygosity, a result also confirmed by the indices highlighted in ITNS and by the genomic structure of the analysed samples. Among the local black-haired breeds, the Mora Romagnola, which has the lowest number of registered breeding animals, showed the highest average inbreeding values. Cosmopolitan breeds exhibited relatively moderate autozygosity, demonstrating adequate inbreeding control despite the high selection pressure.

As regards the identification of continuous heterozygosity segments, the two methods used have shown almost completely overlapping results. The parameters calculated with the two methods were very close, with the exception of $S_{ROH,}$ for which CR in some cases showed double values compared to SW. Pietrain, Large White, Landrace, and Duroc showed a high degree of heterozygosity, suggesting that selective breeding nonetheless preserved potentially advantageous regions of allelic diversity. Nero Siciliano and Casertana also showed relatively high values of continuous allelic diversity, which give value to the efforts of conservation programmes.

ROH and ROHet islands

Out of 18 investigated chromosomes, 12 reported ROH islands in all the breeds except for Duroc. SSC8 and SSC14 showed the largest number of islands per breed. Nero Siciliano had ROH islands on SSC8 and SSC14 that are in common with ITNS and wild boar, which is, in a way, expected. On the same SSC8 region, Muñoz et al. (2019) reported an outlier FST-windows region (92.95–96.07 Mbp), shared by a wild boar sample and five other breeds including Calabrese and Cinta Senese, which is linked to the *SCLT1* gene and the Ciliogenesis function. The homozygosity island on SSC14 (71.08–72.16 Mbp), found in Nero Siciliano, ITNS and wild boar, seems to involve haematological characters related to the functionality of the *STOCX1* and *DDX21* genes. Out of 18 markers that cover this hotspot region, 13 were reported in an association study for 30 haematological and blood clinical-biochemical traits in Large White pigs and were identified within a large basophil QTL region (66.4–73.4 Mbp) (Bovo et al., 2019). The SSC11 island spanning between 33.8 and 36.5 Mbp was a specific region of Nero Siciliano that partially overlapped the ROH island reported by Schiavo, Bovo, Bertolini, et al. (2020) in another sample of this breed and Italian Landrace. ITNS showed a homozygous hotspot on SSC4, one of the most studied chromosomes in the pig genome (Ma et al., 2010), which harbour several QTL linked to pork quality and carcass traits (Slawinska et al., 2009). SSC4 island in ITNS (97.3–99.0 Mbp) was already reported in Large White and Nero Siciliano (Schiavo, Bovo, Bertolini, et al., 2020) and counted 18 genes with different functions. Nonneman et al. (2013) reported highly significant markers in the same region and indicated

RORC as a candidate gene for intramuscular fat content in a Landrace–Duroc–Yorkshire composite population. Moreover, the genes *TUFT1*, *PI4KB*, *ZNF687*, *PSMD4*, *PIP5K1A*, *ARNT*, *VPS45*, *MCL1*, and *ENSA* were reported to be involved in the mesoderm tissue differentiation in a co-association network analysis underpinning transcription factor (*PPARγ*, *ELF1*, and *PRDM16*) on a crossed population (Iberian × Landrace). *PIP5K1A* was co-associated with the *PRDM16* gene, which is a key transcriptional factor determining adiposity (Puig-Oliveras et al., 2014).

Calabrese, Casertana, and ITNS partially shared a homozygous hotspot in SSC14 that was already reported to encompass markers associated with the teat number (Rohrer & Nonneman, 2017) and the intramuscular fat content QTL (Won et al., 2018) in a crossbred population and Berkshire pigs, respectively.

Cinta Sense and Mora Romagnola showed ROH islands already found in previous studies (Schiavo, Bovo, Bertolini, et al., 2020; Wang et al., 2022), including genes related to productive traits (Puig-Oliveras et al., 2014) and reproductive traits (Chen et al., 2019; He et al., 2017; Nonneman et al., 2016; Schneider et al., 2015). Large White and Pietrain also evidenced ROH islands previously reported with a high percentage in the same breeds (Gorssen et al., 2019; Schiavo, Bovo, Bertolini, et al., 2020).

ROHet islands were observed in SSC2, SSC6, SSC8, and SSC13. Some of the identified heterozygosity islands have been previously identified by other surveys as segments of fixed homozygosity. This finding underlines the general diversity among breeds, linked to the different breeding programmes, to the different productive orientations rather than to the internal homogeneity. No ROHet islands were identified in wild boar and ITNS. Nero Siciliano reported a single island on SSC1 composed of 22 markers, which is linked to production-related QTL and harboured six genes. Among the annotated genes, the *NR2E1* was reported as a candidate gene for the feed conversion ratio in Maxgro boars (Pietrain-based terminal line) (Horodyska et al., 2017) and the modulation of carcass and meat quality in cattle (Ramayo-Caldas et al., 2014).

It is worth highlighting that most observed heterozygosity-rich regions were found in the cosmopolitan pig breeds, then assuming a rational control of their degree of inbreeding. Several heterozygosity islands harboured genes previously reported in association with immune response, rather than with health-related QTL. Duroc and Large White shared the ROHet island on SSC2 (91.8–93.63 Mb), which was reported as ROH island by Schiavo, Bovo, Bertolini, et al. (2020) in Calabrese breed and partially overlapped the ROHet island detected in a Duroc sample (Ruan et al., 2022). This island harbours the *EDIL3*, an immune-related gene that was reported 3 times upregulated in umbilical-hernia-affected Landrace pigs (Souza et al., 2020). Chromosome 13 showed a rich heterozygosity hotspot (95.3–96.4 Mb) shared by Duroc, Landrace, Large White, and Calabrese breeds. This region mapped the gene *PLCH1*, known to be associated with the Ca2+ level related to QTL for haematological parameters (Bovo et al., 2016).

The Calabrese breed also showed the genes *ADAMTS3* and *ALB* mapped on SSC8 (68.6–69.6 Mb) that were related with QTL for haematological parameters in an intercross population (Large White boars × Minzhu sows) (Luo et al., 2012). The *CD96*, mapped in the ROHet island on SSC13 (147.7–148.9 Mb) in Duroc, was reported as an important gene involved in the placental development associated with QTL for mummified foetuses at birth trait in Large White × Landrace intercross (Onteru et al., 2012). In the Large White, the ROHet island on SSC13 (134.0–135.8 Mb) is associated with disease susceptibility and reproductive QTL. In particular, 11 markers out of 16 resulted in an association with the F4 enterotoxigenic *Escherichia coli* diarrhoeal disease susceptibility (Goetstouwers et al., 2014). Among the genes harbouring this region, the IQ motif containing G (*IQCG*) was identified as a candidate gene for sperm motility in Large White and Landrace pig breeds (Marques et al., 2018). The ROHet island located on SSC14 (46.5–47.9 Mb) in Pietrain and Cinta Senese breeds was reported as a ROH island in Duroc (Schiavo, Bovo, Bertolini, et al., 2020) and partially overlapped the ROHet island detected in a Duroc population by Ruan et al. (2022). In particular, this region evidenced 11 markers previously reported in association with umbilical hernia QTL (Grindflek et al., 2018). In Pietrain, the ROHet island on SSC15 (103.6–104.8 Mb) highlighted ALGA0086568 (rs81454185), which is located in the intron 6 of gene *NIF3L1* and resulted as the most significant marker responsible for the skin thickness in an intercross population (White Duroc × Erhualian F2) and Chinese Sutai pig (Ai et al., 2014). This region also annotated the *SPATS2L*, which was reported as a candidate gene for increasing litter size in pigs (Yang & Liu, 2015). Interestingly, the Pietrain breed that is irregularly black and white spotted also showed a ROHet island on SSC13 (51.3–52.7 Mb) harbouring the markers ALGA0070134 (rs81445596), ALGA0070129 (rs81445590) and SIRI0000807 (rs346246161) which were found to be significantly related to white spot patterns in Chinese pigs (Wang et al., 2015).

Finally, Cinta Senese, Casertana, Mora Romagnola, and cosmopolitan breeds outlined ROHet islands on SSC2, SSC6, SSC14, and SSC15 that mapped genes and QTL linked to productive traits (Howard et al., 2015; Kim et al., 2021; Li et al., 2011; Qi et al., 2022; Zappaterra et al., 2021; Zhang et al., 2015). In particular, the heterozygosity-rich segment, observed on SSC6 (54.2–55.9 Mb) in Duroc, Landrace, Large White, and Pietrain, correlates with QTL for average daily gain, daily feed intake, fat androstenone level, mean corpuscular haemoglobin concentration and meat colour, and partially overlapped a ROHet island detected by Ruan et al. (2022) in a Duroc sample.

CONCLUSIONS

This study reported an in-depth investigation of the genomic structure in an economically important breed such as the Nero Siciliano pig. The genetic diversity indices showed a moderate level of variability. The results from the several statistical approaches indicated a defined genetic background for Nero Siciliano and highlighted its genetic relationships with Italian local breeds and the wild boar. This is probably to be attributed to the semi-extensive breeding system that exposed Nero Siciliano pigs to natural cross-mating with wild boars. The genetic heterogeneity shown by this local breed was reflected in its relatively low inbreeding which is comparable to that of cosmopolitan breeds. The internal sub-structuring of the analysed sample has probably determined the moderate fixation of regions of homozygosity and the almost null fixation of regions of heterozygosity. Furthermore, the islands of homozygosity found in Nero Siciliano emphasised its relations with the wild boar and the other Italian breeds and highlighted the fixation of genes related to productive traits which are probably a trace of the past link with commercial breeds. The investigation of heterozygosity-rich regions highlighted several immune-related hotspots in the cosmopolitan pig breeds, demonstrating the adequate planning of matings which did not compromise their fitness response. This study confirms the historical and scientific reports that had previously described the Nero Siciliano. Furthermore, it underlines the need for adequate mating plans to make its genomic structure homogeneous and therefore enhance the breed and its productions.

ACKNOWLEDGMENTS

This work has been possible thanks to a collaboration between "Dipartimento of Agricoltura, Alimentazione e Ambiente—Di3A (Università di Catania)," and the components of the REDSUS PROJECT (PSR SICILIA 2014/2020). Part of this research was funded by the project "QUALIGEN"; Linea 2—Piano di Incentivi per la Ricerca di Ateneo 2020/2022; P.I. Giuseppe Luciano.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data supporting this study's findings have been submitted to a public repository and are available at the following link: https://figshare.com/s/ef170a6d2d365f7b261a—DOI: 10.6084/m9.figshare.21286566.

ORCID

Salvatore Bordonaro https://orcid.org/0000-0002-8723-9457
Giorgio Chessari https://orcid.org/0000-0002-9841-4820
Salvatore Mastrangelo https://orcid.org/0000-0001-6511-1981
Serena Tumino https://orcid.org/0000-0002-8221-2490
Andrea Criscione https://orcid.org/0000-0002-6701-1906

REFERENCES

Ai, H., Xiao, S., Zhang, Z., Yang, B., Li, L., Guo, Y. et al. (2014) Three novel quantitative trait loci for skin thickness in swine identified by linkage and genome-wide association studies. *Animal Genetics*, 45, 524–533.

Alexander, D.H., Novembre, J. & Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.

ANAS. (2022) SUIS, Nero Siciliano. Available from: https://suis.anas.it/html/NS_scheda.html. [Accessed 16 January 2023]

Beynon, S.E., Slavov, G.T., Farre, M., Sunduimijid, B., Waddams, K., Davies, B. et al. (2015) Population structure and history of the welsh sheep breeds determined by whole genome genotyping. *BMC Genetics*, 16, 65.

Bigi, D. & Zanon, A. (2008) *Suini. In: Atlante delle razze autoctone*. Milano: Edagricole. pp. 424–450.

Biscarini, F., Cozzi, P., Gaspa, G. & Marras, G. (2018) detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. *CRAN (the Comprehensive R Archive Network)*.

Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G. & Ciampolini, R. (2020) Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in Maremmana semi-feral cattle using pedigree and genomic data. *Animals*, 10, 2285.

Bizarria Dos Santos, W., Pimenta, S.G., Fonseca, M.G., Pereira, G.L., Loyola Chardulo, L.A., Rodrigues Machado Neto, O. et al. (2021) Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. *Journal of Animal Breeding and Genetics*, 138, 161–173.

Bovo, S., Mazzoni, G., Bertolini, F., Schiavo, G., Galimberti, G., Gallo, M. et al. (2019) Genome-wide association studies for 30 haematological and blood clinical-biochemical traits in large white pigs reveal genomic regions affecting intermediate phenotypes. *Scientific Reports*, 9, 7003.

Bovo, S., Ribani, A., Munoz, M., Alves, E., Araujo, J.P., Bozzi, R. et al. (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, 33.

Bovo, S., Ribani, A., Munoz, M., Alves, E., Araujo, J.P., Bozzi, R. et al. (2020) Genome-wide detection of copy number variants in European autochthonous and commercial pig breeds by whole-genome sequencing of DNA pools identified breed-characterising copy number states. *Animal Genetics*, 51, 541–556.

Bovo, S., Schiavo, G., Mazzoni, G., Dall'Olio, S., Galimberti, G., Calò, D.G. et al. (2016) Genome-wide association study for the level of serum electrolytes in Italian large white pigs. *Animal Genetics*, 47, 597–602.

Candek-Potokar, M. & Nieto Linan, R.M. (2019) *European local pig breeds—diversity and performance. A study of project TREASURE*. London: IntechOpen.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4, 7.

Chen, Z., Ye, S., Teng, J., Diao, S., Yuan, X., Chen, Z. et al. (2019) Genome-wide association studies for the number of animals born alive and dead in duroc pigs. *Theriogenology*, 139, 36–42.

Chen, Z., Zhang, Z., Wang, Z., Zhang, Z., Wang, Q. & Pan, Y. (2022) Heterozygosity and homozygosity regions affect reproductive success and the loss of reproduction: a case study with litter traits in pigs. *Computational and Structural Biotechnology Journal*, 20, 4060–4071.

D'Alessandro, E., Sottile, G., Sardina, M.T., Criscione, A., Bordonaro, S., Sutera, A.M. et al. (2019) Genome-wide analyses reveal the regions involved in the phenotypic diversity in Sicilian pigs. *Animal Genetics*, 51, 101–105.

Dixit, S.P., Singh, S., Ganguly, I., Bhatia, A.K., Sharma, A., Kumar, N.A. et al. (2020) Genome-wide runs of homozygosity revealed selection signatures in Bos indicus. *Frontiers in Genetics*, 11, 92.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J. & Ovenden, J.R. (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14, 209–214.

Excoffier, L. & Lischer, H.E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, 10, 564–567.

Fitak, R.R. (2021) OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, 6, 1–6.

Franci, O. & Pugliese, C. (2007) Italian autochthonous pigs: progress report and research perspectives. *Italian Journal of Animal Science*, 6, 663–671.

Giuffra, E., Kijas, J.M.H., Amarger, V., Carlborg, O., Jeon, J.-T. & Andersson, L. (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, 154, 1785–1791.

Goedbloed, D.J., Megens, H.J., Van Hooft, P., Herrero-Medrano, J.M., Lutz, W., Alexandri, P. et al. (2013) Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Molecular Ecology*, 22, 856–866.

Goetstouwers, T., Van Poucke, M., Coppieters, W., Nguyen, V.U., Melkebeek, V., Coddens, A. et al. (2014) Refined candidate region for F4ab/ac enterotoxigenic Escherichia coli susceptibility situated proximal to MUC13 in pigs. *PLoS One*, 9, e105013.

Gorssen, W., Meyermans, R., Buys, N. & Janssens, S. (2019) SNP genotypes reveal breed substructure, selection signatures and highly inbred regions in Pietrain pigs. *Animal Genetics*, 51, 32–42.

Grindflek, E., Hansen, M.H.S., Lien, S. & van Son, M. (2018) Genome-wide association study reveals a QTL and strong candidate genes for umbilical hernia in pigs on SSC14. *BMC Genomics*, 19, 412.

Guastella, A.M., Criscione, A., Marletta, D., Zuccaro, A., Chies, L. & Bordonaro, S. (2010) Molecular characterization and genetic structure of the Nero Siciliano pig breed. *Genetics and Molecular Biology*, 33, 650–656.

He, L.C., Li, P.H., Ma, X., Sui, S.P., Gao, S., Kim, S.W. et al. (2017) Identification of new single nucleotide polymorphisms affecting total number born and candidate genes related to ovulation rate in Chinese Erhualian pigs. *Animal Genetics*, 48, 48–54.

Hedrick, P.W. (2015) Heterozygote advantage: the effect of artificial selection in livestock and pets. *The Journal of Heredity*, 106, 141–154.

Herrero-Medrano, J.M., Megens, H.J., Groenen, M.A., Bosse, M., Pérez-Enciso, M. & Crooijmans, A.R.P. (2014) Whole-genome sequence analysis reveals differences in population management and selection of European low-input pig breeds. *BMC Genomics*, 15, 601.

Horodyska, J., Hamill, R.M., Varley, P.F., Reyer, H. & Wimmers, K. (2017) Genome-wide association analysis and functional annotation of positional candidate genes for feed conversion efficiency and growth rate in pigs. *PLoS One*, 12, e0173482.

Howard, J.T., Jiao, S., Tiezzi, F., Huang, Y., Gray, K.A. & Maltecca, C. (2015) Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc boars. *BMC Genetics*, 16, 59.

Huang da, W., Sherman, B.T. & Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4, 44–57.

Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267.

Iacolina, L., Scandura, M., Goedbloed, D.J., Alexandri, P., Crooijmans, R.P., Larson, G. et al. (2016) Genomic diversity and differentiation of a managed Island wild boar population. *Heredity*, 116, 60–67.

Katoh, H., Sekizuka, T., Nakatsu, Y., Nakagawa, R., Nao, N., Sakata, M. et al. (2019) The R2TP complex regulates paramyxovirus RNA synthesis. *PLoS Pathogens*, 15, e1007749.

Kim, K., Kang, J.K., Jung, Y.H., Lee, S.B., Rametta, R., Dongiovanni, P. et al. (2021) Adipocyte PHLPP2 inhibition prevents obesity-induced fatty liver. *Nature Communications*, 12, 1822.

Li, X., Kim, S.W., Do, K.T., Ha, Y.K., Lee, Y.M., Yoon, S.H. et al. (2011) Analyses of porcine public SNPs in coding-gene regions by re-sequencing and phenotypic association studies. *Molecular Biology Reports*, 38, 3805–3820.

Luo, W., Chen, S., Cheng, D., Wang, L., Li, Y., Ma, X. et al. (2012) Genome-wide association study of porcine hematological parameters in a large white × Minzhu F2 resource population. *International Journal of Biological Sciences*, 8, 870–881.

Ma, J.G., Chang, T.C., Yasue, H., Farmer, A.D., Crow, J.A., Eyer, K. et al. (2010) A high-resolution comparative map of porcine chromosome 4 (SSC4). *Animal Genetics*, 42, 440–444.

Marques, D.B.D., Bastiaansen, J.W.M., Broekhuijse, M.L.W.J., Lopes, M.S., Knol, E.F., Harlizius, B. et al. (2018) Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. *Genetics, Selection, Evolution*, 50, 40.

Mascheroni, E. (1927) *Zootecnica speciale*. Torino: UTET.

Meuwissen, T. (2009) Genetic management of small populations: a review. *Acta Agriculturae Scandinavica, Section A — Animal Science*, 59, 71–79.

Milanesi, M., Capomaccio, S., Vajana, E., Bomba, L., Garcia, J.F., Ajmone-Marsan, P. et al. (2017) BITE: An R Package for Biodiversity Analyses. *bioRxiv* 181610.

Mulim, H.A., Brito, L.F., Pinto, L.F.B., Ferraz, J.B.S., Grigoletto, L., Silva, M.R. et al. (2022) Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. *BMC Genomics*, 23, 209.

Muñoz, M., Bozzi, R., Garcia, F., Nunez, Y., Geraci, C., Crovetti, A. et al. (2018) Diversity across major and candidate genes in European local pig breeds. *PLoS One*, 13, e0207475.

Muñoz, M., Bozzi, R., Garcia-Casco, J., Nuñez, Y., Ribani, A., Franci, O. et al. (2019) Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. *Scientific Reports*, 9, 13546.

Nonneman, D.J., Schneider, J.F., Lents, C.A., Wiedmann, R.T., Vallet, J.L. & Rohrer, G.A. (2016) Genome-wide association and identification of candidate genes for age at puberty in swine. *BMC Genetics*, 17, 50.

Nonneman, D.J., Shackelford, S.D., King, D.A., Wheeler, T.L., Wiedmann, R.T., Snelling, W.M. et al. (2013) Genome-wide association of meat quality traits and tenderness in swine. *Journal of Animal Science*, 91, 4043–4050.

Onteru, S.K., Fan, B., Du, Z.Q., Garrick, D.J., Stalder, K.J. & Rothschild, M.F. (2012) A whole-genome association study for pig reproductive traits. *Animal Genetics*, 43, 18–26.

Pickrell, J.K. & Pritchard, J.K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8, e1002967.

Puig-Oliveras, A., Ballester, M., Corominas, J., Revilla, M., Estelle, J., Fernandez, A.I. et al. (2014) A co-association network analysis of the genetic determination of pig conformation, growth and fatness. *PLoS One*, 9, e114862.

Qi, K., Liu, Y., Li, C., Li, X., Li, X., Wang, K. et al. (2022) Construction of circRNA-related ceRNA networks in longissimus dorsi muscle of Queshan black and large white pigs. *Molecular Genetics and Genomics*, 297, 101–112.

R Core Team (2020) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing https://www.R–project.org/

Ramayo-Caldas, Y., Fortes, M.R., Hudson, N.J., Porto-Neto, L.R., Bolormaa, S., Barendse, W. et al. (2014) A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G, and FOXP3 in intramuscular fat deposition of beef cattle. *Journal of Animal Science*, 92, 2832–2845.

Ramos, A.M., Crooijmans, R.P., Affara, N.A., Amaral, A.J., Archibald, A.L., Beever, J.E. et al. (2009) Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One*, 4, e6524.

Rohrer, G.A. & Nonneman, D.J. (2017) Genetic analysis of teat number in pigs reveals some developmental pathways independent of vertebra number and several loci which only affect a specific side. *Genetics, Selection, Evolution*, 49, 4.

Ruan, D., Yang, J., Zhuang, Z., Ding, R., Huang, J., Quan, J. et al. (2022) Assessment of heterozygosity and genome-wide analysis of heterozygosity regions in two Duroc pig populations. *Frontiers in Genetics*, 12, 812456.

Russo, V., Fontanesi, L., Davoli, R., Chiofalo, L., Liotta, L. & Zumbo, A. (2004) Analysis of single nucleotide polymorphisms in major and candidate genes for production traits in Nero Siciliano pig breed. *Italian Journal of Animal Science*, 3, 19–29.

Sanglard, L.P., Huang, Y., Gray, K.A., Linhares, D.C.L., Dekkers, J.C.M., Niederwerder, M.C. et al. (2021) Further host-genomic characterization of total antibody response to PRRSV vaccination and its relationship with reproductive performance in commercial sows: genome-wide haplotype and zygosity analyses. *Genetics, Selection, Evolution*, 53, 91.

Santiago, E., Novo, I., Pardinas, A.F., Saura, M., Wang, J. & Caballero, A. (2020) Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Molecular Biology and Evolution*, 37, 3642–3653.

Schiavo, G., Bovo, S., Bertolini, F., Dall'Olio, S., Nanni, C.L., Tinarelli, S. et al. (2020) Runs of homozygosity islands in Italian cosmopolitan and autochthonous pig breeds identify selection signatures in the porcine genome. *Livestock Science*, 240, 104219.

Schiavo, G., Bovo, S., Munoz, M., Ribani, A., Alves, E., Araujo, J.P. et al. (2021) Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds. *Animal Genetics*, 52, 155–170.

Schiavo, G., Bovo, S., Tinarelli, S., Gallo, M., Dall'Olio, S. & Fontanesi, L. (2020) Genome-wide association analyses for coat colour patterns in the autochthonous Nero Siciliano pig breed. *Livestock Science*, 236, 104015.

Schneider, J.F., Miles, J.R., Brown-Brandl, T.M., Nienaber, J.A., Rohrer, G.A. & Vallet, J.L. (2015) Genomewide association analysis for average birth interval and stillbirth in swine. *Journal of Animal Science*, 93, 529–540.

Selli, A., Ventura, R.V., Fonseca, P.A.S., Buzanskas, M.E., Andrietta, L.T., Balieiro, J.C.C. et al. (2021) Detection and visualization of heterozygosity-rich regions and runs of homozygosity in worldwide sheep populations. *Animals*, 11, 2696.

Slawinska, A., Siwek, M., Knol, E.F., Roelofs-Prins, D.T., van Wijk, H.J., Dibbits, B. et al. (2009) Validation of the QTL on SSC4 for meat and carcass quality traits in a commercial crossbred pig population. *Journal of Animal Breeding and Genetics*, 126, 43–51.

Souza, M.R., Ibelli, A.M.G., Savoldi, I.R., Cantao, M.E., Peixoto, J.O., Mores, M.A.Z. et al. (2020) Transcriptome analysis identifies genes involved with the development of umbilical hernias in pigs. *PLoS One*, 15, e0232542.

Szmatola, T., Jasielczuk, I., Semik-Gurgul, E., Szyndler-Nedza, M., Blicharski, T., Szulc, K. et al. (2020) Detection of runs of homozygosity in conserved and commercial pig breeds in Poland. *Journal of Animal Breeding and Genetics*, 137, 571–580.

Tinarelli, S., Ribani, A., Utzeri, V.J., Taurisano, V., Bovo, C., Dall'Olio, S. et al. (2021) Redefinition of the Mora Romagnola pig breed herd book standard based on DNA markers useful to authenticate its "mono-breed" products: an example of sustainable conservation of a livestock genetic resource. *Animals*, 11, 526.

Wang, C., Wang, H., Zhang, Y., Tang, Z., Li, K. & Liu, B.J.M.E.R. (2015) Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in C hinese domestic pigs. *Molecular Ecology Resources*, 15, 414–424.

Wang, X., Li, G., Ruan, D., Zhuang, Z., Ding, R., Quan, J. et al. (2022) Runs of homozygosity uncover potential functional-altering mutation associated with body weight and length in two Duroc pig lines. *Front Vet Sci*, 9, 832633.

Waples, R.S. & Do, C. (2010) Linkage disequilibrium estimates of contemporary N e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, 3, 244–262.

Won, S., Jung, J., Park, E. & Kim, H. (2018) Identification of genes related to intramuscular fat content of pigs using genome-wide association study. *Asian-Australasian Journal of Animal Sciences*, 31, 157–162.

Xie, R., Shi, L., Liu, J., Deng, T., Wang, L., Liu, Y. et al. (2019) Genome-wide scan for runs of homozygosity identifies candidate genes in three pig breeds. *Animals*, 9, 518.

Yang, B., Cui, L., Perez-Enciso, M., Traspov, A., Crooijmans, R., Zinovieva, N. et al. (2017) Genome-wide SNP data unveils the globalization of domesticated pigs. *Genetics, Selection, Evolution*, 49, 71.

Yang, D. & Liu, Y.J.A.A.B. (2015) Molecular cloning, sequence identification, polymorphism and association of the porcine SPATS2L gene. *Archives Animal Breeding*, 58, 445–449.

Zappaterra, M., Gioiosa, S., Chillemi, G., Zambonelli, P. & Davoli, R. (2021) Dissecting the gene expression networks associated with variations in the major components of the fatty acid semimembranosus muscle profile in large white heavy pigs. *Animals*, 11, 628.

Zhang, R., Große-Brinkhaus, C., Heidt, H., Uddin, M.J., Cinar, M.U., Tesfaye, D. et al. (2015) Polymorphisms and expression analysis of SOX-6 in relation to porcine growth, carcass, and meat quality traits. *Meat Science*, 107, 26–32.

Zumbo, A., Sutera, A.M., Tardiolo, G. & D'Alessandro, E. (2020) Sicilian black pig: an overview. *Animals*, 10, 2326.

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: Bordonaro, S., Chessari, G., Mastrangelo, S., Senczuk, G., Chessa, S., Castiglioni, B. et al. (2023) Genome-wide population structure, homozygosity, and heterozygosity patterns of Nero Siciliano pig in the framework of Italian and cosmopolitan breeds. *Animal Genetics*, 00, 1–15. Available from: https://doi.org/10.1111/age.13344